Modelling Cellular Perturbations with the Sparse Additive Mechanism Shift Variational Autoencoder
Michael Bereket, Theofanis Karaletsos
Generative models of observations under interventions have been a vibrant topic of interest across machine learning and the sciences in recent years. For example, in drug discovery, there is a need to model the effects of diverse interventions on cells in order to characterize unknown biological mechanisms of action. We propose the Sparse Additive Mechanism Shift Variational Autoencoder, SAMS-VAE, to combine compositionality, disentanglement, and interpretability for perturbation models. SAMS-VAE models the latent state of a perturbed sample as the sum of a local latent variable capturing sample-specific variation and sparse global variables of latent intervention effects. Crucially, SAMS-VAE sparsifies these global latent variables for individual perturbations to identify disentangled, perturbation-specific latent subspaces that are flexibly composable. We evaluate SAMS-VAE both quantitatively and qualitatively on a range of tasks using two popular single cell sequencing datasets.In order to measure perturbation-specific model-properties, we also introduce a framework for evaluation of perturbation models based on average treatment effects with links to posterior predictive checks. SAMS-VAE outperforms comparable models in terms of generalization across in-distribution and out-of-distribution tasks, including a combinatorial reasoning task under resource paucity, and yields interpretable latent structures which correlate strongly to known biological mechanisms. Our results suggest SAMS-VAE is an interesting addition to the modeling toolkit for machine learning-driven scientific discovery.
****
Cross-Episodic Curriculum for Transformer Agents
Lucy Xiaoyang Shi, Yunfan Jiang, Jake Grigsby, Linxi Fan, Yuke Zhu
We present a new algorithm, Cross-Episodic Curriculum (CEC), to boost the learning efficiency and generalization of Transformer agents. Central to CEC is the placement of cross-episodic experiences into a Transformer’s context, which forms the basis of a curriculum. By sequentially structuring online learning trials and mixed-quality demonstrations, CEC constructs curricula that encapsulate learning progression and proficiency increase across episodes. Such synergy combined with the potent pattern recognition capabilities of Transformer models delivers a powerful cross-episodic attention mechanism. The effectiveness of CEC is demonstrated under two representative scenarios: one involving multi-task reinforcement learning with discrete control, such as in DeepMind Lab, where the curriculum captures the learning progression in both individual and progressively complex settings; and the other involving imitation learning with mixed-quality data for continuous control, as seen in RoboMimic, where the curriculum captures the improvement in demonstrators' expertise. In all instances, policies resulting from CEC exhibit superior performance and strong generalization. Code is open-sourced on the project website https://cec-agent.github.io/ to facilitate research on Transformer agent learning.
****
PaintSeg: Painting Pixels for Training-free Segmentation
Xiang Li, Chung-Ching Lin, Yinpeng Chen, Zicheng Liu, Jinglu Wang, Rita Singh, Bhiksha Raj
The paper introduces PaintSeg, a new unsupervised method for segmenting objects without any training. We propose an adversarial masked contrastive painting (AMCP) process, which creates a contrast between the original image and a painted image in which a masked area is painted using off-the-shelf generative models. During the painting process, inpainting and outpainting are alternated, with the former masking the foreground and filling in the background, and the latter masking the background while recovering the missing part of the foreground object. Inpainting and outpainting, also referred to as I-step and O-step, allow our method to gradually advance the target segmentation mask toward the ground truth without supervision or training. PaintSeg can be configured to work with a variety of prompts, e.g. coarse masks, boxes, scribbles, and points. Our experimental results demonstrate that PaintSeg outperforms existing approaches in coarse mask-pro

mpt, box-prompt, and point-prompt segmentation tasks, providing a training-free solution suitable for unsupervised segmentation. Code: https://github.com/lxa9867/PaintSeg.
****
Bootstrapping Vision-Language Learning with Decoupled Language Pre-training
Yiren Jian, Chongyang Gao, Soroush Vosoughi
We present a novel methodology aimed at optimizing the application of frozen large language models (LLMs) for resource-intensive vision-language (VL) pre-training. The current paradigm uses visual features as prompts to guide language models, with a focus on determining the most relevant visual features for corresponding text. Our approach diverges by concentrating on the language component, specifically identifying the optimal prompts to align with visual features. We introduce the Prompt-Transformer (P-Former), a model that predicts these ideal prompts, which is trained exclusively on linguistic data, bypassing the need for image-text pairings. This strategy subtly bifurcates the end-to-end VL training process into an additional, separate stage. Our experiments reveal that our framework significantly enhances the performance of a robust image-to-text baseline (BLIP-2), and effectively narrows the performance gap between models trained with either 4M or 129M image-text pairs. Importantly, our framework is modality-agnostic and flexible in terms of architectural design, as validated by its successful application in a video learning task using varied base modules. The code will be made available at https://github.com/yiren-jian/BLIText.
****
Path following algorithms for ℓ_2-regularized M-estimation with approximation guarantee
Yunzhang Zhu, Renxiong Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
PDF: Point Diffusion Implicit Function for Large-scale Scene Neural Representation
Yuhan Ding, Fukun Yin, Jiayuan Fan, Hui Li, Xin Chen, Wen Liu, Chongshan Lu, Gang Yu, Tao Chen
Recent advances in implicit neural representations have achieved impressive results by sampling and fusing individual points along sampling rays in the sampling space. However, due to the explosively growing sampling space, finely representing and synthesizing detailed textures remains a challenge for unbounded large-scale outdoor scenes. To alleviate the dilemma of using individual points to perceive the entire colossal space, we explore learning the surface distribution of the scene to provide structural priors and reduce the samplable space and propose a Point Diffusion implicit Function, PDF, for large-scale scene neural representation. The core of our method is a large-scale point cloud super-resolution diffusion module that enhances the sparse point cloud reconstructed from several training images into a dense point cloud as an explicit prior. Then in the rendering stage, only sampling points with prior points within the sampling radius are retained. That is, the sampling space is reduced from the unbounded space to the scene surface. Meanwhile, to fill in the background of the scene that cannot be provided by point clouds, the region sampling based on Mip-NeRF 360 is employed to model the background representation. Expensive experiments have demonstrated the effectiveness of our method for large-scale scene novel view synthesis, which outperforms relevant state-of-the-art baselines.
****
Natural Actor-Critic for Robust Reinforcement Learning with Function Approximation
Ruida Zhou, Tao Liu, Min Cheng, Dileep Kalathil, P. R. Kumar, Chao Tian
We study robust reinforcement learning (RL) with the goal of determining a well-performing policy that is robust against model mismatch between the training simulator and the testing environment. Previous policy-based robust RL algorithms m

ainly focus on the tabular setting under uncertainty sets that facilitate robust policy evaluation, but are no longer tractable when the number of states scales up. To this end, we propose two novel uncertainty set formulations, one based on double sampling and the other on an integral probability metric. Both make large-scale robust RL tractable even when one only has access to a simulator. We propose a robust natural actor-critic (RNAC) approach that incorporates the new uncertainty sets and employs function approximation. We provide finite-time convergence guarantees for the proposed RNAC algorithm to the optimal robust policy within the function approximation error. Finally, we demonstrate the robust performance of the policy learned by our proposed RNAC approach in multiple MuJoCo environments and a real-world TurtleBot navigation task.
****
Adaptive Selective Sampling for Online Prediction with Experts
Rui Castro, Fredrik Hellström, Tim van Erven
We consider online prediction of a binary sequence with expert advice. For this setting, we devise label-efficient forecasting algorithms, which use a selective sampling scheme that enables collecting much fewer labels than standard procedures. For the general case without a perfect expert, we prove best-of-both-worlds guarantees, demonstrating that the proposed forecasting algorithm always queries sufficiently many labels in the worst case to obtain optimal regret guarantees, while simultaneously querying much fewer labels in more benign settings. Specifically, for a scenario where one expert is strictly better than the others in expectation, we show that the label complexity of the label-efficient forecaster is roughly upper-bounded by the square root of the number of rounds. Finally, we present numerical experiments empirically showing that the normalized regret of the label-efficient forecaster can asymptotically match known minimax rates for pool-based active learning, suggesting it can optimally adapt to benign settings.
****
Gigastep - One Billion Steps per Second Multi-agent Reinforcement Learning
Mathias Lechner, lianhao yin, Tim Seyde, Tsun-Hsuan Johnson Wang, Wei Xiao, Ramin Hasani, Joshua Rountree, Daniela Rus
Multi-agent reinforcement learning (MARL) research is faced with a trade-off: it either uses complex environments requiring large compute resources, which makes it inaccessible to researchers with limited resources, or relies on simpler dynamics for faster execution, which makes the transferability of the results to more realistic tasks challenging. Motivated by these challenges, we present Gigastep, a fully vectorizable, MARL environment implemented in JAX, capable of executing up to one billion environment steps per second on consumer-grade hardware. Its design allows for comprehensive MARL experimentation, including a complex, high-dimensional space defined by 3D dynamics, stochasticity, and partial observations. Gigastep supports both collaborative and adversarial tasks, continuous and discrete action spaces, and provides RGB image and feature vector observations, allowing the evaluation of a wide range of MARL algorithms. We validate Gigastep's usability through an extensive set of experiments, underscoring its role in widening participation and promoting inclusivity in the MARL research community.
****
Attentive Transfer Entropy to Exploit Transient Emergence of Coupling Effect
Xiaolei Ru, XINYA ZHANG, Zijia Liu, Jack Murdoch Moore, Gang Yan
We consider the problem of reconstructing coupled networks (e.g., biological neural networks) connecting large numbers of variables (e.g.,nerve cells), of which state evolution is governed by dissipative dynamics consisting of strong self-drive (dominants the evolution) and weak coupling-drive. The core difficulty is sparseness of coupling effect that emerges (the coupling force is significant) only momentarily and otherwise remains quiescent in time series (e.g., neuronal activity sequence). Here we learn the idea from attention mechanism to guide the classifier to make inference focusing on the critical regions of time series data where coupling effect may manifest. Specifically, attention coefficients are assigned autonomously by artificial neural networks trained to maximise the Attentive Transfer Entropy (ATEn), which is a novel generalization of the iconic trans

fer entropy metric. Our results show that, without any prior knowledge of dynamics, ATEn explicitly identifies areas where the strength of coupling-drive is distinctly greater than zero. This innovation substantially improves reconstruction performance for both synthetic and real directed coupling networks using data generated by neuronal models widely used in neuroscience.
****
PopSign ASL v1.0: An Isolated American Sign Language Dataset Collected via Smartphones
Thad Starner, Sean Forbes, Matthew So, David Martin, Rohit Sridhar, Gururaj Deshpande, Sam Sepah, Sahir Shahryar, Khushi Bhardwaj, Tyler Kwok, Daksh Sehgal, Saad Hassan, Bill Neubauer, Sofia Vempala, Alec Tan, Jocelyn Heath, Unnathi Kumar, Priyanka Mosur, Tavenner Hall, Rajandeep Singh, Christopher Cui, Glenn Cameron, Sohier Dane, Garrett Tanzer
PopSign is a smartphone-based bubble-shooter game that helps hearing parentsof deaf infants learn sign language. To help parents practice their ability to sign, PopSign is integrating sign language recognition as part of its gameplay. Fortraining the recognizer, we introduce the PopSign ASL v1.0 dataset that collectsexamples of 250 isolated American Sign Language (ASL) signs using Pixel 4Asmartphone selfie cameras in a variety of environments. It is the largest publiclyavailable, isolated sign dataset by number of examples and is the first dataset tofocus on one-handed, smartphone signs. We collected over 210,000 examplesat 1944x2592 resolution made by 47 consenting Deaf adult signers for whomAmerican Sign Language is their primary language. We manually reviewed 217,866of these examples, of which 175,023 (approximately 700 per sign) were the signintended for the educational game. 39,304 examples were recognizable as a signbut were not the desired variant or were a different sign. We provide a training setof 31 signers, a validation set of eight signers, and a test set of eight signers. Abaseline LSTM model for the 250-sign vocabulary achieves 82.1% accuracy (81.9%class-weighted F1 score) on the validation set and 84.2% (83.9% class-weightedF1 score) on the test set. Gameplay suggests that accuracy will be sufficient forcreating educational games involving sign language recognition.
****
Provable Adversarial Robustness for Group Equivariant Tasks: Graphs, Point Clouds, Molecules, and More
Jan Schuchardt, Yan Scholten, Stephan Günnemann
A machine learning model is traditionally considered robust if its prediction remains (almost) constant under input perturbations with small norm. However, real-world tasks like molecular property prediction or point cloud segmentation have inherent equivariances, such as rotation or permutation equivariance. In such tasks, even perturbations with large norm do not necessarily change an input's semantic content. Furthermore, there are perturbations for which a model's prediction explicitly needs to change. For the first time, we propose a sound notion of adversarial robustness that accounts for task equivariance. We then demonstrate that provable robustness can be achieved by (1) choosing a model that matches the task's equivariances (2) certifying traditional adversarial robustness. Certification methods are, however, unavailable for many models, such as those with continuous equivariances. We close this gap by developing the framework of equivariance-preserving randomized smoothing, which enables architecture-agnostic certification. We additionally derive the first architecture-specific graph edit distance certificates, i.e. sound robustness guarantees for isomorphism equivariant tasks like node classification. Overall, a sound notion of robustness is an important prerequisite for future work at the intersection of robust and geometric machine learning.
****
Self-Supervised Motion Magnification by Backpropagating Through Optical Flow
Zhaoying Pan, Daniel Geng, Andrew Owens
This paper presents a simple, self-supervised method for magnifying subtle motions in video: given an input video and a magnification factor, we manipulate the video such that its new optical flow is scaled by the desired amount. To train our model, we propose a loss function that estimates the optical flow of the gene

rated video and penalizes how far if deviates from the given magnification facto r. Thus, training involves differentiating through a pretrained optical flow net work. Since our model is self-supervised, we can further improve its performance through test-time adaptation, by finetuning it on the input video. It can also be easily extended to magnify the motions of only user-selected objects. Our app roach avoids the need for synthetic magnification datasets that have been used t o train prior learning-based approaches. Instead, it leverages the existing capa bilities of off-the-shelf motion estimators. We demonstrate the effectiveness of our method through evaluations of both visual quality and quantitative metrics on a range of real-world and synthetic videos, and we show our method works for both supervised and unsupervised optical flow methods.
****
TexQ: Zero-shot Network Quantization with Texture Feature Distribution Calibrati on
Xinrui Chen, Yizhi Wang, Renao YAN, Yiqing Liu, Tian Guan, Yonghong He
Quantization is an effective way to compress neural networks. By reducing the bi t width of the parameters, the processing efficiency of neural network models at edge devices can be notably improved. Most conventional quantization methods ut ilize real datasets to optimize quantization parameters and fine-tune. Due to th e inevitable privacy and security issues of real samples, the existing real-data -driven methods are no longer applicable. Thus, a natural method is to introduce synthetic samples for zero-shot quantization (ZSQ). However, the conventional s ynthetic samples fail to retain the detailed texture feature distributions, whic h severely limits the knowledge transfer and performance of the quantized model. In this paper, a novel ZSQ method, TexQ is proposed to address this issue. We f irst synthesize a calibration image and extract its calibration center for each class with a texture feature energy distribution calibration method. Then, the c alibration centers are used to guide the generator to synthesize samples. Finall y, we introduce the mixup knowledge distillation module to diversify synthetic s amples for fine-tuning. Extensive experiments on CIFAR10/100 and ImageNet show t hat TexQ is observed to perform state-of-the-art in ultra-low bit width quantiza tion. For example, when ResNet-18 is quantized to 3-bit, TexQ achieves a 12.18% top-1 accuracy increase on ImageNet compared to state-of-the-art methods. Code a t https://github.com/dangsingrue/TexQ.
****
Ambient Diffusion: Learning Clean Distributions from Corrupted Data
Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alex Dimakis, Adam Kl ivans
We present the first diffusion-based framework that can learn an unknown distrib ution using only highly-corrupted samples. This problem arises in scientific app lications where access to uncorrupted samples is impossible or expensive to acqu ire. Another benefit of our approach is the ability to train generative models t hat are less likely to memorize any individual training sample, since they never observe clean training data. Our main idea is to introduce additional measureme nt distortion during the diffusion process and require the model to predict the original corrupted image from the further corrupted image. We prove that our me thod leads to models that learn the conditional expectation of the full uncorrup ted image given this additional measurement corruption. This holds for any corr uption process that satisfies some technical conditions (and in particular inclu des inpainting and compressed sensing). We train models on standard benchmarks (CelebA, CIFAR-10 and AFHQ) and show that we can learn the distribution even whe n all the training samples have 90\% of their pixels missing. We also show that we can finetune foundation models on small corrupted datasets (e.g. MRI scans wi th block corruptions) and learn the clean distribution without memorizing the tr aining set.
****
Scalable Membership Inference Attacks via Quantile Regression
Martin Bertran, Shuai Tang, Aaron Roth, Michael Kearns, Jamie H. Morgenstern, St even Z. Wu
Membership inference attacks are designed to determine, using black box access t

o trained models, whether a particular example was used in training or not. Membership inference can be formalized as a hypothesis testing problem. The most effective existing attacks estimate the distribution of some test statistic (usually the model's confidence on the true label) on points that were (and were not) used in training by training many \emph{shadow models}---i.e. models of the same architecture as the model being attacked, trained on a random subsample of data. While effective, these attacks are extremely computationally expensive, especially when the model under attack is large. \footnotetext[0]{Martin and Shuai are the lead authors, and other authors are ordered alphabetically. {maberlop,shuat}@amazon.com}We introduce a new class of attacks based on performing quantile regression on the distribution of confidence scores induced by the model under attack on points that are not used in training. We show that our method is competitive with state-of-the-art shadow model attacks, while requiring substantially less compute because our attack requires training only a single model. Moreover, unlike shadow model attacks, our proposed attack does not require any knowledge of the architecture of the model under attack and is therefore truly ``black-box". We show the efficacy of this approach in an extensive series of experiments on various datasets and model architectures. Our code is available at \href{https://github.com/amazon-science/quantile-mia}{github.com/amazon-science/quantile-mia.}
****
ESSEN: Improving Evolution State Estimation for Temporal Networks using Von Neumann Entropy
Qiyao Huang, Yingyue Zhang, Zhihong Zhang, Edwin Hancock
Temporal networks are widely used as abstract graph representations for real-world dynamic systems. Indeed, recognizing the network evolution states is crucial in understanding and analyzing temporal networks. For instance, social networks will generate the clustering and formation of tightly-knit groups or communities over time, relying on the triadic closure theory. However, the existing methods often struggle to account for the time-varying nature of these network structures, hindering their performance when applied to networks with complex evolution states. To mitigate this problem, we propose a novel framework called ESSEN, an Evolution StateS awarE Network, to measure temporal network evolution using von Neumann entropy and thermodynamic temperature. The developed framework utilizes a von Neumann entropy aware attention mechanism and network evolution state contrastive learning in the graph encoding. In addition, it employs a unique decoder the so-called Mixture of Thermodynamic Experts (MoTE) for decoding. ESSEN extracts local and global network evolution information using thermodynamic features and adaptively recognizes the network evolution states. Moreover, the proposed method is evaluated on link prediction tasks under both transductive and inductive settings, with the corresponding results demonstrating its effectiveness compared to various state-of-the-art baselines.
****
Label Correction of Crowdsourced Noisy Annotations with an Instance-Dependent Noise Transition Model
Hui GUO, Boyu Wang, Grace Yi
The predictive ability of supervised learning algorithms hinges on the quality of annotated examples, whose labels often come from multiple crowdsourced annotators with diverse expertise. To aggregate noisy crowdsourced annotations, many existing methods employ an annotator-specific instance-independent noise transition matrix to characterize the labeling skills of each annotator. Learning an instance-dependent noise transition model, however, is challenging and remains relatively less explored. To address this problem, in this paper, we formulate the noise transition model in a Bayesian framework and subsequently design a new label correction algorithm. Specifically, we approximate the instance-dependent noise transition matrices using a Bayesian network with a hierarchical spike and slab prior. To theoretically characterize the distance between the noise transition model and the true instance-dependent noise transition matrix, we provide a posterior-concentration theorem that ensures the posterior consistency in terms of the Hellinger distance. We further formulate the label correction process as a h

ypothesis testing problem and propose a novel algorithm to infer the true label from the noisy annotations based on the pairwise likelihood ratio test. Moreover, we establish an information-theoretic bound on the Bayes error for the proposed method. We validate the effectiveness of our approach through experiments on benchmark and real-world datasets.
****
Diffused Task-Agnostic Milestone Planner
Mineui Hong, Minjae Kang, Songhwai Oh
Addressing decision-making problems using sequence modeling to predict future trajectories shows promising results in recent years.In this paper, we take a step further to leverage the sequence predictive method in wider areas such as long-term planning, vision-based control, and multi-task decision-making.To this end, we propose a method to utilize a diffusion-based generative sequence model to plan a series of milestones in a latent space and to have an agent to follow the milestones to accomplish a given task.The proposed method can learn control-relevant, low-dimensional latent representations of milestones, which makes it possible to efficiently perform long-term planning and vision-based control.Furthermore, our approach exploits generation flexibility of the diffusion model, which makes it possible to plan diverse trajectories for multi-task decision-making.We demonstrate the proposed method across offline reinforcement learning (RL) benchmarks and an visual manipulation environment.The results show that our approach outperforms offline RL methods in solving long-horizon, sparse-reward tasks and multi-task problems,while also achieving the state-of-the-art performance on the most challenging vision-based manipulation benchmark.
****
Task-aware Distributed Source Coding under Dynamic Bandwidth
Po-han Li, Sravan Kumar Ankireddy, Ruihan (Philip) Zhao, Hossein Nourkhiz Mahjoub, Ehsan Moradi Pari, Ufuk Topcu, Sandeep Chinchali, Hyeji Kim
Efficient compression of correlated data is essential to minimize communication overload in multi-sensor networks. In such networks, each sensor independently compresses the data and transmits them to a central node. A decoder at the central node decompresses and passes the data to a pre-trained machine learning-based task model to generate the final output. Due to limited communication bandwidth, it is important for the compressor to learn only the features that are relevant to the task. Additionally, the final performance depends heavily on the total available bandwidth. In practice, it is common to encounter varying availability in bandwidth. Since higher bandwidth results in better performance, it is essential for the compressor to dynamically take advantage of the maximum available bandwidth at any instant. In this work, we propose a novel distributed compression framework composed of independent encoders and a joint decoder, which we call neural distributed principal component analysis (NDPCA). NDPCA flexibly compresses data from multiple sources to any available bandwidth with a single model, reducing compute and storage overhead. NDPCA achieves this by learning low-rank task representations and efficiently distributing bandwidth among sensors, thus providing a graceful trade-off between performance and bandwidth. Experiments show that NDPCA improves the success rate of multi-view robotic arm manipulation by 9% and the accuracy of object detection tasks on satellite imagery by 14% compared to an autoencoder with uniform bandwidth allocation.
****
BubbleML: A Multiphase Multiphysics Dataset and Benchmarks for Machine Learning
Sheikh Md Shakeel Hassan, Arthur Feeney, Akash Dhruv, Jihoon Kim, Youngjoon Suh, Jaiyoung Ryu, Yoonjin Won, Aparna Chandramowlishwaran
In the field of phase change phenomena, the lack of accessible and diverse datasets suitable for machine learning (ML) training poses a significant challenge. Existing experimental datasets are often restricted, with limited availability and sparse ground truth, impeding our understanding of this complex multiphysics phenomena. To bridge this gap, we present the BubbleML dataset which leverages physics-driven simulations to provide accurate ground truth information for various boiling scenarios, encompassing nucleate pool boiling, flow boiling, and sub-cooled boiling. This extensive dataset covers a wide range of parameters, includi

ng varying gravity conditions, flow rates, sub-cooling levels, and wall superheat, comprising 79 simulations. BubbleML is validated against experimental observations and trends, establishing it as an invaluable resource for ML research. Furthermore, we showcase its potential to facilitate the exploration of diverse downstream tasks by introducing two benchmarks: (a) optical flow analysis to capture bubble dynamics, and (b) neural PDE solvers for learning temperature and flow dynamics. The BubbleML dataset and its benchmarks aim to catalyze progress in ML-driven research on multiphysics phase change phenomena, providing robust baselines for the development and comparison of state-of-the-art techniques and models.
****
ANTN: Bridging Autoregressive Neural Networks and Tensor Networks for Quantum Many-Body Simulation
Zhuo Chen, Laker Newhouse, Eddie Chen, Di Luo, Marin Soljacic
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Causal Effect Identification in Uncertain Causal Networks
Sina Akbari, Fateme Jamshidi, Ehsan Mokhtarian, Matthew Vowels, Jalal Etesami, Negar Kiyavash
Causal identification is at the core of the causal inference literature, where complete algorithms have been proposed to identify causal queries of interest. The validity of these algorithms hinges on the restrictive assumption of having access to a correctly specified causal structure. In this work, we study the setting where a probabilistic model of the causal structure is available. Specifically, the edges in a causal graph exist with uncertainties which may, for example, represent degree of belief from domain experts. Alternatively, the uncertainty about an edge may reflect the confidence of a particular statistical test. The question that naturally arises in this setting is: Given such a probabilistic graph and a specific causal effect of interest, what is the subgraph which has the highest plausibility and for which the causal effect is identifiable? We show that answering this question reduces to solving an NP-hard combinatorial optimization problem which we call the edge ID problem. We propose efficient algorithms to approximate this problem and evaluate them against both real-world networks and randomly generated graphs.
****
FAST: a Fused and Accurate Shrinkage Tree for Heterogeneous Treatment Effects Estimation
Jia Gu, Caizhi Tang, Han Yan, Qing Cui, Longfei Li, Jun Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Characterizing Graph Datasets for Node Classification: Homophily-Heterophily Dichotomy and Beyond
Oleg Platonov, Denis Kuznedelev, Artem Babenko, Liudmila Prokhorenkova
Homophily is a graph property describing the tendency of edges to connect similar nodes; the opposite is called heterophily. It is often believed that heterophilous graphs are challenging for standard message-passing graph neural networks (GNNs), and much effort has been put into developing efficient methods for this setting. However, there is no universally agreed-upon measure of homophily in the literature. In this work, we show that commonly used homophily measures have critical drawbacks preventing the comparison of homophily levels across different datasets. For this, we formalize desirable properties for a proper homophily measure and verify which measures satisfy which properties. In particular, we show that a measure that we call adjusted homophily satisfies more desirable properties than other popular homophily measures while being rarely used in graph machin

e learning literature. Then, we go beyond the homophily-heterophily dichotomy and propose a new characteristic that allows one to further distinguish different sorts of heterophily. The proposed label informativeness (LI) characterizes how much information a neighbor's label provides about a node's label. We prove that this measure satisfies important desirable properties. We also observe empirically that LI better agrees with GNN performance compared to homophily measures, which confirms that it is a useful characteristic of the graph structure.
****
Equivariant Flow Matching with Hybrid Probability Transport for 3D Molecule Generation
Yuxuan Song, Jingjing Gong, Minkai Xu, Ziyao Cao, Yanyan Lan, Stefano Ermon, Hao Zhou, Wei-Ying Ma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Hyperbolic VAE via Latent Gaussian Distributions
Seunghyuk Cho, Juyong Lee, Dongwoo Kim
We propose a Gaussian manifold variational auto-encoder (GM-VAE) whose latent space consists of a set of Gaussian distributions. It is known that the set of the univariate Gaussian distributions with the Fisher information metric form a hyperbolic space, which we call a Gaussian manifold. To learn the VAE endowed with the Gaussian manifolds, we propose a pseudo-Gaussian manifold normal distribution based on the Kullback-Leibler divergence, a local approximation of the squared Fisher-Rao distance, to define a density over the latent space. We demonstrate the efficacy of GM-VAE on two different tasks: density estimation of image datasets and state representation learning for model-based reinforcement learning. GM-VAE outperforms the other variants of hyperbolic- and Euclidean-VAEs on density estimation tasks and shows competitive performance in model-based reinforcement learning. We observe that our model provides strong numerical stability, addressing a common limitation reported in previous hyperbolic-VAEs. The implementation is available at https://github.com/ml-postech/GM-VAE.
****
A Simple Solution for Offline Imitation from Observations and Examples with Possibly Incomplete Trajectories
Kai Yan, Alex Schwing, Yu-Xiong Wang
Offline imitation from observations aims to solve MDPs where only task-specific expert states and task-agnostic non-expert state-action pairs are available. Offline imitation is useful in real-world scenarios where arbitrary interactions are costly and expert actions are unavailable. The state-of-the-art ‘DIstribution Correction Estimation’ (DICE) methods minimize divergence of state occupancy between expert and learner policies and retrieve a policy with weighted behavior cloning; however, their results are unstable when learning from incomplete trajectories, due to a non-robust optimization in the dual domain. To address the issue, in this paper, we propose Trajectory-Aware Imitation Learning from Observations (TAILO). TAILO uses a discounted sum along the future trajectory as the weight for weighted behavior cloning. The terms for the sum are scaled by the output of a discriminator, which aims to identify expert states. Despite simplicity, TAILO works well if there exist trajectories or segments of expert behavior in the task-agnostic data, a common assumption in prior work. In experiments across multiple testbeds, we find TAILO to be more robust and effective, particularly with incomplete trajectories.
****
Defending against Data-Free Model Extraction by Distributionally Robust Defensive Training
Zhenyi Wang, Li Shen, Tongliang Liu, Tiehang Duan, Yanjun Zhu, Donglin Zhan, DAVID DOERMANN, Mingchen Gao
Data-Free Model Extraction (DFME) aims to clone a black-box model without knowing its original training data distribution, making it much easier for attackers t

o steal commercial models. Defense against DFME faces several challenges: (i) ef fectiveness; (ii) efficiency; (iii) no prior on the attacker's query data distri bution and strategy. However, existing defense methods: (1) are highly computati on and memory inefficient; or (2) need strong assumptions about attack data dist ribution; or (3) can only delay the attack or prove a model theft after the mode l stealing has happened. In this work, we propose a Memory and Computation effic ient defense approach, named MeCo, to prevent DFME from happening while maintain ing the model utility simultaneously by distributionally robust defensive traini ng on the target victim model. Specifically, we randomize the input so that it: (1) causes a mismatch of the knowledge distillation loss for attackers; (2) dist urbs the zeroth-order gradient estimation; (3) changes the label prediction for the attack query data. Therefore, the attacker can only extract misleading infor mation from the black-box model. Extensive experiments on defending against both decision-based and score-based DFME demonstrate that MeCo can significantly red uce the effectiveness of existing DFME methods and substantially improve running efficiency.
****
Large language models transition from integrating across position-yoked, exponen tial windows to structure-yoked, power-law windows
David Skrill, Samuel Norman-Haignere
Modern language models excel at integrating across long temporal scales needed t o encode linguistic meaning and show non-trivial similarities to biological neur al systems. Prior work suggests that human brain responses to language exhibit h ierarchically organized "integration windows" that substantially constrain the o verall influence of an input token (e.g., a word) on the neural response. Howeve r, little prior work has attempted to use integration windows to characterize co mputations in large language models (LLMs). We developed a simple word-swap proc edure for estimating integration windows from black-box language models that doe s not depend on access to gradients or knowledge of the model architecture (e.g. , attention weights). Using this method, we show that trained LLMs exhibit stere otyped integration windows that are well-fit by a convex combination of an expon ential and a power-law function, with a partial transition from exponential to p ower-law dynamics across network layers. We then introduce a metric for quantify ing the extent to which these integration windows vary with structural boundarie s (e.g., sentence boundaries), and using this metric, we show that integration w indows become increasingly yoked to structure at later network layers. None of t hese findings were observed in an untrained model, which as expected integrated uniformly across its input. These results suggest that LLMs learn to integrate i nformation in natural language using a stereotyped pattern: integrating across p osition-yoked, exponential windows at early layers, followed by structure-yoked, power-law windows at later layers. The methods we describe in this paper provid e a general-purpose toolkit for understanding temporal integration in language m odels, facilitating cross-disciplinary research at the intersection of biologica l and artificial intelligence.
****
Where are we in the search for an Artificial Visual Cortex for Embodied Intellig ence?
Arjun Majumdar, Karmesh Yadav, Sergio Arnaud, Jason Ma, Claire Chen, Sneha Silwa l, Aryan Jain, Vincent-Pierre Berges, Tingfan Wu, Jay Vakil, Pieter Abbeel, Jite ndra Malik, Dhruv Batra, Yixin Lin, Oleksandr Maksymets, Aravind Rajeswaran, Fra nziska Meier
We present the largest and most comprehensive empirical study of pre-trained vis ual representations (PVRs) or visual ‘foundation models’ for Embodied AI. First, we curate CortexBench, consisting of 17 different tasks spanning locomotion, na vigation, dexterous, and mobile manipulation. Next, we systematically evaluate e xisting PVRs and find that none are universally dominant. To study the effect of pre-training data size and diversity, we combine over 4,000 hours of egocentric videos from 7 different sources (over 4.3M images) and ImageNet to train differ ent-sized vision transformers using Masked Auto-Encoding (MAE) on slices of this data. Contrary to inferences from prior work, we find that scaling dataset size

and diversity does not improve performance universally (but does so on average). Our largest model, named VC-1, outperforms all prior PVRs on average but does not universally dominate either. Next, we show that task- or domain-specific adaptation of VC-1 leads to substantial gains, with VC-1 (adapted) achieving competitive or superior performance than the best known results on all of the benchmarks in CortexBench. Finally, we present real-world hardware experiments, in which VC-1 and VC-1 (adapted) outperform the strongest pre-existing PVR. Overall, this paper presents no new techniques but a rigorous systematic evaluation, a broad set of findings about PVRs (that in some cases, refute those made in narrow domains in prior work), and open-sourced code and models (that required over 10,000 GPU-hours to train) for the benefit of the research community.
****
Belief Projection-Based Reinforcement Learning for Environments with Delayed Feedback
Jangwon Kim, Hangyeol Kim, Jiwook Kang, Jongchan Baek, Soohee Han
We present a novel actor-critic algorithm for an environment with delayed feedback, which addresses the state-space explosion problem of conventional approaches. Conventional approaches use an augmented state constructed from the last observed state and actions executed since visiting the last observed state. Using the augmented state space, the correct Markov decision process for delayed environments can be constructed; however, this causes the state space to explode as the number of delayed timesteps increases, leading to slow convergence. Our proposed algorithm, called Belief-Projection-Based Q-learning (BPQL), addresses the state-space explosion problem by evaluating the values of the critic for which the input state size is equal to the original state-space size rather than that of the augmented one. We compare BPQL to traditional approaches in continuous control tasks and demonstrate that it significantly outperforms other algorithms in terms of asymptotic performance and sample efficiency. We also show that BPQL solves long-delayed environments, which conventional approaches are unable to do.
****
Batchnorm Allows Unsupervised Radial Attacks
Amur Ghose, Apurv Gupta, Yaoliang Yu, Pascal Poupart
The construction of adversarial examples usually requires the existence of soft or hard labels for each instance, with respect to which a loss gradient provides the signal for construction of the example. We show that for batch normalized deep image recognition architectures, intermediate latents that are produced after a batch normalization step by themselves suffice to produce adversarial examples using an intermediate loss solely utilizing angular deviations, without relying on any label. We motivate our loss through the geometry of batch normed representations and their concentration of norm on a hypersphere and distributional proximity to Gaussians. Our losses expand intermediate latent based attacks that usually require labels. The success of our method implies that leakage of intermediate representations may create a security breach for deployed models, which persists even when the model is transferred to downstream usage. Removal of batch norm weakens our attack, indicating it contributes to this vulnerability. Our attacks also succeed against LayerNorm empirically, thus being relevant for transformer architectures, most notably vision transformers which we analyze.
****
Detecting Any Human-Object Interaction Relationship: Universal HOI Detector with Spatial Prompt Learning on Foundation Models
Yichao Cao, Qingfei Tang, Xiu Su, Song Chen, Shan You, Xiaobo Lu, Chang Xu
Human-object interaction (HOI) detection aims to comprehend the intricate relationships between humans and objects, predicting triplets, and serving as the foundation for numerous computer vision tasks. The complexity and diversity of human-object interactions in the real world, however, pose significant challenges for both annotation and recognition, particularly in recognizing interactions within an open world context. This study explores the universal interaction recognition in an open-world setting through the use of Vision-Language (VL) foundation models and large language models (LLMs). The proposed method is dubbed as UniHOI. We conduct a deep analysis of the three hierarchical features inherent in visu

al HOI detectors and propose a method for high-level relation extraction aimed at VL foundation models, which we call HO prompt-based learning. Our design includes an HO Prompt-guided Decoder (HOPD), facilitates the association of high-level relation representations in the foundation model with various HO pairs within the image. Furthermore, we utilize a LLM (i.e. GPT) for interaction interpretation, generating a richer linguistic understanding for complex HOIs. For open-category interaction recognition, our method supports either of two input types: interaction phrase or interpretive sentence. Our efficient architecture design and learning methods effectively unleash the potential of the VL foundation models and LLMs, allowing UniHOI to surpass all existing methods with a substantial margin, under both supervised and zero-shot settings. The code and pre-trained weights will be made publicly available.
****
Smoothing the Landscape Boosts the Signal for SGD: Optimal Sample Complexity for Learning Single Index Models
Alex Damian, Eshaan Nichani, Rong Ge, Jason D. Lee
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Scale-Invariant Sorting Criterion to Find a Causal Order in Additive Noise Models
Alexander Reisach, Myriam Tami, Christof Seiler, Antoine Chambaz, Sebastian Weichwald
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
PROTES: Probabilistic Optimization with Tensor Sampling
Anastasiia Batsheva, Andrei Chertkov, Gleb Ryzhakov, Ivan Oseledets
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Perturbation Towards Easy Samples Improves Targeted Adversarial Transferability
Junqi Gao, Biqing Qi, Yao Li, Zhichang Guo, Dong Li, Yuming Xing, Dazhi Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
AllSim: Simulating and Benchmarking Resource Allocation Policies in Multi-User Systems
Jeroen Berrevoets, Daniel Jarrett, Alex Chan, Mihaela van der Schaar
Numerous real-world systems, ranging from healthcare to energy grids, involve users competing for finite and potentially scarce resources. Designing policies for resource allocation in such real-world systems is challenging for many reasons, including the changing nature of user types and their (possibly urgent) need for resources. Researchers have developed numerous machine learning solutions for determining resource allocation policies in these challenging settings. However, a key limitation has been the absence of good methods and test-beds for benchmarking these policies; almost all resource allocation policies are benchmarked in environments which are either completely synthetic or do not allow any deviation from historical data. In this paper we introduce AllSim, which is a benchmarking environment for realistically simulating the impact and utility of policies for resource allocation in systems in which users compete for such scarce resources. Building such a benchmarking environment is challenging because it needs to

successfully take into account the entire collective of potential users and the impact a resource allocation policy has on all the other users in the system. AllSim's benchmarking environment is modular (each component being parameterized individually), learnable (informed by historical data), and customizable (adaptable to changing conditions). These, when interacting with an allocation policy, produce a dataset of simulated outcomes for evaluation and comparison of such policies. We believe AllSim is an essential step towards a more systematic evaluation of policies for scarce resource allocation compared to current approaches for benchmarking such methods.
****
AVIS: Autonomous Visual Information Seeking with Large Language Model Agent
Ziniu Hu, Ahmet Iscen, Chen Sun, Kai-Wei Chang, Yizhou Sun, David Ross, Cordelia Schmid, Alireza Fathi
In this paper, we propose an autonomous information seeking visual question answering framework, AVIS. Our method leverages a Large Language Model (LLM) to dynamically strategize the utilization of external tools and to investigate their outputs via tree search, thereby acquiring the indispensable knowledge needed to provide answers to the posed questions. Responding to visual questions that necessitate external knowledge, such as "What event is commemorated by the building depicted in this image?", is a complex task. This task presents a combinatorial search space that demands a sequence of actions, including invoking APIs, analyzing their responses, and making informed decisions. We conduct a user study to collect a variety of instances of human decision-making when faced with this task. This data is then used to design a system comprised of three components: an LLM-powered planner that dynamically determines which tool to use next, an LLM-powered reasoner that analyzes and extracts key information from the tool outputs, and a working memory component that retains the acquired information throughout the process. The collected user behavior serves as a guide for our system in two key ways. First, we create a transition graph by analyzing the sequence of decisions made by users. This graph delineates distinct states and confines the set of actions available at each state. Second, we use examples of user decision-making to provide our LLM-powered planner and reasoner with relevant contextual instances, enhancing their capacity to make informed decisions. We show that AVIS achieves state-of-the-art results on knowledge-based visual question answering benchmarks such as Infoseek and OK-VQA.
****
Conformal Prediction Sets for Ordinal Classification
Prasenjit Dey, Srujana Merugu, Sivaramakrishnan R Kaveri
Ordinal classification (OC), i.e., labeling instances along classes with a natural ordering, is common in multiple applications such as size or budget based recommendations and disease severity labeling. Often in practical scenarios, it is desirable to obtain a small set of likely classes with a guaranteed high chance of including the true class. Recent works on conformal prediction (CP) address this problem for the classification setting with non-ordered labels but the resulting prediction sets (PS) are often non-contiguous and unsuitable for ordinal classification. In this work, we propose a framework to adapt existing CP methods to generate contiguous sets with guaranteed coverage and minimal cardinality. Our framework employs a novel non-parametric approach for modeling unimodal distributions. Empirical results on both synthetic and real-world datasets demonstrate our method outperforms SOTA baselines by 4% on Accuracy@K and 8% on PS size.
****
Minimax-Optimal Location Estimation
Shivam Gupta, Jasper Lee, Eric Price, Paul Valiant
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Tight Bounds for Volumetric Spanners and Applications
Aditya Bhaskara, Sepideh Mahabadi, Ali Vakilian

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Wyze Rule: Federated Rule Dataset for Rule Recommendation Benchmarking

Mohammad Mahdi Kamani, Yuhang Yao, Hanjia Lyu, Zhongwei Cheng, Lin Chen, Liangju Li, Carlee Joe-Wong, Jiebo Luo

In the rapidly evolving landscape of smart home automation, the potential of IoT devices is vast. In this realm, rules are the main tool utilized for this automation, which are predefined conditions or triggers that establish connections between devices, enabling seamless automation of specific processes. However, one significant challenge researchers face is the lack of comprehensive datasets to explore and advance the field of smart home rule recommendations. These datasets are essential for developing and evaluating intelligent algorithms that can effectively recommend rules for automating processes while preserving the privacy of the users, as it involves personal information about users' daily lives. To bridge this gap, we present the Wyze Rule Dataset, a large-scale dataset designed specifically for smart home rule recommendation research. Wyze Rule encompasses over 1 million rules gathered from a diverse user base of 300,000 individuals from Wyze Labs, offering an extensive and varied collection of real-world data. With a focus on federated learning, our dataset is tailored to address the unique challenges of a cross-device federated learning setting in the recommendation domain, featuring a large-scale number of clients with widely heterogeneous data. To establish a benchmark for comparison and evaluation, we have meticulously implemented multiple baselines in both centralized and federated settings. Researchers can leverage these baselines to gauge the performance and effectiveness of their rule recommendation systems, driving advancements in the domain. The Wyze Rule Dataset is publicly accessible through HuggingFace's dataset API.
****
Learning better with Dale’s Law: A Spectral Perspective

Pingsheng Li, Jonathan Cornford, Arna Ghosh, Blake Richards

Most recurrent neural networks (RNNs) do not include a fundamental constraint of real neural circuits: Dale's Law, which implies that neurons must be excitatory (E) or inhibitory (I). Dale's Law is generally absent from RNNs because simply partitioning a standard network's units into E and I populations impairs learning. However, here we extend a recent feedforward bio-inspired EI network architecture, named Dale's ANNs, to recurrent networks, and demonstrate that good performance is possible while respecting Dale's Law. This begs the question: What makes some forms of EI network learn poorly and others learn well? And, why does the simple approach of incorporating Dale's Law impair learning? Historically the answer was thought to be the sign constraints on EI network parameters, and this was a motivation behind Dale's ANNs. However, here we show the spectral properties of the recurrent weight matrix at initialisation are more impactful on network performance than sign constraints. We find that simple EI partitioning results in a singular value distribution that is multimodal and dispersed, whereas standard RNNs have an unimodal, more clustered singular value distribution, as do recurrent Dale's ANNs. We also show that the spectral properties and performance of partitioned EI networks are worse for small networks with fewer I units, and we present normalised SVD entropy as a measure of spectrum pathology that correlates with performance. Overall, this work sheds light on a long-standing mystery in neuroscience-inspired AI and computational neuroscience, paving the way for greater alignment between neural networks and biology.
****
Dense-Exponential Random Features: Sharp Positive Estimators of the Gaussian Kernel

Valerii Likhosherstov, Krzysztof M Choromanski, Kumar Avinava Dubey, Frederick Liu, Tamas Sarlos, Adrian Weller

The problem of efficient approximation of a linear operator induced by the Gaussian or softmax kernel is often addressed using random features (RFs) which yield

an unbiased approximation of the operator's result. Such operators emerge in important applications ranging from kernel methods to efficient Transformers. We propose parameterized, positive, non-trigonometric RFs which approximate Gaussian and softmax-kernels. In contrast to traditional RF approximations, parameters of these new methods can be optimized to reduce the variance of the approximation, and the optimum can be expressed in closed form. We show that our methods lead to variance reduction in practice (e^{10}-times smaller variance and beyond) and outperform previous methods in a kernel regression task. Using our proposed mechanism, we also present FAVOR#, a method for self-attention approximation in Transformers. We show that FAVOR# outperforms other random feature methods in speech modelling and natural language processing.
****
Projection-Free Online Convex Optimization via Efficient Newton Iterations
Khashayar Gatmiry, Zak Mhammedi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Read and Reap the Rewards: Learning to Play Atari with the Help of Instruction Manuals
Yue Wu, Yewen Fan, Paul Pu Liang, Amos Azaria, Yuanzhi Li, Tom M. Mitchell
High sample complexity has long been a challenge for RL. On the other hand, humans learn to perform tasks not only from interaction or demonstrations, but also by reading unstructured text documents, e.g., instruction manuals. Instruction manuals and wiki pages are among the most abundant data that could inform agents of valuable features and policies or task-specific environmental dynamics and reward structures. Therefore, we hypothesize that the ability to utilize human-written instruction manuals to assist learning policies for specific tasks should lead to a more efficient and better-performing agent. We propose the Read and Reward framework. Read and Reward speeds up RL algorithms on Atari games by reading manuals released by the Atari game developers. Our framework consists of a QA Extraction module that extracts and summarizes relevant information from the manual and a Reasoning module that evaluates object-agent interactions based on information from the manual. An auxiliary reward is then provided to a standard A2C RL agent, when interaction is detected. Experimentally, various RL algorithms obtain significant improvement in performance and training speed when assisted by our design. Code at github.com/Holmeswww/RnR
****
Sharpness Minimization Algorithms Do Not Only Minimize Sharpness To Achieve Better Generalization
Kaiyue Wen, Zhiyuan Li, Tengyu Ma
Despite extensive studies, the underlying reason as to why overparameterizedneural networks can generalize remains elusive. Existing theory shows that common stochastic optimizers prefer flatter minimizers of the training loss, and thusa natural potential explanation is that flatness implies generalization. This workcritically examines this explanation. Through theoretical and empirical investigation, we identify the following three scenarios for two-layer ReLU networks: (1)flatness provably implies generalization; (2) there exist non-generalizing flatte stmodels and sharpness minimization algorithms fail to generalize poorly, and (3)perhaps most strikingly, there exist non-generalizing flattest models, but sharpnessminimization algorithms still generalize. Our results suggest that the relationshipbetween sharpness and generalization subtly depends on the data distributionsand the model architectures and sharpness minimization algorithms do not onlyminimize sharpness to achieve better generalization. This calls for the search forother explanations for the generalization of over-parameterized neural networks
****
Feature-Learning Networks Are Consistent Across Widths At Realistic Scales
Nikhil Vyas, Alexander Atanasov, Blake Bordelon, Depen Morwani, Sabarish Sainath

an, Cengiz Pehlevan
We study the effect of width on the dynamics of feature-learning neural networks across a variety of architectures and datasets. Early in training, wide neural networks trained on online data have not only identical loss curves but also agree in their point-wise test predictions throughout training. For simple tasks such as CIFAR-5m this holds throughout training for networks of realistic widths. We also show that structural properties of the models, including internal representations, preactivation distributions, edge of stability phenomena, and large learning rate effects are consistent across large widths. This motivates the hypothesis that phenomena seen in realistic models can be captured by infinite-width, feature-learning limits. For harder tasks (such as ImageNet and language modeling), and later training times, finite-width deviations grow systematically. Two distinct effects cause these deviations across widths. First, the network output has an initialization-dependent variance scaling inversely with width, which can be removed by ensembling networks. We observe, however, that ensembles of narrower networks perform worse than a single wide network. We call this the bias of narrower width. We conclude with a spectral perspective on the origin of this finite-width bias.
****
Taylor TD-learning
Michele Garibbo, Maxime Robeyns, Laurence Aitchison
Many reinforcement learning approaches rely on temporal-difference (TD) learning to learn a critic.However, TD-learning updates can be high variance.Here, we introduce a model-based RL framework, Taylor TD, which reduces this variance in continuous state-action settings. Taylor TD uses a first-order Taylor series expansion of TD updates.This expansion allows Taylor TD to analytically integrate over stochasticity in the action-choice, and some stochasticity in the state distribution for the initial state and action of each TD update.We include theoretical and empirical evidence that Taylor TD updates are indeed lower variance than standard TD updates. Additionally, we show Taylor TD has the same stable learning guarantees as standard TD-learning with linear function approximation under a reasonable assumption.Next, we combine Taylor TD with the TD3 algorithm, forming TaTD3.We show TaTD3 performs as well, if not better, than several state-of-the art model-free and model-based baseline algorithms on a set of standard benchmark tasks.
****
Calibrating Neural Simulation-Based Inference with Differentiable Coverage Probability
Maciej Falkiewicz, Naoya Takeishi, Imahn Shekhzadeh, Antoine Wehenkel, Arnaud Delaunoy, Gilles Louppe, Alexandros Kalousis
Bayesian inference allows expressing the uncertainty of posterior belief under a probabilistic model given prior information and the likelihood of the evidence. Predominantly, the likelihood function is only implicitly established by a simulator posing the need for simulation-based inference (SBI). However, the existing algorithms can yield overconfident posteriors (Hermans et al., 2022) defeating the whole purpose of credibility if the uncertainty quantification is inaccurate. We propose to include a calibration term directly into the training objective of the neural model in selected amortized SBI techniques. By introducing a relaxation of the classical formulation of calibration error we enable end-to-end backpropagation. The proposed method is not tied to any particular neural model and brings moderate computational overhead compared to the profits it introduces. It is directly applicable to existing computational pipelines allowing reliable black-box posterior inference. We empirically show on six benchmark problems that the proposed method achieves competitive or better results in terms of coverage and expected posterior density than the previously existing approaches.
****
Agnostic Multi-Group Active Learning
Nicholas Rittler, Kamalika Chaudhuri
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Self-Weighted Contrastive Learning among Multiple Views for Mitigating Represent
ation Degeneration
Jie Xu, Shuo Chen, Yazhou Ren, Xiaoshuang Shi, Hengtao Shen, Gang Niu, Xiaofeng
Zhu
Recently, numerous studies have demonstrated the effectiveness of contrastive le
arning (CL), which learns feature representations by pulling in positive samples
 while pushing away negative samples. Many successes of CL lie in that there exi
sts semantic consistency between data augmentations of the same instance. In mul
ti-view scenarios, however, CL might cause representation degeneration when the
collected multiple views inherently have inconsistent semantic information or th
eir representations subsequently do not capture sufficient discriminative inform
ation. To address this issue, we propose a novel framework called SEM: SElf-weig
hted Multi-view contrastive learning with reconstruction regularization. Specifi
cally, SEM is a general framework where we propose to first measure the discrepa
ncy between pairwise representations and then minimize the corresponding self-we
ighted contrastive loss, and thus making SEM adaptively strengthen the useful pa
irwise views and also weaken the unreliable pairwise views. Meanwhile, we impose
 a self-supervised reconstruction term to regularize the hidden features of enco
ders, to assist CL in accessing sufficient discriminative information of data. E
xperiments on public multi-view datasets verified that SEM can mitigate represen
tation degeneration in existing CL methods and help them achieve significant per
formance improvements. Ablation studies also demonstrated the effectiveness of S
EM with different options of weighting strategies and reconstruction terms.
****
Neural Polarizer: A Lightweight and Effective Backdoor Defense via Purifying Poi
soned Features
Mingli Zhu, Shaokui Wei, Hongyuan Zha, Baoyuan Wu
Recent studies have demonstrated the susceptibility of deep neural networks to b
ackdoor attacks. Given a backdoored model, its prediction of a poisoned sample w
ith trigger will be dominated by the trigger information, though trigger informa
tion and benign information coexist. Inspired by the mechanism of the optical po
larizer that a polarizer could pass light waves with particular polarizations wh
ile filtering light waves with other polarizations, we propose a novel backdoor
defense method by inserting a learnable neural polarizer into the backdoored mod
el as an intermediate layer, in order to purify the poisoned sample via filterin
g trigger information while maintaining benign information. The neural polarizer
 is instantiated as one lightweight linear transformation layer, which is learne
d through solving a well designed bi-level optimization problem, based on a limi
ted clean dataset. Compared to other fine-tuning-based defense methods which oft
en adjust all parameters of the backdoored model, the proposed method only needs
 to learn one additional layer, such that it is more efficient and requires less
 clean data. Extensive experiments demonstrate the effectiveness and efficiency
of our method in removing backdoors across various neural network architectures
and datasets, especially in the case of very limited clean data. Codes are avail
able at \href{https://github.com/SCLBD/BackdoorBench}{https://github.com/SCLBD/B
ackdoorBench} (PyTorch) and \href{https://github.com/JulieCarlon/NPD-MindSpore}{
https://github.com/JulieCarlon/NPD-MindSpore} (MindSpore).
****
Tools for Verifying Neural Models' Training Data
Dami Choi, Yonadav Shavit, David K. Duvenaud
It is important that consumers and regulators can verify the provenance of large
 neural models to evaluate their capabilities and risks. We introduce the concep
t of a "Proof-of-Training-Data": any protocol that allows a model trainer to con
vince a Verifier of the training data that produced a set of model weights. Such
 protocols could verify the amount and kind of data and compute used to train th
e model, including whether it was trained on specific harmful or beneficial data
 sources. We explore efficient verification strategies for Proof-of-Training-Dat

a that are compatible with most current large-model training procedures. These include a method for the model-trainer to verifiably pre-commit to a random seed used in training, and a method that exploits models' tendency to temporarily overfit to training data in order to detect whether a given data-point was included in training. We show experimentally that our verification procedures can catch a wide variety of attacks, including all known attacks from the Proof-of-Learning literature.
****
Towards Higher Ranks via Adversarial Weight Pruning
Yuchuan Tian, Hanting Chen, Tianyu Guo, Chao Xu, Yunhe Wang
Convolutional Neural Networks (CNNs) are hard to deploy on edge devices due to its high computation and storage complexities. As a common practice for model compression, network pruning consists of two major categories: unstructured and structured pruning, where unstructured pruning constantly performs better. However, unstructured pruning presents a structured pattern at high pruning rates, which limits its performance. To this end, we propose a Rank-based PruninG (RPG) method to maintain the ranks of sparse weights in an adversarial manner. In each step, we minimize the low-rank approximation error for the weight matrices using singular value decomposition, and maximize their distance by pushing the weight matrices away from its low rank approximation. This rank-based optimization objective guides sparse weights towards a high-rank topology. The proposed method is conducted in a gradual pruning fashion to stabilize the change of rank during training. Experimental results on various datasets and different tasks demonstrate the effectiveness of our algorithm in high sparsity. The proposed RPG outperforms the state-of-the-art performance by 1.13\% top-1 accuracy on ImageNet in ResNet-50 with 98\% sparsity. The codes are available at https://github.com/huawei-noah/Efficient-Computing/tree/master/Pruning/RPG and https://gitee.com/mindspore/models/tree/master/research/cv/RPG.
****
On the Overlooked Pitfalls of Weight Decay and How to Mitigate Them: A Gradient-Norm Perspective
Zeke Xie, Zhiqiang Xu, Jingzhao Zhang, Issei Sato, Masashi Sugiyama
Weight decay is a simple yet powerful regularization technique that has been very widely used in training of deep neural networks (DNNs). While weight decay has attracted much attention, previous studies fail to discover some overlooked pitfalls on large gradient norms resulted by weight decay. In this paper, we discover that, weight decay can unfortunately lead to large gradient norms at the final phase (or the terminated solution) of training, which often indicates bad convergence and poor generalization. To mitigate the gradient-norm-centered pitfalls, we present the first practical scheduler for weight decay, called the Scheduled Weight Decay (SWD) method that can dynamically adjust the weight decay strength according to the gradient norm and significantly penalize large gradient norms during training. Our experiments also support that SWD indeed mitigates large gradient norms and often significantly outperforms the conventional constant weight decay strategy for Adaptive Moment Estimation (Adam).
****
Leveraging Early-Stage Robustness in Diffusion Models for Efficient and High-Quality Image Synthesis
Yulhwa Kim, Dongwon Jo, Hyesung Jeon, Taesu Kim, Daehyun Ahn, Hyungjun Kim, jae-joon kim
While diffusion models have demonstrated exceptional image generation capabilities, the iterative noise estimation process required for these models is compute-intensive and their practical implementation is limited by slow sampling speeds. In this paper, we propose a novel approach to speed up the noise estimation network by leveraging the robustness of early-stage diffusion models. Our findings indicate that inaccurate computation during the early-stage of the reverse diffusion process has minimal impact on the quality of generated images, as this stage primarily outlines the image while later stages handle the finer details that require more sensitive information. To improve computational efficiency, we combine our findings with post-training quantization (PTQ) to introduce a method tha

t utilizes low-bit activation for the early reverse diffusion process while main
taining high-bit activation for the later stages. Experimental results show that
 the proposed method can accelerate the early-stage computation without sacrific
ing the quality of the generated images.
****
Adversarial Model for Offline Reinforcement Learning
Mohak Bhardwaj, Tengyang Xie, Byron Boots, Nan Jiang, Ching-An Cheng
We propose a novel model-based offline Reinforcement Learning (RL) framework, ca
lled Adversarial Model for Offline Reinforcement Learning (ARMOR), which can rob
ustly learn policies to improve upon an arbitrary reference policy regardless of
 data coverage. ARMOR is designed to optimize policies for the worst-case perfor
mance relative to the reference policy through adversarially training a Markov d
ecision process model. In theory, we prove that ARMOR, with a well-tuned hyperpa
rameter, can compete with the best policy within data coverage when the referenc
e policy is supported by the data. At the same time, ARMOR is robust to hyperpar
ameter choices: the policy learned by ARMOR, with any admissible hyperparameter,
 would never degrade the performance of the reference policy, even when the refe
rence policy is not covered by the dataset. To validate these properties in prac
tice, we design a scalable implementation of ARMOR, which by adversarial trainin
g, can optimize policies without using model ensembles in contrast to typical mo
del-based methods. We show that ARMOR achieves competent performance with both s
tate-of-the-art offline model-free and model-based RL algorithms and can robustl
y improve the reference policy over various hyperparameter choices.
****
Training Your Image Restoration Network Better with Random Weight Network as Op
timization Function
man zhou, Naishan Zheng, Yuan Xu, Chun-Le Guo, Chongyi Li
The blooming progress made in deep learning-based image restoration has been lar
gely attributed to the availability of high-quality, large-scale datasets and ad
vanced network structures. However, optimization functions such as L1 and L2 are
 still de facto. In this study, we propose to investigate new optimization func
tions to improve image restoration performance. Our key insight is that ``rando
m weight network can be acted as a constraint for training better image restorat
ion networks''. However, not all random weight networks are suitable as constrai
nts. We draw inspiration from Functional theory and show that alternative random
 weight networks should be represented in the form of a strict mathematical mani
fold. We explore the potential of our random weight network prototypes that sati
sfy this requirement: Taylor's unfolding network, invertible neural network, ce
ntral difference convolution, and zero-order filtering. We investigate these pro
totypes from four aspects: 1)   random weight strategies, 2) network architectu
res, 3)   network depths, and 4) combinations of random weight networks. Further
more, we devise the random weight in two variants: the weights are randomly ini
tialized only once during the entire training procedure, and the weights are ra
ndomly initialized in each training epoch. Our approach can be directly integrat
ed into existing networks without incurring additional training and testing comp
utational costs. We perform extensive experiments across multiple image restorat
ion tasks, including image denoising, low-light image enhancement, and guided im
age super-resolution to demonstrate the consistent performance gains achieved by
 our method. Upon acceptance of this paper, we will release the code.
****
Passive learning of active causal strategies in agents and language models
Andrew Lampinen, Stephanie Chan, Ishita Dasgupta, Andrew Nam, Jane Wang
What can be learned about causality and experimentation from passive data? This
question is salient given recent successes of passively-trained language models
in interactive domains such as tool use. Passive learning is inherently limited.
 However, we show that purely passive learning can in fact allow an agent to lea
rn generalizable strategies for determining and using causal structures, as long
 as the agent can intervene at test time. We formally illustrate that learning a
 strategy of first experimenting, then seeking goals, can allow generalization f
rom passive learning in principle. We then show empirically that agents trained

via imitation on expert data can indeed generalize at test time to infer and use causal links which are never present in the training data; these agents can also generalize experimentation strategies to novel variable sets never observed in training.We then show that strategies for causal intervention and exploitation can be generalized from passive data even in a more complex environment with high-dimensional observations, with the support of natural language explanations. Explanations can even allow passive learners to generalize out-of-distribution from perfectly-confounded training data. Finally, we show that language models, trained only on passive next-word prediction, can generalize causal intervention strategies from a few-shot prompt containing explanations and reasoning. These results highlight the surprising power of passive learning of active causal strategies, and have implications for understanding the behaviors and capabilities of language models.
****
Zero-Regret Performative Prediction Under Inequality Constraints
Wenjing YAN, Xuanyu Cao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Towards Free Data Selection with General-Purpose Models
Yichen Xie, Mingyu Ding, Masayoshi TOMIZUKA, Wei Zhan
A desirable data selection algorithm can efficiently choose the most informative samples to maximize the utility of limited annotation budgets. However, current approaches, represented by active learning methods, typically follow a cumbersome pipeline that iterates the time-consuming model training and batch data selection repeatedly. In this paper, we challenge this status quo by designing a distinct data selection pipeline that utilizes existing general-purpose models to select data from various datasets with a single-pass inference without the need for additional training or supervision. A novel free data selection (FreeSel) method is proposed following this new pipeline. Specifically, we define semantic patterns extracted from inter-mediate features of the general-purpose model to capture subtle local information in each image. We then enable the selection of all data samples in a single pass through distance-based sampling at the fine-grained semantic pattern level. FreeSel bypasses the heavy batch selection process, achieving a significant improvement in efficiency and being 530x faster than existing active learning methods. Extensive experiments verify the effectiveness of FreeSel on various computer vision tasks.
****
Communication-Efficient Federated Bilevel Optimization with Global and Local Lower Level Problems
Junyi Li, Feihu Huang, Heng Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Partial Multi-Label Learning with Probabilistic Graphical Disambiguation
Jun-Yi Hang, Min-Ling Zhang
In partial multi-label learning (PML), each training example is associated with a set of candidate labels, among which only some labels are valid. As a common strategy to tackle PML problem, disambiguation aims to recover the ground-truth labeling information from such inaccurate annotations. However, existing approaches mainly rely on heuristics or ad-hoc rules to disambiguate candidate labels, which may not be universal enough in complicated real-world scenarios. To provide a principled way for disambiguation, we make a first attempt to explore the probabilistic graphical model for PML problem, where a directed graph is tailored to infer latent ground-truth labeling information from the generative process of partial multi-label data. Under the framework of stochastic gradient variational

Bayes, a unified variational lower bound is derived for this graphical model, which is further relaxed probabilistically so that the desired prediction model can be induced with simultaneously identified ground-truth labeling information. Comprehensive experiments on multiple synthetic and real-world data sets show that our approach outperforms the state-of-the-art counterparts.
****
Reward Scale Robustness for Proximal Policy Optimization via DreamerV3 Tricks
Ryan Sullivan, Akarsh Kumar, Shengyi Huang, John Dickerson, Joseph Suarez
Most reinforcement learning methods rely heavily on dense, well-normalized environment rewards. DreamerV3 recently introduced a model-based method with a number of tricks that mitigate these limitations, achieving state-of-the-art on a wide range of benchmarks with a single set of hyperparameters. This result sparked discussion about the generality of the tricks, since they appear to be applicable to other reinforcement learning algorithms. Our work applies DreamerV3's tricks to PPO and is the first such empirical study outside of the original work. Surprisingly, we find that the tricks presented do not transfer as general improvements to PPO. We use a high quality PPO reference implementation and present extensive ablation studies totaling over 10,000 A100 hours on the Arcade Learning Environment and the DeepMind Control Suite. Though our experiments demonstrate that these tricks do not generally outperform PPO, we identify cases where they succeed and offer insight into the relationship between the implementation tricks. In particular, PPO with these tricks performs comparably to PPO on Atari games with reward clipping and significantly outperforms PPO without reward clipping.
****
Emergent Correspondence from Image Diffusion
Luming Tang, Menglin Jia, Qianqian Wang, Cheng Perng Phoo, Bharath Hariharan
Finding correspondences between images is a fundamental problem in computer vision. In this paper, we show that correspondence emerges in image diffusion models without any explicit supervision. We propose a simple strategy to extract this implicit knowledge out of diffusion networks as image features, namely DIffusion FeaTures (DIFT), and use them to establish correspondences between real images. Without any additional fine-tuning or supervision on the task-specific data or annotations, DIFT is able to outperform both weakly-supervised methods and competitive off-the-shelf features in identifying semantic, geometric, and temporal correspondences. Particularly for semantic correspondence, DIFT from Stable Diffusion is able to outperform DINO and OpenCLIP by 19 and 14 accuracy points respectively on the challenging SPair-71k benchmark. It even outperforms the state-of-the-art supervised methods on 9 out of 18 categories while remaining on par for the overall performance. Project page: https://diffusionfeatures.github.io.
****
Robust Learning with Progressive Data Expansion Against Spurious Correlation
Yihe Deng, Yu Yang, Baharan Mirzasoleiman, Quanquan Gu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Multiclass Boosting: Simple and Intuitive Weak Learning Criteria
Nataly Brukhim, Amit Daniely, Yishay Mansour, Shay Moran
We study a generalization of boosting to the multiclass setting.We introduce a weak learning condition for multiclass classification that captures the original notion of weak learnability as being “slightly better than random guessing”. We give a simple and efficient boosting algorithm, that does not require realizability assumptions and its sample and oracle complexity bounds are independent of the number of classes. In addition, we utilize our new boosting technique in several theoretical applications within the context of List PAC Learning. First, we establish an equivalence to weak PAC learning. Furthermore, we present a new result on boosting for list learners, as well as provide a novel proof for the characterization of multiclass PAC learning and List PAC learning. Notably, our technique gives rise to simplified algorithms and analysis compared to previous work

s.
****
Approximate Heavy Tails in Offline (Multi-Pass) Stochastic Gradient Descent
Kruno Lehman, Alain Durmus, Umut Simsekli
A recent line of empirical studies has demonstrated that SGD might exhibit a heavy-tailed behavior in practical settings, and the heaviness of the tails might correlate with the overall performance. In this paper, we investigate the emergence of such heavy tails. Previous works on this problem only considered, up to our knowledge, online (also called single-pass) SGD, in which the emergence of heavy tails in theoretical findings is contingent upon access to an infinite amount of data. Hence, the underlying mechanism generating the reported heavy-tailed behavior in practical settings, where the amount of training data is finite, is still not well-understood. Our contribution aims to fill this gap. In particular, we show that the stationary distribution of offline (also called multi-pass) SGD exhibits ‘approximate’ power-law tails and the approximation error is controlled by how fast the empirical distribution of the training data converges to the true underlying data distribution in the Wasserstein metric. Our main takeaway is that, as the number of data points increases, offline SGD will behave increasingly ‘power-law-like’. To achieve this result, we first prove nonasymptotic Wasserstein convergence bounds for offline SGD to online SGD as the number of data points increases, which can be interesting on their own. Finally, we illustrate our theory on various experiments conducted on synthetic data and neural networks.
****
Uncovering Neural Scaling Laws in Molecular Representation Learning
Dingshuo Chen, Yanqiao Zhu, Jieyu Zhang, Yuanqi Du, Zhixun Li, Qiang Liu, Shu Wu, Liang Wang
Molecular Representation Learning (MRL) has emerged as a powerful tool for drug and materials discovery in a variety of tasks such as virtual screening and inverse design. While there has been a surge of interest in advancing model-centric techniques, the influence of both data quantity and quality on molecular representations is not yet clearly understood within this field. In this paper, we delve into the neural scaling behaviors of MRL from a data-centric viewpoint, examining four key dimensions: (1) data modalities, (2) dataset splitting, (3) the role of pre-training, and (4) model capacity.Our empirical studies confirm a consistent power-law relationship between data volume and MRL performance across these dimensions. Additionally, through detailed analysis, we identify potential avenues for improving learning efficiency.To challenge these scaling laws, we adapt seven popular data pruning strategies to molecular data and benchmark their performance. Our findings underline the importance of data-centric MRL and highlight possible directions for future research.
****
FlowCam: Training Generalizable 3D Radiance Fields without Camera Poses via Pixel-Aligned Scene Flow
Cameron Smith, Yilun Du, Ayush Tewari, Vincent Sitzmann
Reconstruction of 3D neural fields from posed images has emerged as a promising method for self-supervised representation learning. The key challenge preventing the deployment of these 3D scene learners on large-scale video data is their dependence on precise camera poses from structure-from-motion, which is prohibitively expensive to run at scale. We propose a method that jointly reconstructs camera poses and 3D neural scene representations online and in a single forward pass. We estimate poses by first lifting frame-to-frame optical flow to 3D scene flow via differentiable rendering, preserving locality and shift-equivariance of the image processing backbone. SE(3) camera pose estimation is then performed via a weighted least-squares fit to the scene flow field. This formulation enables us to jointly supervise pose estimation and a generalizable neural scene representation via re-rendering the input video, and thus, train end-to-end and fully self-supervised on real-world video datasets. We demonstrate that our method performs robustly on diverse, real-world video, notably on sequences traditionally challenging to optimization-based pose estimation techniques.

****
Minimum Description Length and Generalization Guarantees for Representation Lear
ning
Milad Sefidgaran, Abdellatif Zaidi, Piotr Krasnowski
A major challenge in designing efficient statistical supervised learning algorit
hms is finding representations that perform well not only on available training
samples but also on unseen data. While the study of representation learning has
spurred much interest, most existing such approaches are heuristic; and very lit
tle is known about theoretical generalization guarantees. For example, the infor
mation bottleneck method seeks a good generalization by finding a minimal descri
ption of the input that is maximally informative about the label variable, where
minimality and informativeness are both measured by Shannon’s mutual informatio
n. In this paper, we establish a compressibility framework that allows us to der
ive upper bounds on the generalization error of a representation learning algori
thm in terms of the ``Minimum Description Length'' (MDL) of the labels or the la
tent variables (representations). Rather than the mutual information between the
encoder’s input and the representation, which is often believed to reflect the
algorithm’s generalization capability in the related literature but in fact, fal
ls short of doing so, our new bounds involve the "multi-letter" relative entropy
between the distribution of the representations (or labels) of the training and
test sets and a fixed prior. In particular, these new bounds reflect the struct
ure of the encoder and are not vacuous for deterministic algorithms. Our compres
sibility approach, which is information-theoretic in nature, builds upon that of
Blum-Langford for PAC-MDL bounds and introduces two essential ingredients: bloc
k-coding and lossy-compression. The latter allows our approach to subsume the so
-called geometrical compressibility as a special case. To the best knowledge of
the authors, the established generalization bounds are the first of their kind f
or Information Bottleneck type encoders and representation learning. Finally, we
partly exploit the theoretical results by introducing a new data-dependent prio
r. Numerical simulations illustrate the advantages of well-chosen such priors ov
er classical priors used in IB.
****
From Discrete Tokens to High-Fidelity Audio Using Multi-Band Diffusion
Robin San Roman, Yossi Adi, Antoine Deleforge, Romain Serizel, Gabriel Synnaeve,
Alexandre Defossez
Deep generative models can generate high-fidelity audio conditioned on variousty
pes of representations (e.g., mel-spectrograms, Mel-frequency Cepstral Coefficie
nts(MFCC)). Recently, such models have been used to synthesize audiowaveforms co
nditioned on highly compressed representations. Although suchmethods produce imp
ressive results, they are prone to generate audible artifactswhen the conditioni
ng is flawed or imperfect. An alternative modeling approach isto use diffusion m
odels. However, these have mainly been used as speech vocoders(i.e., conditioned
on mel-spectrograms) or generating relatively low samplingrate signals. In this
work, we propose a high-fidelity multi-band diffusion-basedframework that gener
ates any type of audio modality (e.g., speech, music, environmentalsounds) from
low-bitrate discrete representations. At equal bit rate,the proposed approach ou
tperforms state-of-the-art generative techniques in termsof perceptual quality.
Training and evaluation code are available on the facebookresearch/audiocraft gi
thub project. Samples are available on the followinglink (https://ai.honu.io/pap
ers/mbd/).
****
Fixing the NTK: From Neural Network Linearizations to Exact Convex Programs
Rajat Vadiraj Dwaraknath, Tolga Ergen, Mert Pilanci
Recently, theoretical analyses of deep neural networks have broadly focused on t
wo directions: 1) Providing insight into neural network training by SGD in the l
imit of infinite hidden-layer width and infinitesimally small learning rate (als
o known as gradient flow) via the Neural Tangent Kernel (NTK), and 2) Globally o
ptimizing the regularized training objective via cone-constrained convex reformu
lations of ReLU networks. The latter research direction also yielded an alternat
ive formulation of the ReLU network, called a gated ReLU network, that is global

ly optimizable via efficient unconstrained convex programs. In this work, we interpret the convex program for this gated ReLU network as a Multiple Kernel Learning (MKL) model with a weighted data masking feature map and establish a connection to the NTK. Specifically, we show that for a particular choice of mask weights that do not depend on the learning targets, this kernel is equivalent to the NTK of the gated ReLU network on the training data. A consequence of this lack of dependence on the targets is that the NTK cannot perform better than the optimal MKL kernel on the training set. By using iterative reweighting, we improve the weights induced by the NTK to obtain the optimal MKL kernel which is equivalent to the solution of the exact convex reformulation of the gated ReLU network. We also provide several numerical simulations corroborating our theory. Additionally, we provide an analysis of the prediction error of the resulting optimal kernel via consistency results for the group lasso.
****

Birth of a Transformer: A Memory Viewpoint

Alberto Bietti, Vivien Cabannes, Diane Bouchacourt, Herve Jegou, Leon Bottou

Large language models based on transformers have achieved great empirical successes. However, as they are deployed more widely, there is a growing need to better understand their internal mechanisms in order to make them more reliable. These models appear to store vast amounts of knowledge from their training data, and to adapt quickly to new information provided in their context or prompt. We study how transformers balance these two types of knowledge by considering a synthetic setup where tokens are generated from either global or context-specific bigram distributions. By a careful empirical analysis of the training process on a simplified two-layer transformer, we illustrate the fast learning of global bigrams and the slower development of an "induction head" mechanism for the in-context bigrams. We highlight the role of weight matrices as associative memories, provide theoretical insights on how gradients enable their learning during training, and study the role of data-distributional properties.
****

A Variational Perspective on High-Resolution ODEs

Hoomaan Maskan, Konstantinos Zygalakis, Alp Yurtsever

We consider unconstrained minimization of smooth convex functions. We propose a novel variational perspective using forced Euler-Lagrange equation that allows for studying high-resolution ODEs. Through this, we obtain a faster convergence rate for gradient norm minimization using Nesterov's accelerated gradient method. Additionally, we show that Nesterov's method can be interpreted as a rate-matching discretization of an appropriately chosen high-resolution ODE. Finally, using the results from the new variational perspective, we propose a stochastic method for noisy gradients. Several numerical experiments compare and illustrate our stochastic algorithm with state of the art methods.
****

What You See is What You Read? Improving Text-Image Alignment Evaluation

Michal Yarom, Yonatan Bitton, Soravit Changpinyo, Roee Aharoni, Jonathan Herzig, Oran Lang, Eran Ofek, Idan Szpektor

Automatically determining whether a text and a corresponding image are semantically aligned is a significant challenge for vision-language models, with applications in generative text-to-image and image-to-text tasks. In this work, we study methods for automatic text-image alignment evaluation. We first introduce SeeTRUE: a comprehensive evaluation set, spanning multiple datasets from both text-to-image and image-to-text generation tasks, with human judgements for whether a given text-image pair is semantically aligned. We then describe two automatic methods to determine alignment: the first involving a pipeline based on question generation and visual question answering models, and the second employing an end-to-end classification approach by finetuning multimodal pretrained models. Both methods surpass prior approaches in various text-image alignment tasks, with significant improvements in challenging cases that involve complex composition or unnatural images. Finally, we demonstrate how our approaches can localize specific misalignments between an image and a given text, and how they can be used to automatically re-rank candidates in text-to-image generation.

****
On the Robustness of Mechanism Design under Total Variation Distance
Anuran Makur, Marios Mertzanidis, Alexandros Psomas, Athina Terzoglou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
$\mathcal{M}^4$: A Unified XAI Benchmark for Faithfulness Evaluation of Feature Attribution Methods across Metrics, Modalities and Models
Xuhong Li, Mengnan Du, Jiamin Chen, Yekun Chai, Himabindu Lakkaraju, Haoyi Xiong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A generative model of the hippocampal formation trained with theta driven local learning rules
Tom M George, Kimberly L. Stachenfeld, Caswell Barry, Claudia Clopath, Tomoki Fukai
Advances in generative models have recently revolutionised machine learning. Meanwhile, in neuroscience, generative models have long been thought fundamental to animal intelligence. Understanding the biological mechanisms that support these processes promises to shed light on the relationship between biological and artificial intelligence. In animals, the hippocampal formation is thought to learn and use a generative model to support its role in spatial and non-spatial memory. Here we introduce a biologically plausible model of the hippocampal formation tantamount to a Helmholtz machine that we apply to a temporal stream of inputs. A novel component of our model is that fast theta-band oscillations (5-10 Hz) gate the direction of information flow throughout the network, training it akin to a high-frequency wake-sleep algorithm. Our model accurately infers the latent state of high-dimensional sensory environments and generates realistic sensory predictions. Furthermore, it can learn to path integrate by developing a ring attractor connectivity structure matching previous theoretical proposals and flexibly transfer this structure between environments. Whereas many models trade-off biological plausibility with generality, our model captures a variety of hippocampal cognitive functions under one biologically plausible local learning rule.
****
Risk-Averse Model Uncertainty for Distributionally Robust Safe Reinforcement Learning
James Queeney, Mouhacine Benosman
Many real-world domains require safe decision making in uncertain environments. In this work, we introduce a deep reinforcement learning framework for approaching this important problem. We consider a distribution over transition models, and apply a risk-averse perspective towards model uncertainty through the use of coherent distortion risk measures. We provide robustness guarantees for this framework by showing it is equivalent to a specific class of distributionally robust safe reinforcement learning problems. Unlike existing approaches to robustness in deep reinforcement learning, however, our formulation does not involve minimax optimization. This leads to an efficient, model-free implementation of our approach that only requires standard data collection from a single training environment. In experiments on continuous control tasks with safety constraints, we demonstrate that our framework produces robust performance and safety at deployment time across a range of perturbed test environments.
****
Optimal approximation using complex-valued neural networks
Paul Geuchen, Felix Voigtlaender
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
BayesDAG: Gradient-Based Posterior Inference for Causal Discovery
Yashas Annadani, Nick Pawlowski, Joel Jennings, Stefan Bauer, Cheng Zhang, Wenbo Gong
Bayesian causal discovery aims to infer the posterior distribution over causal models from observed data, quantifying epistemic uncertainty and benefiting downstream tasks. However, computational challenges arise due to joint inference over combinatorial space of Directed Acyclic Graphs (DAGs) and nonlinear functions. Despite recent progress towards efficient posterior inference over DAGs, existing methods are either limited to variational inference on node permutation matrices for linear causal models, leading to compromised inference accuracy, or continuous relaxation of adjacency matrices constrained by a DAG regularizer, which cannot ensure resulting graphs are DAGs. In this work, we introduce a scalable Bayesian causal discovery framework based on a combination of stochastic gradient Markov Chain Monte Carlo (SG-MCMC) and Variational Inference (VI) that overcomes these limitations. Our approach directly samples DAGs from the posterior without requiring any DAG regularization, simultaneously draws function parameter samples and is applicable to both linear and nonlinear causal models. To enable our approach, we derive a novel equivalence to the permutation-based DAG learning, which opens up possibilities of using any relaxed gradient estimator defined over permutations. To our knowledge, this is the first framework applying gradient-based MCMC sampling for causal discovery. Empirical evaluation on synthetic and real-world datasets demonstrate our approach's effectiveness compared to state-of-the-art baselines.
****
Bounce: Reliable High-Dimensional Bayesian Optimization for Combinatorial and Mixed Spaces
Leonard Papenmeier, Luigi Nardi, Matthias Poloczek
Impactful applications such as materials discovery, hardware design, neural architecture search, or portfolio optimization require optimizing high-dimensional black-box functions with mixed and combinatorial input spaces.While Bayesian optimization has recently made significant progress in solving such problems, an in-depth analysis reveals that the current state-of-the-art methods are not reliable. Their performances degrade substantially when the unknown optima of the function do not have a certain structure. To fill the need for a reliable algorithm for combinatorial and mixed spaces, this paper proposes Bounce that relies on a novel map of various variable types into nested embeddings of increasing dimensionality.Comprehensive experiments show that Bounce reliably achieves and often even improves upon state-of-the-art performance on a variety of high-dimensional problems.
****
Uniform-in-Time Wasserstein Stability Bounds for (Noisy) Stochastic Gradient Descent
Lingjiong Zhu, Mert Gurbuzbalaban, Anant Raj, Umut Simsekli
Algorithmic stability is an important notion that has proven powerful for deriving generalization bounds for practical algorithms. The last decade has witnessed an increasing number of stability bounds for different algorithms applied on different classes of loss functions. While these bounds have illuminated various properties of optimization algorithms, the analysis of each case typically required a different proof technique with significantly different mathematical tools. In this study, we make a novel connection between learning theory and applied probability and introduce a unified guideline for proving Wasserstein stability bounds for stochastic optimization algorithms. We illustrate our approach on stochastic gradient descent (SGD) and we obtain time-uniform stability bounds (i.e., the bound does not increase with the number of iterations) for strongly convex losses and non-convex losses with additive noise, where we recover similar results to the prior art or extend them to more general cases by using a single proof technique. Our approach is flexible and can be generalizable to other popular optimizers, as it mainly requires developing Lyapunov functions, which are often

readily available in the literature. It also illustrates that ergodicity is an i
mportant component for obtaining time-uniform bounds -- which might not be achi
eved for convex or non-convex losses unless additional noise is injected to the
iterates. Finally, we slightly stretch our analysis technique and prove time-uni
form bounds for SGD under convex and non-convex losses (without additional addit
ive noise), which, to our knowledge, is novel.
****
Towards Generic Semi-Supervised Framework for Volumetric Medical Image Segmentat
ion
Haonan Wang, Xiaomeng Li
Volume-wise labeling in 3D medical images is a time-consuming task that requires
 expertise. As a result, there is growing interest in using semi-supervised lear
ning (SSL) techniques to train models with limited labeled data. However, the ch
allenges and practical applications extend beyond SSL to settings such as unsupe
rvised domain adaptation (UDA) and semi-supervised domain generalization (SemiDG
). This work aims to develop a generic SSL framework that can handle all three s
ettings. We identify two main obstacles to achieving this goal in the existing S
SL framework: 1) the weakness of capturing distribution-invariant features; and
2) the tendency for unlabeled data to be overwhelmed by labeled data, leading to
 over-fitting to the labeled data during training. To address these issues, we p
ropose an Aggregating & Decoupling framework. The aggregating part consists of a
 Diffusion encoder that constructs a "common knowledge set" by extracting distri
bution-invariant features from aggregated information from multiple distribution
s/domains. The decoupling part consists of three decoders that decouple the trai
ning process with labeled and unlabeled data, thus avoiding over-fitting to labe
led data, specific domains and classes. We evaluate our proposed framework on fo
ur benchmark datasets for SSL, Class-imbalanced SSL, UDA and SemiDG. The results
 showcase notable improvements compared to state-of-the-art methods across all f
our settings, indicating the potential of our framework to tackle more challengi
ng SSL scenarios. Code and models are available at: https://github.com/xmed-lab/
GenericSSL.
****
Stochastic Distributed Optimization under Average Second-order Similarity: Algor
ithms and Analysis
Dachao Lin, Yuze Han, Haishan Ye, Zhihua Zhang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
PolyDiffuse: Polygonal Shape Reconstruction via Guided Set Diffusion Models
Jiacheng Chen, Ruizhi Deng, Yasutaka Furukawa
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Can You Rely on Your Model Evaluation? Improving Model Evaluation with Synthetic
 Test Data
Boris van Breugel, Nabeel Seedat, Fergus Imrie, Mihaela van der Schaar
Evaluating the performance of machine learning models on diverse and underrepres
ented subgroups is essential for ensuring fairness and reliability in real-world
 applications. However, accurately assessing model performance becomes challengi
ng due to two main issues: (1) a scarcity of test data, especially for small sub
groups, and (2) possible distributional shifts in the model's deployment setting
, which may not align with the available test data. In this work, we introduce
3S Testing, a deep generative modeling framework to facilitate model evaluation
by generating synthetic test sets for small subgroups and simulating distributio
nal shifts. Our experiments demonstrate that 3S-Testing outperforms traditional
baselines---including real test data alone---in estimating model performance on

minority subgroups and under plausible distributional shifts. In addition, 3S of
fers intervals around its performance estimates, exhibiting superior coverage of
 the ground truth compared to existing approaches. Overall, these results raise
 the question of whether we need a paradigm shift away from limited real test da
ta towards synthetic test data.
****
Rethinking the Backward Propagation for Adversarial Transferability
Wang Xiaosen, Kangheng Tong, Kun He
Transfer-based attacks generate adversarial examples on the surrogate model, whi
ch can mislead other black-box models without access, making it promising to att
ack real-world applications. Recently, several works have been proposed to boost
 adversarial transferability, in which the surrogate model is usually overlooked
. In this work, we identify that non-linear layers (e.g., ReLU, max-pooling, etc
.) truncate the gradient during backward propagation, making the gradient w.r.t.
 input image imprecise to the loss function. We hypothesize and empirically vali
date that such truncation undermines the transferability of adversarial examples
. Based on these findings, we propose a novel method called Backward Propagation
 Attack (BPA) to increase the relevance between the gradient w.r.t. input image
and loss function so as to generate adversarial examples with higher transferabi
lity. Specifically, BPA adopts a non-monotonic function as the derivative of ReL
U and incorporates softmax with temperature to smooth the derivative of max-pool
ing, thereby mitigating the information loss during the backward propagation of
gradients. Empirical results on the ImageNet dataset demonstrate that not only d
oes our method substantially boost the adversarial transferability, but it is al
so general to existing transfer-based attacks. Code is available at https://gith
ub.com/Trustworthy-AI-Group/RPA.
****
Bullying10K: A Large-Scale Neuromorphic Dataset towards Privacy-Preserving Bully
ing Recognition
Yiting Dong, Yang Li, Dongcheng Zhao, Guobin Shen, Yi Zeng
The prevalence of violence in daily life poses significant threats to individual
s' physical and mental well-being. Using surveillance cameras in public spaces h
as proven effective in proactively deterring and preventing such incidents. Howe
ver, concerns regarding privacy invasion have emerged due to their widespread de
ployment.To address the problem, we leverage Dynamic Vision Sensors (DVS) camera
s to detect violent incidents and preserve privacy since it captures pixel brigh
tness variations instead of static imagery. We introduce the Bullying10K dataset
, encompassing various actions, complex movements, and occlusions from real-life
 scenarios. It provides three benchmarks for evaluating different tasks: action
recognition, temporal action localization, and pose estimation. With 10,000 even
t segments, totaling 12 billion events and 255 GB of data, Bullying10K contribut
es significantly by balancing violence detection and personal privacy perseverin
g. And it also poses a challenge to the neuromorphic dataset. It will serve as a
 valuable resource for training and developing privacy-protecting video systems.
 The Bullying10K opens new possibilities for innovative approaches in these doma
ins.
****
Compression with Bayesian Implicit Neural Representations
Zongyu Guo, Gergely Flamich, Jiajun He, Zhibo Chen, José Miguel Hernández-Lobato
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Towards Unbounded Machine Unlearning
Meghdad Kurmanji, Peter Triantafillou, Jamie Hayes, Eleni Triantafillou
Deep machine unlearning is the problem of 'removing' from a trained neural netwo
rk a subset of its training set. This problem is very timely and has many applic
ations, including the key tasks of removing biases (RB), resolving confusion (RC
) (caused by mislabelled data in trained models), as well as allowing users to e

xercise their 'right to be forgotten' to protect User Privacy (UP). This paper is the first, to our knowledge, to study unlearning for different applications (RB, RC, UP), with the view that each has its own desiderata, definitions for 'forgetting' and associated metrics for forget quality. For UP, we propose a novel adaptation of a strong Membership Inference Attack for unlearning. We also propose SCRUB, a novel unlearning algorithm, which is the only method that is consistently a top performer for forget quality across the different application-dependent metrics for RB, RC, and UP. At the same time, SCRUB is also consistently a top performer on metrics that measure model utility (i.e. accuracy on retained data and generalization), and is more efficient than previous work. The above are substantiated through a comprehensive empirical evaluation against previous state-of-the-art.
****
Collaborative Learning via Prediction Consensus
Dongyang Fan, Celestine Mendler-Dünner, Martin Jaggi
We consider a collaborative learning setting where the goal of each agent is to improve their own model by leveraging the expertise of collaborators, in addition to their own training data. To facilitate the exchange of expertise among agents, we propose a distillation-based method leveraging shared unlabeled auxiliary data, which is pseudo-labeled by the collective. Central to our method is a trust weighting scheme that serves to adaptively weigh the influence of each collaborator on the pseudo-labels until a consensus on how to label the auxiliary data is reached. We demonstrate empirically that our collaboration scheme is able to significantly boost individual models’ performance in the target domain from which the auxiliary data is sampled. At the same time, it can provably mitigate the negative impact of bad models on the collective. By design, our method adeptly accommodates heterogeneity in model architectures and substantially reduces communication overhead compared to typical collaborative learning methods.
****
Identification of Nonlinear Latent Hierarchical Models
Lingjing Kong, Biwei Huang, Feng Xie, Eric Xing, Yuejie Chi, Kun Zhang
Identifying latent variables and causal structures from observational data is essential to many real-world applications involving biological data, medical data, and unstructured data such as images and languages. However, this task can be highly challenging, especially when observed variables are generated by causally related latent variables and the relationships are nonlinear. In this work, we investigate the identification problem for nonlinear latent hierarchical causal models in which observed variables are generated by a set of causally related latent variables, and some latent variables may not have observed children. We show that the identifiability of causal structures and latent variables (up to invertible transformations) can be achieved under mild assumptions: on causal structures, we allow for multiple paths between any pair of variables in the graph, which relaxes latent tree assumptions in prior work; on structural functions, we permit general nonlinearity and multi-dimensional continuous variables, alleviating existing work's parametric assumptions. Specifically, we first develop an identification criterion in the form of novel identifiability guarantees for an elementary latent variable model. Leveraging this criterion, we show that both causal structures and latent variables of the hierarchical model can be identified asymptotically by explicitly constructing an estimation procedure. To the best of our knowledge, our work is the first to establish identifiability guarantees for both causal structures and latent variables in nonlinear latent hierarchical models.
****
Sample Efficient Reinforcement Learning in Mixed Systems through Augmented Samples and Its Applications to Queueing Networks
Honghao Wei, Xin Liu, Weina Wang, Lei Ying
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****
Temporal Graph Benchmark for Machine Learning on Temporal Graphs
Shenyang Huang, Farimah Poursafaei, Jacob Danovitch, Matthias Fey, Weihua Hu, Emanuele Rossi, Jure Leskovec, Michael Bronstein, Guillaume Rabusseau, Reihaneh Rabbany
We present the Temporal Graph Benchmark (TGB), a collection of challenging and diverse benchmark datasets for realistic, reproducible, and robust evaluation of machine learning models on temporal graphs. TGB datasets are of large scale, spanning years in duration, incorporate both node and edge-level prediction tasks and cover a diverse set of domains including social, trade, transaction, and transportation networks. For both tasks, we design evaluation protocols based on realistic use-cases. We extensively benchmark each dataset and find that the performance of common models can vary drastically across datasets. In addition, on dynamic node property prediction tasks, we show that simple methods often achieve superior performance compared to existing temporal graph models. We believe that these findings open up opportunities for future research on temporal graphs. Finally, TGB provides an automated machine learning pipeline for reproducible and accessible temporal graph research, including data loading, experiment setup and performance evaluation. TGB will be maintained and updated on a regular basis and welcomes community feedback. TGB datasets, data loaders, example codes, evaluation setup, and leaderboards are publicly available at https://tgb.complexdatalab.com/.
****
Navigating Data Heterogeneity in Federated Learning: A Semi-Supervised Federated Object Detection
Taehyeon Kim, Eric Lin, Junu Lee, Christian Lau, Vaikkunth Mugunthan
Federated Learning (FL) has emerged as a potent framework for training models across distributed data sources while maintaining data privacy. Nevertheless, it faces challenges with limited high-quality labels and non-IID client data, particularly in applications like autonomous driving. To address these hurdles, we navigate the uncharted waters of Semi-Supervised Federated Object Detection (SSFOD). We present a pioneering SSFOD framework, designed for scenarios where labeled data reside only at the server while clients possess unlabeled data. Notably, our method represents the inaugural implementation of SSFOD for clients with 0% labeled non-IID data, a stark contrast to previous studies that maintain some subset of labels at each client. We propose FedSTO, a two-stage strategy encompassing Selective Training followed by Orthogonally enhanced full-parameter training, to effectively address data shift (e.g. weather conditions) between server and clients. Our contributions include selectively refining the backbone of the detector to avert overfitting, orthogonality regularization to boost representation divergence, and local EMA-driven pseudo label assignment to yield high-quality pseudo labels. Extensive validation on prominent autonomous driving datasets (BDD100K, Cityscapes, and SODA10M) attests to the efficacy of our approach, demonstrating state-of-the-art results. Remarkably, FedSTO, using just 20-30% of labels, performs nearly as well as fully-supervised centralized training methods.
****
On the Generalization Properties of Diffusion Models
Puheng Li, Zhong Li, Huishuai Zhang, Jiang Bian
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Regularized Behavior Cloning for Blocking the Leakage of Past Action Information
Seokin Seo, HyeongJoo Hwang, Hongseok Yang, Kee-Eung Kim
For partially observable environments, imitation learning with observation histories (ILOH) assumes that control-relevant information is sufficiently captured in the observation histories for imitating the expert actions. In the offline setting wherethe agent is required to learn to imitate without interaction with the environment, behavior cloning (BC) has been shown to be a simple yet effective

method for imitation learning. However, when the information about the actions executed in the past timesteps leaks into the observation histories, ILOH via BC often ends up imitating its own past actions. In this paper, we address this catastrophic failure by proposing a principled regularization for BC, which we name Past Action Leakage Regularization (PALR). The main idea behind our approach is to leverage the classical notion of conditional independence to mitigate the leakage. We compare different instances of our framework with natural choices of conditional independence metric and its estimator. The result of our comparison advocates the use of a particular kernel-based estimator for the conditional independence metric. We conduct an extensive set of experiments on benchmark datasets in order to assess the effectiveness of our regularization method. The experimental results show that our method significantly outperforms prior related approaches, highlighting its potential to successfully imitate expert actions when the past action information leaks into the observation histories.
****
The Distortion of Binomial Voting Defies Expectation
Yannai A. Gonczarowski, Gregory Kehne, Ariel D. Procaccia, Ben Schiffer, Shirley Zhang
In computational social choice, the distortion of a voting rule quantifies the degree to which the rule overcomes limited preference information to select a socially desirable outcome. This concept has been investigated extensively, but only through a worst-case lens. Instead, we study the expected distortion of voting rules with respect to an underlying distribution over voter utilities. Our main contribution is the design and analysis of a novel and intuitive rule, binomial voting, which provides strong distribution-independent guarantees for both expected distortion and expected welfare.
****
UP-DP: Unsupervised Prompt Learning for Data Pre-Selection with Vision-Language Models
Xin Li, Sima Behpour, Thang Long Doan, Wenbin He, Liang Gou, Liu Ren
In this study, we investigate the task of data pre-selection, which aims to select instances for labeling from an unlabeled dataset through a single pass, thereby optimizing performance for undefined downstream tasks with a limited annotation budget. Previous approaches to data pre-selection relied solely on visual features extracted from foundation models, such as CLIP and BLIP-2, but largely ignored the powerfulness of text features. In this work, we argue that, with proper design, the joint feature space of both vision and text can yield a better representation for data pre-selection. To this end, we introduce UP-DP, a simple yet effective unsupervised prompt learning approach that adapts vision-language models, like BLIP-2, for data pre-selection. Specifically, with the BLIP-2 parameters frozen, we train text prompts to extract the joint features with improved representation, ensuring a diverse cluster structure that covers the entire dataset. We extensively compare our method with the state-of-the-art using seven benchmark datasets in different settings, achieving up to a performance gain of 20\%. Interestingly, the prompts learned from one dataset demonstrate significant generalizability and can be applied directly to enhance the feature extraction of BLIP-2 from other datasets. To the best of our knowledge, UP-DP is the first work to incorporate unsupervised prompt learning in a vision-language model for data pre-selection.
****
Optimistic Rates for Multi-Task Representation Learning
Austin Watkins, Enayat Ullah, Thanh Nguyen-Tang, Raman Arora
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Patch n’ Pack: NaViT, a Vision Transformer for any Aspect Ratio and Resolution
Mostafa Dehghani, Basil Mustafa, Josip Djolonga, Jonathan Heek, Matthias Minderer, Mathilde Caron, Andreas Steiner, Joan Puigcerver, Robert Geirhos, Ibrahim M.

Alabdulmohsin, Avital Oliver, Piotr Padlewski, Alexey Gritsenko, Mario Lucic, Ne
il Houlsby
The ubiquitous and demonstrably suboptimal choice of resizing images to a fixed
resolution before processing them with computer vision models has not yet been s
uccessfully challenged. However, models such as the Vision Transformer (ViT) off
er flexible sequence-based modeling, and hence varying input sequence lengths.
We take advantage of this with NaViT (Native Resolution ViT) which uses sequence
 packing during training to process inputs of arbitrary resolutions and aspect r
atios. Alongside flexible model usage, we demonstrate improved training efficien
cy for large-scale supervised and contrastive image-text pretraining.NaViT can b
e efficiently transferred to standard tasks such as image and video classificati
on, object detection, and semantic segmentation and leads to improved results on
 robustness and fairness benchmarks. At inference time, the input resolution fle
xibility can be used to smoothly navigate the test-time cost-performance trade-o
ff. We believe that NaViTmarks a departure from the standard, CNN-designed, inpu
t and modelling pipeline used by most computer vision models, and represents a p
romising direction for ViTs.
****
The Benefits of Being Distributional: Small-Loss Bounds for Reinforcement Learni
ng
Kaiwen Wang, Kevin Zhou, Runzhe Wu, Nathan Kallus, Wen Sun
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Honesty Is the Best Policy: Defining and Mitigating AI Deception
Francis Ward, Francesca Toni, Francesco Belardinelli, Tom Everitt
Deceptive agents are a challenge for the safety, trustworthiness, and cooperatio
n of AI systems. We focus on the problem that agents might deceive in order to a
chieve their goals (for instance, in our experiments with language models, the g
oal of being evaluated as truthful).There are a number of existing definitions o
f deception in the literature on game theory and symbolic AI, but there is no ov
erarching theory of deception for learning agents in games. We introduce a forma
ldefinition of deception in structural causal games, grounded in the philosophyl
iterature, and applicable to real-world machine learning systems.Several example
s and results illustrate that our formal definition aligns with the philosophica
l and commonsense meaning of deception.Our main technical result is to provide g
raphical criteria for deception. We show, experimentally, that these results can
 be used to mitigate deception in reinforcement learning agents and language mod
els.
****
Improving *day-ahead* Solar Irradiance Time Series Forecasting by Leveraging Spa
tio-Temporal Context
Oussama Boussif, Ghait Boukachab, Dan Assouline, Stefano Massaroli, Tianle Yuan,
 Loubna Benabbou, Yoshua Bengio
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Uncovering and Quantifying Social Biases in Code Generation
Yan Liu, Xiaokang Chen, Yan Gao, Zhe Su, Fengji Zhang, Daoguang Zan, Jian-Guang
Lou, Pin-Yu Chen, Tsung-Yi Ho
With the popularity of automatic code generation tools, such as Copilot, the stu
dy of the potential hazards of these tools is gaining importance. In this work,
we explore the social bias problem in pre-trained code generation models. We pro
pose a new paradigm to construct code prompts and successfully uncover social bi
ases in code generation models. To quantify the severity of social biases in gen
erated code, we develop a dataset along with three metrics to evaluate the overa

ll social bias and fine-grained unfairness across different demographics. Experimental results on three pre-trained code generation models (Codex, InCoder, and CodeGen) with varying sizes, reveal severe social biases. Moreover, we conduct analysis to provide useful insights for further choice of code generation models with low social bias.
****
A Bounded Ability Estimation for Computerized Adaptive Testing
Yan Zhuang, Qi Liu, Guanhao Zhao, Zhenya Huang, Weizhe Huang, Zachary Pardos, Enhong Chen, Jinze Wu, Xin Li
Computerized adaptive testing (CAT), as a tool that can efficiently measure student's ability, has been widely used in various standardized tests (e.g., GMAT and GRE). The adaptivity of CAT refers to the selection of the most informative questions for each student, reducing test length. Existing CAT methods do not explicitly target ability estimation accuracy since there is no student's true ability as ground truth; therefore, these methods cannot be guaranteed to make the estimate converge to the true with such limited responses. In this paper, we analyze the statistical properties of estimation and find a theoretical approximation of the true ability: the ability estimated by full responses to question bank. Based on this, a Bounded Ability Estimation framework for CAT (BECAT) is proposed in a data-summary manner, which selects a question subset that closely matches the gradient of the full responses. Thus, we develop an expected gradient difference approximation to design a simple greedy selection algorithm, and show the rigorous theoretical and error upper-bound guarantees of its ability estimate. Experiments on both real-world and synthetic datasets, show that it can reach the same estimation accuracy using 15\% less questions on average, significantly reducing test length.
****
ForecastPFN: Synthetically-Trained Zero-Shot Forecasting
Samuel Dooley, Gurnoor Singh Khurana, Chirag Mohapatra, Siddartha V Naidu, Colin White
The vast majority of time-series forecasting approaches require a substantial training dataset. However, many real-life forecasting applications have very little initial observations, sometimes just 40 or fewer. Thus, the applicability of most forecasting methods is restricted in data-sparse commercial applications. While there is recent work in the setting of very limited initial data (so-called `zero-shot' forecasting), its performance is inconsistent depending on the data used for pretraining. In this work, we take a different approach and devise ForecastPFN, the first zero-shot forecasting model trained purely on a novel synthetic data distribution. ForecastPFN is a prior-data fitted network, trained to approximate Bayesian inference, which can make predictions on a new time series dataset in a single forward pass. Through extensive experiments, we show that zero-shot predictions made by ForecastPFN are more accurate and faster compared to state-of-the-art forecasting methods, even when the other methods are allowed to train on hundreds of additional in-distribution data points.
****
Exact Bayesian Inference on Discrete Models via Probability Generating Functions: A Probabilistic Programming Approach
Fabian Zaiser, Andrzej Murawski, Chih-Hao Luke Ong
We present an exact Bayesian inference method for discrete statistical models, which can find exact solutions to a large class of discrete inference problems, even with infinite support and continuous priors.To express such models, we introduce a probabilistic programming language that supports discrete and continuous sampling, discrete observations, affine functions, (stochastic) branching, and conditioning on discrete events.Our key tool is probability generating functions: they provide a compact closed-form representation of distributions that are definable by programs, thus enabling the exact computation of posterior probabilities, expectation, variance, and higher moments.Our inference method is provably correct and fully automated in a tool called Genfer, which uses automatic differentiation (specifically, Taylor polynomials), but does not require computer algebra.Our experiments show that Genfer is often faster than the existing exact infer

ence tools PSI, Dice, and Prodigy.On a range of real-world inference problems that none of these exact tools can solve, Genfer's performance is competitive with approximate Monte Carlo methods, while avoiding approximation errors.
****

$SE(3)$ Equivariant Convolution and Transformer in Ray Space
Yinshuang Xu, Jiahui Lei, Kostas Daniilidis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Principle-Driven Self-Alignment of Language Models from Scratch with Minimal Human Supervision
Zhiqing Sun, Yikang Shen, Qinhong Zhou, Hongxin Zhang, Zhenfang Chen, David Cox, Yiming Yang, Chuang Gan
Recent AI-assistant agents, such as ChatGPT, predominantly rely on supervised fine-tuning (SFT) with human annotations and reinforcement learning from human feedback (RLHF) to align the output of large language models (LLMs) with human intentions, ensuring they are helpful, ethical, and reliable. However, this dependence can significantly constrain the true potential of AI-assistant agents due to the high cost of obtaining human supervision and the related issues on quality, reliability, diversity, self-consistency, and undesirable biases. To address these challenges, we propose a novel approach called SELF-ALIGN, which combines principle-driven reasoning and the generative power of LLMs for the self-alignment of AI agents with minimal human supervision. Our approach encompasses four stages: first, we use an LLM to generate synthetic prompts, and a topic-guided method to augment the prompt diversity; second, we use a small set of human-written principles for AI models to follow, and guide the LLM through in-context learning from demonstrations (of principles application) to produce helpful, ethical, and reliable responses to user's queries; third, we fine-tune the original LLM with the high-quality self-aligned responses so that the resulting model can generate desirable responses for each query directly without the principle set and the demonstrations anymore; and finally, we offer a refinement step to address the issues of overly-brief or indirect responses. Applying SELF-ALIGN to the LLaMA-65b base language model, we develop an AI assistant named Dromedary. With fewer than 300 lines of human annotations (including < 200 seed prompts, 16 generic principles, and 5 exemplars for in-context learning). Dromedary significantly surpasses the performance of several state-of-the-art AI systems, including Text-Davinci-003 and Alpaca, on benchmark datasets with various settings.
****

Prototypical Variational Autoencoder for 3D Few-shot Object Detection
Weiliang Tang, Biqi YANG, Xianzhi Li, Yun-Hui Liu, Pheng-Ann Heng, Chi-Wing Fu
Few-Shot 3D Point Cloud Object Detection (FS3D) is a challenging task, aiming to detect 3D objects of novel classes using only limited annotated samples for training. Considering that the detection performance highly relies on the quality of the latent features, we design a VAE-based prototype learning scheme, named prototypical VAE (P-VAE), to learn a probabilistic latent space for enhancing the diversity and distinctiveness of the sampled features. The network encodes a multi-center GMM-like posterior, in which each distribution centers at a prototype. For regularization, P-VAE incorporates a reconstruction task to preserve geometric information. To adopt P-VAE for the detection framework, we formulate Geometric-informative Prototypical VAE (GP-VAE) to handle varying geometric components and Class-specific Prototypical VAE (CP-VAE) to handle varying object categories. In the first stage, we harness GP-VAE to aid feature extraction from the input scene. In the second stage, we cluster the geometric-informative features into per-instance features and use CP-VAE to refine each instance feature with category-level guidance. Experimental results show the top performance of our approach over the state of the arts on two FS3D benchmarks. Quantitative ablations and qualitative prototype analysis further demonstrate that our probabilistic modeling can significantly boost prototype learning for FS3D.

****
Double Gumbel Q-Learning
David Yu-Tung Hui, Aaron C. Courville, Pierre-Luc Bacon
We show that Deep Neural Networks introduce two heteroscedastic Gumbel noise sou rces into Q-Learning. To account for these noise sources, we propose Double Gum bel Q-Learning, a Deep Q-Learning algorithm applicable for both discrete and con tinuous control. In discrete control, we derive a closed-form expression for th e loss function of our algorithm. In continuous control, this loss function is intractable and we therefore derive an approximation with a hyperparameter whose value regulates pessimism in Q-Learning. We present a default value for our pe ssimism hyperparameter that enables DoubleGum to outperform DDPG, TD3, SAC, XQL, quantile regression, and Mixture-of-Gaussian Critics in aggregate over 33 tasks from DeepMind Control, MuJoCo, MetaWorld, and Box2D and show that tuning this h yperparameter may further improve sample efficiency.
****
Mutual-Information Regularized Multi-Agent Policy Iteration
Wang, Deheng Ye, Zongqing Lu
Despite the success of cooperative multi-agent reinforcement learning algorithms , most of them focus on a single team composition, which prevents them from bein g used in more realistic scenarios where dynamic team composition is possible. W hile some studies attempt to solve this problem via multi-task learning in a fix ed set of team compositions, there is still a risk of overfitting to the trainin g set, which may lead to catastrophic performance when facing dramatically varyi ng team compositions during execution. To address this problem, we propose to us e mutual information (MI) as an augmented reward to prevent individual policies from relying too much on team-related information and encourage agents to learn policies that are robust in different team compositions. Optimizing this MI-augm ented objective in an off-policy manner can be intractable due to the existence of dynamic marginal distribution. To alleviate this problem, we first propose a multi-agent policy iteration algorithm with a fixed marginal distribution and pr ove its convergence and optimality. Then, we propose to employ the Blahut–Arimot o algorithm and an imaginary team composition distribution for optimization with approximate marginal distribution as the practical implementation. Empirically, our method demonstrates strong zero-shot generalization to dynamic team composi tions in complex cooperative tasks.
****
An Efficient End-to-End Training Approach for Zero-Shot Human-AI Coordination
Xue Yan, Jiaxian Guo, Xingzhou Lou, Jun Wang, Haifeng Zhang, Yali Du
The goal of zero-shot human-AI coordination is to develop an agent that can coll aborate with humans without relying on human data. Prevailing two-stage populati on-based methods require a diverse population of mutually distinct policies to s imulate diverse human behaviors. The necessity of such populations severely limi ts their computational efficiency. To address this issue, we propose E3T, an Eff icient End-to-End Training approach for zero-shot human-AI coordination. E3T emp loys a mixture of ego policy and random policy to construct the partner policy, making it both coordination-skilled and diverse. In this way, the ego agent is e nd-to-end trained with this mixture policy without the need of a pre-trained pop ulation, thus significantly improving the training efficiency. In addition, a p artner modeling module is proposed to predict the partner's action from historic al information. With the predicted partner's action, the ego policy is able to a dapt its policy and take actions accordingly when collaborating with humans of d ifferent behavior patterns. Empirical results on the Overcooked environment show that our method significantly improves the training efficiency while preserving comparable or superior performance than the population-based baselines. Demo vi deos are available at https://sites.google.com/view/e3t-overcooked.
****
Computing Optimal Equilibria and Mechanisms via Learning in Zero-Sum Extensive-F orm Games
Brian Zhang, Gabriele Farina, Ioannis Anagnostides, Federico Cacciamani, Stephen McAleer, Andreas Haupt, Andrea Celli, Nicola Gatti, Vincent Conitzer, Tuomas Sa

ndholm
We introduce a new approach for computing optimal equilibria via learning in gam
es. It applies to extensive-form settings with any number of players, including
mechanism design, information design, and solution concepts such as correlated,
communication, and certification equilibria. We observe that optimal equilibria
are minimax equilibrium strategies of a player in an extensive-form zero-sum gam
e. This reformulation allows to apply techniques for learning in zero-sum games,
yielding the first learning dynamics that converge to optimal equilibria, not o
nly in empirical averages, but also in iterates. We demonstrate the practical sc
alability and flexibility of our approach by attaining state-of-the-art performa
nce in benchmark tabular games, and by computing an optimal mechanism for a sequ
ential auction design problem using deep reinforcement learning.
****
Parts of Speech–Grounded Subspaces in Vision-Language Models
James Oldfield, Christos Tzelepis, Yannis Panagakis, Mihalis Nicolaou, Ioannis P
atras
Latent image representations arising from vision-language models have proved imm
ensely useful for a variety of downstream tasks. However, their utility is limit
ed by their entanglement with respect to different visual attributes. For instan
ce, recent work has shown that CLIP image representations are often biased towar
d specific visual properties (such as objects or actions) in an unpredictable ma
nner. In this paper, we propose to separate representations of the different vis
ual modalities in CLIP’s joint vision-language space by leveraging the associati
on between parts of speech and specific visual modes of variation (e.g. nouns re
late to objects, adjectives describe appearance). This is achieved by formulatin
g an appropriate component analysis model that learns subspaces capturing variab
ility corresponding to a specific part of speech, while jointly minimising varia
bility to the rest. Such a subspace yields disentangled representations of the d
ifferent visual properties of an image or text in closed form while respecting t
he underlying geometry of the manifold on which the representations lie. What’s
more, we show the proposed model additionally facilitates learning subspaces cor
responding to specific visual appearances (e.g. artists’ painting styles), which
enables the selective removal of entire visual themes from CLIP-based text-to-i
mage synthesis. We validate the model both qualitatively, by visualising the sub
space projections with a text-to-image model and by preventing the imitation of
artists’ styles, and quantitatively, through class invariance metrics and improv
ements to baseline zero-shot classification.
****
Searching for Optimal Per-Coordinate Step-sizes with Multidimensional Backtracki
ng
Frederik Kunstner, Victor Sanches Portella, Mark Schmidt, Nicholas Harvey
The backtracking line-search is an effective technique to automatically tune the
step-size in smooth optimization. It guarantees similar performance to using th
e theoretically optimal step-size. Many approaches have been developed to instea
d tune per-coordinate step-sizes, also known as diagonal preconditioners, but no
ne of the existing methods are provably competitive with the optimal per-coordin
ate step-sizes. We propose multidimensional backtracking, an extension of the ba
cktracking line-search to find good diagonal preconditioners for smooth convex p
roblems. Our key insight is that the gradient with respect to the step-sizes, al
so known as hyper-gradients, yields separating hyperplanes that let us search fo
r good preconditioners using cutting-plane methods. As black-box cutting-plane a
pproaches like the ellipsoid method are computationally prohibitive, we develop
an efficient algorithm tailored to our setting. Multidimensional backtracking is
provably competitive with the best diagonal preconditioner and requires no manu
al tuning.
****
Estimating the Rate-Distortion Function by Wasserstein Gradient Descent
Yibo Yang, Stephan Eckstein, Marcel Nutz, Stephan Mandt
Requests for name changes in the electronic proceedings will be accepted with no
questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Epistemic Neural Networks
Ian Osband, Zheng Wen, Seyed Mohammad Asghari, Vikranth Dwaracherla, MORTEZA IBRAHIMI, Xiuyuan Lu, Benjamin Van Roy
Intelligence relies on an agent's knowledge of what it does not know.This capability can be assessed based on the quality of joint predictions of labels across multiple inputs.In principle, ensemble-based approaches can produce effective joint predictions, but the computational costs of large ensembles become prohibitive.We introduce the epinet: an architecture that can supplement any conventional neural network, including large pretrained models, and can be trained with modest incremental computation to estimate uncertainty.With an epinet, conventional neural networks outperform very large ensembles, consisting of hundreds or more particles, with orders of magnitude less computation.The epinet does not fit the traditional framework of Bayesian neural networks.To accommodate development of approaches beyond BNNs, such as the epinet, we introduce the epistemic neural network (ENN) as a general interface for models that produce joint predictions.
****
HotBEV: Hardware-oriented Transformer-based Multi-View 3D Detector for BEV Perception
Peiyan Dong, Zhenglun Kong, Xin Meng, Pinrui Yu, Yifan Gong, Geng Yuan, Hao Tang, Yanzhi Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Mip-Grid: Anti-aliased Grid Representations for Neural Radiance Fields
Seungtae Nam, Daniel Rho, Jong Hwan Ko, Eunbyung Park
Despite the remarkable achievements of neural radiance fields (NeRF) in representing 3D scenes and generating novel view images, the aliasing issue, rendering 'jaggies' or 'blurry' images at varying camera distances, remains unresolved in most existing approaches. The recently proposed mip-NeRF has effectively addressed this challenge by introducing integrated positional encodings (IPE). However, it relies on MLP architecture to represent the radiance fields, missing out on the fast training speed offered by the latest grid-based methods. In this work, we present mip-Grid, a novel approach that integrates anti-aliasing techniques into grid-based representations for radiance fields, mitigating the aliasing artifacts while enjoying fast training time. Notably, the proposed method uses a single-scale shared grid representation and a single-sampling approach, which only introduces minimal additions to the model parameters and computational costs. To handle scale ambiguity, mip-Grid generates multiple grids by applying simple convolution operations over the shared grid and uses the scale-aware coordinate to retrieve the appropriate features from the generated multiple grids. To test the effectiveness, we incorporated the proposed approach into the two recent representative grid-based methods, TensoRF and K-Planes. The experimental results demonstrated that mip-Grid greatly improved the rendering performance of both methods and showed comparable performance to mip-NeRF on multi-scale datasets while achieving significantly faster training time.
****
Theoretically Guaranteed Bidirectional Data Rectification for Robust Sequential Recommendation
Yatong Sun, Bin Wang, Zhu Sun, Xiaochun Yang, Yan Wang
Sequential recommender systems (SRSs) are typically trained to predict the next item as the target given its preceding (and succeeding) items as the input. Such a paradigm assumes that every input-target pair is reliable for training. However, users can be induced to click on items that are inconsistent with their true preferences, resulting in unreliable instances, i.e., mismatched input-target pairs. Current studies on mitigating this issue suffer from two limitations: (i)

they discriminate instance reliability according to models trained with unreliab
le data, yet without theoretical guarantees that such a seemingly contradictory
solution can be effective; and (ii) most methods can only tackle either unreliab
le input or targets but fail to handle both simultaneously. To fill the gap, we
theoretically unveil the relationship between SRS predictions and instance relia
bility, whereby two error-bounded strategies are proposed to rectify unreliable
targets and input, respectively. On this basis, we devise a model-agnostic Bidir
ectional Data Rectification (BirDRec) framework, which can be flexibly implement
ed with most existing SRSs for robust training against unreliable data. Addition
ally, a rectification sampling strategy is devised and a self-ensemble mechanism
 is adopted to reduce the (time and space) complexity of BirDRec. Extensive expe
riments on four real-world datasets verify the generality, effectiveness, and ef
ficiency of our proposed BirDRec.
****
Consistent Aggregation of Objectives with Diverse Time Preferences Requires Non-
Markovian Rewards
Silviu Pitis
As the capabilities of artificial agents improve, they are being increasingly de
ployed to service multiple diverse objectives and stakeholders. However, the com
position of these objectives is often performed ad hoc, with no clear justificat
ion. This paper takes a normative approach to multi-objective agency: from a set
 of intuitively appealing axioms, it is shown that Markovian aggregation of Mark
ovian reward functions is not possible when the time preference (discount factor
) for each objective may vary. It follows that optimal multi-objective agents mu
st admit rewards that are non-Markovian with respect to the individual objective
s. To this end, a practical non-Markovian aggregation scheme is proposed, which
overcomes the impossibility with only one additional parameter for each objectiv
e. This work offers new insights into sequential, multi-objective agency and int
ertemporal choice, and has practical implications for the design of AI systems d
eployed to serve multiple generations of principals with varying time preference
.
****
Diffusion-Based Adversarial Sample Generation for Improved Stealthiness and Cont
rollability
Haotian Xue, Alexandre Araujo, Bin Hu, Yongxin Chen
Neural networks are known to be susceptible to adversarial samples: small variat
ions of natural examples crafted to deliberatelymislead the models. While they c
an be easily generated using gradient-based techniques in digital and physical s
cenarios, they often differ greatly from the actual data distribution of natural
 images, resulting in a trade-off between strength and stealthiness. In this pap
er, we propose a novel framework dubbed Diffusion-Based Projected Gradient Desce
nt (Diff-PGD) for generating realistic adversarial samples. By exploiting a grad
ient guided by a diffusion model, Diff-PGD ensures that adversarial samples rema
in close to the original data distribution while maintaining their effectiveness
. Moreover, our framework can be easily customized for specific tasks such as di
gital attacks, physical-world attacks, and style-based attacks. Compared with ex
isting methods for generating natural-style adversarial samples, our framework e
nables the separation of optimizing adversarial loss from other surrogate losses
 (e.g. content/smoothness/style loss), making it more stable and controllable. F
inally, we demonstrate that the samples generated using Diff-PGD have better tra
nsferability and anti-purification power than traditional gradient-based methods
.
****
InstanT: Semi-supervised Learning with Instance-dependent Thresholds
Muyang Li, Runze Wu, Haoyu Liu, Jun Yu, Xun Yang, Bo Han, Tongliang Liu
Semi-supervised learning (SSL) has been a fundamental challenge in machine learn
ing for decades. The primary family of SSL algorithms, known as pseudo-labeling,
 involves assigning pseudo-labels to confident unlabeled instances and incorpora
ting them into the training set. Therefore, the selection criteria of confident
instances are crucial to the success of SSL. Recently, there has been growing in

terest in the development of SSL methods that use dynamic or adaptive thresholds. Yet, these methods typically apply the same threshold to all samples, or use class-dependent thresholds for instances belonging to a certain class, while neglecting instance-level information. In this paper, we propose the study of instance-dependent thresholds, which has the highest degree of freedom compared with existing methods. Specifically, we devise a novel instance-dependent threshold function for all unlabeled instances by utilizing their instance-level ambiguity and the instance-dependent error rates of pseudo-labels, so instances that are more likely to have incorrect pseudo-labels will have higher thresholds. Furthermore, we demonstrate that our instance-dependent threshold function provides a bounded probabilistic guarantee for the correctness of the pseudo-labels it assigns.
****

Neural Lyapunov Control for Discrete-Time Systems

Junlin Wu, Andrew Clark, Yiannis Kantaros, Yevgeniy Vorobeychik

While ensuring stability for linear systems is well understood, it remains a major challenge for nonlinear systems. A general approach in such cases is to compute a combination of a Lyapunov function and an associated control policy. However, finding Lyapunov functions for general nonlinear systems is a challenging task. To address this challenge, several methods have been proposed that represent Lyapunov functions using neural networks. However, such approaches either focus on continuous-time systems, or highly restricted classes of nonlinear dynamics. We propose the first approach for learning neural Lyapunov control in a broad class of discrete-time systems. Three key ingredients enable us to effectively learn provably stable control policies. The first is a novel mixed-integer linear programming approach for verifying the discrete-time Lyapunov stability conditions, leveraging the particular structure of these conditions. The second is a novel approach for computing verified sublevel sets. The third is a heuristic gradient-based method for quickly finding counterexamples to significantly speed up Lyapunov function learning. Our experiments on four standard benchmarks demonstrate that our approach significantly outperforms state-of-the-art baselines. For example, on the path tracking benchmark, we outperform recent neural Lyapunov control baselines by an order of magnitude in both running time and the size of the region of attraction, and on two of the four benchmarks (cartpole and PVTOL), ours is the first automated approach to return a provably stable controller. Our code is available at: https://github.com/jlwu002/nlc_discrete.
****

Information Maximization Perspective of Orthogonal Matching Pursuit with Applications to Explainable AI

Aditya Chattopadhyay, Ryan Pilgrim, Rene Vidal

Information Pursuit (IP) is a classical active testing algorithm for predicting an output by sequentially and greedily querying the input in order of information gain. However, IP is computationally intensive since it involves estimating mutual information in high-dimensional spaces. This paper explores Orthogonal Matching Pursuit (OMP) as an alternative to IP for greedily selecting the queries. OMP is a classical signal processing algorithm for sequentially encoding a signal in terms of dictionary atoms chosen in order of correlation gain. In each iteration, OMP selects the atom that is most correlated with the signal residual (the signal minus its reconstruction thus far). Our first contribution is to establish a fundamental connection between IP and OMP, where we prove that IP with random projections of dictionary atoms as queries ``almost'' reduces to OMP, with the difference being that IP selects atoms in order of normalized correlation gain. We call this version IP-OMP and present simulations indicating that this difference does not have any appreciable effect on the sparse code recovery rate of IP-OMP compared to that of OMP for random Gaussian dictionaries. Inspired by this connection, our second contribution is to explore the utility of IP-OMP for generating explainable predictions, an area in which IP has recently gained traction. More specifically, we propose a simple explainable AI algorithm which encodes an image as a sparse combination of semantically meaningful dictionary atoms that are defined as text embeddings of interpretable concepts. The final predictio

n is made using the weights of this sparse combination, which serve as an explan ation. Empirically, our proposed algorithm is not only competitive with existing explainability methods but also computationally less expensive.
****
Evolving Connectivity for Recurrent Spiking Neural Networks
Guan Wang, Yuhao Sun, Sijie Cheng, Sen Song
Recurrent spiking neural networks (RSNNs) hold great potential for advancing art ificial general intelligence, as they draw inspiration from the biological nervo us system and show promise in modeling complex dynamics.However, the widely-used surrogate gradient-based training methods for RSNNs are inherently inaccurate a nd unfriendly to neuromorphic hardware.To address these limitations, we propose the evolving connectivity (EC) framework, an inference-only method for training RSNNs.The EC framework reformulates weight-tuning as a search into parameterized connection probability distributions, and employs Natural Evolution Strategies (NES) for optimizing these distributions.Our EC framework circumvents the need f or gradients and features hardware-friendly characteristics, including sparse bo olean connections and high scalability.We evaluate EC on a series of standard ro botic locomotion tasks, where it achieves comparable performance with deep neura l networks and outperforms gradient-trained RSNNs, even solving the complex 17-D oF humanoid task.Additionally, the EC framework demonstrates a two to three fold speedup in efficiency compared to directly evolving parameters.By providing a p erformant and hardware-friendly alternative, the EC framework lays the groundwor k for further energy-efficient applications of RSNNs and advances the developmen t of neuromorphic devices.Our code is publicly available at https://github.com/i moneoi/EvolvingConnectivity.
****
Bayesian Optimization with Cost-varying Variable Subsets
Sebastian Tay, Chuan Sheng Foo, Daisuke Urano, Richalynn Leong, Bryan Kian Hsian g Low
We introduce the problem of Bayesian optimization with cost-varying variable sub sets (BOCVS) where in each iteration, the learner chooses a subset of query vari ables and specifies their values while the rest are randomly sampled. Each chose n subset has an associated cost. This presents the learner with the novel challe nge of balancing between choosing more informative subsets for more directed lea rning versus leaving some variables to be randomly sampled to reduce incurred co sts. This paper presents a novel Gaussian process upper confidence bound-based a lgorithm for solving the BOCVS problem that is provably no-regret. We analyze ho w the availability of cheaper control sets helps in exploration and reduces over all regret. We empirically show that our proposed algorithm can find significant ly better solutions than comparable baselines with the same budget.
****
Transformed Low-Rank Parameterization Can Help Robust Generalization for Tensor Neural Networks
Andong Wang, Chao Li, Mingyuan Bai, Zhong Jin, Guoxu Zhou, Qibin Zhao
Multi-channel learning has gained significant attention in recent applications, where neural networks with t-product layers (t-NNs) have shown promising perform ance through novel feature mapping in the transformed domain. However, despite t he practical success of t-NNs, the theoretical analysis of their generalization remains unexplored. We address this gap by deriving upper bounds on the generali zation error of t-NNs in both standard and adversarial settings. Notably, it rev eals that t-NNs compressed with exact transformed low-rank parameterization can achieve tighter adversarial generalization bounds compared to non-compressed mod els. While exact transformed low-rank weights are rare in practice, the analysis demonstrates that through adversarial training with gradient flow, highly over-parameterized t-NNs with the ReLU activation can be implicitly regularized towar ds a transformed low-rank parameterization under certain conditions. Moreover, t his paper establishes sharp adversarial generalization bounds for t-NNs with app roximately transformed low-rank weights. Our analysis highlights the potential o f transformed low-rank parameterization in enhancing the robust generalization o f t-NNs, offering valuable insights for further research and development.

****
Testing the General Deductive Reasoning Capacity of Large Language Models Using OOD Examples
Abulhair Saparov, Richard Yuanzhe Pang, Vishakh Padmakumar, Nitish Joshi, Mehran Kazemi, Najoung Kim, He He
Given the intractably large size of the space of proofs, any model that is capable of general deductive reasoning must generalize to proofs of greater complexity. Recent studies have shown that large language models (LLMs) possess some abstract deductive reasoning ability given chain-of-thought prompts. However, they have primarily been tested on proofs using modus ponens or of a specific size, and from the same distribution as the in-context examples. To measure the general deductive reasoning ability of LLMs, we test on a broad set of deduction rules and measure their ability to generalize to more complex proofs from simpler demonstrations from multiple angles: depth-, width-, and compositional generalization. To facilitate systematic exploration, we construct a new synthetic and programmable reasoning dataset that enables control over deduction rules and proof complexity. Our experiments on four LLMs of various sizes and training objectives show that they are able to generalize to compositional proofs. However, they have difficulty generalizing to longer proofs, and they require explicit demonstrations to produce hypothetical subproofs, specifically in proof by cases and proof by contradiction.
****
MosaicBERT: A Bidirectional Encoder Optimized for Fast Pretraining
Jacob Portes, Alexander Trott, Sam Havens, DANIEL KING, Abhinav Venigalla, Moin Nadeem, Nikhil Sardana, Daya Khudia, Jonathan Frankle
Although BERT-style encoder models are heavily used in NLP research, many researchers do not pretrain their own BERTs from scratch due to the high cost of training. In the past half-decade since BERT first rose to prominence, many advances have been made with other transformer architectures and training configurations that have yet to be systematically incorporated into BERT. Here, we introduce MosaicBERT, a BERT-style encoder architecture and training recipe that is empirically optimized for fast pretraining. This efficient architecture incorporates FlashAttention, Attention with Linear Biases (ALiBi), Gated Linear Units (GLU), a module to dynamically remove padded tokens, and low precision LayerNorm into the classic transformer encoder block. The training recipe includes a 30% masking ratio for the Masked Language Modeling (MLM) objective, bfloat16 precision, and vocabulary size optimized for GPU throughput, in addition to best-practices from RoBERTa and other encoder models. When pretrained from scratch on the C4 dataset, this base model achieves a downstream average GLUE (dev) score of 79.6 in 1.13 hours on 8 A100 80 GB GPUs at a cost of roughly $20. We plot extensive accuracy vs. pretraining speed Pareto curves and show that MosaicBERT base and large are consistently Pareto optimal when compared to a competitive BERT base and large. This empirical speed up in pretraining enables researchers and engineers to pretrain custom BERT-style models at low cost instead of finetune on existing generic models. We open source our model weights and code.
****
GraphMP: Graph Neural Network-based Motion Planning with Efficient Graph Search
Xiao Zang, Miao Yin, Jinqi Xiao, Saman Zonouz, Bo Yuan
Motion planning, which aims to find a high-quality collision-free path in the configuration space, is a fundamental task in robotic systems. Recently, learning-based motion planners, especially the graph neural network-powered, have shown promising planning performance. However, though the state-of-the-art GNN planner can efficiently extract and learn graph information, its inherent mechanism is not well suited for graph search process, hindering its further performance improvement. To address this challenge and fully unleash the potential of GNN in motion planning, this paper proposes GraphMP, a neural motion planner for both low and high-dimensional planning tasks. With the customized model architecture and training mechanism design, GraphMP can simultaneously perform efficient graph pattern extraction and graph search processing, leading to strong planning performance. Experiments on a variety of environments, ranging from 2D Maze to 14D dual

KUKA robotic arm, show that our proposed GraphMP achieves significant improvement on path quality and planning speed over the state-of-the-art learning-based and classical planners; while preserving the competitive success rate.
****

Accountability in Offline Reinforcement Learning: Explaining Decisions with a Corpus of Examples

Hao Sun, Alihan Hüyük, Daniel Jarrett, Mihaela van der Schaar

Learning controllers with offline data in decision-making systems is an essential area of research due to its potential to reduce the risk of applications in real-world systems. However, in responsibility-sensitive settings such as healthcare, decision accountability is of paramount importance, yet has not been adequately addressed by the literature.This paper introduces the Accountable Offline Controller (AOC) that employs the offline dataset as the Decision Corpus and performs accountable control based on a tailored selection of examples, referred to as the Corpus Subset. AOC operates effectively in low-data scenarios, can be extended to the strictly offline imitation setting, and displays qualities of both conservation and adaptability.We assess AOC's performance in both simulated and real-world healthcare scenarios, emphasizing its capability to manage offline control tasks with high levels of performance while maintaining accountability.
****

Synthcity: a benchmark framework for diverse use cases of tabular synthetic data

Zhaozhi Qian, Rob Davis, Mihaela van der Schaar

Accessible high-quality data is the bread and butter of machine learning research, and the demand for data has exploded as larger and more advanced ML models are built across different domains. Yet, real data often contain sensitive information, are subject to various biases, and are costly to acquire, which compromise their quality and accessibility. Synthetic data have thus emerged as a complement to, sometimes even a replacement for, real data for ML training. However, the landscape of synthetic data research has been fragmented due to the diverse range of data modalities, such as tabular, time series, and images, and the wide array of use cases, including privacy preservation, fairness considerations, and data augmentation. This fragmentation poses practical challenges when comparing and selecting synthetic data generators in for different problem settings. To this end, we develop Synthcity, an open-source Python library that allows researchers and practitioners to perform one-click benchmarking of synthetic data generators across data modalities and use cases. Beyond benchmarking, Synthcity serves as a centralized toolkit for accessing cutting-edge data generators. In addition, Synthcity’s flexible plug-in style API makes it easy to incorporate additional data generators into the framework. Using examples of tabular data generation and data augmentation, we illustrate the general applicability of Synthcity, and the insight one can obtain.
****

SOAR: Improved Indexing for Approximate Nearest Neighbor Search

Philip Sun, David Simcha, Dave Dopson, Ruiqi Guo, Sanjiv Kumar

This paper introduces SOAR: Spilling with Orthogonality-Amplified Residuals, a novel data indexing technique for approximate nearest neighbor (ANN) search. SOAR extends upon previous approaches to ANN search, such as spill trees, that utilize multiple redundant representations while partitioning the data to reduce the probability of missing a nearest neighbor during search. Rather than training and computing these redundant representations independently, however, SOAR uses an orthogonality-amplified residual loss, which optimizes each representation to compensate for cases where other representations perform poorly. This drastically improves the overall index quality, resulting in state-of-the-art ANN benchmark performance while maintaining fast indexing times and low memory consumption.
****

Type-to-Track: Retrieve Any Object via Prompt-based Tracking

Pha Nguyen, Kha Gia Quach, Kris Kitani, Khoa Luu

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Finding Counterfactually Optimal Action Sequences in Continuous State Spaces
Stratis Tsirtsis, Manuel Rodriguez
Whenever a clinician reflects on the efficacy of a sequence of treatment decisions for a patient, they may try to identify critical time steps where, had they made different decisions, the patient's health would have improved. While recent methods at the intersection of causal inference and reinforcement learning promise to aid human experts, as the clinician above, to retrospectively analyze sequential decision making processes, they have focused on environments with finitely many discrete states. However, in many practical applications, the state of the environment is inherently continuous in nature. In this paper, we aim to fill this gap. We start by formally characterizing a sequence of discrete actions and continuous states using finite horizon Markov decision processes and a broad class of bijective structural causal models. Building upon this characterization, we formalize the problem of finding counterfactually optimal action sequences and show that, in general, we cannot expect to solve it in polynomial time. Then, we develop a search method based on the A* algorithm that, under a natural form of Lipschitz continuity of the environment’s dynamics, is guaranteed to return the optimal solution to the problem. Experiments on real clinical data show that our method is very efficient in practice, and it has the potential to offer interesting insights for sequential decision making tasks.
****
Reusing Pretrained Models by Multi-linear Operators for Efficient Training
Yu Pan, Ye Yuan, Yichun Yin, Zenglin Xu, Lifeng Shang, Xin Jiang, Qun Liu
Training large models from scratch usually costs a substantial amount of resources. Towards this problem, recent studies such as bert2BERT and LiGO have reused small pretrained models to initialize a large model (termed the ``target model''), leading to a considerable acceleration in training. Despite the successes of these previous studies, they grew pretrained models by mapping partial weights only, ignoring potential correlations across the entire model. As we show in this paper, there are inter- and intra-interactions among the weights of both the pretrained and the target models. As a result, the partial mapping may not capture the complete information and lead to inadequate growth. In this paper, we propose a method that linearly correlates each weight of the target model to all the weights of the pretrained model to further enhance acceleration ability. We utilize multi-linear operators to reduce computational and spacial complexity, enabling acceptable resource requirements. Experiments demonstrate that our method can save 76\% computational costs on DeiT-base transferred from DeiT-small, which outperforms bert2BERT by +12\% and LiGO by +21\%, respectively.
****
Tartarus: A Benchmarking Platform for Realistic And Practical Inverse Molecular Design
AkshatKumar Nigam, Robert Pollice, Gary Tom, Kjell Jorner, John Willes, Luca Thiede, Anshul Kundaje, Alan Aspuru-Guzik
The efficient exploration of chemical space to design molecules with intended properties enables the accelerated discovery of drugs, materials, and catalysts, and is one of the most important outstanding challenges in chemistry. Encouraged by the recent surge in computer power and artificial intelligence development, many algorithms have been developed to tackle this problem. However, despite the emergence of many new approaches in recent years, comparatively little progress has been made in developing realistic benchmarks that reflect the complexity of molecular design for real-world applications. In this work, we develop a set of practical benchmark tasks relying on physical simulation of molecular systems mimicking real-life molecular design problems for materials, drugs, and chemical reactions. Additionally, we demonstrate the utility and ease of use of our new benchmark set by demonstrating how to compare the performance of several well-established families of algorithms. Overall, we believe that our benchmark suite will help move the field towards more realistic molecular design benchmarks, and move the development of inverse molecular design algorithms closer to the practice

of designing molecules that solve existing problems in both academia and industry alike.
****
DreamSparse: Escaping from Plato’s Cave with 2D Diffusion Model Given Sparse Views
Paul Yoo, Jiaxian Guo, Yutaka Matsuo, Shixiang (Shane) Gu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Sample Complexity Bounds for Score-Matching: Causal Discovery and Generative Modeling
Zhenyu Zhu, Francesco Locatello, Volkan Cevher
This paper provides statistical sample complexity bounds for score-matching and its applications in causal discovery. We demonstrate that accurate estimation of the score function is achievable by training a standard deep ReLU neural network using stochastic gradient descent. We establish bounds on the error rate of recovering causal relationships using the score-matching-based causal discovery method of Rolland et al. [2022], assuming a sufficiently good estimation of the score function. Finally, we analyze the upper bound of score-matching estimation within the score-based generative modeling, which has been applied for causal discovery but is also of independent interest within the domain of generative models.
****
Adversarial Robustness in Graph Neural Networks: A Hamiltonian Approach
Kai Zhao, Qiyu Kang, Yang Song, Rui She, Sijie Wang, Wee Peng Tay
Graph neural networks (GNNs) are vulnerable to adversarial perturbations, including those that affect both node features and graph topology. This paper investigates GNNs derived from diverse neural flows, concentrating on their connection to various stability notions such as BIBO stability, Lyapunov stability, structural stability, and conservative stability. We argue that Lyapunov stability, despite its common use, does not necessarily ensure adversarial robustness. Inspired by physics principles, we advocate for the use of conservative Hamiltonian neural flows to construct GNNs that are robust to adversarial attacks. The adversarial robustness of different neural flow GNNs is empirically compared on several benchmark datasets under a variety of adversarial attacks. Extensive numerical experiments demonstrate that GNNs leveraging conservative Hamiltonian flows with Lyapunov stability substantially improve robustness against adversarial perturbations. The implementation code of experiments is available at \url{https://github.com/zknus/NeurIPS-2023-HANG-Robustness}.
****
A Path to Simpler Models Starts With Noise
Lesia Semenova, Harry Chen, Ronald Parr, Cynthia Rudin
The Rashomon set is the set of models that perform approximately equally well on a given dataset, and the Rashomon ratio is the fraction of all models in a given hypothesis space that are in the Rashomon set. Rashomon ratios are often large for tabular datasets in criminal justice, healthcare, lending, education, and in other areas, which has practical implications about whether simpler models can attain the same level of accuracy as more complex models. An open question is why Rashomon ratios often tend to be large. In this work, we propose and study a mechanism of the data generation process, coupled with choices usually made by the analyst during the learning process, that determines the size of the Rashomon ratio. Specifically, we demonstrate that noisier datasets lead to larger Rashomon ratios through the way that practitioners train models. Additionally, we introduce a measure called pattern diversity, which captures the average difference in predictions between distinct classification patterns in the Rashomon set, and motivate why it tends to increase with label noise. Our results explain a key aspect of why simpler models often tend to perform as well as black box models on complex, noisier datasets.

****
Winner-Take-All Column Row Sampling for Memory Efficient Adaptation of Language Model
Zirui Liu, Guanchu Wang, Shaochen (Henry) Zhong, Zhaozhuo Xu, Daochen Zha, Ruixiang (Ryan) Tang, Zhimeng (Stephen) Jiang, Kaixiong Zhou, Vipin Chaudhary, Shuai Xu, Xia Hu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Zeroth-Order Methods for Nondifferentiable, Nonconvex, and Hierarchical Federated Optimization
Yuyang Qiu, Uday Shanbhag, Farzad Yousefian
Federated learning (FL) has emerged as an enabling framework for communication-efficient decentralized training. We study three broadly applicable problem classes in FL: (i) Nondifferentiable nonconvex federated optimization; (ii) Federated bilevel optimization; (iii) Federated minimax problems. Notably, in an implicit sense, both (ii) and (iii) are instances of (i). However, the hierarchical problems in (ii) and (iii) are often complicated by the absence of a closed-form expression for the implicit objective function. Unfortunately, research on these problems has been limited and afflicted by reliance on strong assumptions, including the need for differentiability and L-smoothness of the implicit function. We address this shortcoming by making the following contributions. In (i), by leveraging convolution-based smoothing and Clarke’s subdifferential calculus, we devise a randomized smoothing-enabled zeroth-order FL method and derive communication and iteration complexity guarantees for computing an approximate Clarke stationary point. To contend with (ii) and (iii), we devise a unified randomized implicit zeroth-order FL framework, equipped with explicit communication and iteration complexities. Importantly, our method utilizes delays during local steps to skip making calls to the inexact lower-level FL oracle. This results in significant reduction in communication overhead when addressing hierarchical problems. We empirically validate the theory on nonsmooth and hierarchical ML problems.
****
Language Model Alignment with Elastic Reset
Michael Noukhovitch, Samuel Lavoie, Florian Strub, Aaron C. Courville
Finetuning language models with reinforcement learning (RL), e.g. from human feedback (HF), is a prominent method for alignment. But optimizing against a reward model can improve on reward while degrading performance in other areas, a phenomenon known as reward hacking, alignment tax, or language drift. First, we argue that commonly-used test metrics are insufficient and instead measure how different algorithms tradeoff between reward and drift. The standard method modified the reward with a Kullback-Lieber (KL) penalty between the online and initial model. We propose Elastic Reset, a new algorithm that achieves higher reward with less drift without explicitly modifying the training objective. We periodically reset the online model to an exponentially moving average (EMA) of itself, then reset the EMA model to the initial model. Through the use of an EMA, our model recovers quickly after resets and achieves higher reward with less drift in the same number of steps. We demonstrate that fine-tuning language models with Elastic Reset leads to state-of-the-art performance on a small scale pivot-translation benchmark, outperforms all baselines in a medium-scale RLHF-like IMDB mock sentiment task and leads to a more performant and more aligned technical QA chatbot with LLaMA-7B. Code available https://github.com/mnoukhov/elastic-reset
****
Resolving the Tug-of-War: A Separation of Communication and Learning in Federated Learning
Junyi Li, Heng Huang
Federated learning (FL) is a promising privacy-preserving machine learning paradigm over distributed data. In this paradigm, each client trains the parameter of a model locally and the server aggregates the parameter from clients periodical

ly. Therefore, we perform the learning and communication over the same set of pa
rameters. However, we find that learning and communication have fundamentally di
vergent requirements for parameter selection, akin to two opposite teams in a tu
g-of-war game. To mitigate this discrepancy, we introduce FedSep, a novel two-la
yer federated learning framework. FedSep consists of separated communication and
 learning layers for each client and the two layers are connected through decode
/encode operations. In particular, the decoding operation is formulated as a min
imization problem. We view FedSep as a federated bilevel optimization problem an
d propose an efficient algorithm to solve it. Theoretically, we demonstrate that
 its convergence matches that of the standard FL algorithms. The separation of c
ommunication and learning in FedSep offers innovative solutions to various chall
enging problems in FL, such as Communication-Efficient FL and Heterogeneous-Mode
l FL. Empirical validation shows the superior performance of FedSep over various
 baselines in these tasks.
****
GlucoSynth: Generating Differentially-Private Synthetic Glucose Traces
Josephine Lamp, Mark Derdzinski, Christopher Hannemann, Joost van der Linden, Lu
 Feng, Tianhao Wang, David Evans
We focus on the problem of generating high-quality, private synthetic glucose tr
aces, a task generalizable to many other time series sources. Existing methods f
or time series data synthesis, such as those using Generative Adversarial Networ
ks (GANs), are not able to capture the innate characteristics of glucose data an
d cannot provide any formal privacy guarantees without severely degrading the ut
ility of the synthetic data. In this paper we present GlucoSynth, a novel privac
y-preserving GAN framework to generate synthetic glucose traces. The core intuit
ion behind our approach is to conserve relationships amongst motifs (glucose eve
nts) within the traces, in addition to temporal dynamics. Our framework incorpor
ates differential privacy mechanisms to provide strong formal privacy guarantees
. We provide a comprehensive evaluation on the real-world utility of the data us
ing 1.2 million glucose traces; GlucoSynth outperforms all previous methods in i
ts ability to generate high-quality synthetic glucose traces with strong privacy
 guarantees.
****
OBJECT 3DIT: Language-guided 3D-aware Image Editing
Oscar Michel, Anand Bhattad, Eli VanderBilt, Ranjay Krishna, Aniruddha Kembhavi,
 Tanmay Gupta
Existing image editing tools, while powerful, typically disregard the underlying
 3D geometry from which the image is projected. As a result, edits made using th
ese tools may become detached from the geometry and lighting conditions that are
 at the foundation of the image formation process; such edits break the portraya
l of a coherent 3D world. 3D-aware generative models are a promising solution, b
ut currently only succeed on small datasets or at the level of a single object.
In this work, we formulate the new task of language-guided 3D-aware editing, whe
re objects in an image should be edited according to a language instruction whil
e remaining consistent with the underlying 3D scene. To promote progress towards
 this goal, we release OBJect: a benchmark dataset of 400K editing examples crea
ted from procedurally generated 3D scenes. Each example consists of an input ima
ge, editing instruction in language, and the edited image. We also introduce 3DI
T: single and multi-task models for four editing tasks. Our models show impressi
ve abilities to understand the 3D composition of entire scenes, factoring in sur
rounding objects, surfaces, lighting conditions, shadows, and physically-plausib
le object configurations. Surprisingly, training on only synthetic scenes from \
dataset, editing capabilities of 3DIT generalize to real-world images.
****
Learning Rule-Induced Subgraph Representations for Inductive Relation Prediction
Tianyu Liu, Qitan Lv, Jie Wang, Shuling Yang, Hanzhu Chen
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.

****
Linguistic Binding in Diffusion Models: Enhancing Attribute Correspondence throu
gh Attention Map Alignment
Royi Rassin, Eran Hirsch, Daniel Glickman, Shauli Ravfogel, Yoav Goldberg, Gal C
hechik
Text-conditioned image generation models often generate incorrect associations b
etween entities and their visual attributes. This reflects an impaired mapping b
etween linguistic binding of entities and modifiers in the prompt and visual bin
ding of the corresponding elements in the generated image. As one example, a que
ry like ``a pink sunflower and a yellow flamingo'' may incorrectly produce an im
age of a yellow sunflower and a pink flamingo. To remedy this issue, we propose
SynGen, an approach which first syntactically analyses the prompt to identify en
tities and their modifiers, and then uses a novel loss function that encourages
the cross-attention maps to agree with the linguistic binding reflected by the s
yntax. Specifically, we encourage large overlap between attention maps of entiti
es and their modifiers, and small overlap with other entities and modifier words
. The loss is optimized during inference, without retraining or fine-tuning the
model. Human evaluation on three datasets, including one new and challenging set
, demonstrate significant improvements of SynGen compared with current state of
the art methods. This work highlights how making use of sentence structure durin
g inference can efficiently and substantially improve the faithfulness of text-t
o-image generation.
****
Optimistic Natural Policy Gradient: a Simple Efficient Policy Optimization Frame
work for Online RL
Qinghua Liu, Gellert Weisz, András György, Chi Jin, Csaba Szepesvari
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Two-Stage Learning to Defer with Multiple Experts
Anqi Mao, Christopher Mohri, Mehryar Mohri, Yutao Zhong
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
A Computationally Efficient Sparsified Online Newton Method
Fnu Devvrit, Sai Surya Duvvuri, Rohan Anil, Vineet Gupta, Cho-Jui Hsieh, Inderji
t Dhillon
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
SparseProp: Efficient Event-Based Simulation and Training of Sparse Recurrent Sp
iking Neural Networks
Rainer Engelken
Spiking Neural Networks (SNNs) are biologically-inspired models that are capable
 of processing information in streams of action potentials. However, simulating
and training SNNs is computationally expensive due to the need to solve large sy
stems of coupled differential equations. In this paper, we propose a novel event
-based algorithm called SparseProp for simulating and training sparse SNNs. Our
algorithm reduces the computational cost of both forward pass and backward pass
operations from O(N) to O(log(N)) per network spike, enabling numerically exact
simulations of large spiking networks and their efficient training using backpro
pagation through time. By exploiting the sparsity of the network, SparseProp avo
ids iterating through all neurons at every spike and uses efficient state update
s. We demonstrate the effectiveness of SparseProp for several classical integrat

e-and-fire neuron models, including simulating a sparse SNN with one million LIF neurons, which is sped up by more than four orders of magnitude compared to previous implementations. Our work provides an efficient and exact solution for training large-scale spiking neural networks and opens up new possibilities for building more sophisticated brain-inspired models.
****

ConRad: Image Constrained Radiance Fields for 3D Generation from a Single Image
Senthil Purushwalkam, Nikhil Naik
We present a novel method for reconstructing 3D objects from a single RGB image. Our method leverages the latest image generation models to infer the hidden 3D structure while remaining faithful to the input image. While existing methods obtain impressive results in generating 3D models from text prompts, they do not provide an easy approach for conditioning on input RGB data. Naive extensions of these methods often lead to improper alignment in appearance between the input image and the 3D reconstructions. We address these challenges by introducing Image Constrained Radiance Fields (ConRad), a novel variant of neural radiance fields. ConRad is an efficient 3D representation that explicitly captures the appearance of an input image in one viewpoint. We propose a training algorithm that leverages the single RGB image in conjunction with pretrained Diffusion Models to optimize the parameters of a ConRad representation. Extensive experiments show that ConRad representations can simplify preservation of image details while producing a realistic 3D reconstruction. Compared to existing state-of-the-art baselines, we show that our 3D reconstructions remain more faithful to the input and produce more consistent 3D models while demonstrating significantly improved quantitative performance on a ShapeNet object benchmark.
****

Fair Canonical Correlation Analysis
Zhuoping Zhou, Davoud Ataee Tarzanagh, Bojian Hou, Boning Tong, Jia Xu, Yanbo Feng, Qi Long, Li Shen
This paper investigates fairness and bias in Canonical Correlation Analysis (CCA), a widely used statistical technique for examining the relationship between two sets of variables. We present a framework that alleviates unfairness by minimizing the correlation disparity error associated with protected attributes. Our approach enables CCA to learn global projection matrices from all data points while ensuring that these matrices yield comparable correlation levels to group-specific projection matrices. Experimental evaluation on both synthetic and real-world datasets demonstrates the efficacy of our method in reducing correlation disparity error without compromising CCA accuracy.
****

DIFUSCO: Graph-based Diffusion Solvers for Combinatorial Optimization
Zhiqing Sun, Yiming Yang
Neural network-based Combinatorial Optimization (CO) methods have shown promising results in solving various NP-complete (NPC) problems without relying on hand-crafted domain knowledge. This paper broadens the current scope of neural solvers for NPC problems by introducing a new graph-based diffusion framework, namely DIFUSCO. It formulates NPC problems into a discrete {0, 1}-vector space and uses graph-based denoising diffusion models to generate high-quality solutions. Specifically, we explore diffusion models with Gaussian and Bernoulli noise, respectively, and also introduce an effective inference schedule to improve the generation quality. We evaluate our methods on two well-studied combinatorial optimization problems: Traveling Salesman Problem (TSP) and Maximal Independent Set (MIS). Experimental results show that DIFUSCO strongly outperforms the previous state-of-the-art neural solvers, improving the performance gap between ground-truth and neural solvers from 1.76% to 0.46% on TSP-500, from 2.46% to 1.17% on TSP-1000, and from 3.19% to 2.58% on TSP-10000. For the MIS problem, DIFUSCO outperforms the previous state-of-the-art neural solver on the challenging SATLIB benchmark. Our code is available at this url.
****

Exposing flaws of generative model evaluation metrics and their unfair treatment of diffusion models

George Stein, Jesse Cresswell, Rasa Hosseinzadeh, Yi Sui, Brendan Ross, Valentin Villecroze, Zhaoyan Liu, Anthony L. Caterini, Eric Taylor, Gabriel Loaiza-Ganem
We systematically study a wide variety of generative models spanning semantically-diverse image datasets to understand and improve the feature extractors and metrics used to evaluate them.Using best practices in psychophysics, we measure human perception of image realism for generated samples by conducting the largest experiment evaluating generative models to date, and find that no existing metric strongly correlates with human evaluations.Comparing to 17 modern metrics for evaluating the overall performance, fidelity, diversity, rarity, and memorization of generative models, we find that the state-of-the-art perceptual realism of diffusion models as judged by humans is not reflected in commonly reported metrics such as FID. This discrepancy is not explained by diversity in generated samples, though one cause is over-reliance on Inception-V3.We address these flaws through a study of alternative self-supervised feature extractors, find that the semantic information encoded by individual networks strongly depends on their training procedure, and show that DINOv2-ViT-L/14 allows for much richer evaluation of generative models. Next, we investigate data memorization, and find that generative models do memorize training examples on simple, smaller datasets like CIFAR10, but not necessarily on more complex datasets like ImageNet. However, our experiments show that current metrics do not properly detect memorization: none in the literature is able to separate memorization from other phenomena such as underfitting or mode shrinkage. To facilitate further development of generative models and their evaluation we release all generated image datasets, human evaluation data, and a modular library to compute 17 common metrics for 9 different encoders at https://github.com/layer6ai-labs/dgm-eval.
****
Online Clustering of Bandits with Misspecified User Models
Zhiyong Wang, Jize Xie, Xutong Liu, Shuai Li, John C.S. Lui
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Temporal Conditioning Spiking Latent Variable Models of the Neural Response to Natural Visual Scenes
Gehua Ma, Runhao Jiang, Rui Yan, Huajin Tang
Developing computational models of neural response is crucial for understanding sensory processing and neural computations. Current state-of-the-art neural network methods use temporal filters to handle temporal dependencies, resulting in an unrealistic and inflexible processing paradigm. Meanwhile, these methods target trial-averaged firing rates and fail to capture important features in spike trains. This work presents the temporal conditioning spiking latent variable models (TeCoS-LVM) to simulate the neural response to natural visual stimuli. We use spiking neurons to produce spike outputs that directly match the recorded trains. This approach helps to avoid losing information embedded in the original spike trains. We exclude the temporal dimension from the model parameter space and introduce a temporal conditioning operation to allow the model to adaptively explore and exploit temporal dependencies in stimuli sequences in a natural paradigm. We show that TeCoS-LVM models can produce more realistic spike activities and accurately fit spike statistics than powerful alternatives. Additionally, learned TeCoS-LVM models can generalize well to longer time scales. Overall, while remaining computationally tractable, our model effectively captures key features of neural coding systems. It thus provides a useful tool for building accurate predictive computational accounts for various sensory perception circuits.
****
Double Auctions with Two-sided Bandit Feedback
Soumya Basu, Abishek Sankararaman
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Evaluating Graph Neural Networks for Link Prediction: Current Pitfalls and New Benchmarking
Juanhui Li, Harry Shomer, Haitao Mao, Shenglai Zeng, Yao Ma, Neil Shah, Jiliang Tang, Dawei Yin
Link prediction attempts to predict whether an unseen edge exists based on only a portion of the graph. A flurry of methods has been created in recent years that attempt to make use of graph neural networks (GNNs) for this task. Furthermore, new and diverse datasets have also been created to better evaluate the effectiveness of these new models. However, multiple limitations currently exist that hinders our ability to properly evaluate these new methods. This includes, but is not limited to: (1) The underreporting of performance on multiple baselines, (2) A lack of a unified data split and evaluation metric on some datasets, (3) An unrealistic evaluation setting that produces negative samples that are easy to classify. To overcome these challenges we first conduct a fair comparison across prominent methods and datasets, utilizing the same dataset settings and hyperparameter settings. We then create a new real-world evaluation setting that samples difficult negative samples via multiple heuristics. The new evaluation setting helps promote new challenges and opportunities in link prediction by aligning the evaluation with real-world situations.
****
EHRXQA: A Multi-Modal Question Answering Dataset for Electronic Health Records with Chest X-ray Images
Seongsu Bae, Daeun Kyung, Jaehee Ryu, Eunbyeol Cho, Gyubok Lee, Sunjun Kweon, Jungwoo Oh, Lei Ji, Eric Chang, Tackeun Kim, Edward Choi
Electronic Health Records (EHRs), which contain patients' medical histories in various multi-modal formats, often overlook the potential for joint reasoning across imaging and table modalities underexplored in current EHR Question Answering (QA) systems. In this paper, we introduce EHRXQA, a novel multi-modal question answering dataset combining structured EHRs and chest X-ray images. To develop our dataset, we first construct two uni-modal resources: 1) The MIMIC- CXR-VQA dataset, our newly created medical visual question answering (VQA) benchmark, specifically designed to augment the imaging modality in EHR QA, and 2) EHRSQL (MIMIC-IV), a refashioned version of a previously established table-based EHR QA dataset. By integrating these two uni-modal resources, we successfully construct a multi-modal EHR QA dataset that necessitates both uni-modal and cross-modal reasoning. To address the unique challenges of multi-modal questions within EHRs, we propose a NeuralSQL-based strategy equipped with an external VQA API. This pioneering endeavor enhances engagement with multi-modal EHR sources and we believe that our dataset can catalyze advances in real-world medical scenarios such as clinical decision-making and research. EHRXQA is available at https://github.com/baeseongsu/ehrxqa.
****
Enhancing Robot Program Synthesis Through Environmental Context
Tianyi Chen, Qidi Wang, Zhen Dong, Liwei Shen, Xin Peng
Program synthesis aims to automatically generate an executable program that conforms to the given specification. Recent advancements have demonstrated that deep neural methodologies and large-scale pretrained language models are highly proficient in capturing program semantics.For robot programming, prior works have facilitated program synthesis by incorporating global environments. However, the assumption of acquiring a comprehensive understanding of the entire environment is often excessively challenging to achieve.In this work, we present a framework that learns to synthesize a program by rectifying potentially erroneous code segments, with the aid of partially observed environments. To tackle the issue of inadequate attention to partial observations, we propose to first learn an environment embedding space that can implicitly evaluate the impacts of each program token based on the precondition. Furthermore, by employing a graph structure, the model can aggregate both environmental and syntactic information flow and furnish smooth program rectification guidance.Extensive experimental evaluations and

ablation studies on the partially observed VizDoom domain authenticate that our method offers superior generalization capability across various tasks and greater robustness when encountering noises.
****
ScenarioNet: Open-Source Platform for Large-Scale Traffic Scenario Simulation and Modeling
Quanyi Li, Zhenghao (Mark) Peng, Lan Feng, Zhizheng Liu, Chenda Duan, Wenjie Mo, Bolei Zhou
Large-scale driving datasets such as Waymo Open Dataset and nuScenes substantially accelerate autonomous driving research, especially for perception tasks such as 3D detection and trajectory forecasting. Since the driving logs in these datasets contain HD maps and detailed object annotations which accurately reflect the real-world complexity of traffic behaviors, we can harvest a massive number of complex traffic scenarios and recreate their digital twins in simulation. Compared to the hand-crafted scenarios often used in existing simulators, data-driven scenarios collected from the real world can facilitate many research opportunities in machine learning and autonomous driving. In this work, we present ScenarioNet, an open-source platform for large-scale traffic scenario modeling and simulation. ScenarioNet defines a unified scenario description format and collects a large-scale repository of real-world traffic scenarios from the heterogeneous data in various driving datasets including Waymo, nuScenes, Lyft L5, and nuPlan datasets. These scenarios can be further replayed and interacted with in multiple views from Bird-Eye-View layout to realistic 3D rendering in MetaDrive simulator. This provides a benchmark for evaluating the safety of autonomous driving stacks in simulation before their real-world deployment. We further demonstrate the strengths of ScenarioNet on large-scale scenario generation, imitation learning, and reinforcement learning in both single-agent and multi-agent settings. Code, demo videos, and website are available at https://github.com/metadriverse/scenarionet
****
Understanding Deep Gradient Leakage via Inversion Influence Functions
Haobo Zhang, Junyuan Hong, Yuyang Deng, Mehrdad Mahdavi, Jiayu Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Joint Learning of Label and Environment Causal Independence for Graph Out-of-Distribution Generalization
Shurui Gui, Meng Liu, Xiner Li, Youzhi Luo, Shuiwang Ji
We tackle the problem of graph out-of-distribution (OOD) generalization. Existing graph OOD algorithms either rely on restricted assumptions or fail to exploit environment information in training data. In this work, we propose to simultaneously incorporate label and environment causal independence (LECI) to fully make use of label and environment information, thereby addressing the challenges faced by prior methods on identifying causal and invariant subgraphs. We further develop an adversarial training strategy to jointly optimize these two properties for casual subgraph discovery with theoretical guarantees. Extensive experiments and analysis show that LECI significantly outperforms prior methods on both synthetic and real-world datasets, establishing LECI as a practical and effective solution for graph OOD generalization.
****
Bayesian Learning of Optimal Policies in Markov Decision Processes with Countably Infinite State-Space
Saghar Adler, Vijay Subramanian
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

CARE: Modeling Interacting Dynamics Under Temporal Environmental Variation
Xiao Luo, Haixin Wang, Zijie Huang, Huiyu Jiang, Abhijeet Gangan, Song Jiang, Yizhou Sun
Modeling interacting dynamical systems, such as fluid dynamics and intermolecular interactions, is a fundamental research problem for understanding and simulating complex real-world systems. Many of these systems can be naturally represented by dynamic graphs, and graph neural network-based approaches have been proposed and shown promising performance. However, most of these approaches assume the underlying dynamics does not change over time, which is unfortunately untrue. For example, a molecular dynamics can be affected by the environment temperature over the time. In this paper, we take an attempt to provide a probabilistic view for time-varying dynamics and propose a model Context-attended Graph ODE (CARE) for modeling time-varying interacting dynamical systems. In our CARE, we explicitly use a context variable to model time-varying environment and construct an encoder to initialize the context variable from historical trajectories. Furthermore, we employ a neural ODE model to depict the dynamic evolution of the context variable inferred from system states. This context variable is incorporated into a coupled ODE to simultaneously drive the evolution of systems. Comprehensive experiments on four datasets demonstrate the effectiveness of our proposed CARE compared with several state-of-the-art approaches.
****
Diffused Redundancy in Pre-trained Representations
Vedant Nanda, Till Speicher, John Dickerson, Krishna Gummadi, Soheil Feizi, Adrian Weller
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
AI for Interpretable Chemistry: Predicting Radical Mechanistic Pathways via Contrastive Learning
Mohammadamin Tavakoli, Pierre Baldi, Ann Marie Carlton, Yin Ting Chiu, Alexander Shmakov, David Van Vranken
Deep learning-based reaction predictors have undergone significant architectural evolution. However, their reliance on reactions from the US Patent Office results in a lack of interpretable predictions and limited generalizability to other chemistry domains, such as radical and atmospheric chemistry. To address these challenges, we introduce a new reaction predictor system, RMechRP, that leverages contrastive learning in conjunction with mechanistic pathways, the most interpretable representation of chemical reactions. Specifically designed for radical reactions, RMechRP provides different levels of interpretation of chemical reactions. We develop and train multiple deep-learning models using RMechDB, a public database of radical reactions, to establish the first benchmark for predicting radical reactions. Our results demonstrate the effectiveness of RMechRP in providing accurate and interpretable predictions of radical reactions, and its potential for various applications in atmospheric chemistry.
****
Randomized Sparse Neural Galerkin Schemes for Solving Evolution Equations with Deep Networks
Jules Berman, Benjamin Peherstorfer
Training neural networks sequentially in time to approximate solution fields of time-dependent partial differential equations can be beneficial for preserving causality and other physics properties; however, the sequential-in-time training is numerically challenging because training errors quickly accumulate and amplify over time. This work introduces Neural Galerkin schemes that update randomized sparse subsets of network parameters at each time step. The randomization avoids overfitting locally in time and so helps prevent the error from accumulating quickly over the sequential-in-time training, which is motivated by dropout that addresses a similar issue of overfitting due to neuron co-adaptation. The sparsity of the update reduces the computational costs of training without losing expr

essiveness because many of the network parameters are redundant locally at each time step. In numerical experiments with a wide range of evolution equations, the proposed scheme with randomized sparse updates is up to two orders of magnitude more accurate at a fixed computational budget and up to two orders of magnitude faster at a fixed accuracy than schemes with dense updates.
****
Handling Data Heterogeneity via Architectural Design for Federated Visual Recognition
Sara Pieri, Jose Restom, Samuel Horváth, Hisham Cholakkal
Federated Learning (FL) is a promising research paradigm that enables the collaborative training of machine learning models among various parties without the need for sensitive information exchange. Nonetheless, retaining data in individual clients introduces fundamental challenges to achieving performance on par with centrally trained models. Our study provides an extensive review of federated learning applied to visual recognition. It underscores the critical role of thoughtful architectural design choices in achieving optimal performance, a factor often neglected in the FL literature. Many existing FL solutions are tested on shallow or simple networks, which may not accurately reflect real-world applications. This practice restricts the transferability of research findings to large-scale visual recognition models. Through an in-depth analysis of diverse cutting-edge architectures such as convolutional neural networks, transformers, and MLP-mixers, we experimentally demonstrate that architectural choices can substantially enhance FL systems' performance, particularly when handling heterogeneous data. We study visual recognition models from five different architectural families on four challenging FL datasets. We also re-investigate the inferior performance convolution-based architectures in the FL setting and analyze the influence of normalization layers on the FL performance. Our findings emphasize the importance of architectural design for computer vision tasks in practical scenarios, effectively narrowing the performance gap between federated and centralized learning.
****
Spatial-frequency channels, shape bias, and adversarial robustness
Ajay Subramanian, Elena Sizikova, Najib Majaj, Denis Pelli
What spatial frequency information do humans and neural networks use to recognize objects? In neuroscience, critical band masking is an established tool that can reveal the frequency-selective filters used for object recognition. Critical band masking measures the sensitivity of recognition performance to noise added at each spatial frequency. Existing critical band masking studies show that humans recognize periodic patterns (gratings) and letters by means of a spatial-frequency filter (or "channel") that has a frequency bandwidth of one octave (doubling of frequency). Here, we introduce critical band masking as a task for network-human comparison and test 14 humans and 76 neural networks on 16-way ImageNet categorization in the presence of narrowband noise. We find that humans recognize objects in natural images using the same one-octave-wide channel that they use for letters and gratings, making it a canonical feature of human object recognition. Unlike humans, the neural network channel is very broad, 2-4 times wider than the human channel. This means that the network channel extends to frequencies higher and lower than those that humans are sensitive to. Thus, noise at those frequencies will impair network performance and spare human performance. Adversarial and augmented-image training are commonly used to increase network robustness and shape bias. Does this training align network and human object recognition channels? Three network channel properties (bandwidth, center frequency, peak noise sensitivity) correlate strongly with shape bias (51% variance explained) and robustness of adversarially-trained networks (66% variance explained). Adversarial training increases robustness but expands the channel bandwidth even further beyond the human bandwidth. Thus, critical band masking reveals that the network channel is more than twice as wide as the human channel, and that adversarial training only makes it worse. Networks with narrower channels might be more robust.
****

Optimality in Mean Estimation: Beyond Worst-Case, Beyond Sub-Gaussian, and Beyond $1+\alpha$ Moments
Trung Dang, Jasper Lee, Maoyuan 'Raymond' Song, Paul Valiant
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Provably Efficient Offline Goal-Conditioned Reinforcement Learning with General Function Approximation and Single-Policy Concentrability
Hanlin Zhu, Amy Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SQ Lower Bounds for Non-Gaussian Component Analysis with Weaker Assumptions
Ilias Diakonikolas, Daniel Kane, Lisheng Ren, Yuxin Sun
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Efficient Equivariant Transfer Learning from Pretrained Models
Sourya Basu, Pulkit Katdare, Prasanna Sattigeri, Vijil Chenthamarakshan, Katherine Driggs-Campbell, Payel Das, Lav R. Varshney
Efficient transfer learning algorithms are key to the success of foundation models on diverse downstream tasks even with limited data. Recent works of Basu et al. (2023) and Kaba et al. (2022) propose group averaging (equitune) and optimization-based methods, respectively, over features from group-transformed inputs to obtain equivariant outputs from non-equivariant neural networks. While Kaba et al. (2022) are only concerned with training from scratch, we find that equitune performs poorly on equivariant zero-shot tasks despite good finetuning results. We hypothesize that this is because pretrained models provide better quality features for certain transformations than others and simply averaging them is deleterious. Hence, we propose λ-equitune that averages the features using importance weights, λs. These weights are learned directly from the data using a small neural network, leading to excellent zero-shot and finetuned results that outperform equitune. Further, we prove that λ-equitune is equivariant and a universal approximator of equivariant functions. Additionally, we show that the method of Kaba et al. (2022) used with appropriate loss functions, which we call equizero, also gives excellent zero-shot and finetuned performance. Both equitune and equizero are special cases of λ- equitune. To show the simplicity and generality of our method, we validate on a wide range of diverse applications and models such as 1) image classification using CLIP, 2) deep Q-learning, 3) fairness in natural language generation (NLG), 4) compositional generalization in languages, and 5) image classification using pretrained CNNs such as Resnet and Alexnet.
****
Kernelized Reinforcement Learning with Order Optimal Regret Bounds
Sattar Vakili, Julia Olkhovskaya
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Domain-Aware Detection Head with Prompt Tuning
Haochen Li, Rui Zhang, Hantao Yao, Xinkai Song, Yifan Hao, Yongwei Zhao, Ling Li, Yunji Chen
Domain adaptive object detection (DAOD) aims to generalize detectors trained on an annotated source domain to an unlabelled target domain. However, existing me

thods focus on reducing the domain bias of the detection backbone by inferring a discriminative visual encoder, while ignoring the domain bias in the detection head. Inspired by the high generalization of vision-language models (VLMs), applying a VLM as the robust detection backbone following a domain-aware detection head is a reasonable way to learn the discriminative detector for each domain, rather than reducing the domain bias in traditional methods. To achieve the above issue, we thus propose a novel DAOD framework named Domain-Aware detection head with Prompt tuning (DA-Pro), which applies the learnable domain-adaptive prompt to generate the dynamic detection head for each domain. Formally, the domain-adaptive prompt consists of the domain-invariant tokens, domain-specific tokens, and the domain-related textual description along with the class label. Furthermore, two constraints between the source and target domains are applied to ensure that the domain-adaptive prompt can capture the domains-shared and domain-specific knowledge. A prompt ensemble strategy is also proposed to reduce the effect of prompt disturbance. Comprehensive experiments over multiple cross-domain adaptation tasks demonstrate that using the domain-adaptive prompt can produce an effectively domain-related detection head for boosting domain-adaptive object detection. Our code is available at https://github.com/Therock90421/DA-Pro.
****
Parallel Sampling of Diffusion Models
Andy Shih, Suneel Belkhale, Stefano Ermon, Dorsa Sadigh, Nima Anari
Diffusion models are powerful generative models but suffer from slow sampling, often taking 1000 sequential denoising steps for one sample. As a result, considerable efforts have been directed toward reducing the number of denoising steps, but these methods hurt sample quality. Instead of reducing the number of denoising steps (trading quality for speed), in this paper we explore an orthogonal approach: can we run the denoising steps in parallel (trading compute for speed)? In spite of the sequential nature of the denoising steps, we show that surprisingly it is possible to parallelize sampling via Picard iterations, by guessing the solution of future denoising steps and iteratively refining until convergence. With this insight, we present ParaDiGMS, a novel method to accelerate the sampling of pretrained diffusion models by denoising multiple steps in parallel. ParaDiGMS is the first diffusion sampling method that enables trading compute for speed and is even compatible with existing fast sampling techniques such as DDIM and DPMSolver. Using ParaDiGMS, we improve sampling speed by 2-4x across a range of robotics and image generation models, giving state-of-the-art sampling speeds of 0.2s on 100-step DiffusionPolicy and 14.6s on 1000-step StableDiffusion-v2 with no measurable degradation of task reward, FID score, or CLIP score.
****
Fractal Landscapes in Policy Optimization
Tao Wang, Sylvia Herbert, Sicun Gao
Policy gradient lies at the core of deep reinforcement learning (RL) in continuous domains. Despite much success, it is often observed in practice that RL training with policy gradient can fail for many reasons, even on standard control problems with known solutions. We propose a framework for understanding one inherent limitation of the policy gradient approach: the optimization landscape in the policy space can be extremely non-smooth or fractal for certain classes of MDPs, such that there does not exist gradient to be estimated in the first place. We draw on techniques from chaos theory and non-smooth analysis, and analyze the maximal Lyapunov exponents and H\"older exponents of the policy optimization objectives. Moreover, we develop a practical method that can estimate the local smoothness of objective function from samples to identify when the training process has encountered fractal landscapes. We show experiments to illustrate how some failure cases of policy optimization can be explained by such fractal landscapes.
****
Moral Responsibility for AI Systems
Sander Beckers
As more and more decisions that have a significant ethical dimension are being outsourced to AI systems, it is important to have a definition of moral responsibility that can be applied to AI systems. Moral responsibility for an outcome of

an agent who performs some action is commonly taken to involve both a causal con dition and an epistemic condition: the action should cause the outcome, and the agent should have been aware - in some form or other - of the possible moral con sequences of their action. This paper presents a formal definition of both condi tions within the framework of causal models. I compare my approach to the existi ng approaches of Braham and van Hees (BvH) and of Halpern and Kleiman-Weiner (HK). I then generalize my definition into a degree of responsibility.
****
Characterizing the Impacts of Semi-supervised Learning for Weak Supervision
Jeffrey Li, Jieyu Zhang, Ludwig Schmidt, Alexander J. Ratner
Labeling training data is a critical and expensive step in producing high accura cy ML models, whether training from scratch or fine-tuning. To make labeling mor e efficient, two major approaches are programmatic weak supervision (WS) and sem i-supervised learning (SSL). More recent works have either explicitly or implici tly used techniques at their intersection, but in various complex and ad hoc way s. In this work, we define a simple, modular design space to study the use of SS L techniques for WS more systematically. Surprisingly, we find that fairly simpl e methods from our design space match the performance of more complex state-of-t he-art methods, averaging a 3 p.p. increase in accuracy/F1-score across 8 standa rd WS benchmarks. Further, we provide practical guidance on when different compo nents are worth their added complexity and training costs. Contrary to current u nderstanding, we find using SSL is not necessary to obtain the best performance on most WS benchmarks but is more effective when: (1) end models are smaller, an d (2) WS provides labels for only a small portion of training examples.
****
Finite-Time Logarithmic Bayes Regret Upper Bounds
Alexia Atsidakou, Branislav Kveton, Sumeet Katariya, Constantine Caramanis, Suja y Sanghavi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Frequency-Enhanced Data Augmentation for Vision-and-Language Navigation
Keji He, Chenyang Si, Zhihe Lu, Yan Huang, Liang Wang, Xinchao Wang
Vision-and-Language Navigation (VLN) is a challenging task that requires an agen t to navigate through complex environments based on natural language instruction s. In contrast to conventional approaches, which primarily focus on the spatial domain exploration, we propose a paradigm shift toward the Fourier domain. This alternative perspective aims to enhance visual-textual matching, ultimately impr oving the agent's ability to understand and execute navigation tasks based on th e given instructions. In this study, we first explore the significance of high-f requency information in VLN and provide evidence that it is instrumental in bols tering visual-textual matching processes. Building upon this insight, we further propose a sophisticated and versatile Frequency-enhanced Data Augmentation (FDA) technique to improve the VLN model's capability of capturing critical high-fre quency information. Specifically, this approach requires the agent to navigate i n environments where only a subset of high-frequency visual information correspo nds with the provided textual instructions, ultimately fostering the agent's abi lity to selectively discern and capture pertinent high-frequency features accord ing to the given instructions. Promising results on R2R, RxR, CVDN and REVERIE d emonstrate that our FDA can be readily integrated with existing VLN approaches, improving performance without adding extra parameters, and keeping models simple and efficient. The code is available at https://github.com/hekj/FDA.
****
Building Socio-culturally Inclusive Stereotype Resources with Community Engageme nt
Sunipa Dev, Jaya Goyal, Dinesh Tewari, Shachi Dave, Vinodkumar Prabhakaran
With rapid development and deployment of generative language models in global se ttings, there is an urgent need to also scale our measurements of harm, not just

in the number and types of harms covered, but also how well they account for local cultural contexts, including marginalized identities and the social biases experienced by them.Current evaluation paradigms are limited in their abilities to address this, as they are not representative of diverse, locally situated but global, socio-cultural perspectives. It is imperative that our evaluation resources are enhanced and calibrated by including people and experiences from different cultures and societies worldwide, in order to prevent gross underestimations or skews in measurements of harm. In this work, we demonstrate a socio-culturally aware expansion of evaluation resources in the Indian societal context, specifically for the harm of stereotyping. We devise a community engaged effort to build a resource which contains stereotypes for axes of disparity that are uniquely present in India. The resultant resource increases the number of stereotypes known for and in the Indian context by over 1000 stereotypes across many unique identities. We also demonstrate the utility and effectiveness of such expanded resources for evaluations of language models.CONTENT WARNING: This paper contains examples of stereotypes that may be offensive.
****
Language Quantized AutoEncoders: Towards Unsupervised Text-Image Alignment
Hao Liu, Wilson Yan, Pieter Abbeel
Recent progress in scaling up large language models has shown impressive capabilities in performing few-shot learning across a wide range of natural language tasks. However, a key limitation is that these language models fundamentally lack grounding to visual perception - a crucial attribute needed to extend to real world tasks such as in visual-question answering and robotics. While prior works have largely connected image to text through pretraining or fine-tuning, learning such alignments are generally costly due to a combination of curating massive datasets and large computational burdens. In order to resolve these limitations, we propose a simple yet effective approach called Language-Quantized AutoEncoder (LQAE), a modification of VQ-VAE that learns to align text-image data in an unsupervised manner by leveraging pretrained language model denoisers (e.g., BERT). Our main idea is to encode images as sequences of text tokens by directly quantizing image embeddings using a pretrained language codebook. We then feed a masked version of the quantized embeddings into a BERT to reconstruct the original input. By doing so, LQAE learns to represent similar images with similar clusters of text tokens, thereby aligning these two modalities without the use of aligned text-image pairs. We show LQAE learns text-aligned image tokens that enable few-shot multi-modal learning with large language models, outperforming baseline methods in tasks such as image classification and VQA while requiring as few as 1-10 image-text pairs.
****
QuIP: 2-Bit Quantization of Large Language Models With Guarantees
Jerry Chee, Yaohui Cai, Volodymyr Kuleshov, Christopher M. De Sa
This work studies post-training parameter quantization in large language models (LLMs). We introduce quantization with incoherence processing (QuIP), a new method based on the insight that quantization benefits from incoherent weight and Hessian matrices, i.e., from the weights being even in magnitude and the directions in which it is important to round them accurately being unaligned with the coordinate axes. QuIP consists of two steps: (1) an adaptive rounding procedure minimizing a quadratic proxy objective; (2) efficient pre- and post-processing that ensures weight and Hessian incoherence via multiplication by random orthogonal matrices. We complement QuIP with the first theoretical analysis for an LLM-scale quantization algorithm, and show that our theory also applies to an existing method, OPTQ. Empirically, we find that our incoherence preprocessing improves several existing quantization algorithms and yields the first LLM quantization methods that produce viable results using only two bits per weight. Our code can be found at https://github.com/Cornell-RelaxML/QuIP.
****
Exploiting Correlated Auxiliary Feedback in Parameterized Bandits
Arun Verma, Zhongxiang Dai, YAO SHU, Bryan Kian Hsiang Low
We study a novel variant of the parameterized bandits problem in which the learn

er can observe additional auxiliary feedback that is correlated with the observe
d reward. The auxiliary feedback is readily available in many real-life applicat
ions, e.g., an online platform that wants to recommend the best-rated services t
o its users can observe the user's rating of service (rewards) and collect addit
ional information like service delivery time (auxiliary feedback). In this paper
, we first develop a method that exploits auxiliary feedback to build a reward e
stimator with tight confidence bounds, leading to a smaller regret. We then char
acterize the regret reduction in terms of the correlation coefficient between re
ward and its auxiliary feedback. Experimental results in different settings also
 verify the performance gain achieved by our proposed method.
****
Multi-modal Queried Object Detection in the Wild
Yifan Xu, Mengdan Zhang, Chaoyou Fu, Peixian Chen, Xiaoshan Yang, Ke Li, Changsh
eng Xu
We introduce MQ-Det, an efficient architecture and pre-training strategy design
to utilize both textual description with open-set generalization and visual exem
plars with rich description granularity as category queries, namely, Multi-modal
 Queried object Detection, for real-world detection with both open-vocabulary ca
tegories and various granularity. MQ-Det incorporates vision queries into existi
ng well-established language-queried-only detectors. A plug-and-play gated class
-scalable perceiver module upon the frozen detector is proposed to augment categ
ory text with class-wise visual information. To address the learning inertia pro
blem brought by the frozen detector, a vision conditioned masked language predic
tion strategy is proposed. MQ-Det's simple yet effective architecture and traini
ng strategy design is compatible with most language-queried object detectors, th
us yielding versatile applications. Experimental results demonstrate that multi-
modal queries largely boost open-world detection. For instance, MQ-Det significa
ntly improves the state-of-the-art open-set detector GLIP by +7.8% AP on the LVI
S benchmark via multi-modal queries without any downstream finetuning, and avera
gely +6.3% AP on 13 few-shot downstream tasks, with merely additional 3% modulat
ing time required by GLIP. Code is available at https://github.com/YifanXu74/MQ-
Det.
****
H-Consistency Bounds: Characterization and Extensions
Anqi Mao, Mehryar Mohri, Yutao Zhong
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Direction-oriented Multi-objective Learning: Simple and Provable Stochastic Algo
rithms
Peiyao Xiao, Hao Ban, Kaiyi Ji
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
DeepfakeBench: A Comprehensive Benchmark of Deepfake Detection
Zhiyuan Yan, Yong Zhang, Xinhang Yuan, Siwei Lyu, Baoyuan Wu
A critical yet frequently overlooked challenge in the field of deepfake detectio
n is the lack of a standardized, unified, comprehensive benchmark. This issue le
ads to unfair performance comparisons and potentially misleading results. Specif
ically, there is a lack of uniformity in data processing pipelines, resulting in
 inconsistent data inputs for detection models. Additionally, there are noticeab
le differences in experimental settings, and evaluation strategies and metrics l
ack standardization. To fill this gap, we present the first comprehensive benchm
ark for deepfake detection, called \textit{DeepfakeBench}, which offers three ke
y contributions: 1) a unified data management system to ensure consistent input
across all detectors, 2) an integrated framework for state-of-the-art methods im

plementation, and 3) standardized evaluation metrics and protocols to promote tr ansparency and reproducibility. Featuring an extensible, modular-based codebase , \textit{DeepfakeBench} contains 15 state-of-the-art detection methods, 9 deepf ake datasets, a series of deepfake detection evaluation protocols and analysis t ools, as well as comprehensive evaluations. Moreover, we provide new insights b ased on extensive analysis of these evaluations from various perspectives (\eg, data augmentations, backbones). We hope that our efforts could facilitate future research and foster innovation in this increasingly critical domain. All codes, evaluations, and analyses of our benchmark are publicly available at \url{https ://github.com/SCLBD/DeepfakeBench}.
****
DreamWaltz: Make a Scene with Complex 3D Animatable Avatars
Yukun Huang, Jianan Wang, Ailing Zeng, He CAO, Xianbiao Qi, Yukai Shi, Zheng-Jun Zha, Lei Zhang
We present DreamWaltz, a novel framework for generating and animating complex 3D avatars given text guidance and parametric human body prior. While recent metho ds have shown encouraging results for text-to-3D generation of common objects, c reating high-quality and animatable 3D avatars remains challenging. To create hi gh-quality 3D avatars, DreamWaltz proposes 3D-consistent occlusion-aware Score D istillation Sampling (SDS) to optimize implicit neural representations with cano nical poses. It provides view-aligned supervision via 3D-aware skeleton conditio ning which enables complex avatar generation without artifacts and multiple face s. For animation, our method learns an animatable 3D avatar representation from abundant image priors of diffusion model conditioned on various poses, which cou ld animate complex non-rigged avatars given arbitrary poses without retraining. Extensive evaluations demonstrate that DreamWaltz is an effective and robust app roach for creating 3D avatars that can take on complex shapes and appearances as well as novel poses for animation. The proposed framework further enables the c reation of complex scenes with diverse compositions, including avatar-avatar, av atar-object and avatar-scene interactions. See https://dreamwaltz3d.github.io/ f or more vivid 3D avatar and animation results.
****
Where2Explore: Few-shot Affordance Learning for Unseen Novel Categories of Artic ulated Objects
Chuanruo Ning, Ruihai Wu, Haoran Lu, Kaichun Mo, Hao Dong
Articulated object manipulation is a fundamental yet challenging task in robotic s. Due to significant geometric and semantic variations across object categories , previous manipulation models struggle to generalize to novel categories. Few-s hot learning is a promising solution for alleviating this issue by allowing robo ts to perform a few interactions with unseen objects. However, extant approaches often necessitate costly and inefficient test-time interactions with each unsee n instance. Recognizing this limitation, we observe that despite their distinct shapes, different categories often share similar local geometries essential for manipulation, such as pullable handles and graspable edges - a factor typically underutilized in previous few-shot learning works. To harness this commonality, we introduce 'Where2Explore', an affordance learning framework that effectively explores novel categories with minimal interactions on a limited number of insta nces. Our framework explicitly estimates the geometric similarity across differe nt categories, identifying local areas that differ from shapes in the training c ategories for efficient exploration while concurrently transferring affordance k nowledge to similar parts of the objects. Extensive experiments in simulated and real-world environments demonstrate our framework's capacity for efficient few-shot exploration and generalization.
****
OpenProteinSet: Training data for structural biology at scale
Gustaf Ahdritz, Nazim Bouatta, Sachin Kadyan, Lukas Jarosch, Dan Berenberg, Ian Fisk, Andrew Watkins, Stephen Ra, Richard Bonneau, Mohammed AlQuraishi
Multiple sequence alignments (MSAs) of proteins encode rich biological informati on and have been workhorses in bioinformatic methods for tasks like protein desi gn and protein structure prediction for decades. Recent breakthroughs like Alpha

Fold2 that use transformers to attend directly over large quantities of raw MSAs have reaffirmed their importance. Generation of MSAs is highly computationally intensive, however, and no datasets comparable to those used to train AlphaFold2 have been made available to the research community, hindering progress in machine learning for proteins. To remedy this problem, we introduce OpenProteinSet, an open-source corpus of more than 16 million MSAs, associated structural homologs from the Protein Data Bank, and AlphaFold2 protein structure predictions. We have previously demonstrated the utility of OpenProteinSet by successfully retraining AlphaFold2 on it. We expect OpenProteinSet to be broadly useful as training and validation data for 1) diverse tasks focused on protein structure, function, and design and 2) large-scale multimodal machine learning research.
****

Counting Distinct Elements in the Turnstile Model with Differential Privacy under Continual Observation
Palak Jain, Iden Kalemaj, Sofya Raskhodnikova, Satchit Sivakumar, Adam Smith
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Demystifying Softmax Gating Function in Gaussian Mixture of Experts
Huy Nguyen, TrungTin Nguyen, Nhat Ho
Understanding the parameter estimation of softmax gating Gaussian mixture of experts has remained a long-standing open problem in the literature. It is mainly due to three fundamental theoretical challenges associated with the softmax gating function: (i) the identifiability only up to the translation of parameters; (ii) the intrinsic interaction via partial differential equations between the softmax gating and the expert functions in the Gaussian density; (iii) the complex dependence between the numerator and denominator of the conditional density of softmax gating Gaussian mixture of experts. We resolve these challenges by proposing novel Voronoi loss functions among parameters and establishing the convergence rates of maximum likelihood estimator (MLE) for solving parameter estimation in these models. When the true number of experts is unknown and over-specified, our findings show a connection between the convergence rate of the MLE and a solvability problem of a system of polynomial equations.
****

Hybrid Policy Optimization from Imperfect Demonstrations
Hanlin Yang, Chao Yu, peng sun, Siji Chen
Exploration is one of the main challenges in Reinforcement Learning (RL), especially in environments with sparse rewards. Learning from Demonstrations (LfD) is a promising approach to solving this problem by leveraging expert demonstrations. However, expert demonstrations of high quality are usually costly or even impossible to collect in real-world applications. In this work, we propose a novel RL algorithm called HYbrid Policy Optimization (HYPO), which uses a small number of imperfect demonstrations to accelerate an agent's online learning process. The key idea is to train an offline guider policy using imitation learning in order to instruct an online agent policy to explore efficiently. Through mutual update of the guider policy and the agent policy, the agent can leverage suboptimal demonstrations for efficient exploration while avoiding the conservative policy caused by imperfect demonstrations. Empirical results show that HYPO significantly outperforms several baselines in various challenging tasks, such as MuJoCo with sparse rewards, Google Research Football, and the AirSim drone simulation.
****

What is Flagged in Uncertainty Quantification? Latent Density Models for Uncertainty Categorization
Hao Sun, Boris van Breugel, Jonathan Crabbé, Nabeel Seedat, Mihaela van der Schaar
Uncertainty quantification (UQ) is essential for creating trustworthy machine learning models. Recent years have seen a steep rise in UQ methods that can flag suspicious examples, however, it is often unclear what exactly these methods iden

tify. In this work, we propose a framework for categorizing uncertain examples flagged by UQ methods. We introduce the confusion density matrix---a kernel-based approximation of the misclassification density---and use this to categorize suspicious examples identified by a given uncertainty method into three classes: out-of-distribution (OOD) examples, boundary (Bnd) examples, and examples in regions of high in-distribution misclassification (IDM). Through extensive experiments, we show that our framework provides a new and distinct perspective for assessing differences between uncertainty quantification methods, thereby forming a valuable assessment benchmark.
****
Datasets and Benchmarks for Nanophotonic Structure and Parametric Design Simulations
Jungtaek Kim, Mingxuan Li, Oliver Hinder, Paul Leu
Nanophotonic structures have versatile applications including solar cells, anti-reflective coatings, electromagnetic interference shielding, optical filters, and light emitting diodes. To design and understand these nanophotonic structures, electrodynamic simulations are essential. These simulations enable us to model electromagnetic fields over time and calculate optical properties. In this work, we introduce frameworks and benchmarks to evaluate nanophotonic structures in the context of parametric structure design problems. The benchmarks are instrumental in assessing the performance of optimization algorithms and identifying an optimal structure based on target optical properties. Moreover, we explore the impact of varying grid sizes in electrodynamic simulations, shedding light on how evaluation fidelity can be strategically leveraged in enhancing structure designs.
****
Efficient Data Subset Selection to Generalize Training Across Models: Transductive and Inductive Networks
Eeshaan Jain, Tushar Nandy, Gaurav Aggarwal, Ashish Tendulkar, Rishabh Iyer, Abir De
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
NIS3D: A Completely Annotated Benchmark for Dense 3D Nuclei Image Segmentation
Wei Zheng, Cheng Peng, Zeyuan Hou, Boyu Lyu, Mengfan Wang, Xuelong Mi, Shuoxuan Qiao, Yinan Wan, Guoqiang Yu
3D segmentation of nuclei images is a fundamental task for many biological studies. Despite the rapid advances of large-volume 3D imaging acquisition methods and the emergence of sophisticated algorithms to segment the nuclei in recent years, a benchmark with all cells completely annotated is still missing, making it hard to accurately assess and further improve the performance of the algorithms. The existing nuclei segmentation benchmarks either worked on 2D only or annotated a small number of 3D cells, perhaps due to the high cost of 3D annotation for large-scale data. To fulfill the critical need, we constructed NIS3D, a 3D, high cell density, large-volume, and completely annotated Nuclei Image Segmentation benchmark, assisted by our newly designed semi-automatic annotation software. NIS3D provides more than 22,000 cells across multiple most-used species in this area. Each cell is labeled by three independent annotators, so we can measure the variability of each annotation. A confidence score is computed for each cell, allowing more nuanced testing and performance comparison. A comprehensive review on the methods of segmenting 3D dense nuclei was conducted. The benchmark was used to evaluate the performance of several selected state-of-the-art segmentation algorithms. The best of current methods is still far away from human-level accuracy, corroborating the necessity of generating such a benchmark. The testing results also demonstrated the strength and weakness of each method and pointed out the directions of further methodological development. The dataset can be downloaded here: https://github.com/yu-lab-vt/NIS3D.
****

HiBug: On Human-Interpretable Model Debug
Muxi Chen, YU LI, Qiang Xu
Machine learning models can frequently produce systematic errors on critical subsets (or slices) of data that share common attributes. Discovering and explaining such model bugs is crucial for reliable model deployment. However, existing bug discovery and interpretation methods usually involve heavy human intervention and annotation, which can be cumbersome and have low bug coverage.In this paper, we propose HiBug, an automated framework for interpretable model debugging. Our approach utilizes large pre-trained models, such as chatGPT, to suggest human-understandable attributes that are related to the targeted computer vision tasks. By leveraging pre-trained vision-language models, we can efficiently identify common visual attributes of underperforming data slices using human-understandable terms. This enables us to uncover rare cases in the training data, identify spurious correlations in the model, and use the interpretable debug results to select or generate new training data for model improvement. Experimental results demonstrate the efficacy of the HiBug framework.
****
A Theoretical Analysis of the Test Error of Finite-Rank Kernel Ridge Regression
Tin Sum Cheng, Aurelien Lucchi, Anastasis Kratsios, Ivan Dokmani■, David Belius
Existing statistical learning guarantees for general kernel regressors often yield loose bounds when used with finite-rank kernels. Yet, finite-rank kernels naturally appear in a number of machine learning problems, e.g. when fine-tuning a pre-trained deep neural network's last layer to adapt it to a novel task when performing transfer learning. We address this gap for finite-rank kernel ridge regression (KRR) by deriving sharp non-asymptotic upper and lower bounds for the KRR test error of any finite-rank KRR. Our bounds are tighter than previously derived bounds on finite-rank KRR and, unlike comparable results, they also remain valid for any regularization parameters.
****
Learning Invariant Representations with a Nonparametric Nadaraya-Watson Head
Alan Wang, Minh Nguyen, Mert Sabuncu
Machine learning models will often fail when deployed in an environment with a data distribution that is different than the training distribution. When multiple environments are available during training, many methods exist that learn representations which are invariant across the different distributions, with the hope that these representations will be transportable to unseen domains. In this work, we present a nonparametric strategy for learning invariant representations based on the recently-proposed Nadaraya-Watson (NW) head. The NW head makes a prediction by comparing the learned representations of the query to the elements of a support set that consists of labeled data. We demonstrate that by manipulating the support set, one can encode different causal assumptions. In particular, restricting the support set to a single environment encourages the model to learn invariant features that do not depend on the environment. We present a causally-motivated setup for our modeling and training strategy and validate on three challenging real-world domain generalization tasks in computer vision.
****
Conformalized matrix completion
Yu Gui, Rina Barber, Cong Ma
Matrix completion aims to estimate missing entries in a data matrix, using the assumption of a low-complexity structure (e.g., low-rankness) so that imputation is possible. While many effective estimation algorithms exist in the literature, uncertainty quantification for this problem has proved to be challenging, and existing methods are extremely sensitive to model misspecification. In this work, we propose a distribution-free method for predictive inference in the matrix completion problem. Our method adapts the framework of conformal prediction, which provides prediction intervals with guaranteed distribution-free validity in the setting of regression, to the problem of matrix completion. Our resulting method, conformalized matrix completion (cmc), offers provable predictive coverage regardless of the accuracy of the low-rank model. Empirical results on simulated and real data demonstrate that cmc is robust to model misspecification while matc

hing the performance of existing model-based methods when the model is correct.
****

Mixture Weight Estimation and Model Prediction in Multi-source Multi-target Domain Adaptation

Yuyang Deng, Ilja Kuzborskij, Mehrdad Mahdavi

We consider a problem of learning a model from multiple sources with the goal to performwell on a new target distribution. Such problem arises inlearning with data collected from multiple sources (e.g. crowdsourcing) orlearning in distributed systems, where the data can be highly heterogeneous. Thegoal of learner is to mix these data sources in a target-distribution aware way andsimultaneously minimize the empirical risk on the mixed source. The literature has made some tangible advancements in establishingtheory of learning on mixture domain. However, there are still two unsolved problems. Firstly, how to estimate the optimal mixture of sources, given a target domain; Secondly, when there are numerous target domains, we have to solve empirical risk minimization for each target on possibly unique mixed source data , which is computationally expensive. In this paper we address both problems efficiently and with guarantees.We cast the first problem, mixture weight estimation as convex-nonconcave compositional minimax, and propose an efficient stochasticalgorithm with provable stationarity guarantees.Next, for the second problem, we identify that for certain regime,solving ERM for each target domain individually can be avoided, and instead parameters for a target optimalmodel can be viewed as a non-linear function ona space of the mixture coefficients.To this end, we show that in offline setting, a GD-trained overparameterized neural network can provably learn such function.Finally, we also consider an online setting and propose an label efficient online algorithm, which predicts parameters for new models given arbitrary sequence of mixing coefficients, while enjoying optimal regret.
****

CELLE-2: Translating Proteins to Pictures and Back with a Bidirectional Text-to-Image Transformer

Emaad Khwaja, Yun Song, Aaron Agarunov, Bo Huang

We present CELL-E 2, a novel bidirectional transformer that can generate images depicting protein subcellular localization from the amino acid sequences (and vice versa). Protein localization is a challenging problem that requires integrating sequence and image information, which most existing methods ignore. CELL-E 2 extends the work of CELL-E, not only capturing the spatial complexity of protein localization and produce probability estimates of localization atop a nucleus image, but also being able to generate sequences from images, enabling de novo protein design. We train and finetune CELL-E 2 on two large-scale datasets of human proteins. We also demonstrate how to use CELL-E 2 to create hundreds of novel nuclear localization signals (NLS). Results and interactive demos are featured at https://bohuanglab.github.io/CELL-E_2/.
****

HeadSculpt: Crafting 3D Head Avatars with Text

Xiao Han, Yukang Cao, Kai Han, Xiatian Zhu, Jiankang Deng, Yi-Zhe Song, Tao Xiang, Kwan-Yee K. Wong

Recently, text-guided 3D generative methods have made remarkable advancements in producing high-quality textures and geometry, capitalizing on the proliferation of large vision-language and image diffusion models. However, existing methods still struggle to create high-fidelity 3D head avatars in two aspects: (1) They rely mostly on a pre-trained text-to-image diffusion model whilst missing the necessary 3D awareness and head priors. This makes them prone to inconsistency and geometric distortions in the generated avatars. (2) They fall short in fine-grained editing. This is primarily due to the inherited limitations from the pre-trained 2D image diffusion models, which become more pronounced when it comes to 3D head avatars. In this work, we address these challenges by introducing a versatile coarse-to-fine pipeline dubbed HeadSculpt for crafting (i.e., generating and editing) 3D head avatars from textual prompts. Specifically, we first equip the diffusion model with 3D awareness by leveraging landmark-based control and a learned textual embedding representing the back view appearance of heads, enablin

g 3D-consistent head avatar generations. We further propose a novel identity-aware editing score distillation strategy to optimize a textured mesh with a high-resolution differentiable rendering technique. This enables identity preservation while following the editing instruction.We showcase HeadSculpt's superior fidelity and editing capabilities through comprehensive experiments and comparisons with existing methods.
****
CBD: A Certified Backdoor Detector Based on Local Dominant Probability
Zhen Xiang, Zidi Xiong, Bo Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SheetCopilot: Bringing Software Productivity to the Next Level through Large Language Models
Hongxin Li, Jingran Su, Yuntao Chen, Qing Li, ZHAO-XIANG ZHANG
Computer end users have spent billions of hours completing daily tasks like tabular data processing and project timeline scheduling. Most of these tasks are repetitive and error-prone, yet most end users lack the skill to automate these burdensome works. With the advent of large language models (LLMs), directing software with natural language user requests become a reachable goal. In this work, we propose a SheetCopilot agent that takes natural language task and control spreadsheet to fulfill the requirements. We propose a set of atomic actions as an abstraction of spreadsheet software functionalities. We further design a state machine-based task planning framework for LLMs to robustly interact with spreadsheets. We curate a representative dataset containing 221 spreadsheet control tasks and establish a fully automated evaluation pipeline for rigorously benchmarking the ability of LLMs in software control tasks. Our SheetCopilot correctly completes 44.3\% of tasks for a single generation, outperforming the strong code generation baseline by a wide margin. Our project page: https://sheetcopilot.github.io/.
****
Beyond Uniform Sampling: Offline Reinforcement Learning with Imbalanced Datasets
Zhang-Wei Hong, Aviral Kumar, Sathwik Karnik, Abhishek Bhandwaldar, Akash Srivastava, Joni Pajarinen, Romain Laroche, Abhishek Gupta, Pulkit Agrawal
Offline reinforcement learning (RL) enables learning a decision-making policy without interaction with the environment. This makes it particularly beneficial in situations where such interactions are costly. However, a known challenge for offline RL algorithms is the distributional mismatch between the state-action distributions of the learned policy and the dataset, which can significantly impact performance. State-of-the-art algorithms address it by constraining the policy to align with the state-action pairs in the dataset. However, this strategy struggles on datasets that predominantly consist of trajectories collected by low-performing policies and only a few trajectories from high-performing ones. Indeed, the constraint to align with the data leads the policy to imitate low-performing behaviors predominating the dataset. Our key insight to address this issue is to constrain the policy to the policy that collected the good parts of the dataset rather than all data. To this end, we optimize the importance sampling weights to emulate sampling data from a data distribution generated by a nearly optimal policy. Our method exhibits considerable performance gains (up to five times better) over the existing approaches in state-of-the-art offline RL algorithms over 72 imbalanced datasets with varying types of imbalance.
****
Variational Weighting for Kernel Density Ratios
Sangwoong Yoon, Frank Park, Gunsu YUN, Iljung Kim, Yung-Kyun Noh
Kernel density estimation (KDE) is integral to a range of generative and discriminative tasks in machine learning. Drawing upon tools from the multidimensional calculus of variations, we derive an optimal weight function that reduces bias in standard kernel density estimates for density ratios, leading to improved esti

mates of prediction posteriors and information-theoretic measures. In the proces
s, we shed light on some fundamental aspects of density estimation, particularly
 from the perspective of algorithms that employ KDEs as their main building bloc
ks.
****
Adversarial Examples Exist in Two-Layer ReLU Networks for Low Dimensional Linear
 Subspaces
Odelia Melamed, Gilad Yehudai, Gal Vardi
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Complexity of Derivative-Free Policy Optimization for Structured $\mathcal{H}_\i
nfty$ Control
Xingang Guo, Darioush Keivan, Geir Dullerud, Peter Seiler, Bin Hu
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Meet in the Middle: A New Pre-training Paradigm
Anh Nguyen, Nikos Karampatziakis, Weizhu Chen
Most language models (LMs) are trained and applied in an autoregressive left-to-
right fashion, predicting the next token from the preceding ones. However, this
ignores that the full sequence is available during training. In this paper, we i
ntroduce ``Meet in the Middle'' (MIM) a new pre-training paradigm that improves
data efficiency by training in two directions, left-to-right and right-to-left,
and encouraging the respective modelsto agree on their token distribution for ea
ch position. While the primary outcome is an improved left-to-right LM,we also o
btain secondary benefits in the infilling task. There, we leverage the two pre-t
rained directions to propose an infilling procedure that builds the completion s
imultaneously from both sides. We conduct extensive experiments on both programm
ing and natural languages and show that MIM significantly surpasses existing pre
-training paradigms, in both left-to-right generation as well as infilling.Code
and models available at https://github.com/microsoft/Meet-in-the-Middle
****
Score-based Source Separation with Applications to Digital Communication Signals
Tejas Jayashankar, Gary C.F. Lee, Alejandro Lancho, Amir Weiss, Yury Polyanskiy,
 Gregory Wornell
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Fair Streaming Principal Component Analysis: Statistical and Algorithmic Viewpoi
nt
Junghyun Lee, Hanseul Cho, Se-Young Yun, Chulhee Yun
Fair Principal Component Analysis (PCA) is a problem setting where we aim to per
form PCA while making the resulting representation fair in that the projected di
stributions, conditional on the sensitive attributes, match one another. However
, existing approaches to fair PCA have two main problems: theoretically, there h
as been no statistical foundation of fair PCA in terms of learnability; practica
lly, limited memory prevents us from using existing approaches, as they explicit
ly rely on full access to the entire data. On the theoretical side, we rigorousl
y formulate fair PCA using a new notion called probably approximately fair and o
ptimal (PAFO) learnability. On the practical side, motivated by recent advances
in streaming algorithms for addressing memory limitation, we propose a new setti
ng called fair streaming PCA along with a memory-efficient algorithm, fair noisy
 power method (FNPM). We then provide its statistical guarantee in terms of PAFO

-learnability, which is the first of its kind in fair PCA literature. We verify our algorithm in the CelebA dataset without any pre-processing; while the existing approaches are inapplicable due to memory limitations, by turning it into a streaming setting, we show that our algorithm performs fair PCA efficiently and effectively.
****

DDCoT: Duty-Distinct Chain-of-Thought Prompting for Multimodal Reasoning in Language Models

Ge Zheng, Bin Yang, Jiajin Tang, Hong-Yu Zhou, Sibei Yang

A long-standing goal of AI systems is to perform complex multimodal reasoning like humans. Recently, large language models (LLMs) have made remarkable strides in such multi-step reasoning on the language modality solely by leveraging the chain of thought (CoT) to mimic human thinking. However, the transfer of these advancements to multimodal contexts introduces heightened challenges, including but not limited to the impractical need for labor-intensive annotation and the limitations in terms of flexibility, generalizability, and explainability. To evoke CoT reasoning in multimodality, this work first conducts an in-depth analysis of these challenges posed by multimodality and presents two key insights: “keeping critical thinking” and “letting everyone do their jobs” in multimodal CoT reasoning. Furthermore, this study proposes a novel DDCoT prompting that maintains a critical attitude through negative-space prompting and incorporates multimodality into reasoning by first dividing the reasoning responsibility of LLMs into reasoning and recognition and then integrating the visual recognition capability of visual models into the joint reasoning process. The rationales generated by DDCoT not only improve the reasoning abilities of both large and small language models in zero-shot prompting and fine-tuning learning, significantly outperforming state-of-the-art methods but also exhibit impressive generalizability and explainability.
****

Adversarially Robust Learning with Uncertain Perturbation Sets

Tosca Lechner, Vinayak Pathak, Ruth Urner

In many real-world settings exact perturbation sets to be used by an adversary are not plausibly available to a learner. While prior literature has studied both scenarios with completely known and completely unknown perturbation sets, we propose an in-between setting of learning with respect to a class of perturbation sets. We show that in this setting we can improve on previous results with completely unknown perturbation sets, while still addressing the concerns of not having perfect knowledge of these sets in real life. In particular, we give the first positive results for the learnability of infinite Littlestone classes when having access to a perfect-attack oracle. We also consider a setting of learning with abstention, where predictions are considered robustness violations, only when the wrong prediction is made within the perturbation set. We show there are classes for which perturbation-set unaware learning without query access is possible, but abstention is required.
****

Common Ground in Cooperative Communication

Xiaoran Hao, Yash Jhaveri, Patrick Shafto

Cooperative communication plays a fundamental role in theories of human-human interaction--cognition, culture, development, language, etc.--as well as human-robot interaction. The core challenge in cooperative communication is the problem of common ground: having enough shared knowledge and understanding to successfully communicate. Prior models of cooperative communication, however, uniformly assume the strongest form of common ground, perfect and complete knowledge sharing, and, therefore, fail to capture the core challenge of cooperative communication. We propose a general theory of cooperative communication that is mathematically principled and explicitly defines a spectrum of common ground possibilities, going well beyond that of perfect and complete knowledge sharing, on spaces that permit arbitrary representations of data and hypotheses. Our framework is a strict generalization of prior models of cooperative communication. After considering a parametric form of common ground and viewing the data selection and hypothes

is inference processes of communication as encoding and decoding, we establish a connection to variational autoencoding, a powerful model in modern machine learning. Finally, we carry out a series of empirical simulations to support and elaborate on our theoretical results.
****

Keep Various Trajectories: Promoting Exploration of Ensemble Policies in Continuous Control

Chao Li, Chen GONG, Qiang He, Xinwen Hou

The combination of deep reinforcement learning (DRL) with ensemble methods has been proved to be highly effective in addressing complex sequential decision-making problems. This success can be primarily attributed to the utilization of multiple models, which enhances both the robustness of the policy and the accuracy of value function estimation. However, there has been limited analysis of the empirical success of current ensemble RL methods thus far. Our new analysis reveals that the sample efficiency of previous ensemble DRL algorithms may be limited by sub-policies that are not as diverse as they could be. Motivated by these findings, our study introduces a new ensemble RL algorithm, termed \textbf{T}rajectories-awar\textbf{E} \textbf{E}nsemble exploratio\textbf{N} (TEEN). The primary goal of TEEN is to maximize the expected return while promoting more diverse trajectories. Through extensive experiments, we demonstrate that TEEN not only enhances the sample diversity of the ensemble policy compared to using sub-policies alone but also improves the performance over ensemble RL algorithms. On average, TEEN outperforms the baseline ensemble DRL algorithms by 41\% in performance on the tested representative environments.
****

ReSync: Riemannian Subgradient-based Robust Rotation Synchronization

Huikang Liu, Xiao Li, Anthony Man-Cho So

This work presents ReSync, a Riemannian subgradient-based algorithm for solving the robust rotation synchronization problem, which arises in various engineering applications. ReSync solves a least-unsquared minimization formulation over the rotation group, which is nonsmooth and nonconvex, and aims at recovering the underlying rotations directly. We provide strong theoretical guarantees for ReSync under the random corruption setting. Specifically, we first show that the initialization procedure of ReSync yields a proper initial point that lies in a local region around the ground-truth rotations. We next establish the weak sharpness property of the aforementioned formulation and then utilize this property to derive the local linear convergence of ReSync to the ground-truth rotations. By combining these guarantees, we conclude that ReSync converges linearly to the ground-truth rotations under appropriate conditions. Experiment results demonstrate the effectiveness of ReSync.
****

On the Exploration of Local Significant Differences For Two-Sample Test

Zhijian Zhou, Jie Ni, Jia-He Yao, Wei Gao

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Fine-Grained Cross-View Geo-Localization Using a Correlation-Aware Homography Estimator

Xiaolong Wang, Runsen Xu, Zhuofan Cui, Zeyu Wan, Yu Zhang

In this paper, we introduce a novel approach to fine-grained cross-view geo-localization. Our method aligns a warped ground image with a corresponding GPS-tagged satellite image covering the same area using homography estimation. We first employ a differentiable spherical transform, adhering to geometric principles, to accurately align the perspective of the ground image with the satellite map. This transformation effectively places ground and aerial images in the same view and on the same plane, reducing the task to an image alignment problem. To address challenges such as occlusion, small overlapping range, and seasonal variations, we propose a robust correlation-aware homography estimator to align similar pa

rts of the transformed ground image with the satellite image. Our method achieves sub-pixel resolution and meter-level GPS accuracy by mapping the center point of the transformed ground image to the satellite image using a homography matrix and determining the orientation of the ground camera using a point above the central axis. Operating at a speed of 30 FPS, our method outperforms state-of-the-art techniques, reducing the mean metric localization error by 21.3\% and 32.4\% in same-area and cross-area generalization tasks on the VIGOR benchmark, respectively, and by 34.4\% on the KITTI benchmark in same-area evaluation.
****

DataPerf: Benchmarks for Data-Centric AI Development
Mark Mazumder, Colby Banbury, Xiaozhe Yao, Bojan Karlaš, William Gaviria Rojas, Sudnya Diamos, Greg Diamos, Lynn He, Alicia Parrish, Hannah Rose Kirk, Jessica Quaye, Charvi Rastogi, Douwe Kiela, David Jurado, David Kanter, Rafael Mosquera, Will Cukierski, Juan Ciro, Lora Aroyo, Bilge Acun, Lingjiao Chen, Mehul Raje, Max Bartolo, Evan Sabri Eyuboglu, Amirata Ghorbani, Emmett Goodman, Addison Howard, Oana Inel, Tariq Kane, Christine R. Kirkpatrick, D. Sculley, Tzu-Sheng Kuo, Jonas W. Mueller, Tristan Thrush, Joaquin Vanschoren, Margaret Warren, Adina Williams, Serena Yeung, Newsha Ardalani, Praveen Paritosh, Ce Zhang, James Y. Zou, Carole-Jean Wu, Cody Coleman, Andrew Ng, Peter Mattson, Vijay Janapa Reddi
Machine learning research has long focused on models rather than datasets, and prominent datasets are used for common ML tasks without regard to the breadth, difficulty, and faithfulness of the underlying problems. Neglecting the fundamental importance of data has given rise to inaccuracy, bias, and fragility in real-world applications, and research is hindered by saturation across existing dataset benchmarks. In response, we present DataPerf, a community-led benchmark suite for evaluating ML datasets and data-centric algorithms. We aim to foster innovation in data-centric AI through competition, comparability, and reproducibility. We enable the ML community to iterate on datasets, instead of just architectures, and we provide an open, online platform with multiple rounds of challenges to support this iterative development. The first iteration of DataPerf contains five benchmarks covering a wide spectrum of data-centric techniques, tasks, and modalities in vision, speech, acquisition, debugging, and diffusion prompting, and we support hosting new contributed benchmarks from the community. The benchmarks, online evaluation platform, and baseline implementations are open source, and the MLCommons Association will maintain DataPerf to ensure long-term benefits to academia and industry.
****

Non-Smooth Weakly-Convex Finite-sum Coupled Compositional Optimization
Quanqi Hu, Dixian Zhu, Tianbao Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Optimal Transport for Treatment Effect Estimation
Hao Wang, Jiajun Fan, Zhichao Chen, Haoxuan Li, Weiming Liu, Tianqiao Liu, Quanyu Dai, Yichao Wang, Zhenhua Dong, Ruiming Tang
Estimating individual treatment effects from observational data is challenging due to treatment selection bias. Prevalent methods mainly mitigate this issue by aligning different treatment groups in the latent space, the core of which is the calculation of distribution discrepancy. However, two issues that are often overlooked can render these methods invalid:(1) mini-batch sampling effects (MSE), where the calculated discrepancy is erroneous in non-ideal mini-batches with outcome imbalance and outliers;(2) unobserved confounder effects (UCE), where the unobserved confounders are not considered in the discrepancy calculation.Both of these issues invalidate the calculated discrepancy, mislead the training of estimators, and thus impede the handling of treatment selection bias.To tackle these issues, we propose Entire Space CounterFactual Regression (ESCFR), which is a new take on optimal transport technology in the context of causality.Specifically, based on the canonical optimal transport framework, we propose a relaxed mass

-preserving regularizer to address the MSE issue and design a proximal factual outcome regularizer to handle the UCE issue.Extensive experiments demonstrate that ESCFR estimates distribution discrepancy accurately, handles the treatment selection bias effectively, and outperforms prevalent competitors significantly.
****

Initialization Matters: Privacy-Utility Analysis of Overparameterized Neural Networks

Jiayuan Ye, Zhenyu Zhu, Fanghui Liu, Reza Shokri, Volkan Cevher

We analytically investigate how over-parameterization of models in randomized machine learning algorithms impacts the information leakage about their training data. Specifically, we prove a privacy bound for the KL divergence between model distributions on worst-case neighboring datasets, and explore its dependence on the initialization, width, and depth of fully connected neural networks. We find that this KL privacy bound is largely determined by the expected squared gradient norm relative to model parameters during training. Notably, for the special setting of linearized network, our analysis indicates that the squared gradient norm (and therefore the escalation of privacy loss) is tied directly to the per-layer variance of the initialization distribution. By using this analysis, we demonstrate that privacy bound improves with increasing depth under certain initializations (LeCun and Xavier), while degrades with increasing depth under other initializations (He and NTK). Our work reveals a complex interplay between privacy and depth that depends on the chosen initialization distribution. We further prove excess empirical risk bounds under a fixed KL privacy budget, and show that the interplay between privacy utility trade-off and depth is similarly affected by the initialization.
****

Cause-Effect Inference in Location-Scale Noise Models: Maximum Likelihood vs. Independence Testing

Xiangyu Sun, Oliver Schulte

A fundamental problem of causal discovery is cause-effect inference, to learn the correct causal direction between two random variables. Significant progress has been made through modelling the effect as a function of its cause and a noise term, which allows us to leverage assumptions about the generating function class. The recently introduced heteroscedastic location-scale noise functional models (LSNMs) combine expressive power with identifiability guarantees. LSNM model selection based on maximizing likelihood achieves state-of-the-art accuracy, when the noise distributions are correctly specified. However, through an extensive empirical evaluation, we demonstrate that the accuracy deteriorates sharply when the form of the noise distribution is misspecified by the user. Our analysis shows that the failure occurs mainly when the conditional variance in the anti-causal direction is smaller than that in the causal direction. As an alternative, we find that causal model selection through residual independence testing is much more robust to noise misspecification and misleading conditional variance.
****

M3Exam: A Multilingual, Multimodal, Multilevel Benchmark for Examining Large Language Models

Wenxuan Zhang, Mahani Aljunied, Chang Gao, Yew Ken Chia, Lidong Bing

Despite the existence of various benchmarks for evaluating natural language processing models, we argue that human exams are a more suitable means of evaluating general intelligence for large language models (LLMs), as they inherently demand a much wider range of abilities such as language understanding, domain knowledge, and problem-solving skills. To this end, we introduce M3Exam, a novel benchmark sourced from real and official human exam questions for evaluating LLMs in a multilingual, multimodal, and multilevel context. M3Exam exhibits three unique characteristics: (1) multilingualism, encompassing questions from multiple countries that require strong multilingual proficiency and cultural knowledge; (2) multimodality, accounting for the multimodal nature of many exam questions to test the model's multimodal understanding capability; and (3) multilevel structure, featuring exams from three critical educational periods to comprehensively assess a model's proficiency at different levels. In total, M3Exam contains 12,317 qu

estions in 9 diverse languages with three educational levels, where about 23\% of the questions require processing images for successful solving. We assess the performance of top-performing LLMs on M3Exam and find that current models, including GPT-4, still struggle with multilingual text, particularly in low-resource and non-Latin script languages. Multimodal LLMs also perform poorly with complex multimodal questions. We believe that M3Exam can be a valuable resource for comprehensively evaluating LLMs by examining their multilingual and multimodal abilities and tracking their development. Data and evaluation code is available at \url{https://github.com/DAMO-NLP-SG/M3Exam}.
****
CROMA: Remote Sensing Representations with Contrastive Radar-Optical Masked Autoencoders
Anthony Fuller, Koreen Millard, James Green
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
OpenAGI: When LLM Meets Domain Experts
Yingqiang Ge, Wenyue Hua, Kai Mei, jianchao ji, Juntao Tan, Shuyuan Xu, Zelong Li, Yongfeng Zhang
Human Intelligence (HI) excels at combining basic skills to solve complex tasks. This capability is vital for Artificial Intelligence (AI) and should be embedded in comprehensive AI Agents, enabling them to harness expert models for complex task-solving towards Artificial General Intelligence (AGI). Large Language Models (LLMs) show promising learning and reasoning abilities, and can effectively use external models, tools, plugins, or APIs to tackle complex problems. In this work, we introduce OpenAGI, an open-source AGI research and development platform designed for solving multi-step, real-world tasks. Specifically, OpenAGI uses a dual strategy, integrating standard benchmark tasks for benchmarking and evaluation, and open-ended tasks including more expandable models, tools, plugins, or APIs for creative problem-solving. Tasks are presented as natural language queries to the LLM, which then selects and executes appropriate models. We also propose a Reinforcement Learning from Task Feedback (RLTF) mechanism that uses task results to improve the LLM's task-solving ability, which creates a self-improving AI feedback loop. While we acknowledge that AGI is a broad and multifaceted research challenge with no singularly defined solution path, the integration of LLMs with domain-specific expert models, inspired by mirroring the blend of general and specialized intelligence in humans, offers a promising approach towards AGI. We are open-sourcing the OpenAGI project's code, dataset, benchmarks, evaluation methods, and the UI demo to foster community involvement in AGI advancement: https://github.com/agiresearch/OpenAGI.
****
Neural Frailty Machine: Beyond proportional hazard assumption in neural survival regressions
Ruofan Wu, Jiawei Qiao, Mingzhe Wu, Wen Yu, Ming Zheng, Tengfei LIU, Tianyi Zhang, Weiqiang Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Non-autoregressive Machine Translation with Probabilistic Context-free Grammar
Shangtong Gui, Chenze Shao, Zhengrui Ma, xishan zhang, Yunji Chen, Yang Feng
Non-autoregressive Transformer(NAT) significantly accelerates the inference of neural machine translation. However, conventional NAT models suffer from limited expression power and performance degradation compared to autoregressive (AT) models due to the assumption of conditional independence among target tokens. To address these limitations, we propose a novel approach called PCFG-NAT, which leverages a specially designed Probabilistic Context-Free Grammar (PCFG) to enhance

the ability of NAT models to capture complex dependencies among output tokens. Experimental results on major machine translation benchmarks demonstrate that PCFG-NAT further narrows the gap in translation quality between NAT and AT models. Moreover, PCFG-NAT facilitates a deeper understanding of the generated sentences, addressing the lack of satisfactory explainability in neural machine translation. Code is publicly available at https://github.com/ictnlp/PCFG-NAT.
****
Constrained Policy Optimization with Explicit Behavior Density For Offline Reinforcement Learning
Jing Zhang, Chi Zhang, Wenjia Wang, Bingyi Jing
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Large Language Models are Fixated by Red Herrings: Exploring Creative Problem Solving and Einstellung Effect using the Only Connect Wall Dataset
Saeid Alavi Naeini, Raeid Saqur, Mozhgan Saeidi, John Giorgi, Babak Taati
The quest for human imitative AI has been an enduring topic in AI research since inception. The technical evolution and emerging capabilities of the latest cohort of large language models (LLMs) have reinvigorated the subject beyond academia to cultural zeitgeist. While recent NLP evaluation benchmark tasks test some aspects of human-imitative behaviour (e.g., BIG-bench's `human-like behavior' tasks), few, if not none, examine creative problem solving abilities. Creative problem solving in humans is a well-studied topic in cognitive neuroscience with standardized tests that predominantly use ability to associate (heterogeneous) connections among clue words as a metric for creativity. Exposure to misleading stimuli --- distractors dubbed red herrings --- impede human performance in such tasks via the fixation effect and Einstellung paradigm. In cognitive neuroscience studies, such fixations are experimentally induced by pre-exposing participants to orthographically similar incorrect words to subsequent word-fragments or clues. The popular British quiz show Only Connect's Connecting Wall segment essentially mimics Mednick's Remote Associates Test (RAT) formulation with built-in, deliberate red herrings, that makes it an ideal proxy dataset to explore and study fixation effect and Einstellung paradigm from cognitive neuroscience in LLMs. In addition to presenting the novel Only Connect Wall (OCW) dataset, we also report results from our evaluation of selected pre-trained language models and LLMs (including OpenAI's GPT series) on creative problem solving tasks like grouping clue words by heterogeneous connections, and identifying correct open knowledge domain connections in respective groups. We synthetically generate two additional datasets: OCW-Randomized, OCW-WordNet to further analyze our red-herrings hypothesis in language models.The code and link to the dataset is available at url.
****
Formalizing locality for normative synaptic plasticity models
Colin Bredenberg, Ezekiel Williams, Cristina Savin, Blake Richards, Guillaume Lajoie
In recent years, many researchers have proposed new models for synaptic plasticity in the brain based on principles of machine learning. The central motivation has been the development of learning algorithms that are able to learn difficult tasks while qualifying as "biologically plausible". However, the concept of a biologically plausible learning algorithm is only heuristically defined as an algorithm that is potentially implementable by biological neural networks. Further, claims that neural circuits could implement any given algorithm typically rest on an amorphous concept of "locality" (both in space and time). As a result, it is unclear what many proposed local learning algorithms actually predict biologically, and which of these are consequently good candidates for experimental investigation. Here, we address this lack of clarity by proposing formal and operational definitions of locality. Specifically, we define different classes of locality, each of which makes clear what quantities cannot be included in a learning rule if an algorithm is to qualify as local with respect to a given (biological)

constraint. We subsequently use this framework to distill testable predictions
from various classes of biologically plausible synaptic plasticity models that a
re robust to arbitrary choices about neural network architecture. Therefore, our
framework can be used to guide claims of biological plausibility and to identif
y potential means of experimentally falsifying a proposed learning algorithm for
the brain.
****
Exact Verification of ReLU Neural Control Barrier Functions
Hongchao Zhang, Junlin Wu, Yevgeniy Vorobeychik, Andrew Clark
Control Barrier Functions (CBFs) are a popular approach for safe control of nonl
inear systems. In CBF-based control, the desired safety properties of the system
are mapped to nonnegativity of a CBF, and the control input is chosen to ensure
that the CBF remains nonnegative for all time. Recently, machine learning metho
ds that represent CBFs as neural networks (neural control barrier functions, or
NCBFs) have shown great promise due to the universal representability of neural
networks. However, verifying that a learned CBF guarantees safety remains a chal
lenging research problem. This paper presents novel exact conditions and algorit
hms for verifying safety of feedforward NCBFs with ReLU activation functions. Th
e key challenge in doing so is that, due to the piecewise linearity of the ReLU
function, the NCBF will be nondifferentiable at certain points, thus invalidatin
g traditional safety verification methods that assume a smooth barrier function.
We resolve this issue by leveraging a generalization of Nagumo's theorem for pr
oving invariance of sets with nonsmooth boundaries to derive necessary and suffi
cient conditions for safety. Based on this condition, we propose an algorithm fo
r safety verification of NCBFs that first decomposes the NCBF into piecewise lin
ear segments and then solves a nonlinear program to verify safety of each segmen
t as well as the intersections of the linear segments. We mitigate the complexi
ty by only considering the boundary of the safe region and by pruning the segme
nts with Interval Bound Propagation (IBP) and linear relaxation. We evaluate our
approach through numerical studies with comparison to state-of-the-art SMT-base
d methods. Our code is available at https://github.com/HongchaoZhang-HZ/exactver
if-reluncbf-nips23.
****
Normalization-Equivariant Neural Networks with Application to Image Denoising
Sébastien Herbreteau, Emmanuel Moebel, Charles Kervrann
In many information processing systems, it may be desirable to ensure that any c
hange of the input, whether by shifting or scaling, results in a corresponding
change in the system response. While deep neural networks are gradually replaci
ng all traditional automatic processing methods, they surprisingly do not guaran
tee such normalization-equivariance (scale + shift) property, which can be detri
mental in many applications. To address this issue, we propose a methodology for
adapting existing neural networks so that normalization-equivariance holds by d
esign. Our main claim is that not only ordinary convolutional layers, but also a
ll activation functions, including the ReLU (rectified linear unit), which are a
pplied element-wise to the pre-activated neurons, should be completely removed f
rom neural networks and replaced by better conditioned alternatives. To this end
, we introduce affine-constrained convolutions and channel-wise sort pooling lay
ers as surrogates and show that these two architectural modifications do preserv
e normalization-equivariance without loss of performance. Experimental results i
n image denoising show that normalization-equivariant neural networks, in additi
on to their better conditioning, also provide much better generalization across
noise levels.
****
Budgeting Counterfactual for Offline RL
Yao Liu, Pratik Chaudhari, Rasool Fakoor
Requests for name changes in the electronic proceedings will be accepted with no
questions asked. However name changes may cause bibliographic tracking issues.
Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****

Federated Conditional Stochastic Optimization
Xidong Wu, Jianhui Sun, Zhengmian Hu, Junyi Li, Aidong Zhang, Heng Huang
Conditional stochastic optimization has found applications in a wide range of machine learning tasks, such as invariant learning, AUPRC maximization, and meta-learning. As the demand for training models with large-scale distributed data grows in these applications, there is an increasing need for communication-efficient distributed optimization algorithms, such as federated learning algorithms. This paper considers the nonconvex conditional stochastic optimization in federated learning and proposes the first federated conditional stochastic optimization algorithm (FCSG) with a conditional stochastic gradient estimator and a momentum-based algorithm (\emph{i.e.}, FCSG-M). To match the lower bound complexity in the single-machine setting, we design an accelerated algorithm (Acc-FCSG-M) via the variance reduction to achieve the best sample and communication complexity. Compared with the existing optimization analysis for Meta-Learning in FL, federated conditional stochastic optimization considers the sample of tasks. Extensive experimental results on various tasks validate the efficiency of these algorithms.
****
LaFTer: Label-Free Tuning of Zero-shot Classifier using Language and Unlabeled Image Collections
Muhammad Jehanzeb Mirza, Leonid Karlinsky, Wei Lin, Horst Possegger, Mateusz Kozinski, Rogerio Feris, Horst Bischof
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Contextually Affinitive Neighborhood Refinery for Deep Clustering
Chunlin Yu, Ye Shi, Jingya Wang
Previous endeavors in self-supervised learning have enlightened the research of deep clustering from an instance discrimination perspective. Built upon this foundation, recent studies further highlight the importance of grouping semantically similar instances. One effective method to achieve this is by promoting the semantic structure preserved by neighborhood consistency. However, the samples in the local neighborhood may be limited due to their close proximity to each other, which may not provide substantial and diverse supervision signals. Inspired by the versatile re-ranking methods in the context of image retrieval, we propose to employ an efficient online re-ranking process to mine more informative neighbors in a Contextually Affinitive (ConAff) Neighborhood, and then encourage the cross-view neighborhood consistency. To further mitigate the intrinsic neighborhood noises near cluster boundaries, we propose a progressively relaxed boundary filtering strategy to circumvent the issues brought by noisy neighbors. Our method can be easily integrated into the generic self-supervised frameworks and outperforms the state-of-the-art methods on several popular benchmarks.
****
Differentiable Blocks World: Qualitative 3D Decomposition by Rendering Primitives
Tom Monnier, Jake Austin, Angjoo Kanazawa, Alexei Efros, Mathieu Aubry
Given a set of calibrated images of a scene, we present an approach that produces a simple, compact, and actionable 3D world representation by means of 3D primitives. While many approaches focus on recovering high-fidelity 3D scenes, we focus on parsing a scene into mid-level 3D representations made of a small set of textured primitives. Such representations are interpretable, easy to manipulate and suited for physics-based simulations. Moreover, unlike existing primitive decomposition methods that rely on 3D input data, our approach operates directly on images through differentiable rendering. Specifically, we model primitives as textured superquadric meshes and optimize their parameters from scratch with an image rendering loss. We highlight the importance of modeling transparency for each primitive, which is critical for optimization and also enables handling varying numbers of primitives. We show that the resulting textured primitives faithfu

lly reconstruct the input images and accurately model the visible 3D points, while providing amodal shape completions of unseen object regions. We compare our approach to the state of the art on diverse scenes from DTU, and demonstrate its robustness on real-life captures from BlendedMVS and Nerfstudio. We also showcase how our results can be used to effortlessly edit a scene or perform physical simulations. Code and video results are available at https://www.tmonnier.com/DBW.

****

Learning Shared Safety Constraints from Multi-task Demonstrations

Konwoo Kim, Gokul Swamy, ZUXIN LIU, DING ZHAO, Sanjiban Choudhury, Steven Z. Wu

Regardless of the particular task we want to perform in an environment, there are often shared safety constraints we want our agents to respect. For example, regardless of whether it is making a sandwich or clearing the table, a kitchen robot should not break a plate. Manually specifying such a constraint can be both time-consuming and error-prone. We show how to learn constraints from expert demonstrations of safe task completion by extending inverse reinforcement learning (IRL) techniques to the space of constraints. Intuitively, we learn constraints that forbid highly rewarding behavior that the expert could have taken but chose not to. Unfortunately, the constraint learning problem is rather ill-posed and typically leads to overly conservative constraints that forbid all behavior that the expert did not take. We counter this by leveraging diverse demonstrations that naturally occur in multi-task setting to learn a tighter set of constraints. We validate our method with simulation experiments on high-dimensional continuous control tasks.

****

Don’t Stop Pretraining? Make Prompt-based Fine-tuning Powerful Learner

Zhengxiang Shi, Aldo Lipani

Language models (LMs) trained on vast quantities of unlabelled data have greatly advanced the field of natural language processing (NLP). In this study, we re-visit the widely accepted notion in NLP that continued pre-training LMs on task-related texts improves the performance of fine-tuning (FT) in downstream tasks. Through experiments on eight single-sentence tasks and eight sentence-pair tasks in both semi-supervised and fully-supervised settings, we find that conventional continued pre-training does not consistently provide benefits and can even be detrimental for sentence-pair tasks or when prompt-based FT is used. To tackle these issues, we propose Prompt-based Continued Pre-training (PCP), which combines the idea of instruction tuning with conventional continued pre-training. Our approach aims to improve the performance of prompt-based FT by presenting both task-related texts and prompt templates to LMs through unsupervised pre-training objectives before fine-tuning for the target task. Our empirical evaluations on 21 benchmarks demonstrate that the PCP consistently improves the performance of state-of-the-art prompt-based FT approaches (up to 20.1% absolute) in both semi-supervised and fully-supervised settings, even with only hundreds of unlabelled examples. Additionally, prompt-based FT with PCP outperforms state-of-the-art semi-supervised approaches with greater simplicity, eliminating the need for an iterative process and extra data augmentation. Our further analysis explores the performance lower bound of the PCP and reveals that the advantages of PCP persist across different sizes of models and datasets.

****

GIMLET: A Unified Graph-Text Model for Instruction-Based Molecule Zero-Shot Learning

Haiteng Zhao, Shengchao Liu, Ma Chang, Hannan Xu, Jie Fu, Zhihong Deng, Lingpeng Kong, Qi Liu

Molecule property prediction has gained significant attention in recent years. The main bottleneck is the label insufficiency caused by expensive lab experiments. In order to alleviate this issue and to better leverage textual knowledge for tasks, this study investigates the feasibility of employing natural language instructions to accomplish molecule-related tasks in a zero-shot setting. We discover that existing molecule-text models perform poorly in this setting due to inadequate treatment of instructions and limited capacity for graphs. To overcome

these issues, we propose GIMLET, which unifies language models for both graph and text data. By adopting generalized position embedding, our model is extended to encode both graph structures and instruction text without additional graph encoding modules. GIMLET also decouples encoding of the graph from tasks instructions in the attention mechanism, enhancing the generalization of graph features across novel tasks. We construct a dataset consisting of more than two thousand molecule tasks with corresponding instructions derived from task descriptions. We pretrain GIMLET on the molecule tasks along with instructions, enabling the model to transfer effectively to a broad range of tasks. Experimental results demonstrate that GIMLET significantly outperforms molecule-text baselines in instruction-based zero-shot learning, even achieving closed results to supervised GNN models on tasks such as toxcast and muv.
****
GEX: A flexible method for approximating influence via Geometric Ensemble
SungYub Kim, Kyungsu Kim, Eunho Yang
Through a deeper understanding of predictions of neural networks, Influence Function (IF) has been applied to various tasks such as detecting and relabeling mislabeled samples, dataset pruning, and separation of data sources in practice. However, we found standard approximations of IF suffer from performance degradation due to oversimplified influence distributions caused by their bilinear approximation, suppressing the expressive power of samples with a relatively strong influence. To address this issue, we propose a new interpretation of existing IF approximations as an average relationship between two linearized losses over parameters sampled from the Laplace approximation (LA). In doing so, we highlight two significant limitations of current IF approximations: the linearity of gradients and the singularity of Hessian. Accordingly, by improving each point, we introduce a new IF approximation method with the following features: i) the removal of linearization to alleviate the bilinear constraint and ii) the utilization of Geometric Ensemble (GE) tailored for non-linear losses. Empirically, our approach outperforms existing IF approximations for downstream tasks with lighter computation, thereby providing new feasibility of low-complexity/nonlinear-based IF design.
****
Offline Reinforcement Learning for Mixture-of-Expert Dialogue Management
Dhawal Gupta, Yinlam Chow, Azamat Tulepbergenov, Mohammad Ghavamzadeh, Craig Boutilier
Reinforcement learning (RL) has shown great promise for developing agents for dialogue management (DM) that are non-myopic, conduct rich conversations, and maximize overall user satisfaction. Despite the advancements in RL and language models (LMs), employing RL to drive conversational chatbots still poses significant challenges. A primary issue stems from RL’s dependency on online exploration for effective learning, a process that can be costly. Moreover, engaging in online interactions with humans during the training phase can raise safety concerns, as the LM can potentially generate unwanted outputs. This issue is exacerbated by the combinatorial action spaces facing these algorithms, as most LM agents generate responses at the word level. We develop various RL algorithms, specialized in dialogue planning, that leverage recent Mixture-of-Expert Language Models (MoE-LMs)---models that capture diverse semantics, generate utterances reflecting different intents, and are amenable for multi-turn DM. By exploiting the MoE-LM structure, our methods significantly reduce the size of the action space and improve the efficacy of RL-based DM. We evaluate our methods in open-domain dialogue to demonstrate their effectiveness with respect to the diversity of intent in generated utterances and overall DM performance.
****
Binary Classification with Confidence Difference
Wei Wang, Lei Feng, Yuchen Jiang, Gang Niu, Min-Ling Zhang, Masashi Sugiyama
Recently, learning with soft labels has been shown to achieve better performance than learning with hard labels in terms of model generalization, calibration, and robustness. However, collecting pointwise labeling confidence for all training examples can be challenging and time-consuming in real-world scenarios. This p

aper delves into a novel weakly supervised binary classification problem called confidence-difference (ConfDiff) classification. Instead of pointwise labeling confidence, we are given only unlabeled data pairs with confidence difference that specifies the difference in the probabilities of being positive. We propose a risk-consistent approach to tackle this problem and show that the estimation error bound achieves the optimal convergence rate. We also introduce a risk correction approach to mitigate overfitting problems, whose consistency and convergence rate are also proven. Extensive experiments on benchmark data sets and a real-world recommender system data set validate the effectiveness of our proposed approaches in exploiting the supervision information of the confidence difference.
****
On student-teacher deviations in distillation: does it pay to disobey?
Vaishnavh Nagarajan, Aditya K. Menon, Srinadh Bhojanapalli, Hossein Mobahi, Sanjiv Kumar
Knowledge distillation (KD) has been widely used to improve the test accuracy of a "student" network, by training it to mimic the soft probabilities of a trained "teacher" network. Yet, it has been shown in recent work that, despite being trained to fit the teacher's probabilities, the student may not only significantly deviate from the teacher probabilities, but may also outdo than the teacher in performance. Our work aims to reconcile this seemingly paradoxical observation. Specifically, we characterize the precise nature of the student-teacher deviations, and argue how they can co-occur with better generalization. First, through experiments on image and language data, we identify that these probability deviations correspond to the student systematically exaggerating the confidence levels of the teacher.Next, we theoretically and empirically establish another form of exaggeration in some simple settings: KD exaggerates the implicit bias of gradient descent in converging faster along the top eigendirections of the data. Finally, we tie these two observations together: we demonstrate that the exaggerated bias of KD can simultaneously result in both (a) the exaggeration of confidence and (b) the improved generalization of the student, thus offering a resolution to the apparent paradox. Our analysis brings existing theory and practice closer by considering the role of gradient descent in KD and by demonstrating the exaggerated bias effect in both theoretical and empirical settings.
****
Resilient Multiple Choice Learning: A learned scoring scheme with application to audio scene analysis
Victor Letzelter, Mathieu Fontaine, Mickael Chen, Patrick Pérez, Slim Essid, Gaël Richard
We introduce Resilient Multiple Choice Learning (rMCL), an extension of the MCL approach for conditional distribution estimation in regression settings where multiple targets may be sampled for each training input.Multiple Choice Learning is a simple framework to tackle multimodal density estimation, using the Winner-Takes-All (WTA) loss for a set of hypotheses. In regression settings, the existing MCL variants focus on merging the hypotheses, thereby eventually sacrificing the diversity of the predictions. In contrast, our method relies on a novel learned scoring scheme underpinned by a mathematical framework based on Voronoi tessellations of the output space, from which we can derive a probabilistic interpretation.After empirically validating rMCL with experiments on synthetic data, we further assess its merits on the sound source localization problem, demonstrating its practical usefulness and the relevance of its interpretation.
****
Graph of Circuits with GNN for Exploring the Optimal Design Space
Aditya Shahane, Saripilli Swapna Manjiri, Ankesh Jain, Sandeep Kumar
The design automation of analog circuits poses significant challenges in terms of the large design space, complex interdependencies between circuit specifications, and resource-intensive simulations. To address these challenges, this paper presents an innovative framework called the Graph of Circuits Explorer (GCX). Leveraging graph structure learning along with graph neural networks, GCX enables the creation of a surrogate model that facilitates efficient exploration of the optimal design space within a semi-supervised learning framework which reduces t

he need for large labelled datasets. The proposed approach comprises three key stages. First, we learn the geometric representation of circuits and enrich it with technology information to create a comprehensive feature vector. Subsequently, integrating feature-based graph learning with few-shot and zero-shot learning enhances the generalizability in predictions for unseen circuits. Finally, we introduce two algorithms namely, EASCO and ASTROG which upon integration with GCX optimize the available samples to yield the optimal circuit configuration meeting the designer's criteria. The effectiveness of the proposed approach is demonstrated through simulated performance evaluation of various circuits, using derived parameters in 180nm CMOS technology. Furthermore, the generalizability of the approach is extended to higher-order topologies and different technology nodes such as 65nm and 45nm CMOS process nodes.
****
Structure-free Graph Condensation: From Large-scale Graphs to Condensed Graph-free Data
Xin Zheng, Miao Zhang, Chunyang Chen, Quoc Viet Hung Nguyen, Xingquan Zhu, Shirui Pan
Graph condensation, which reduces the size of a large-scale graph by synthesizing a small-scale condensed graph as its substitution, has immediate benefits for various graph learning tasks.However, existing graph condensation methods rely on the joint optimization of nodes and structures in the condensed graph, and overlook critical issues in effectiveness and generalization ability.In this paper, we advocate a new Structure-Free Graph Condensation paradigm, named SFGC, to distill a large-scale graph into a small-scale graph node set without explicit graph structures, i.e., graph-free data.Our idea is to implicitly encode topology structure information into the node attributes in the synthesized graph-free data, whose topology is reduced to an identity matrix.Specifically, SFGC contains two collaborative components: (1) a training trajectory meta-matching scheme for effectively synthesizing small-scale graph-free data;(2) a graph neural feature score metric for dynamically evaluating the quality of the condensed data. Through training trajectory meta-matching, SFGC aligns the long-term GNN learning behaviors between the large-scale graph and the condensed small-scale graph-free data, ensuring comprehensive and compact transfer of informative knowledge to the graph-free data.Afterward, the underlying condensed graph-free data would be dynamically evaluated with the graph neural feature score, which is a closed-form metric for ensuring the excellent expressiveness of the condensed graph-free data. Extensive experiments verify the superiority of SFGC across different condensation ratios.
****
Visual Programming for Step-by-Step Text-to-Image Generation and Evaluation
Jaemin Cho, Abhay Zala, Mohit Bansal
As large language models have demonstrated impressive performance in many domains, recent works have adopted language models (LMs) as controllers of visual modules for vision-and-language tasks. While existing work focuses on equipping LMs with visual understanding, we propose two novel interpretable/explainable visual programming frameworks for text-to-image (T2I) generation and evaluation. First, we introduce VPGen, an interpretable step-by-step T2I generation framework that decomposes T2I generation into three steps: object/count generation, layout generation, and image generation. We employ an LM to handle the first two steps (object/count generation and layout generation), by finetuning it on text-layout pairs. Our step-by-step T2I generation framework provides stronger spatial control than end-to-end models, the dominant approach for this task. Furthermore, we leverage the world knowledge of pretrained LMs, overcoming the limitation of previous layout-guided T2I works that can only handle predefined object classes. We demonstrate that our VPGen has improved control in counts/spatial relations/scales of objects than state-of-the-art T2I generation models. Second, we introduce VPEval, an interpretable and explainable evaluation framework for T2I generation based on visual programming. Unlike previous T2I evaluations with a single scoring model that is accurate in some skills but unreliable in others, VPEval produces evaluation programs that invoke a set of visual modules that are experts in

different skills, and also provides visual+textual explanations of the evaluation results. Our analysis shows that VPEval provides a more human-correlated evaluation for skill-specific and open-ended prompts than widely used single model-based evaluation. We hope that our work encourages future progress on interpretable/explainable generation and evaluation for T2I models.
****
Auditing Fairness by Betting
Ben Chugg, Santiago Cortes-Gomez, Bryan Wilder, Aaditya Ramdas
We provide practical, efficient, and nonparametric methods for auditing the fairness of deployed classification and regression models. Whereas previous work relies on a fixed-sample size, our methods are sequential and allow for the continuous monitoring of incoming data, making them highly amenable to tracking the fairness of real-world systems. We also allow the data to be collected by a probabilistic policy as opposed to sampled uniformly from the population. This enables auditing to be conducted on data gathered for another purpose. Moreover, this policy may change over time and different policies may be used on different subpopulations. Finally, our methods can handle distribution shift resulting from either changes to the model or changes in the underlying population. Our approach is based on recent progress in anytime-valid inference and game-theoretic statistics---the ``testing by betting'' framework in particular. These connections ensure that our methods are interpretable, fast, and easy to implement. We demonstrate the efficacy of our approach on three benchmark fairness datasets.
****
Truly Scale-Equivariant Deep Nets with Fourier Layers
Md Ashiqur Rahman, Raymond A. Yeh
In computer vision, models must be able to adapt to changes in image resolution to effectively carry out tasks such as image segmentation; This is known as scale-equivariance. Recent works have made progress in developing scale-equivariant convolutional neural networks, e.g., through weight-sharing and kernel resizing. However, these networks are not truly scale-equivariant in practice. Specifically, they do not consider anti-aliasing as they formulate the down-scaling operation in the continuous domain. To address this shortcoming, we directly formulate down-scaling in the discrete domain with consideration of anti-aliasing. We then propose a novel architecture based on Fourier layers to achieve truly scale-equivariant deep nets, i.e., absolute zero equivariance-error. Following prior works, we test this model on MNIST-scale and STL-10 datasets. Our proposed model achieves competitive classification performance while maintaining zero equivariance-error.
****
Projection-Free Methods for Stochastic Simple Bilevel Optimization with Convex Lower-level Problem
Jincheng Cao, Ruichen Jiang, Nazanin Abolfazli, Erfan Yazdandoost Hamedani, Aryan Mokhtari
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On the Implicit Bias of Linear Equivariant Steerable Networks
Ziyu Chen, Wei Zhu
We study the implicit bias of gradient flow on linear equivariant steerable networks in group-invariant binary classification. Our findings reveal that the parameterized predictor converges in direction to the unique group-invariant classifier with a maximum margin defined by the input group action. Under a unitary assumption on the input representation, we establish the equivalence between steerable networks and data augmentation. Furthermore, we demonstrate the improved margin and generalization bound of steerable networks over their non-invariant counterparts.
****
Memory-Constrained Algorithms for Convex Optimization

Moise Blanchard, Junhui Zhang, Patrick Jaillet
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Nonparametric Boundary Geometry in Physics Informed Deep Learning
Scott Cameron, Arnu Pretorius, S Roberts
Engineering design problems frequently require solving systems ofpartial differential equations with boundary conditions specified onobject geometries in the form of a triangular mesh. These boundarygeometries are provided by a designer and are problem dependent.The efficiency of the design process greatly benefits from fast turnaroundtimes when repeatedly solving PDEs on various geometries. However,most current work that uses machine learning to speed up the solutionprocess relies heavily on a fixed parameterization of the geometry, whichcannot be changed after training. This severely limits the possibility ofreusing a trained model across a variety of design problems.In this work, we propose a novel neural operator architecture which acceptsboundary geometry, in the form of triangular meshes, as input and produces anapproximate solution to a given PDE as output. Once trained, the model can beused to rapidly estimate the PDE solution over a new geometry, without the need forretraining or representation of the geometry to a pre-specified parameterization.
****
Tracking Most Significant Shifts in Nonparametric Contextual Bandits
Joe Suk, Samory Kpotufe
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Empowering Collaborative Filtering with Principled Adversarial Contrastive Loss
An Zhang, Leheng Sheng, Zhibo Cai, Xiang Wang, Tat-Seng Chua
Contrastive Learning (CL) has achieved impressive performance in self-supervised learning tasks, showing superior generalization ability. Inspired by the success, adopting CL into collaborative filtering (CF) is prevailing in semi-supervised topK recommendations. The basic idea is to routinely conduct heuristic-based data augmentation and apply contrastive losses (e.g., InfoNCE) on the augmented views. Yet, some CF-tailored challenges make this adoption suboptimal, such as the issue of out-of-distribution, the risk of false negatives, and the nature of top-K evaluation. They necessitate the CL-based CF scheme to focus more on mining hard negatives and distinguishing false negatives from the vast unlabeled user-item interactions, for informative contrast signals. Worse still, there is limited understanding of contrastive loss in CF methods, especially w.r.t. its generalization ability. To bridge the gap, we delve into the reasons underpinning the success of contrastive loss in CF, and propose a principled Adversarial InfoNCE loss (AdvInfoNCE), which is a variant of InfoNCE, specially tailored for CF methods. AdvInfoNCE adaptively explores and assigns hardness to each negative instance in an adversarial fashion and further utilizes a fine-grained hardness-aware ranking criterion to empower the recommender’s generalization ability. Training CF models with AdvInfoNCE, we validate the effectiveness of AdvInfoNCE on both synthetic and real-world benchmark datasets, thus showing its generalization ability to mitigate out-of-distribution problems. Given the theoretical guarantees and empirical superiority of AdvInfoNCE over most contrastive loss functions, we advocate its adoption as a standard loss in recommender systems, particularly for the out-of-distribution tasks. Codes are available at https://github.com/LehengTHU/AdvInfoNCE.
****
The Rashomon Importance Distribution: Getting RID of Unstable, Single Model-based Variable Importance
Jon Donnelly, Srikar Katta, Cynthia Rudin, Edward Browne

Quantifying variable importance is essential for answering high-stakes questions in fields like genetics, public policy, and medicine. Current methods generally calculate variable importance for a given model trained on a given dataset. However, for a given dataset, there may be many models that explain the target outcome equally well; without accounting for all possible explanations, different researchers may arrive at many conflicting yet equally valid conclusions given the same data. Additionally, even when accounting for all possible explanations for a given dataset, these insights may not generalize because not all good explanations are stable across reasonable data perturbations. We propose a new variable importance framework that quantifies the importance of a variable across the set of all good models and is stable across the data distribution. Our framework is extremely flexible and can be integrated with most existing model classes and global variable importance metrics. We demonstrate through experiments that our framework recovers variable importance rankings for complex simulation setups where other methods fail. Further, we show that our framework accurately estimates the true importance of a variable for the underlying data distribution. We provide theoretical guarantees on the consistency and finite sample error rates for our estimator. Finally, we demonstrate its utility with a real-world case study exploring which genes are important for predicting HIV load in persons with HIV, highlighting an important gene that has not previously been studied in connection with HIV.
****
Model-Based Control with Sparse Neural Dynamics
Ziang Liu, Genggeng Zhou, Jeff He, Tobia Marcucci, Fei-Fei Li, Jiajun Wu, Yunzhu Li
Learning predictive models from observations using deep neural networks (DNNs) is a promising new approach to many real-world planning and control problems. However, common DNNs are too unstructured for effective planning, and current control methods typically rely on extensive sampling or local gradient descent. In this paper, we propose a new framework for integrated model learning and predictive control that is amenable to efficient optimization algorithms. Specifically, we start with a ReLU neural model of the system dynamics and, with minimal losses in prediction accuracy, we gradually sparsify it by removing redundant neurons. This discrete sparsification process is approximated as a continuous problem, enabling an end-to-end optimization of both the model architecture and the weight parameters. The sparsified model is subsequently used by a mixed-integer predictive controller, which represents the neuron activations as binary variables and employs efficient branch-and-bound algorithms. Our framework is applicable to a wide variety of DNNs, from simple multilayer perceptrons to complex graph neural dynamics. It can efficiently handle tasks involving complicated contact dynamics, such as object pushing, compositional object sorting, and manipulation of deformable objects. Numerical and hardware experiments show that, despite the aggressive sparsification, our framework can deliver better closed-loop performance than existing state-of-the-art methods.
****
AmadeusGPT: a natural language interface for interactive animal behavioral analysis
Shaokai Ye, Jessy Lauer, Mu Zhou, Alexander Mathis, Mackenzie Mathis
The process of quantifying and analyzing animal behavior involves translating the naturally occurring descriptive language of their actions into machine-readable code. Yet, codifying behavior analysis is often challenging without deep understanding of animal behavior and technical machine learning knowledge. To limit this gap, we introduce AmadeusGPT: a natural language interface that turns natural language descriptions of behaviors into machine-executable code. Large-language models (LLMs) such as GPT3.5 and GPT4 allow for interactive language-based queries that are potentially well suited for making interactive behavior analysis. However, the comprehension capability of these LLMs is limited by the context window size, which prevents it from remembering distant conversations. To overcome the context window limitation, we implement a novel dual-memory mechanism to allow communication between short-term and long-term memory using symbols as conte

xt pointers for retrieval and saving. Concretely, users directly use language-based definitions of behavior and our augmented GPT develops code based on the core AmadeusGPT API, which contains machine learning, computer vision, spatio-temporal reasoning, and visualization modules. Users then can interactively refine results, and seamlessly add new behavioral modules as needed. We used the MABe 2022 behavior challenge tasks to benchmark AmadeusGPT and show excellent performance. Note, an end-user would not need to write any code to achieve this. Thus, collectively AmadeusGPT presents a novel way to merge deep biological knowledge, large-language models, and core computer vision modules into a more naturally intelligent system. Code and demos can be found at: https://github.com/AdaptiveMotorControlLab/AmadeusGPT
****

Provably Efficient Algorithm for Nonstationary Low-Rank MDPs

Yuan Cheng, Jing Yang, Yingbin Liang

Reinforcement learning (RL) under changing environment models many real-world applications via nonstationary Markov Decision Processes (MDPs), and hence gains considerable interest. However, theoretical studies on nonstationary MDPs in the literature have mainly focused on tabular and linear (mixture) MDPs, which do not capture the nature of unknown representation in deep RL. In this paper, we make the first effort to investigate nonstationary RL under episodic low-rank MDPs, where both transition kernels and rewards may vary over time, and the low-rank model contains unknown representation in addition to the linear state embedding function. We first propose a parameter-dependent policy optimization algorithm called PORTAL,and further improve PORTAL to its parameter-free version of Ada-PORTAL, which is able to tune its hyper-parameters adaptively without any prior knowledge of nonstationarity. For both algorithms, we provide upper bounds on the average dynamic suboptimality gap, which show that as long as the nonstationarity is not significantly large, PORTAL and Ada-PORTAL are sample-efficient and can achieve arbitrarily small average dynamic suboptimality gap with polynomial sample complexity.
****

Time-uniform confidence bands for the CDF under nonstationarity

Paul Mineiro, Steven Howard

Estimation of a complete univariate distribution from a sequence of observations is a useful primitive for both manual and automated decision making. This problem has received extensive attention in the i.i.d. setting, but the arbitrary data dependent setting remains largely unaddressed. We present computationally felicitous time-uniform and value-uniform bounds on the CDF of the running averaged conditional distribution of a sequence of real-valued random variables. Consistent with known impossibility results, our CDF bounds are always valid but sometimes trivial when the instance is too hard, and we give an instance-dependent convergence guarantee. The importance-weighted extension is appropriate for estimating complete counterfactual distributions of rewards given data from a randomized experiment, e.g., from an A/B test or a contextual bandit.
****

Risk-Averse Active Sensing for Timely Outcome Prediction under Cost Pressure

Yuchao Qin, Mihaela van der Schaar, Changhee Lee

Timely outcome prediction is essential in healthcare to enable early detection and intervention of adverse events. However, in longitudinal follow-ups to patients' health status, cost-efficient acquisition of patient covariates is usually necessary due to the significant expense involved in screening and lab tests. To balance the timely and accurate outcome predictions with acquisition costs, an effective active sensing strategy is crucial. In this paper, we propose a novel risk-averse active sensing approach RAS that addresses the composite decision problem of when to conduct the acquisition and which measurements to make. Our approach decomposes the policy into two sub-policies: acquisition scheduler and feature selector, respectively. Moreover, we introduce a novel risk-aversion training strategy to focus on the underrepresented subgroup of high-risk patients for whom timely and accurate prediction of disease progression is of greater value. Our method outperforms baseline active sensing approaches in experiments with bot

h synthetic and real-world datasets, and we illustrate the significance of our policy decomposition and the necessity of a risk-averse sensing policy through case studies.
****
Single-Pass Pivot Algorithm for Correlation Clustering. Keep it simple!
Konstantin Makarychev, Sayak Chakrabarty
We show that a simple single-pass semi-streaming variant of the Pivot algorithm for Correlation Clustering gives a (3+eps)-approximation using O(n/eps) words of memory. This is a slight improvement over the recent results of Cambus, Kuhn, Lindy, Pai, and Uitto, who gave a (3+eps)-approximation using O(n log n) words of memory, and Behnezhad, Charikar, Ma, and Tan, who gave a 5-approximation using O(n) words of memory. One of the main contributions of our paper is that the algorithm and its analysis are simple and easy to understand.
****
SPACE: Single-round Participant Amalgamation for Contribution Evaluation in Federated Learning
Yi-Chung Chen, Hsi-Wen Chen, Shun-Gui Wang, Ming-syan Chen
The evaluation of participant contribution in federated learning (FL) has recently gained significant attention due to its applicability in various domains, such as incentive mechanisms, robustness enhancement, and client selection. Previous approaches have predominantly relied on the widely adopted Shapley value for participant evaluation. However, the computation of the Shapley value is expensive, despite using techniques like gradient-based model reconstruction and truncating unnecessary evaluations. Therefore, we present an efficient approach called Single-round Participants Amalgamation for Contribution Evaluation (SPACE). SPACE incorporates two novel components, namely Federated Knowledge Amalgamation and Prototype-based Model Evaluation to reduce the evaluation effort by eliminating the dependence on the size of the validation set and enabling participant evaluation within a single communication round. Experimental results demonstrate that SPACE outperforms state-of-the-art methods in terms of both running time and Pearson’s Correlation Coefficient (PCC). Furthermore, extensive experiments conducted on applications, client reweighting, and client selection highlight the effectiveness of SPACE. The code is available at https://github.com/culiver/SPACE.
****
SAME: Uncovering GNN Black Box with Structure-aware Shapley-based Multipiece Explanations
Ziyuan Ye, Rihan Huang, Qilin Wu, Quanying Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Federated Learning with Client Subsampling, Data Heterogeneity, and Unbounded Smoothness: A New Algorithm and Lower Bounds
Michael Crawshaw, Yajie Bao, Mingrui Liu
We study the problem of Federated Learning (FL) under client subsampling and data heterogeneity with an objective function that has potentially unbounded smoothness. This problem is motivated by empirical evidence that the class of relaxed smooth functions, where the Lipschitz constant of the gradient scales linearly with the gradient norm, closely resembles the loss functions of certain neural networks such as recurrent neural networks (RNNs) with possibly exploding gradient. We introduce EPISODE++, the first algorithm to solve this problem. It maintains historical statistics for each client to construct control variates and decide clipping behavior for sampled clients in the current round. We prove that EPISODE++ achieves linear speedup in the number of participating clients, reduced communication rounds, and resilience to data heterogeneity. Our upper bound proof relies on novel techniques of recursively bounding the client updates under unbounded smoothness and client subsampling, together with a refined high probability analysis. In addition, we prove a lower bound showing that the convergence rate of a special case of clipped minibatch SGD (without randomness in the stochasti

c gradient and with randomness in client subsampling) suffers from an explicit dependence on the maximum gradient norm of the objective in a sublevel set, which may be large. This effectively demonstrates that applying gradient clipping to minibatch SGD in our setting does not eliminate the problem of exploding gradients. Our lower bound is based on new constructions of hard instances tailored to client subsampling and a novel analysis of the trajectory of the algorithm in the presence of clipping. Lastly, we provide an experimental evaluation of EPISODE++ when training RNNs on federated text classification tasks, demonstrating that EPISODE++ outperforms strong baselines in FL. The code is available at https://github.com/MingruiLiu-ML-Lab/episode_plusplus.
****
NeuroGraph: Benchmarks for Graph Machine Learning in Brain Connectomics
Anwar Said, Roza Bayrak, Tyler Derr, Mudassir Shabbir, Daniel Moyer, Catie Chang, Xenofon Koutsoukos
Machine learning provides a valuable tool for analyzing high-dimensional functional neuroimaging data, and is proving effective in predicting various neurological conditions, psychiatric disorders, and cognitive patterns. In functional magnetic resonance imaging (MRI) research, interactions between brain regions are commonly modeled using graph-based representations. The potency of graph machine learning methods has been established across myriad domains, marking a transformative step in data interpretation and predictive modeling. Yet, despite their promise, the transposition of these techniques to the neuroimaging domain has been challenging due to the expansive number of potential preprocessing pipelines and the large parameter search space for graph-based dataset construction. In this paper, we introduce NeuroGraph, a collection of graph-based neuroimaging datasets, and demonstrated its utility for predicting multiple categories of behavioral and cognitive traits. We delve deeply into the dataset generation search space by crafting 35 datasets that encompass static and dynamic brain connectivity, running in excess of 15 baseline methods for benchmarking. Additionally, we provide generic frameworks for learning on both static and dynamic graphs. Our extensive experiments lead to several key observations. Notably, using correlation vectors as node features, incorporating larger number of regions of interest, and employing sparser graphs lead to improved performance. To foster further advancements in graph-based data driven neuroimaging analysis, we offer a comprehensive open-source Python package that includes the benchmark datasets, baseline implementations, model training, and standard evaluation.
****
Quantifying the Cost of Learning in Queueing Systems
Daniel Freund, Thodoris Lykouris, Wentao Weng
Queueing systems are widely applicable stochastic models with use cases in communication networks, healthcare, service systems, etc. Although their optimal control has been extensively studied, most existing approaches assume perfect knowledge of the system parameters. Of course, this assumption rarely holds in practice where there is parameter uncertainty, thus motivating a recent line of work on bandit learning for queueing systems. This nascent stream of research focuses on the asymptotic performance of the proposed algorithms. In this paper, we argue that an asymptotic metric, which focuses on late-stage performance, is insufficient to capture the intrinsic statistical complexity of learning in queueing systems which typically occurs in the early stage. Instead, we propose the Cost of Learning in Queueing (CLQ), a new metric that quantifies the maximum increase in time-averaged queue length caused by parameter uncertainty.We characterize the CLQ of a single-queue multi-server system, and then extend these results to multi-queue multi-server systems and networks of queues. In establishing our results, we propose a unified analysis framework for CLQ that bridges Lyapunov and bandit analysis, provides guarantees for a wide range of algorithms, and could be of independent interest.
****
One-Line-of-Code Data Mollification Improves Optimization of Likelihood-based Generative Models
Ba-Hien Tran, Giulio Franzese, Pietro Michiardi, Maurizio Filippone

Generative Models (GMs) have attracted considerable attention due to their tremendous success in various domains, such as computer vision where they are capable to generate impressive realistic-looking images. Likelihood-based GMs are attractive due to the possibility to generate new data by a single model evaluation. However, they typically achieve lower sample quality compared to state-of-the-art score-based Diffusion Models (DMs). This paper provides a significant step in the direction of addressing this limitation. The idea is to borrow one of the strengths of score-based DMs, which is the ability to perform accurate density estimation in low-density regions and to address manifold overfitting by means of data mollification. We propose a view of data mollification within likelihood-based GMs as a continuation method, whereby the optimization objective smoothly transitions from simple-to-optimize to the original target. Crucially, data mollification can be implemented by adding one line of code in the optimization loop, and we demonstrate that this provides a boost in generation quality of likelihood-based GMs, without computational overheads. We report results on real-world image data sets and UCI benchmarks with popular likelihood-based GMs, including variants of variational autoencoders and normalizing flows, showing large improvements in FID score and density estimation.

****

FLSL: Feature-level Self-supervised Learning

Qing Su, Anton Netchaev, Hai Li, Shihao Ji

Current self-supervised learning (SSL) methods (e.g., SimCLR, DINO, VICReg, MOCOv3) target primarily on representations at instance level and do not generalize well to dense prediction tasks, such as object detection and segmentation. Towards aligning SSL with dense predictions, this paper demonstrates for the first time the underlying mean-shift clustering process of Vision Transformers (ViT), which aligns well with natural image semantics (e.g., a world of objects and stuffs). By employing transformer for joint embedding and clustering, we propose a bi-level feature clustering SSL method, coined Feature-Level Self-supervised Learning (FLSL). We present the formal definition of the FLSL problem and construct the objectives from the mean-shift and k-means perspectives. We show that FLSL promotes remarkable semantic cluster representations and learns an embedding scheme amenable to intra-view and inter-view feature clustering. Experiments show that FLSL yields significant improvements in dense prediction tasks, achieving 44.9 (+2.8)% AP and 46.5% AP in object detection, as well as 40.8 (+2.3)% AP and 42.1% AP in instance segmentation on MS-COCO, using Mask R-CNN with ViT-S/16 and ViT-S/8 as backbone, respectively. FLSL consistently outperforms existing SSL methods across additional benchmarks, including UAV object detection on UAVDT, and video instance segmentation on DAVIS 2017. We conclude by presenting visualization and various ablation studies to better understand the success of FLSL. The source code is available at https://github.com/ISL-CV/FLSL.

****

FeCAM: Exploiting the Heterogeneity of Class Distributions in Exemplar-Free Continual Learning

Dipam Goswami, Yuyang Liu, Bart■omiej Twardowski, Joost van de Weijer

Exemplar-free class-incremental learning (CIL) poses several challenges since it prohibits the rehearsal of data from previous tasks and thus suffers from catastrophic forgetting. Recent approaches to incrementally learning the classifier by freezing the feature extractor after the first task have gained much attention. In this paper, we explore prototypical networks for CIL, which generate new class prototypes using the frozen feature extractor and classify the features based on the Euclidean distance to the prototypes. In an analysis of the feature distributions of classes, we show that classification based on Euclidean metrics is successful for jointly trained features. However, when learning from non-stationary data, we observe that the Euclidean metric is suboptimal and that feature distributions are heterogeneous. To address this challenge, we revisit the anisotropic Mahalanobis distance for CIL. In addition, we empirically show that modeling the feature covariance relations is better than previous attempts at sampling features from normal distributions and training a linear classifier. Unlike existing methods, our approach generalizes to both many- and few-shot CIL settings,

as well as to domain-incremental settings. Interestingly, without updating the
backbone network, our method obtains state-of-the-art results on several standar
d continual learning benchmarks. Code is available at https://github.com/dipamgo
swami/FeCAM.
****
Learning non-Markovian Decision-Making from State-only Sequences
Aoyang Qin, Feng Gao, Qing Li, Song-Chun Zhu, Sirui Xie
Requests for name changes in the electronic proceedings will be accepted with no
questions asked. However name changes may cause bibliographic tracking issues.
Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Spectral Invariant Learning for Dynamic Graphs under Distribution Shifts
Zeyang Zhang, Xin Wang, Ziwei Zhang, Zhou Qin, Weigao Wen, Hui Xue', Haoyang Li,
Wenwu Zhu
Dynamic graph neural networks (DyGNNs) currently struggle with handling distribu
tion shifts that are inherent in dynamic graphs.Existing work on DyGNNs with out
-of-distribution settings only focuses on the time domain, failing to handle cas
es involving distribution shifts in the spectral domain. In this paper, we disco
ver that there exist cases with distribution shifts unobservable in the time dom
ain while observable in the spectral domain, and propose to study distribution s
hifts on dynamic graphs in the spectral domain for the first time.However, this
investigation poses two key challenges: i) it is non-trivial to capture differen
t graph patterns that are driven by various frequency components entangled in th
e spectral domain; and ii) it remains unclear how to handle distribution shifts
with the discovered spectral patterns. To address these challenges, we propose S
pectral Invariant Learning for Dynamic Graphs under Distribution Shifts (SILD),
which can handle distribution shifts on dynamic graphs by capturing and utilizin
g invariant and variant spectral patterns. Specifically, we first design a DyGNN
with Fourier transform to obtain the ego-graph trajectory spectrums, allowing t
he mixed dynamic graph patterns to be transformed into separate frequency compon
ents. We then develop a disentangled spectrum mask to filter graph dynamics from
various frequency components and discover the invariant and variant spectral pa
tterns. Finally, we propose invariant spectral filtering, which encourages the m
odel to rely on invariant patterns for generalization under distribution shifts.
Experimental results on synthetic and real-world dynamic graph datasets demonst
rate the superiority of our method for both node classification and link predict
ion tasks under distribution shifts.
****
Efficient Activation Function Optimization through Surrogate Modeling
Garrett Bingham, Risto Miikkulainen
Carefully designed activation functions can improve the performance of neural ne
tworks in many machine learning tasks. However, it is difficult for humans to c
onstruct optimal activation functions, and current activation function search al
gorithms are prohibitively expensive. This paper aims to improve the state of t
he art through three steps: First, the benchmark datasets Act-Bench-CNN, Act-Ben
ch-ResNet, and Act-Bench-ViT were created by training convolutional, residual, a
nd vision transformer architectures from scratch with 2,913 systematically gener
ated activation functions. Second, a characterization of the benchmark space was
developed, leading to a new surrogate-based method for optimization. More speci
fically, the spectrum of the Fisher information matrix associated with the model
's predictive distribution at initialization and the activation function's outpu
t distribution were found to be highly predictive of performance. Third, the sur
rogate was used to discover improved activation functions in several real-world
tasks, with a surprising finding: a sigmoidal design that outperformed all other
activation functions was discovered, challenging the status quo of always using
rectifier nonlinearities in deep learning. Each of these steps is a contributi
on in its own right; together they serve as a practical and theoretical foundati
on for further research on activation function optimization.
****

Data Market Design through Deep Learning
Sai Srivatsa Ravindranath, Yanchen Jiang, David C. Parkes
The data market design problem is a problem in economic theory to find a set of signaling schemes (statistical experiments) to maximize expected revenue to the information seller, where each experiment reveals some of the information known to a seller and has a corresponding price. Each buyer has their own decision to make in a world environment, and their subjective expected value for the information associated with a particular experiment comes from the improvement in this decision and depends on their prior and value for different outcomes. In a setting with multiple buyers, a buyer's expected value for an experiment may also depend on the information sold to others. We introduce the application of deep learning for the design of revenue-optimal data markets, looking to expand the frontiers of what can be understood and achieved. Relative to earlier work on deep learning for auction design, we must learn signaling schemes rather than allocation rules and handle obedience constraints — these arising from modeling the downstream actions of buyers — in addition to incentive constraints on bids. Our experiments demonstrate that this new deep learning framework can almost precisely replicate all known solutions from theory, expand to more complex settings, and be used to establish the optimality of new designs for data markets and make conjectures in regard to the structure of optimal designs.
****
When Visual Prompt Tuning Meets Source-Free Domain Adaptive Semantic Segmentation
Xinhong Ma, Yiming Wang, Hao Liu, Tianyu Guo, Yunhe Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Benchmarking and Analyzing 3D-aware Image Synthesis with a Modularized Codebase
Qiuyu Wang, Zifan Shi, Kecheng Zheng, Yinghao Xu, Sida Peng, Yujun Shen
Despite the rapid advance of 3D-aware image synthesis, existing studies usually adopt a mixture of techniques and tricks, leaving it unclear how each part contributes to the final performance in terms of generality. Following the most popular and effective paradigm in this field, which incorporates a neural radiance field (NeRF) into the generator of a generative adversarial network (GAN), we build a well-structured codebase through modularizing the generation process. Such a design allows researchers to develop and replace each module independently, and hence offers an opportunity to fairly compare various approaches and recognize their contributions from the module perspective. The reproduction of a range of cutting-edge algorithms demonstrates the availability of our modularized codebase. We also perform a variety of in-depth analyses, such as the comparison across different types of point feature, the necessity of the tailing upsampler in the generator, the reliance on the camera pose prior, etc., which deepen our understanding of existing methods and point out some further directions of the research work. Code and models will be made publicly available to facilitate the development and evaluation of this field.
****
RL-ViGen: A Reinforcement Learning Benchmark for Visual Generalization
Zhecheng Yuan, Sizhe Yang, Pu Hua, Can Chang, Kaizhe Hu, Huazhe Xu
Visual Reinforcement Learning (Visual RL), coupled with high-dimensional observations, has consistently confronted the long-standing challenge of out-of-distribution generalization. Despite the focus on algorithms aimed at resolving visual generalization problems, we argue that the devil is in the existing benchmarks as they are restricted to isolated tasks and generalization categories, undermining a comprehensive evaluation of agents' visual generalization capabilities. To bridge this gap, we introduce RL-ViGen: a novel Reinforcement Learning Benchmark for Visual Generalization, which contains diverse tasks and a wide spectrum of generalization types, thereby facilitating the derivation of more reliable conclusions. Furthermore, RL-ViGen incorporates the latest generalization visual RL a

lgorithms into a unified framework, under which the experiment results indicate that no single existing algorithm has prevailed universally across tasks. Our aspiration is that Rl-ViGen will serve as a catalyst in this area, and lay a foundation for the future creation of universal visual generalization RL agents suitable for real-world scenarios. Access to our code and implemented algorithms is provided at https://gemcollector.github.io/RL-ViGen/.
****
DoWG Unleashed: An Efficient Universal Parameter-Free Gradient Descent Method
Ahmed Khaled, Konstantin Mishchenko, Chi Jin
This paper proposes a new easy-to-implement parameter-free gradient-based optimizer: DoWG (Distance over Weighted Gradients). We prove that DoWG is efficient---matching the convergence rate of optimally tuned gradient descent in convex optimization up to a logarithmic factor without tuning any parameters, and universal---automatically adapting to both smooth and nonsmooth problems. While popular algorithms following the AdaGrad framework compute a running average of the squared gradients, DoWG maintains a new distance-based weighted version of the running average, which is crucial to achieve the desired properties. To complement our theory, we also show empirically that DoWG trains at the edge of stability, and validate its effectiveness on practical machine learning tasks.
****
Multitask Learning with No Regret: from Improved Confidence Bounds to Active Learning
Pier Giuseppe Sessa, Pierre Laforgue, Nicolò Cesa-Bianchi, Andreas Krause
Multitask learning is a powerful framework that enables one to simultaneously learn multiple related tasks by sharing information between them. Quantifying uncertainty in the estimated tasks is of pivotal importance for many downstream applications, such as online or active learning. In this work, we provide novel confidence intervals for multitask regression in the challenging agnostic setting, i.e., when neither the similarity between tasks nor the tasks' features are available to the learner. The obtained intervals do not require i.i.d. data and can be directly applied to bound the regret in online learning. Through a refined analysis of the multitask information gain, we obtain new regret guarantees that, depending on a task similarity parameter, can significantly improve over treating tasks independently. We further propose a novel online learning algorithm that achieves such improved regret without knowing this parameter in advance, i.e., automatically adapting to task similarity. As a second key application of our results, we introduce a novel multitask active learning setup where several tasks must be simultaneously optimized, but only one of them can be queried for feedback by the learner at each round. For this problem, we design a no-regret algorithm that uses our confidence intervals to decide which task should be queried. Finally, we empirically validate our bounds and algorithms on synthetic and real-world (drug discovery) data.
****
Posterior Sampling with Delayed Feedback for Reinforcement Learning with Linear Function Approximation
Nikki Lijing Kuang, Ming Yin, Mengdi Wang, Yu-Xiang Wang, Yian Ma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Macro Placement by Wire-Mask-Guided Black-Box Optimization
Yunqi Shi, Ke Xue, Song Lei, Chao Qian
The development of very large-scale integration (VLSI) technology has posed new challenges for electronic design automation (EDA) techniques in chip floorplanning. During this process, macro placement is an important subproblem, which tries to determine the positions of all macros with the aim of minimizing half-perimeter wirelength (HPWL) and avoiding overlapping. Previous methods include packing-based, analytical and reinforcement learning methods. In this paper, we propose a new black-box optimization (BBO) framework (called WireMask-BBO) for macro pl

acement, by using a wire-mask-guided greedy procedure for objective evaluation. Equipped with different BBO algorithms, WireMask-BBO empirically achieves significant improvements over previous methods, i.e., achieves significantly shorter HPWL by using much less time. Furthermore, it can fine-tune existing placements by treating them as initial solutions, which can bring up to 50% improvement in HPWL. WireMask-BBO has the potential to significantly improve the quality and efficiency of chip floorplanning, which makes it appealing to researchers and practitioners in EDA and will also promote the application of BBO. Our code is available at https://github.com/lamda-bbo/WireMask-BBO.
****
Reconciling Competing Sampling Strategies of Network Embedding
Yuchen Yan, Baoyu Jing, Lihui Liu, Ruijie Wang, Jinning Li, Tarek Abdelzaher, Hanghang Tong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Zero-shot causal learning
Hamed Nilforoshan, Michael Moor, Yusuf Roohani, Yining Chen, Anja Šurina, Michihiro Yasunaga, Sara Oblak, Jure Leskovec
Predicting how different interventions will causally affect a specific individual is important in a variety of domains such as personalized medicine, public policy, and online marketing. There are a large number of methods to predict the effect of an existing intervention based on historical data from individuals who received it. However, in many settings it is important to predict the effects of novel interventions (e.g., a newly invented drug), which these methods do not address.Here, we consider zero-shot causal learning: predicting the personalized effects of a novel intervention. We propose CaML, a causal meta-learning framework which formulates the personalized prediction of each intervention's effect as a task. CaML trains a single meta-model across thousands of tasks, each constructed by sampling an intervention, its recipients, and its nonrecipients. By leveraging both intervention information (e.g., a drug's attributes) and individual features (e.g., a patient's history), CaML is able to predict the personalized effects of novel interventions that do not exist at the time of training. Experimental results on real world datasets in large-scale medical claims and cell-line perturbations demonstrate the effectiveness of our approach. Most strikingly, CaML's zero-shot predictions outperform even strong baselines trained directly on data from the test interventions.
****
Learning Modulated Transformation in GANs
Ceyuan Yang, Qihang Zhang, Yinghao Xu, Jiapeng Zhu, Yujun Shen, Bo Dai
The success of style-based generators largely benefits from style modulation,which helps take care of the cross-instance variation within data. However, theinstance-wise stochasticity is typically introduced via regular convolution, wherekernels interact with features at some fixed locations, limiting its capacity formodeling geometric variation. To alleviate this problem, we equip the generatorin generative adversarial networks (GANs) with a plug-and-play module, termedas modulated transformation module (MTM). This module predicts spatial offsetsunder the control of latent codes, based on which the convolution operation canbe applied at variable locations for different instances, and hence offers the modelan additional degree of freedom to handle geometry deformation. Extensiveexperiments suggest that our approach can be faithfully generalized to variousgenerative tasks, including image generation, 3D-aware image synthesis, andvideo generation, and get compatible with state-of-the-art frameworks withoutany hyper-parameter tuning. It is noteworthy that, towards human generation onthe challenging TaiChi dataset, we improve the FID of StyleGAN3 from 21.36 to13.60, demonstrating the efficacy of learning modulated geometry transformation.Code and models are available at https://github.com/limbo0000/mtm.
****

Active Negative Loss Functions for Learning with Noisy Labels
Xichen Ye, Xiaoqiang Li, songmin dai, Tong Liu, Yan Sun, Weiqin Tong
Robust loss functions are essential for training deep neural networks in the presence of noisy labels. Some robust loss functions use Mean Absolute Error (MAE) as its necessary component. For example, the recently proposed Active Passive Loss (APL) uses MAE as its passive loss function. However, MAE treats every sample equally, slows down the convergence and can make training difficult. In this work, we propose a new class of theoretically robust passive loss functions different from MAE, namely Normalized Negative Loss Functions (NNLFs), which focus more on memorized clean samples. By replacing the MAE in APL with our proposed NNLFs, we improve APL and propose a new framework called Active Negative Loss (ANL). Experimental results on benchmark and real-world datasets demonstrate that the new set of loss functions created by our ANL framework can outperform state-of-the-art methods. The code is available athttps://github.com/Virusdoll/Active-Negative-Loss.
****
Compositional Generalization from First Principles
Thaddäus Wiedemer, Prasanna Mayilvahanan, Matthias Bethge, Wieland Brendel
Leveraging the compositional nature of our world to expedite learning and facilitate generalization is a hallmark of human perception. In machine learning, on the other hand, achieving compositional generalization has proven to be an elusive goal, even for models with explicit compositional priors. To get a better handle on compositional generalization, we here approach it from the bottom up: Inspired by identifiable representation learning, we investigate compositionality as a property of the data-generating process rather than the data itself. This reformulation enables us to derive mild conditions on only the support of the training distribution and the model architecture, which are sufficient for compositional generalization. We further demonstrate how our theoretical framework applies to real-world scenarios and validate our findings empirically. Our results set the stage for a principled theoretical study of compositional generalization.
****
PanoGRF: Generalizable Spherical Radiance Fields for Wide-baseline Panoramas
Zheng Chen, Yan-Pei Cao, Yuan-Chen Guo, Chen Wang, Ying Shan, Song-Hai Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Heat Diffusion Perspective on Geodesic Preserving Dimensionality Reduction
Guillaume Huguet, Alexander Tong, Edward De Brouwer, Yanlei Zhang, Guy Wolf, Ian Adelstein, Smita Krishnaswamy
Diffusion-based manifold learning methods have proven useful in representation learning and dimensionality reduction of modern high dimensional, high throughput, noisy datasets. Such datasets are especially present in fields like biology and physics. While it is thought that these methods preserve underlying manifold structure of data by learning a proxy for geodesic distances, no specific theoretical links have been established. Here, we establish such a link via results in Riemannian geometry explicitly connecting heat diffusion to manifold distances. In this process, we also formulate a more general heat kernel based manifold embedding method that we call heat geodesic embeddings. This novel perspective makes clearer the choices available in manifold learning and denoising. Results show that our method outperforms existing state of the art in preserving ground truth manifold distances, and preserving cluster structure in toy datasets. We also showcase our method on single cell RNA-sequencing datasets with both continuum and cluster structure, where our method enables interpolation of withheld timepoints of data. Finally, we show that parameters of our more general method can be configured to give results similar to PHATE (a state-of-the-art diffusion based manifold learning method) as well as SNE (an attraction/repulsion neighborhood based method that forms the basis of t-SNE).
****

Finite-Time Analysis of Single-Timescale Actor-Critic
Xuyang Chen, Lin Zhao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
VanillaNet: the Power of Minimalism in Deep Learning
Hanting Chen, Yunhe Wang, Jianyuan Guo, Dacheng Tao
At the heart of foundation models is the philosophy of "more is different", exemplified by the astonishing success in computer vision and natural language processing. However, the challenges of optimization and inherent complexity of transformer models call for a paradigm shift towards simplicity. In this study, we introduce VanillaNet, a neural network architecture that embraces elegance in design. By avoiding high depth, shortcuts, and intricate operations like self-attention, VanillaNet is refreshingly concise yet remarkably powerful. Each layer is carefully crafted to be compact and straightforward, with nonlinear activation functions pruned after training to restore the original architecture. VanillaNet overcomes the challenges of inherent complexity, making it ideal for resource-constrained environments. Its easy-to-understand and highly simplified architecture opens new possibilities for efficient deployment. Extensive experimentation demonstrates that VanillaNet delivers performance on par with renowned deep neural networks and vision transformers, showcasing the power of minimalism in deep learning. This visionary journey of VanillaNet has significant potential to redefine the landscape and challenge the status quo of foundation model, setting a new path for elegant and effective model design. Pre-trained models and codes are available at https://github.com/huawei-noah/VanillaNet and https://gitee.com/mindspore/models/tree/master/research/cv/vanillanet
****
Probabilistic inverse optimal control for non-linear partially observable systems disentangles perceptual uncertainty and behavioral costs
Dominik Straub, Matthias Schultheis, Heinz Koeppl, Constantin A. Rothkopf
Inverse optimal control can be used to characterize behavior in sequential decision-making tasks. Most existing work, however, is limited to fully observable or linear systems, or requires the action signals to be known. Here, we introduce a probabilistic approach to inverse optimal control for partially observable stochastic non-linear systems with unobserved action signals, which unifies previous approaches to inverse optimal control with maximum causal entropy formulations. Using an explicit model of the noise characteristics of the sensory and motor systems of the agent in conjunction with local linearization techniques, we derive an approximate likelihood function for the model parameters, which can be computed within a single forward pass. We present quantitative evaluations on stochastic and partially observable versions of two classic control tasks and two human behavioral tasks. Importantly, we show that our method can disentangle perceptual factors and behavioral costs despite the fact that epistemic and pragmatic actions are intertwined in sequential decision-making under uncertainty, such as in active sensing and active learning. The proposed method has broad applicability, ranging from imitation learning to sensorimotor neuroscience.
****
TIES-Merging: Resolving Interference When Merging Models
Prateek Yadav, Derek Tam, Leshem Choshen, Colin A. Raffel, Mohit Bansal
Transfer learning – i.e., further fine-tuning a pre-trained model on a downstream task – can confer significant advantages, including improved downstream performance, faster convergence, and better sample efficiency. These advantages have led to a proliferation of task-specific fine-tuned models, which typically can only perform a single task and do not benefit from one another. Recently, model merging techniques have emerged as a solution to combine multiple task-specific models into a single multitask model without performing additional training. However, existing merging methods often ignore the interference between parameters of different models, resulting in large performance drops when merging multiple mo

dels. In this paper, we demonstrate that prior merging techniques inadvertently lose valuable information due to two major sources of interference: (a) interference due to redundant parameter values and (b) disagreement on the sign of a given parameter's values across models. To address this, we propose our method, TrIm, Elect Sign & Merge (TIES-Merging), which introduces three novel steps when merging models: (1) resetting parameters that only changed a small amount during fine-tuning, (2) resolving sign conflicts, and (3) merging only the parameters that are in alignment with the final agreed-upon sign. We find that TIES-Merging outperforms existing methods in diverse settings covering a range of modalities, domains, number of tasks, model sizes, architectures, and fine-tuning settings. We further analyze the impact of different types of interference on model parameters, highlight the importance of signs, and show that estimating the signs using the validation data could further improve performance.
****

3D-IntPhys: Towards More Generalized 3D-grounded Visual Intuitive Physics under Challenging Scenes
Haotian Xue, Antonio Torralba, Josh Tenenbaum, Dan Yamins, Yunzhu Li, Hsiao-Yu Tung
Given a visual scene, humans have strong intuitions about how a scene can evolve over time under given actions. The intuition, often termed visual intuitive physics, is a critical ability that allows us to make effective plans to manipulate the scene to achieve desired outcomes without relying on extensive trial and error. In this paper, we present a framework capable of learning 3D-grounded visual intuitive physics models from videos of complex scenes with fluids. Our method is composed of a conditional Neural Radiance Field (NeRF)-style visual frontend and a 3D point-based dynamics prediction backend, using which we can impose strong relational and structural inductive bias to capture the structure of the underlying environment. Unlike existing intuitive point-based dynamics works that rely on the supervision of dense point trajectory from simulators, we relax the requirements and only assume access to multi-view RGB images and (imperfect) instance masks acquired using color prior. This enables the proposed model to handle scenarios where accurate point estimation and tracking are hard or impossible. We generate datasets including three challenging scenarios involving fluid, granular materials, and rigid objects in the simulation. The datasets do not include any dense particle information so most previous 3D-based intuitive physics pipelines can barely deal with that. We show our model can make long-horizon future predictions by learning from raw images and significantly outperforms models that do not employ an explicit 3D representation space. We also show that once trained, our model can achieve strong generalization in complex scenarios under extrapolate settings.
****

Entropy-based Training Methods for Scalable Neural Implicit Samplers
Weijian Luo, Boya Zhang, Zhihua Zhang
Efficiently sampling from un-normalized target distributions is a fundamental problem in scientific computing and machine learning. Traditional approaches such as Markov Chain Monte Carlo (MCMC) guarantee asymptotically unbiased samples from such distributions but suffer from computational inefficiency, particularly when dealing with high-dimensional targets, as they require numerous iterations to generate a batch of samples. In this paper, we introduce an efficient and scalable neural implicit sampler that overcomes these limitations. The implicit sampler can generate large batches of samples with low computational costs by leveraging a neural transformation that directly maps easily sampled latent vectors to target samples without the need for iterative procedures. To train the neural implicit samplers, we introduce two novel methods: the KL training method and the Fisher training method. The former method minimizes the Kullback-Leibler divergence, while the latter minimizes the Fisher divergence between the sampler and the target distributions. By employing the two training methods, we effectively optimize the neural implicit samplers to learn and generate from the desired target distribution. To demonstrate the effectiveness, efficiency, and scalability of our proposed samplers, we evaluate them on three sampling benchmarks with diffe

rent scales. These benchmarks include sampling from 2D targets, Bayesian inference, and sampling from high-dimensional energy-based models (EBMs). Notably, in the experiment involving high-dimensional EBMs, our sampler produces samples that are comparable to those generated by MCMC-based methods while being more than 100 times more efficient, showcasing the efficiency of our neural sampler. Besides the theoretical contributions and strong empirical performances, the proposed neural samplers and corresponding training methods will shed light on further research on developing efficient samplers for various applications beyond the ones explored in this study.
****
Direct Diffusion Bridge using Data Consistency for Inverse Problems
Hyungjin Chung, Jeongsol Kim, Jong Chul Ye
Diffusion model-based inverse problem solvers have shown impressive performance, but are limited in speed, mostly as they require reverse diffusion sampling starting from noise. Several recent works have tried to alleviate this problem by building a diffusion process, directly bridging the clean and the corrupted for specific inverse problems. In this paper, we first unify these existing works under the name Direct Diffusion Bridges (DDB), showing that while motivated by different theories, the resulting algorithms only differ in the choice of parameters. Then, we highlight a critical limitation of the current DDB framework, namely that it does not ensure data consistency. To address this problem, we propose a modified inference procedure that imposes data consistency without the need for fine-tuning. We term the resulting method data Consistent DDB (CDDB), which outperforms its inconsistent counterpart in terms of both perception and distortion metrics, thereby effectively pushing the Pareto-frontier toward the optimum. Our proposed method achieves state-of-the-art results on both evaluation criteria, showcasing its superiority over existing methods. Code is open-sourced here.
****
Mask Propagation for Efficient Video Semantic Segmentation
Yuetian Weng, Mingfei Han, Haoyu He, Mingjie Li, Lina Yao, Xiaojun Chang, Bohan Zhuang
Video Semantic Segmentation (VSS) involves assigning a semantic label to each pixel in a video sequence. Prior work in this field has demonstrated promising results by extending image semantic segmentation models to exploit temporal relationships across video frames; however, these approaches often incur significant computational costs. In this paper, we propose an efficient mask propagation framework for VSS, called MPVSS. Our approach first employs a strong query-based image segmentor on sparse key frames to generate accurate binary masks and class predictions. We then design a flow estimation module utilizing the learned queries to generate a set of segment-aware flow maps, each associated with a mask prediction from the key frame. Finally, the mask-flow pairs are warped to serve as the mask predictions for the non-key frames. By reusing predictions from key frames, we circumvent the need to process a large volume of video frames individually with resource-intensive segmentors, alleviating temporal redundancy and significantly reducing computational costs. Extensive experiments on VSPW and Cityscapes demonstrate that our mask propagation framework achieves SOTA accuracy and efficiency trade-offs. For instance, our best model with Swin-L backbone outperforms the SOTA MRCFA using MiT-B5 by 4.0% mIoU, requiring only 26% FLOPs on the VSPW dataset. Moreover, our framework reduces up to 4× FLOPs compared to the per-frame Mask2Former baseline with only up to 2% mIoU degradation on the Cityscapes validation set. Code is available at https://github.com/ziplab/MPVSS.
****
Private Distribution Learning with Public Data: The View from Sample Compression
Shai Ben-David, Alex Bie, Clément L Canonne, Gautam Kamath, Vikrant Singhal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ChessGPT: Bridging Policy Learning and Language Modeling

Xidong Feng, Yicheng Luo, Ziyan Wang, Hongrui Tang, Mengyue Yang, Kun Shao, David Mguni, Yali Du, Jun Wang
When solving decision-making tasks, humans typically depend on information from two key sources: (1) Historical policy data, which provides interaction replay from the environment, and (2) Analytical insights in natural language form, exposing the invaluable thought process or strategic considerations. Despite this, the majority of preceding research focuses on only one source: they either use historical replay exclusively to directly learn policy or value functions, or engaged in language model training utilizing mere language corpus. In this paper, we argue that a powerful autonomous agent should cover both sources. Thus, we propose ChessGPT, a GPT model bridging policy learning and language modeling by integrating data from these two sources in Chess games. Specifically, we build a large-scale game and language dataset related to chess. Leveraging the dataset, we showcase two model examples ChessCLIP and ChessGPT, integrating policy learning and language modeling. Finally, we propose a full evaluation framework for evaluating language model's chess ability. Experimental results validate our model and dataset's effectiveness. We open source our code, model, and dataset at https://github.com/waterhorse1/ChessGPT.
****
Fitting trees to ℓ_1-hyperbolic distances
Joon-Hyeok Yim, Anna Gilbert
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Robust Statistics for Simulation-based Inference under Model Misspecification
Daolang Huang, Ayush Bharti, Amauri Souza, Luigi Acerbi, Samuel Kaski
Simulation-based inference (SBI) methods such as approximate Bayesian computation (ABC), synthetic likelihood, and neural posterior estimation (NPE) rely on simulating statistics to infer parameters of intractable likelihood models. However, such methods are known to yield untrustworthy and misleading inference outcomes under model misspecification, thus hindering their widespread applicability. In this work, we propose the first general approach to handle model misspecification that works across different classes of SBI methods. Leveraging the fact that the choice of statistics determines the degree of misspecification in SBI, we introduce a regularized loss function that penalizes those statistics that increase the mismatch between the data and the model. Taking NPE and ABC as use cases, we demonstrate the superior performance of our method on high-dimensional time-series models that are artificially misspecified. We also apply our method to real data from the field of radio propagation where the model is known to be misspecified. We show empirically that the method yields robust inference in misspecified scenarios, whilst still being accurate when the model is well-specified.
****
Block-State Transformers
Jonathan Pilault, Mahan Fathi, Orhan Firat, Chris Pal, Pierre-Luc Bacon, Ross Goroshin
State space models (SSMs) have shown impressive results on tasks that require modeling long-range dependencies and efficiently scale to long sequences owing to their subquadratic runtime complexity.Originally designed for continuous signals, SSMs have shown superior performance on a plethora of tasks, in vision and audio; however, SSMs still lag Transformer performance in Language Modeling tasks.In this work, we propose a hybrid layer named Block-State Transformer (BST), that internally combines an SSM sublayer for long-range contextualization, and a Block Transformer sublayer for short-term representation of sequences.We study three different, and completely parallelizable, variants that integrate SSMs and block-wise attention.We show that our model outperforms similar Transformer-based architectures on language modeling perplexity and generalizes to longer sequences. In addition, the Block-State Transformer demonstrates a more than tenfold incr

ease in speed at the layer level compared to the Block-Recurrent Transformer whe n model parallelization is employed.
****
Explaining Predictive Uncertainty with Information Theoretic Shapley Values
David Watson, Joshua O'Hara, Niek Tax, Richard Mudd, Ido Guy
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Learning to Taste: A Multimodal Wine Dataset
Thoranna Bender, Simon Sørensen, Alireza Kashani, Kristjan Eldjarn Hjorleifsson, Grethe Hyldig, Søren Hauberg, Serge Belongie, Frederik Warburg
We present WineSensed, a large multimodal wine dataset for studying the relation s between visual perception, language, and flavor. The dataset encompasses 897k images of wine labels and 824k reviews of wines curated from the Vivino platform . It has over 350k unique vintages, annotated with year, region, rating, alcohol percentage, price, and grape composition. We obtained fine-grained flavor annot ations on a subset by conducting a wine-tasting experiment with 256 participants who were asked to rank wines based on their similarity in flavor, resulting in more than 5k pairwise flavor distances. We propose a low-dimensional concept emb edding algorithm that combines human experience with automatic machine similarit y kernels. We demonstrate that this shared concept embedding space improves upon separate embedding spaces for coarse flavor classification (alcohol percentage , country, grape, price, rating) and representing human perception of flavor.
****
CADet: Fully Self-Supervised Out-Of-Distribution Detection With Contrastive Lear ning
Charles Guille-Escuret, Pau Rodriguez, David Vazquez, Ioannis Mitliagkas, Joao M onteiro
Handling out-of-distribution (OOD) samples has become a major stake in the real-world deployment of machine learning systems. This work explores the use of self -supervised contrastive learning to the simultaneous detection of two types of O OD samples: unseen classes and adversarial perturbations. First, we pair self-su pervised contrastive learning with the maximum mean discrepancy (MMD) two-sample test. This approach enables us to robustly test whether two independent sets of samples originate from the same distribution, and we demonstrate its effectiven ess by discriminating between CIFAR-10 and CIFAR-10.1 with higher confidence tha n previous work. Motivated by this success, we introduce CADet (Contrastive Anom aly Detection), a novel method for OOD detection of single samples. CADet draws inspiration from MMD, but leverages the similarity between contrastive transform ations of a same sample. CADet outperforms existing adversarial detection method s in identifying adversarially perturbed samples on ImageNet and achieves compar able performance to unseen label detection methods on two challenging benchmarks : ImageNet-O and iNaturalist. Significantly, CADet is fully self-supervised and requires neither labels for in-distribution samples nor access to OOD examples.
****
PriorBand: Practical Hyperparameter Optimization in the Age of Deep Learning
Neeratyoy Mallik, Edward Bergman, Carl Hvarfner, Danny Stoll, Maciej Janowski, M arius Lindauer, Luigi Nardi, Frank Hutter
Hyperparameters of Deep Learning (DL) pipelines are crucial for their downstream performance. While a large number of methods for Hyperparameter Optimization (H PO) have been developed, their incurred costs are often untenable for modern DL. Consequently, manual experimentation is still the most prevalent approach to opt imize hyperparameters, relying on the researcher's intuition, domain knowledge, and cheap preliminary explorations.To resolve this misalignment between HPO algo rithms and DL researchers, we propose PriorBand, an HPO algorithm tailored to DL , able to utilize both expert beliefs and cheap proxy tasks. Empirically, we dem onstrate PriorBand's efficiency across a range of DL benchmarks and show its gai ns under informative expert input and robustness against poor expert beliefs.

****
Towards Efficient Image Compression Without Autoregressive Models
Muhammad Salman Ali, Yeongwoong Kim, Maryam Qamar, Sung-Chang Lim, Donghyun Kim, Chaoning Zhang, Sung-Ho Bae, Hui Yong Kim
Recently, learned image compression (LIC) has garnered increasing interest with its rapidly improving performance surpassing conventional codecs. A key ingredient of LIC is a hyperprior-based entropy model, where the underlying joint probability of the latent image features is modeled as a product of Gaussian distributions from each latent element. Since latents from the actual images are not spatially independent, autoregressive (AR) context based entropy models were proposed to handle the discrepancy between the assumed distribution and the actual distribution. Though the AR-based models have proven effective, the computational complexity is significantly increased due to the inherent sequential nature of the algorithm. In this paper, we present a novel alternative to the AR-based approach that can provide a significantly better trade-off between performance and complexity. To minimize the discrepancy, we introduce a correlation loss that forces the latents to be spatially decorrelated and better fitted to the independent probability model. Our correlation loss is proved to act as a general plug-in for the hyperprior (HP) based learned image compression methods. The performance gain from our correlation loss is ‘free’ in terms of computation complexity for both inference time and decoding time. To our knowledge, our method gives the best trade-off between the complexity and performance: combined with the Checkerboard-CM, it attains 90% and when combined with ChARM-CM, it attains 98% of the AR-based BD-Rate gains yet is around 50 times and 30 times faster than AR-based methods respectively
****
De novo Drug Design using Reinforcement Learning with Multiple GPT Agents
Xiuyuan Hu, Guoqing Liu, Yang Zhao, Hao Zhang
De novo drug design is a pivotal issue in pharmacology and a new area of focus in AI for science research. A central challenge in this field is to generate molecules with specific properties while also producing a wide range of diverse candidates. Although advanced technologies such as transformer models and reinforcement learning have been applied in drug design, their potential has not been fully realized. Therefore, we propose MolRL-MGPT, a reinforcement learning algorithm with multiple GPT agents for drug molecular generation. To promote molecular diversity, we encourage the agents to collaborate in searching for desirable molecules in diverse directions. Our algorithm has shown promising results on the GuacaMol benchmark and exhibits efficacy in designing inhibitors against SARS-CoV-2 protein targets. The codes are available at: https://github.com/HXYfighter/MolRL-MGPT.
****
Pointwise uncertainty quantification for sparse variational Gaussian process regression with a Brownian motion prior
Luke Travis, Kolyan Ray
We study pointwise estimation and uncertainty quantification for a sparse variational Gaussian process method with eigenvector inducing variables. For a rescaled Brownian motion prior, we derive theoretical guarantees and limitations for the frequentist size and coverage of pointwise credible sets. For sufficiently many inducing variables, we precisely characterize the asymptotic frequentist coverage, deducing when credible sets from this variational method are conservative and when overconfident/misleading. We numerically illustrate the applicability of our results and discuss connections with other common Gaussian process priors.
****
Few-shot Generation via Recalling Brain-Inspired Episodic-Semantic Memory
Zhibin Duan, Zhiyi Lv, Chaojie Wang, Bo Chen, Bo An, Mingyuan Zhou
Aimed at adapting a generative model to a novel generation task with only a few given data samples, the capability of few-shot generation is crucial for many real-world applications with limited data, \emph{e.g.}, artistic domains.Instead of training from scratch, recent works tend to leverage the prior knowledge stored in previous datasets, which is quite similar to the memory mechanism of human

intelligence, but few of these works directly imitate the memory-recall mechanism that humans make good use of in accomplishing creative tasks, \emph{e.g.}, painting and writing.Inspired by the memory mechanism of human brain, in this work, we carefully design a variational structured memory module (VSM), which can simultaneously store both episodic and semantic memories to assist existing generative models efficiently recall these memories during sample generation.Meanwhile, we introduce a bionic memory updating strategy for the conversion between episodic and semantic memories, which can also model the uncertainty during conversion.Then, we combine the developed VSM with various generative models under the Bayesian framework, and evaluate these memory-augmented generative models with few-shot generation tasks, demonstrating the effectiveness of our methods.
****
Balancing memorization and generalization in RNNs for high performance brain-machine Interfaces
Joseph Costello, Hisham Temmar, Luis Cubillos, Matthew Mender, Dylan Wallace, Matt Willsey, Parag Patil, Cynthia Chestek
Brain-machine interfaces (BMIs) can restore motor function to people with paralysis but are currently limited by the accuracy of real-time decoding algorithms. Recurrent neural networks (RNNs) using modern training techniques have shown promise in accurately predicting movements from neural signals but have yet to be rigorously evaluated against other decoding algorithms in a closed-loop setting. Here we compared RNNs to other neural network architectures in real-time, continuous decoding of finger movements using intracortical signals from nonhuman primates. Across one and two finger online tasks, LSTMs (a type of RNN) outperformed convolutional and transformer-based neural networks, averaging 18% higher throughput than the convolution network. On simplified tasks with a reduced movement set, RNN decoders were allowed to memorize movement patterns and matched able-bodied control. Performance gradually dropped as the number of distinct movements increased but did not go below fully continuous decoder performance. Finally, in a two-finger task where one degree-of-freedom had poor input signals, we recovered functional control using RNNs trained to act both like a movement classifier and continuous decoder. Our results suggest that RNNs can enable functional real-time BMI control by learning and generating accurate movement patterns.
****
Saddle-to-Saddle Dynamics in Diagonal Linear Networks
Scott Pesme, Nicolas Flammarion
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Encoding Human Behavior in Information Design through Deep Learning
Guanghui Yu, Wei Tang, Saumik Narayanan, Chien-Ju Ho
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Collaboratively Learning Linear Models with Structured Missing Data
Chen Cheng, Gary Cheng, John C. Duchi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Generating Behaviorally Diverse Policies with Latent Diffusion Models
Shashank Hegde, Sumeet Batra, K.R. Zentner, Gaurav Sukhatme
Recent progress in Quality Diversity Reinforcement Learning (QD-RL) has enabled learning a collection of behaviorally diverse, high performing policies. However, these methods typically involve storing thousands of policies, which results i

n high space-complexity and poor scaling to additional behaviors. Condensing the archive into a single model while retaining the performance and coverage of the original collection of policies has proved challenging. In this work, we propose using diffusion models to distill the archive into a single generative model over policy parameters. We show that our method achieves a compression ratio of 13x while recovering 98% of the original rewards and 89% of the original humanoid archive coverage. Further, the conditioning mechanism of diffusion models allows for flexibly selecting and sequencing behaviors, including using language. Project website: https://sites.google.com/view/policydiffusion/home.
****

Incentives in Private Collaborative Machine Learning

Rachael Sim, Yehong Zhang, Nghia Hoang, Xinyi Xu, Bryan Kian Hsiang Low, Patrick Jaillet

Collaborative machine learning involves training models on data from multiple parties but must incentivize their participation. Existing data valuation methods fairly value and reward each party based on shared data or model parameters but neglect the privacy risks involved. To address this, we introduce differential privacy (DP) as an incentive. Each party can select its required DP guarantee and perturb its sufficient statistic (SS) accordingly. The mediator values the perturbed SS by the Bayesian surprise it elicits about the model parameters. As our valuation function enforces a privacy-valuation trade-off, parties are deterred from selecting excessive DP guarantees that reduce the utility of the grand coalition's model. Finally, the mediator rewards each party with different posterior samples of the model parameters. Such rewards still satisfy existing incentives like fairness but additionally preserve DP and a high similarity to the grand coalition's posterior. We empirically demonstrate the effectiveness and practicality of our approach on synthetic and real-world datasets.
****

VideoComposer: Compositional Video Synthesis with Motion Controllability

Xiang Wang, Hangjie Yuan, Shiwei Zhang, Dayou Chen, Jiuniu Wang, Yingya Zhang, Yujun Shen, Deli Zhao, Jingren Zhou

The pursuit of controllability as a higher standard of visual content creation has yielded remarkable progress in customizable image synthesis. However, achieving controllable video synthesis remains challenging due to the large variation of temporal dynamics and the requirement of cross-frame temporal consistency. Based on the paradigm of compositional generation, this work presents VideoComposer that allows users to flexibly compose a video with textual conditions, spatial conditions, and more importantly temporal conditions. Specifically, considering the characteristic of video data, we introduce the motion vector from compressed videos as an explicit control signal to provide guidance regarding temporal dynamics. In addition, we develop a Spatio-Temporal Condition encoder (STC-encoder) that serves as a unified interface to effectively incorporate the spatial and temporal relations of sequential inputs, with which the model could make better use of temporal conditions and hence achieve higher inter-frame consistency. Extensive experimental results suggest that VideoComposer is able to control the spatial and temporal patterns simultaneously within a synthesized video in various forms, such as text description, sketch sequence, reference video, or even simply hand-crafted motions. The code and models are publicly available athttps://videocomposer.github.io.
****

Look Beneath the Surface: Exploiting Fundamental Symmetry for Sample-Efficient Offline RL

Peng Cheng, Xianyuan Zhan, zhihao wu, Wenjia Zhang, Youfang Lin, Shou cheng Song, Han Wang, Li Jiang

Offline reinforcement learning (RL) offers an appealing approach to real-world tasks by learning policies from pre-collected datasets without interacting with the environment. However, the performance of existing offline RL algorithms heavily depends on the scale and state-action space coverage of datasets. Real-world data collection is often expensive and uncontrollable, leading to small and narrowly covered datasets and posing significant challenges for practical deployment

s of offline RL. In this paper, we provide a new insight that leveraging the fundamental symmetry of system dynamics can substantially enhance offline RL performance under small datasets. Specifically, we propose a Time-reversal symmetry (T-symmetry) enforced Dynamics Model (TDM), which establishes consistency between a pair of forward and reverse latent dynamics. TDM provides both well-behaved representations for small datasets and a new reliability measure for OOD samples based on compliance with the T-symmetry. These can be readily used to construct a new offline RL algorithm (TSRL) with less conservative policy constraints and a reliable latent space data augmentation procedure. Based on extensive experiments, we find TSRL achieves great performance on small benchmark datasets with as few as 1% of the original samples, which significantly outperforms the recent offline RL algorithms in terms of data efficiency and generalizability. Code is available at:https://github.com/pcheng2/TSRL
****

Initialization-Dependent Sample Complexity of Linear Predictors and Neural Networks

Roey Magen, Ohad Shamir

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Incentivizing Honesty among Competitors in Collaborative Learning and Optimization

Florian E. Dorner, Nikola Konstantinov, Georgi Pashaliev, Martin Vechev

Collaborative learning techniques have the potential to enable training machine learning models that are superior to models trained on a single entity’s data. However, in many cases, potential participants in such collaborative schemes are competitors on a downstream task, such as firms that each aim to attract customers by providing the best recommendations. This can incentivize dishonest updates that damage other participants' models, potentially undermining the benefits of collaboration. In this work, we formulate a game that models such interactions and study two learning tasks within this framework: single-round mean estimation and multi-round SGD on strongly-convex objectives. For a natural class of player actions, we show that rational clients are incentivized to strongly manipulate their updates, preventing learning. We then propose mechanisms that incentivize honest communication and ensure learning quality comparable to full cooperation. Lastly, we empirically demonstrate the effectiveness of our incentive scheme on a standard non-convex federated learning benchmark. Our work shows that explicitly modeling the incentives and actions of dishonest clients, rather than assuming them malicious, can enable strong robustness guarantees for collaborative learning.
****

SNAP: Self-Supervised Neural Maps for Visual Positioning and Semantic Understanding

Paul-Edouard Sarlin, Eduard Trulls, Marc Pollefeys, Jan Hosang, Simon Lynen

Semantic 2D maps are commonly used by humans and machines for navigation purposes, whether it's walking or driving. However, these maps have limitations: they lack detail, often contain inaccuracies, and are difficult to create and maintain, especially in an automated fashion. Can we use raw imagery to automatically create better maps that can be easily interpreted by both humans and machines? We introduce SNAP, a deep network that learns rich 2D neural maps from ground-level and overhead images. We train our model to align neural maps estimated from different inputs, supervised only with camera poses over tens of millions of StreetView images. SNAP can resolve the location of challenging image queries beyond the reach of traditional methods, outperforming the state of the art in localization by a large margin. Moreover, our neural maps encode not only geometry and appearance but also high-level semantics, discovered without explicit supervision. This enables effective pre-training for data-efficient semantic scene understanding, with the potential to unlock cost-efficient creation of more detailed map

s.
****
Waymax: An Accelerated, Data-Driven Simulator for Large-Scale Autonomous Driving Research
Cole Gulino, Justin Fu, Wenjie Luo, George Tucker, Eli Bronstein, Yiren Lu, Jean Harb, Xinlei Pan, Yan Wang, Xiangyu Chen, John Co-Reyes, Rishabh Agarwal, Rebecca Roelofs, Yao Lu, Nico Montali, Paul Mougin, Zoey Yang, Brandyn White, Aleksandra Faust, Rowan McAllister, Dragomir Anguelov, Benjamin Sapp
Simulation is an essential tool to develop and benchmark autonomous vehicle planning software in a safe and cost-effective manner. However, realistic simulation requires accurate modeling of multi-agent interactive behaviors to be trustworthy, behaviors which can be highly nuanced and complex. To address these challenges, we introduce Waymax, a new data-driven simulator for autonomous driving in multi-agent scenes, designed for large-scale simulation and testing. Waymax uses publicly-released, real-world driving data (e.g., the Waymo Open Motion Dataset) to initialize or play back a diverse set of multi-agent simulated scenarios. It runs entirely on hardware accelerators such as TPUs/GPUs and supports in-graph simulation for training, making it suitable for modern large-scale, distributed machine learning workflows. To support online training and evaluation, Waymax includes several learned and hard-coded behavior models that allow for realistic interaction within simulation. To supplement Waymax, we benchmark a suite of popular imitation and reinforcement learning algorithms with ablation studies on different design decisions, where we highlight the effectiveness of routes as guidance for planning agents and the ability of RL to overfit against simulated agents.
****
Equal Opportunity of Coverage in Fair Regression
Fangxin Wang, Lu Cheng, Ruocheng Guo, Kay Liu, Philip S Yu
We study fair machine learning (ML) under predictive uncertainty to enable reliable and trustworthy decision-making. The seminal work of 'equalized coverage' proposed an uncertainty-aware fairness notion. However, it does not guarantee equal coverage rates across more fine-grained groups (e.g., low-income females) conditioning on the true label and is biased in the assessment of uncertainty. To tackle these limitations, we propose a new uncertainty-aware fairness -- Equal Opportunity of Coverage (EOC) -- that aims to achieve two properties: (1) coverage rates for different groups with similar outcomes are close, and (2) the coverage rate for the entire population remains at a predetermined level. Further, the prediction intervals should be narrow to be informative. We propose Binned Fair Quantile Regression (BFQR), a distribution-free post-processing method to improve EOC with reasonable width for any trained ML models. It first calibrates a hold-out set to bound deviation from EOC, then leverages conformal prediction to maintain EOC on a test set, meanwhile optimizing prediction interval width. Experimental results demonstrate the effectiveness of our method in improving EOC.
****
Nonparametric Teaching for Multiple Learners
Chen Zhang, Xiaofeng Cao, Weiyang Liu, Ivor Tsang, James Kwok
We study the problem of teaching multiple learners simultaneously in the nonparametric iterative teaching setting, where the teacher iteratively provides examples to the learner for accelerating the acquisition of a target concept. This problem is motivated by the gap between current single-learner teaching setting and the real-world scenario of human instruction where a teacher typically imparts knowledge to multiple students. Under the new problem formulation, we introduce a novel framework -- Multi-learner Nonparametric Teaching (MINT). In MINT, the teacher aims to instruct multiple learners, with each learner focusing on learning a scalar-valued target model. To achieve this, we frame the problem as teaching a vector-valued target model and extend the target model space from a scalar-valued reproducing kernel Hilbert space used in single-learner scenarios to a vector-valued space. Furthermore, we demonstrate that MINT offers significant teaching speed-up over repeated single-learner teaching, particularly when the multiple learners can communicate with each other. Lastly, we conduct extensive experi

ments to validate the practicality and efficiency of MINT.
****
EvoPrompting: Language Models for Code-Level Neural Architecture Search
Angelica Chen, David Dohan, David So
Given the recent impressive accomplishments of language models (LMs) for code generation, we explore the use of LMs as general adaptive mutation and crossover operators for an evolutionary neural architecture search (NAS) algorithm.While NAS still proves too difficult a task for LMs to succeed at solely through prompting, we find that the combination of evolutionary prompt engineering with soft prompt-tuning, a method we term EvoPrompting, consistently finds diverse and high performing models. We first demonstrate that EvoPrompting is effective on the computationally efficient MNIST-1D dataset, where EvoPrompting produces convolutional architecture variants that outperform both those designed by human experts and naive few-shot prompting in terms of accuracy and model size. We then apply our method to searching for graph neural networks on the CLRS Algorithmic Reasoning Benchmark, where EvoPrompting is able to design novel architectures that outperform current state-of-the-art models on 21 out of 30 algorithmic reasoning tasks while maintaining similar model size. EvoPrompting is successful at designing accurate and efficient neural network architectures across a variety of machine learning tasks, while also being general enough for easy adaptation to other tasks beyond neural network design.
****
Global-correlated 3D-decoupling Transformer for Clothed Avatar Reconstruction
Zechuan Zhang, Li Sun, Zongxin Yang, Ling Chen, Yi Yang
Reconstructing 3D clothed human avatars from single images is a challenging task, especially when encountering complex poses and loose clothing. Current methods exhibit limitations in performance, largely attributable to their dependence on insufficient 2D image features and inconsistent query methods. Owing to this, we present the Global-correlated 3D-decoupling Transformer for clothed Avatar reconstruction (GTA), a novel transformer-based architecture that reconstructs clothed human avatars from monocular images. Our approach leverages transformer architectures by utilizing a Vision Transformer model as an encoder for capturing global-correlated image features. Subsequently, our innovative 3D-decoupling decoder employs cross-attention to decouple tri-plane features, using learnable embeddings as queries for cross-plane generation. To effectively enhance feature fusion with the tri-plane 3D feature and human body prior, we propose a hybrid prior fusion strategy combining spatial and prior-enhanced queries, leveraging the benefits of spatial localization and human body prior knowledge. Comprehensive experiments on CAPE and THuman2.0 datasets illustrate that our method outperforms state-of-the-art approaches in both geometry and texture reconstruction, exhibiting high robustness to challenging poses and loose clothing, and producing higher-resolution textures. Codes are available at https://github.com/River-Zhang/GTA.
****
TopP&R: Robust Support Estimation Approach for Evaluating Fidelity and Diversity in Generative Models
Pum Jun Kim, Yoojin Jang, Jisu Kim, Jaejun Yoo
We propose a robust and reliable evaluation metric for generative models called Topological Precision and Recall (TopP&R, pronounced “topper”), which systematically estimates supports by retaining only topologically and statistically significant features with a certain level of confidence. Existing metrics, such as Inception Score (IS), Frechet Inception Distance (FID), and various Precision and Recall (P&R) variants, rely heavily on support estimates derived from sample features. However, the reliability of these estimates has been overlooked, even though the quality of the evaluation hinges entirely on their accuracy. In this paper, we demonstrate that current methods not only fail to accurately assess sample quality when support estimation is unreliable, but also yield inconsistent results. In contrast, TopP&R reliably evaluates the sample quality and ensures statistical consistency in its results. Our theoretical and experimental findings reveal that TopP&R provides a robust evaluation, accurately capturing the true tren

d of change in samples, even in the presence of outliers and non-independent and identically distributed (Non-IID) perturbations where other methods result in inaccurate support estimations. To our knowledge, TopP&R is the first evaluation metric specifically focused on the robust estimation of supports, offering statistical consistency under noise conditions.
****
A Unified Detection Framework for Inference-Stage Backdoor Defenses
Xun Xian, Ganghua Wang, Jayanth Srinivasa, Ashish Kundu, Xuan Bi, Mingyi Hong, Jie Ding
Backdoor attacks involve inserting poisoned samples during training, resulting in a model containing a hidden backdoor that can trigger specific behaviors without impacting performance on normal samples. These attacks are challenging to detect, as the backdoored model appears normal until activated by the backdoor trigger, rendering them particularly stealthy. In this study, we devise a unified inference-stage detection framework to defend against backdoor attacks. We first rigorously formulate the inference-stage backdoor detection problem, encompassing various existing methods, and discuss several challenges and limitations. We then propose a framework with provable guarantees on the false positive rate or the probability of misclassifying a clean sample. Further, we derive the most powerful detection rule to maximize the detection power, namely the rate of accurately identifying a backdoor sample, given a false positive rate under classical learning scenarios. Based on the theoretically optimal detection rule, we suggest a practical and effective approach for real-world applications based on the latent representations of backdoored deep nets. We extensively evaluate our method on 14 different backdoor attacks using Computer Vision (CV) and Natural Language Processing (NLP) benchmark datasets. The experimental findings align with our theoretical results. We significantly surpass the state-of-the-art methods, e.g., up to 300\% improvement on the detection power as evaluated by AUCROC, over the state-of-the-art defense against advanced adaptive backdoor attacks.
****
Non-Stationary Bandits with Auto-Regressive Temporal Dependency
Qinyi Chen, Negin Golrezaei, Djallel Bouneffouf
Traditional multi-armed bandit (MAB) frameworks, predominantly examined under stochastic or adversarial settings, often overlook the temporal dynamics inherent in many real-world applications such as recommendation systems and online advertising. This paper introduces a novel non-stationary MAB framework that captures the temporal structure of these real-world dynamics through an auto-regressive (AR) reward structure. We propose an algorithm that integrates two key mechanisms: (i) an alternation mechanism adept at leveraging temporal dependencies to dynamically balance exploration and exploitation, and (ii) a restarting mechanism designed to discard out-of-date information. Our algorithm achieves a regret upper bound that nearly matches the lower bound, with regret measured against a robust dynamic benchmark. Finally, via a real-world case study on tourism demand prediction, we demonstrate both the efficacy of our algorithm and the broader applicability of our techniques to more complex, rapidly evolving time series.
****
Globally solving the Gromov-Wasserstein problem for point clouds in low dimensional Euclidean spaces
Martin Ryner, Jan Kronqvist, Johan Karlsson
This paper presents a framework for computing the Gromov-Wasserstein problem between two sets of points in low dimensional spaces, where the discrepancy is the squared Euclidean norm.The Gromov-Wasserstein problem is a generalization of the optimal transport problem that finds the assignment between two sets preserving pairwise distances as much as possible. This can be used to quantify the similarity between two formations or shapes, a common problem in AI and machine learning.The problem can be formulated as a Quadratic Assignment Problem (QAP), which is in general computationally intractable even for small problems. Our framework addresses this challenge by reformulating the QAP as an optimization problem with a low-dimensional domain, leveraging the fact that the problem can be expressed as a concave quadratic optimization problem with low rank. The method scales

well with the number of points, and it can be used to find the global solution for large-scale problems with thousands of points.We compare the computational complexity of our approach with state-of-the-art methods on synthetic problems and apply it to a near-symmetrical problem which is of particular interest in computational biology.
****
Combinatorial Optimization with Policy Adaptation using Latent Space Search
Felix Chalumeau, Shikha Surana, Clément Bonnet, Nathan Grinsztajn, Arnu Pretorius, Alexandre Laterre, Tom Barrett
Combinatorial Optimization underpins many real-world applications and yet, designing performant algorithms to solve these complex, typically NP-hard, problems remains a significant research challenge. Reinforcement Learning (RL) provides a versatile framework for designing heuristics across a broad spectrum of problem domains. However, despite notable progress, RL has not yet supplanted industrial solvers as the go-to solution. Current approaches emphasize pre-training heuristics that construct solutions, but often rely on search procedures with limited variance, such as stochastically sampling numerous solutions from a single policy, or employing computationally expensive fine-tuning of the policy on individual problem instances. Building on the intuition that performant search at inference time should be anticipated during pre-training, we propose COMPASS, a novel RL approach that parameterizes a distribution of diverse and specialized policies conditioned on a continuous latent space. We evaluate COMPASS across three canonical problems - Travelling Salesman, Capacitated Vehicle Routing, and Job-Shop Scheduling - and demonstrate that our search strategy (i) outperforms state-of-the-art approaches in 9 out of 11 standard benchmarking tasks and (ii) generalizes better, surpassing all other approaches on a set of 18 procedurally transformed instance distributions.
****
SubseasonalClimateUSA: A Dataset for Subseasonal Forecasting and Benchmarking
Soukayna Mouatadid, Paulo Orenstein, Genevieve Flaspohler, Miruna Oprescu, Judah Cohen, Franklyn Wang, Sean Knight, Maria Geogdzhayeva, Sam Levang, Ernest Fraenkel, Lester Mackey
Subseasonal forecasting of the weather two to six weeks in advance is critical for resource allocation and climate adaptation but poses many challenges for the forecasting community. At this forecast horizon, physics-based dynamical models have limited skill, and the targets for prediction depend in a complex manner on both local weather variables and global climate variables. Recently, machine learning methods have shown promise in advancing the state of the art but only at the cost of complex data curation, integrating expert knowledge with aggregation across multiple relevant data sources, file formats, and temporal and spatial resolutions.To streamline this process and accelerate future development, we introduce SubseasonalClimateUSA, a curated dataset for training and benchmarking subseasonal forecasting models in the United States. We use this dataset to benchmark a diverse suite of models, including operational dynamical models, classical meteorological baselines, and ten state-of-the-art machine learning and deep learning-based methods from the literature. Overall, our benchmarks suggest simple and effective ways to extend the accuracy of current operational models. SubseasonalClimateUSA is regularly updated and accessible via the https://github.com/microsoft/subseasonal_data/ Python package.
****
RenderMe-360: A Large Digital Asset Library and Benchmarks Towards High-fidelity Head Avatars
Dongwei Pan, Long Zhuo, Jingtan Piao, Huiwen Luo, Wei Cheng, Yuxin WANG, Siming Fan, Shengqi Liu, Lei Yang, Bo Dai, Ziwei Liu, Chen Change Loy, Chen Qian, Wayne Wu, Dahua Lin, Kwan-Yee Lin
Synthesizing high-fidelity head avatars is a central problem for computer vision and graphics. While head avatar synthesis algorithms have advanced rapidly, the best ones still face great obstacles in real-world scenarios. One of the vital causes is the inadequate datasets -- 1) current public datasets can only support researchers to explore high-fidelity head avatars in one or two task direction

s; 2) these datasets usually contain digital head assets with limited data volume, and narrow distribution over different attributes, such as expressions, ages, and accessories. In this paper, we present RenderMe-360, a comprehensive 4D human head dataset to drive advance in head avatar algorithms across different scenarios. It contains massive data assets, with 243+ million complete head frames and over 800k video sequences from 500 different identities captured by multi-view cameras at 30 FPS. It is a large-scale digital library for head avatars with three key attributes: 1) High Fidelity: all subjects are captured in 360 degrees via 60 synchronized, high-resolution 2K cameras. 2) High Diversity: The collected subjects vary from different ages, eras, ethnicities, and cultures, providing abundant materials with distinctive styles in appearance and geometry. Moreover, each subject is asked to perform various dynamic motions, such as expressions and head rotations, which further extend the richness of assets. 3) Rich Annotations: the dataset provides annotations with different granularities: cameras' parameters, background matting, scan, 2D/3D facial landmarks, FLAME fitting, and text description.   Based on the dataset, we build a comprehensive benchmark for head avatar research, with 16 state-of-the-art methods performed on five main tasks: novel view synthesis, novel expression synthesis, hair rendering, hair editing, and talking head generation. Our experiments uncover the strengths and flaws of state-of-the-art methods. RenderMe-360 opens the door for future exploration in modern head avatars. All of the data, code, and models will be publicly available at https://renderme-360.github.io/.
****
Amazon-M2: A Multilingual Multi-locale Shopping Session Dataset for Recommendation and Text Generation
Wei Jin, Haitao Mao, Zheng Li, Haoming Jiang, Chen Luo, Hongzhi Wen, Haoyu Han, Hanqing Lu, Zhengyang Wang, Ruirui Li, Zhen Li, Monica Cheng, Rahul Goutam, Haiyang Zhang, Karthik Subbian, Suhang Wang, Yizhou Sun, Jiliang Tang, Bing Yin, Xianfeng Tang
Modeling customer shopping intentions is a crucial task for e-commerce, as it directly impacts user experience and engagement. Thus, accurately understanding customer preferences is essential for providing personalized recommendations. Session-based recommendation, which utilizes customer session data to predict their next interaction, has become increasingly popular. However, existing session datasets have limitations in terms of item attributes, user diversity, and dataset scale. As a result, they cannot comprehensively capture the spectrum of user behaviors and preferences.To bridge this gap, we present the Amazon Multilingual Multi-locale Shopping Session Dataset, namely Amazon-M2. It is the first multilingual dataset consisting of millions of user sessions from six different locales, where the major languages of products are English, German, Japanese, French, Italian, and Spanish.Remarkably, the dataset can help us enhance personalization and understanding of user preferences, which can benefit various existing tasks as well as enable new tasks. To test the potential of the dataset, we introduce three tasks in this work:(1) next-product recommendation, (2) next-product recommendation with domain shifts, and (3) next-product title generation.With the above tasks, we benchmark a range of algorithms on our proposed dataset, drawing new insights for further research and practice. In addition, based on the proposed dataset and tasks, we hosted a competition in the KDD CUP 2023 https://www.aicrowd.com/challenges/amazon-kdd-cup-23-multilingual-recommendation-challenge and have attracted thousands of users and submissions. The winning solutions and the associated workshop can be accessed at our website~https://kddcup23.github.io/.
****
Adversarial Resilience in Sequential Prediction via Abstention
Surbhi Goel, Steve Hanneke, Shay Moran, Abhishek Shetty
We study the problem of sequential prediction in the stochastic setting with an adversary that is allowed to inject clean-label adversarial (or out-of-distribution) examples. Algorithms designed to handle purely stochastic data tend to fail in the presence of such adversarial examples, often leading to erroneous predictions. This is undesirable in many high-stakes applications such as medical recommendations, where abstaining from predictions on adversarial examples is prefer

able to misclassification. On the other hand, assuming fully adversarial data leads to very pessimistic bounds that are often vacuous in practice. To move away from these pessimistic guarantees, we propose a new model of sequential prediction that sits between the purely stochastic and fully adversarial settings by allowing the learner to abstain from making a prediction at no cost on adversarial examples, thereby asking the learner to make predictions with certainty. Assuming access to the marginal distribution on the non-adversarial examples, we design a learner whose error scales with the VC dimension (mirroring the stochastic setting) of the hypothesis class, as opposed to the Littlestone dimension which characterizes the fully adversarial setting. Furthermore, we design learners for VC dimension~1 classes and the class of axis-aligned rectangles, which work even in the absence of access to the marginal distribution. Our key technical contribution is a novel measure for quantifying uncertainty for learning VC classes, which may be of independent interest.
****

Simplicity Bias in 1-Hidden Layer Neural Networks
Depen Morwani, Jatin Batra, Prateek Jain, Praneeth Netrapalli
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

AVOIDDS: Aircraft Vision-based Intruder Detection Dataset and Simulator
Elysia Smyers, Sydney Katz, Anthony Corso, Mykel J Kochenderfer
Designing robust machine learning systems remains an open problem, and there is a need for benchmark problems that cover both environmental changes and evaluation on a downstream task. In this work, we introduce AVOIDDS, a realistic object detection benchmark for the vision-based aircraft detect-and-avoid problem. We provide a labeled dataset consisting of 72,000 photorealistic images of intruder aircraft with various lighting conditions, weather conditions, relative geometries, and geographic locations. We also provide an interface that evaluates trained models on slices of this dataset to identify changes in performance with respect to changing environmental conditions. Finally, we implement a fully-integrated, closed-loop simulator of the vision-based detect-and-avoid problem to evaluate trained models with respect to the downstream collision avoidance task. This benchmark will enable further research in the design of robust machine learning systems for use in safety-critical applications. The AVOIDDS dataset and code are publicly available at https://purl.stanford.edu/hj293cv5980 and https://github.com/sisl/VisionBasedAircraftDAA, respectively.
****

Temporally Disentangled Representation Learning under Unknown Nonstationarity
Xiangchen Song, Weiran Yao, Yewen Fan, Xinshuai Dong, Guangyi Chen, Juan Carlos Niebles, Eric Xing, Kun Zhang
In unsupervised causal representation learning for sequential data with time-delayed latent causal influences, strong identifiability results for the disentanglement of causally-related latent variables have been established in stationary settings by leveraging temporal structure.However, in nonstationary setting, existing work only partially addressed the problem by either utilizing observed auxiliary variables (e.g., class labels and/or domain indexes) as side information or assuming simplified latent causal dynamics. Both constrain the method to a limited range of scenarios.In this study, we further explored the Markov Assumption under time-delayed causally related process in nonstationary setting and showed that under mild conditions, the independent latent components can be recovered from their nonlinear mixture up to a permutation and a component-wise transformation, without the observation of auxiliary variables. We then introduce NCTRL, a principled estimation framework, to reconstruct time-delayed latent causal variables and identify their relations from measured sequential data only.Empirical evaluations demonstrated the reliable identification of time-delayed latent causal influences, with our methodology substantially outperforming existing baselines that fail to exploit the nonstationarity adequately and then, consequently, c

annot distinguish distribution shifts.
****
Accelerated Quasi-Newton Proximal Extragradient: Faster Rate for Smooth Convex Optimization
Ruichen Jiang, Aryan Mokhtari
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Conditional Adapters: Parameter-efficient Transfer Learning with Fast Inference
Tao Lei, Junwen Bai, Siddhartha Brahma, Joshua Ainslie, Kenton Lee, Yanqi Zhou, Nan Du, Vincent Zhao, Yuexin Wu, Bo Li, Yu Zhang, Ming-Wei Chang
We propose Conditional Adapter (CoDA), a parameter-efficient transfer learning method that also improves inference efficiency. CoDA generalizes beyond standard adapter approaches to enable a new way of balancing speed and accuracy using conditional computation.Starting with an existing dense pretrained model, CoDA adds sparse activation together with a small number of new parameters and a light-weight training phase.Our experiments demonstrate that the CoDA approach provides an unexpectedly efficient way to transfer knowledge.Across a variety of language, vision, and speech tasks, CoDA achieves a 2x to 8x inference speed-up compared to the state-of-the-art Adapter approaches with moderate to no accuracy loss and the same parameter efficiency.
****
Time-Independent Information-Theoretic Generalization Bounds for SGLD
Futoshi Futami, Masahiro Fujisawa
We provide novel information-theoretic generalization bounds for stochastic gradient Langevin dynamics (SGLD) under the assumptions of smoothness and dissipativity, which are widely used in sampling and non-convex optimization studies.Our bounds are time-independent and decay to zero as the sample size increases, regardless of the number of iterations and whether the step size is fixed.Unlike previous studies, we derive the generalization error bounds by focusing on the time evolution of the Kullback--Leibler divergence, which is related to the stability of datasets and is the upper bound of the mutual information between output parameters and an input dataset.Additionally, we establish the first information-theoretic generalization bound when the training and test loss are the same by showing that a loss function of SGLD is sub-exponential.This bound is also time-independent and removes the problematic step size dependence in existing work, leading to an improved excess risk bound by combining our analysis with the existing non-convex optimization error bounds.
****
Topology-Aware Uncertainty for Image Segmentation
Saumya Gupta, Yikai Zhang, Xiaoling Hu, Prateek Prasanna, Chao Chen
Segmentation of curvilinear structures such as vasculature and road networks is challenging due to relatively weak signals and complex geometry/topology. To facilitate and accelerate large scale annotation, one has to adopt semi-automatic approaches such as proofreading by experts. In this work, we focus on uncertainty estimation for such tasks, so that highly uncertain, and thus error-prone structures can be identified for human annotators to verify. Unlike most existing works, which provide pixel-wise uncertainty maps, we stipulate it is crucial to estimate uncertainty in the units of topological structures, e.g., small pieces of connections and branches. To achieve this, we leverage tools from topological data analysis, specifically discrete Morse theory (DMT), to first capture the structures, and then reason about their uncertainties. To model the uncertainty, we (1) propose a joint prediction model that estimates the uncertainty of a structure while taking the neighboring structures into consideration (inter-structural uncertainty); (2) propose a novel Probabilistic DMT to model the inherent uncertainty within each structure (intra-structural uncertainty) by sampling its representations via a perturb-and-walk scheme. On various 2D and 3D datasets, our method produces better structure-wise uncertainty maps compared to existing works.

Code available at: https://github.com/Saumya-Gupta-26/struct-uncertainty
****
Multiplication-Free Transformer Training via Piecewise Affine Operations
Atli Kosson, Martin Jaggi
Multiplications are responsible for most of the computational cost involved in neural network training and inference. Recent research has thus looked for ways to reduce the cost associated with them. Inspired by Mogami 2020, we replace multiplication with a cheap piecewise affine approximation that is achieved by adding the bit representation of the floating point numbers together as integers. We show that transformers can be trained with the resulting modified matrix multiplications on both vision and language tasks with little to no performance impact, and without changes to the training hyperparameters. We further replace all non-linearities in the networks making them fully and jointly piecewise affine in both inputs and weights. Finally, we show that we can eliminate all multiplications in the entire training process, including operations in the forward pass, backward pass and optimizer update, demonstrating the first successful training of modern neural network architectures in a fully multiplication-free fashion.
****
A Unified Framework for Uniform Signal Recovery in Nonlinear Generative Compressed Sensing
Junren Chen, Jonathan Scarlett, Michael Ng, Zhaoqiang Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Tempo Adaptation in Non-stationary Reinforcement Learning
Hyunin Lee, Yuhao Ding, Jongmin Lee, Ming Jin, Javad Lavaei, Somayeh Sojoudi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Unsupervised Semantic Correspondence Using Stable Diffusion
Eric Hedlin, Gopal Sharma, Shweta Mahajan, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, Kwang Moo Yi
Text-to-image diffusion models are now capable of generating images that are often indistinguishable from real images. To generate such images, these models must understand the semantics of the objects they are asked to generate. In this work we show that, without any training, one can leverage this semantic knowledge within diffusion models to find semantic correspondences – locations in multiple images that have the same semantic meaning. Specifically, given an image, we optimize the prompt embeddings of these models for maximum attention on the regions of interest. These optimized embeddings capture semantic information about the location, which can then be transferred to another image. By doing so we obtain results on par with the strongly supervised state of the art on the PF-Willow dataset and significantly outperform (20.9% relative for the SPair-71k dataset) any existing weakly- or unsupervised method on PF-Willow, CUB-200 and SPair-71k datasets.
****
Efficient Subgame Refinement for Extensive-form Games
Zhenxing Ge, Zheng Xu, Tianyu Ding, Wenbin Li, Yang Gao
Subgame solving is an essential technique in addressing large imperfect information games, with various approaches developed to enhance the performance of refined strategies in the abstraction of the target subgame. However, directly applying existing subgame solving techniques may be difficult, due to the intricate nature and substantial size of many real-world games. To overcome this issue, recent subgame solving methods allow for subgame solving on limited knowledge order subgames, increasing their applicability in large games; yet this may still face obstacles due to extensive information set sizes. To address this challenge, we

propose a generative subgame solving (GS2) framework, which utilizes a generation function to identify a subset of the earliest-reached nodes, reducing the size of the subgame. Our method is supported by a theoretical analysis and employs a diversity-based generation function to enhance safety. Experiments conducted on medium-sized games as well as the challenging large game of GuanDan demonstrate a significant improvement over the blueprint.
****
NeRF-IBVS: Visual Servo Based on NeRF for Visual Localization and Navigation
Yuanze Wang, Yichao Yan, Dianxi Shi, Wenhan Zhu, Jianqiang Xia, Tan Jeff, Songchang Jin, KE GAO, XIAOBO LI, Xiaokang Yang
Visual localization is a fundamental task in computer vision and robotics. Training existing visual localization methods requires a large number of posed images to generalize to novel views, while state-of-the-art methods generally require dense ground truth 3D labels for supervision. However, acquiring a large number of posed images and dense 3D labels in the real world is challenging and costly. In this paper, we present a novel visual localization method that achieves accurate localization while using only a few posed images compared to other localization methods. To achieve this, we first use a few posed images with coarse pseudo-3D labels provided by NeRF to train a coordinate regression network. Then a coarse pose is estimated from the regression network with PNP. Finally, we use the image-based visual servo (IBVS) with the scene prior provided by NeRF for pose optimization. Furthermore, our method can provide effective navigation prior, which enable navigation based on IBVS without using custom markers and depth sensor. Extensive experiments on 7-Scenes and 12-Scenes datasets demonstrate that our method outperforms state-of-the-art methods under the same setting, with only 5\% to 25\% training data. Furthermore, our framework can be naturally extended to the visual navigation task based on IBVS, and its effectiveness is verified in simulation experiments.
****
How Does Adaptive Optimization Impact Local Neural Network Geometry?
Kaiqi Jiang, Dhruv Malik, Yuanzhi Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Are Diffusion Models Vision-And-Language Reasoners?
Benno Krojer, Elinor Poole-Dayan, Vikram Voleti, Chris Pal, Siva Reddy
Text-conditioned image generation models have recently shown immense qualitative success using denoising diffusion processes. However, unlike discriminative vision-and-language models, it is a non-trivial task to subject these diffusion-based generative models to automatic fine-grained quantitative evaluation of high-level phenomena such as compositionality.Towards this goal, we perform two innovations. First, we transform diffusion-based models (in our case, Stable Diffusion) for any image-text matching (ITM) task using a novel method called DiffusionITM.Second, we introduce the Generative-Discriminative Evaluation Benchmark (GDBench) benchmark with 7 complex vision-and-language tasks, bias evaluation and detailed analysis.We find that Stable Diffusion + DiffusionITM is competitive on many tasks and outperforms CLIP on compositional tasks like like CLEVR and Winoground.We further boost its compositional performance with a transfer setup by fine-tuning on MS-COCO while retaining generative capabilities. We also measure the stereotypical bias in diffusion models, and find that Stable Diffusion 2.1 is, for the most part, less biased than Stable Diffusion 1.5.Overall, our results point in an exciting direction bringing discriminative and generative model evaluation closer. We will release code and benchmark setup soon.
****
ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation
Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan LI, Hang Su, Jun Zhu
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

SAMoSSA: Multivariate Singular Spectrum Analysis with Stochastic Autoregressive Noise

Abdullah Alomar, Munther Dahleh, Sean Mann, Devavrat Shah

The well-established practice of time series analysis involves estimating deterministic, non-stationary trend and seasonality components followed by learning the residual stochastic, stationary components. Recently, it has been shown that one can learn the deterministic non-stationary components accurately using multivariate Singular Spectrum Analysis (mSSA) in the absence of a correlated stationary component; meanwhile, in the absence of deterministic non-stationary components, the Autoregressive (AR) stationary component can also be learnt readily, e.g. via Ordinary Least Squares (OLS). However, a theoretical underpinning of multi-stage learning algorithms involving both deterministic and stationary components has been absent in the literature despite its pervasiveness. We resolve this open question by establishing desirable theoretical guarantees for a natural two-stage algorithm, where mSSA is first applied to estimate the non-stationary components despite the presence of a correlated stationary AR component, which is subsequently learned from the residual time series. We provide a finite-sample forecasting consistency bound for the proposed algorithm, SAMoSSA, which is data-driven and thus requires minimal parameter tuning. To establish theoretical guarantees, we overcome three hurdles: (i) we characterize the spectra of Page matrices of stable AR processes, thus extending the analysis of mSSA; (ii) we extend the analysis of AR process identification in the presence of arbitrary bounded perturbations; (iii) we characterize the out-of-sample or forecasting error, as opposed to solely considering model identification. Through representative empirical studies, we validate the superior performance of SAMoSSA compared to existing baselines. Notably, SAMoSSA's ability to account for AR noise structure yields improvements ranging from 5% to 37% across various benchmark datasets.
****

Hierarchical Vector Quantized Transformer for Multi-class Unsupervised Anomaly Detection

Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, Ruimin Hu

Unsupervised image Anomaly Detection (UAD) aims to learn robust and discriminative representations of normal samples. While separate solutions per class endow expensive computation and limited generalizability, this paper focuses on building a unified framework for multiple classes. Under such a challenging setting, popular reconstruction-based networks with continuous latent representation assumption always suffer from the "identical shortcut" issue, where both normal and abnormal samples can be well recovered and difficult to distinguish. To address this pivotal issue, we propose a hierarchical vector quantized prototype-oriented Transformer under a probabilistic framework. First, instead of learning the continuous representations, we preserve the typical normal patterns as discrete iconic prototypes, and confirm the importance of Vector Quantization in preventing the model from falling into the shortcut. The vector quantized iconic prototypes are integrated into the Transformer for reconstruction, such that the abnormal data point is flipped to a normal data point. Second, we investigate an exquisite hierarchical framework to relieve the codebook collapse issue and replenish frail normal patterns. Third, a prototype-oriented optimal transport method is proposed to better regulate the prototypes and hierarchically evaluate the abnormal score. By evaluating on MVTec-AD and VisA datasets, our model surpasses the state-of-the-art alternatives and possesses good interpretability. The code is available at https://github.com/RuiyingLu/HVQ-Trans.
****

MCUFormer: Deploying Vision Tranformers on Microcontrollers with Limited Memory

Yinan Liang, Ziwei Wang, Xiuwei Xu, Yansong Tang, Jie Zhou, Jiwen Lu

Due to the high price and heavy energy consumption of GPUs, deploying deep models on IoT devices such as microcontrollers makes significant contributions for ec

ological AI. Conventional methods successfully enable convolutional neural network inference of high resolution images on microcontrollers, while the framework for vision transformers that achieve the state-of-the-art performance in many vision applications still remains unexplored. In this paper, we propose a hardware-algorithm co-optimizations method called MCUFormer to deploy vision transformers on microcontrollers with extremely limited memory, where we jointly design transformer architecture and construct the inference operator library to fit the memory resource constraint. More specifically, we generalize the one-shot network architecture search (NAS) to discover the optimal architecture with highest task performance given the memory budget from the microcontrollers, where we enlarge the existing search space of vision transformers by considering the low-rank decomposition dimensions and patch resolution for memory reduction. For the construction of the inference operator library of vision transformers, we schedule the memory buffer during inference through operator integration, patch embedding decomposition, and token overwriting, allowing the memory buffer to be fully utilized to adapt to the forward pass of the vision transformer. Experimental results demonstrate that our MCUFormer achieves 73.62\% top-1 accuracy on ImageNet for image classification with 320KB memory on STM32F746 microcontroller. Code is available at https://github.com/liangyn22/MCUFormer.
****

Towards Accelerated Model Training via Bayesian Data Selection
Zhijie Deng, Peng Cui, Jun Zhu
Mislabeled, duplicated, or biased data in real-world scenarios can lead to prolonged training and even hinder model convergence. Traditional solutions prioritizing easy or hard samples lack the flexibility to handle such a variety simultaneously. Recent work has proposed a more reasonable data selection principle by examining the data's impact on the model's generalization loss. However, its practical adoption relies on less principled approximations and additional holdout data. This work solves these problems by leveraging a lightweight Bayesian treatment and incorporating off-the-shelf zero-shot predictors built on large-scale pre-trained models. The resulting algorithm is efficient and easy to implement. We perform extensive empirical studies on challenging benchmarks with considerable data noise and imbalance in the online batch selection scenario, and observe superior training efficiency over competitive baselines. Notably, on the challenging WebVision benchmark, our method can achieve similar predictive performance with significantly fewer training iterations than leading data selection methods.
****

CSOT: Curriculum and Structure-Aware Optimal Transport for Learning with Noisy Labels
Wanxing Chang, Ye Shi, Jingya Wang
Learning with noisy labels (LNL) poses a significant challenge in training a well-generalized model while avoiding overfitting to corrupted labels.Recent advances have achieved impressive performance by identifying clean labels and correcting corrupted labels for training.However, the current approaches rely heavily on the model’s predictions and evaluate each sample independently without considering either the global or local structure of the sample distribution.These limitations typically result in a suboptimal solution for the identification and correction processes, which eventually leads to models overfitting to incorrect labels.In this paper, we propose a novel optimal transport (OT) formulation, called Curriculum and Structure-aware Optimal Transport (CSOT). CSOT concurrently considers the inter- and intra-distribution structure of the samples to construct a robust denoising and relabeling allocator.During the training process, the allocator incrementally assigns reliable labels to a fraction of the samples with the highest confidence. These labels have both global discriminability and local coherence.Notably, CSOT is a new OT formulation with a nonconvex objective function and curriculum constraints, so it is not directly compatible with classical OT solvers. Here, we develop a lightspeed computational method that involves a scaling iteration within a generalized conditional gradient framework to solve CSOT efficiently.Extensive experiments demonstrate the superiority of our method over the current state-of-the-arts in LNL.

****
In-Context Learning Unlocked for Diffusion Models
Zhendong Wang, Yifan Jiang, Yadong Lu, yelong shen, Pengcheng He, Weizhu Chen, Zhangyang "Atlas" Wang, Mingyuan Zhou
We present Prompt Diffusion, a framework for enabling in-context learning in diffusion-based generative models. Given a pair of task-specific example images, such as depth from/to image and scribble from/to image, and a text guidance, our model automatically understands the underlying task and performs the same task on a new query image following the text guidance. To achieve this, we propose a vision-language prompt that can model a wide range of vision-language tasks and a diffusion model that takes it as input. The diffusion model is trained jointly on six different tasks using these prompts. The resulting Prompt Diffusion model becomes the first diffusion-based vision-language foundation model capable of in-context learning. It demonstrates high-quality in-context generation for the trained tasks and effectively generalizes to new, unseen vision tasks using their respective prompts. Our model also shows compelling text-guided image editing results. Our framework aims to facilitate research into in-context learning for computer vision. We share our code and pre-trained models at https://github.com/Zhendong-Wang/Prompt-Diffusion.
****
Object-Centric Slot Diffusion
Jindong Jiang, Fei Deng, Gautam Singh, Sungjin Ahn
The recent success of transformer-based image generative models in object-centric learning highlights the importance of powerful image generators for handling complex scenes. However, despite the high expressiveness of diffusion models in image generation, their integration into object-centric learning remains largely unexplored in this domain. In this paper, we explore the feasibility and potential of integrating diffusion models into object-centric learning and investigate the pros and cons of this approach. We introduce Latent Slot Diffusion (LSD), a novel model that serves dual purposes: it is the first object-centric learning model to replace conventional slot decoders with a latent diffusion model conditioned on object slots, and it is also the first unsupervised compositional conditional diffusion model that operates without the need for supervised annotations like text. Through experiments on various object-centric tasks, including the first application of the FFHQ dataset in this field, we demonstrate that LSD significantly outperforms state-of-the-art transformer-based decoders, particularly in more complex scenes, and exhibits superior unsupervised compositional generation quality. In addition, we conduct a preliminary investigation into the integration of pre-trained diffusion models in LSD and demonstrate its effectiveness in real-world image segmentation and generation. Project page is available at https://latentslotdiffusion.github.io
****
NAS-X: Neural Adaptive Smoothing via Twisting
Dieterich Lawson, Michael Li, Scott Linderman
Sequential latent variable models (SLVMs) are essential tools in statistics and machine learning, with applications ranging from healthcare to neuroscience. As their flexibility increases, analytic inference and model learning can become challenging, necessitating approximate methods. Here we introduce neural adaptive smoothing via twisting (NAS-X), a method that extends reweighted wake-sleep (RWS) to the sequential setting by using smoothing sequential Monte Carlo (SMC) to estimate intractable posterior expectations. Combining RWS and smoothing SMC allows NAS-X to provide low-bias and low-variance gradient estimates, and fit both discrete and continuous latent variable models. We illustrate the theoretical advantages of NAS-X over previous methods and explore these advantages empirically in a variety of tasks, including a challenging application to mechanistic models of neuronal dynamics. These experiments show that NAS-X substantially outperforms previous VI- and RWS-based methods in inference and model learning, achieving lower parameter error and tighter likelihood bounds.
****
Reflexion: language agents with verbal reinforcement learning

Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, Shunyu Yao
Large language models (LLMs) have been increasingly used to interact with extern al environments (e.g., games, compilers, APIs) as goal-driven agents. However, i t remains challenging for these language agents to quickly and efficiently learn from trial-and-error as traditional reinforcement learning methods require exte nsive training samples and expensive model fine-tuning. We propose \emph{Reflexi on}, a novel framework to reinforce language agents not by updating weights, but instead through linguistic feedback. Concretely, Reflexion agents verbally refl ect on task feedback signals, then maintain their own reflective text in an epis odic memory buffer to induce better decision-making in subsequent trials. Reflex ion is flexible enough to incorporate various types (scalar values or free-form language) and sources (external or internally simulated) of feedback signals, an d obtains significant improvements over a baseline agent across diverse tasks (s equential decision-making, coding, language reasoning). For example, Reflexion a chieves a 91\% pass@1 accuracy on the HumanEval coding benchmark, surpassing the previous state-of-the-art GPT-4 that achieves 80\%. We also conduct ablation an d analysis studies using different feedback signals, feedback incorporation meth ods, and agent types, and provide insights into how they affect performance. We release all code, demos, and datasets at \url{https://github.com/noahshinn024/re flexion}.
****
Demographic Parity Constrained Minimax Optimal Regression under Linear Model
Kazuto Fukuchi, Jun Sakuma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
GeoCLIP: Clip-Inspired Alignment between Locations and Images for Effective Worl dwide Geo-localization
Vicente Vivanco Cepeda, Gaurav Kumar Nayak, Mubarak Shah
Worldwide Geo-localization aims to pinpoint the precise location of images taken anywhere on Earth. This task has considerable challenges due to the immense var iation in geographic landscapes. The image-to-image retrieval-based approaches f ail to solve this problem on a global scale as it is not feasible to construct a large gallery of images covering the entire world. Instead, existing approaches divide the globe into discrete geographic cells, transforming the problem into a classification task. However, their performance is limited by the predefined c lasses and often results in inaccurate localizations when an image's location si gnificantly deviates from its class center. To overcome these limitations, we pr opose GeoCLIP, a novel CLIP-inspired Image-to-GPS retrieval approach that enforc es alignment between the image and its corresponding GPS locations. GeoCLIP's lo cation encoder models the Earth as a continuous function by employing positional encoding through random Fourier features and constructing a hierarchical repres entation that captures information at varying resolutions to yield a semanticall y rich high-dimensional feature suitable to use even beyond geo-localization. To the best of our knowledge, this is the first work employing GPS encoding for ge o-localization. We demonstrate the efficacy of our method via extensive experime nts and ablations on benchmark datasets. We achieve competitive performance with just 20% of training data, highlighting its effectiveness even in limited-data settings. Furthermore, we qualitatively demonstrate geo-localization using a tex t query by leveraging the CLIP backbone of our image encoder. The project webpag e is available at: https://vicentevivan.github.io/GeoCLIP
****
RECESS Vaccine for Federated Learning: Proactive Defense Against Model Poisoning Attacks
Haonan Yan, Wenjing Zhang, Qian Chen, Xiaoguang Li, Wenhai Sun, HUI LI, Xiaodong Lin
Model poisoning attacks greatly jeopardize the application of federated learning (FL). The effectiveness of existing defenses is susceptible to the latest model

poisoning attacks, leading to a decrease in prediction accuracy. Besides, these defenses are intractable to distinguish benign outliers from malicious gradients, which further compromises the model generalization. In this work, we propose a novel defense including detection and aggregation, named RECESS, to serve as a "vaccine" for FL against model poisoning attacks. Different from the passive analysis in previous defenses, RECESS proactively queries each participating client with a delicately constructed aggregation gradient, accompanied by the detection of malicious clients according to their responses with higher accuracy. Further, RECESS adopts a newly proposed trust scoring based mechanism to robustly aggregate gradients. Rather than previous methods of scoring in each iteration, RECESS takes into account the correlation of clients' performance over multiple iterations to estimate the trust score, bringing in a significant increase in detection fault tolerance. Finally, we extensively evaluate RECESS on typical model architectures and four datasets under various settings including white/black-box, cross-silo/device FL, etc. Experimental results show the superiority of RECESS in terms of reducing accuracy loss caused by the latest model poisoning attacks over five classic and two state-of-the-art defenses.
****
Minimum norm interpolation by perceptra: Explicit regularization and implicit bias
Jiyoung Park, Ian Pelakh, Stephan Wojtowytsch
We investigate how shallow ReLU networks interpolate between known regions. Our analysis shows that empirical risk minimizers converge to a minimum norm interpolant as the number of data points and parameters tends to infinity when a weight decay regularizer is penalized with a coefficient which vanishes at a precise rate as the network width and the number of data points grow. With and without explicit regularization, we numerically study the implicit bias of common optimization algorithms towards known minimum norm interpolants.
****
Spectral Co-Distillation for Personalized Federated Learning
Zihan Chen, Howard Yang, Tony Quek, Kai Fong Ernest Chong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DVSOD: RGB-D Video Salient Object Detection
Jingjing Li, Wei Ji, Size Wang, Wenbo Li, Li cheng
Salient object detection (SOD) aims to identify standout elements in a scene, with recent advancements primarily focused on integrating depth data (RGB-D) or temporal data from videos to enhance SOD in complex scenes. However, the unison of two types of crucial information remains largely underexplored due to data constraints. To bridge this gap, we in this work introduce the DViSal dataset, fueling further research in the emerging field of RGB-D video salient object detection (DVSOD). Our dataset features 237 diverse RGB-D videos alongside comprehensive annotations, including object and instance-level markings, as well as bounding boxes and scribbles. These resources enable a broad scope for potential research directions. We also conduct benchmarking experiments using various SOD models, affirming the efficacy of multimodal video input for salient object detection. Lastly, we highlight some intriguing findings and promising future research avenues. To foster growth in this field, our dataset and benchmark results are publicly accessible at: https://dvsod.github.io/.
****
Gradient Informed Proximal Policy Optimization
Sanghyun Son, Laura Zheng, Ryan Sullivan, Yi-Ling Qiao, Ming Lin
We introduce a novel policy learning method that integrates analytical gradients from differentiable environments with the Proximal Policy Optimization (PPO) algorithm. To incorporate analytical gradients into the PPO framework, we introduce the concept of an α-policy that stands as a locally superior policy. By adaptively modifying the α value, we can effectively manage the influence of analytica

l policy gradients during learning. To this end, we suggest metrics for assessin
g the variance and bias of analytical gradients, reducing dependence on these gr
adients when high variance or bias is detected. Our proposed approach outperform
s baseline algorithms in various scenarios, such as function optimization, physi
cs simulations, and traffic control environments. Our code can be found online:
https://github.com/SonSang/gippo.
****
SAMRS: Scaling-up Remote Sensing Segmentation Dataset with Segment Anything Mode
l
Di Wang, Jing Zhang, Bo Du, Minqiang Xu, Lin Liu, Dacheng Tao, Liangpei Zhang
The success of the Segment Anything Model (SAM) demonstrates the significance of
 data-centric machine learning. However, due to the difficulties and high costs
associated with annotating Remote Sensing (RS) images, a large amount of valuabl
e RS data remains unlabeled, particularly at the pixel level. In this study, we
leverage SAM and existing RS object detection datasets to develop an efficient p
ipeline for generating a large-scale RS segmentation dataset, dubbed SAMRS. SAMR
S totally possesses 105,090 images and 1,668,241 instances, surpassing existing
high-resolution RS segmentation datasets in size by several orders of magnitude.
 It provides object category, location, and instance information that can be use
d for semantic segmentation, instance segmentation, and object detection, either
 individually or in combination. We also provide a comprehensive analysis of SAM
RS from various aspects. Moreover, preliminary experiments highlight the import
ance of conducting segmentation pre-training with SAMRS to address task discrepa
ncies and alleviate the limitations posed by limited training data during fine-t
uning. The code and dataset will be available at https://github.com/ViTAE-Transf
ormer/SAMRS
****
Blockwise Parallel Transformers for Large Context Models
Hao Liu, Pieter Abbeel
Transformers have emerged as the cornerstone of state-of-the-art natural languag
e processing models, showcasing exceptional performance across a wide range of A
I applications. However, the memory demands posed by the self-attention mechanis
m and the large feedforward network in Transformers limit their ability to handl
e long sequences, thereby creating challenges for tasks involving multiple long
sequences or long-term dependencies. We present a distinct approach, Blockwise P
arallel Transformer (BPT), that leverages blockwise computation of self-attentio
n and feedforward network fusion to minimize memory costs. By processing longer
input sequences while maintaining memory efficiency, BPT enables training sequen
ces 32 times longer than vanilla Transformers and up to 4 times longer than prev
ious memory-efficient methods. Extensive experiments on language modeling and re
inforcement learning tasks demonstrate the effectiveness of BPT in reducing memo
ry requirements and improving performance.
****
Neural Combinatorial Optimization with Heavy Decoder: Toward Large Scale General
ization
Fu Luo, Xi Lin, Fei Liu, Qingfu Zhang, Zhenkun Wang
Neural combinatorial optimization (NCO) is a promising learning-based approach f
or solving challenging combinatorial optimization problems without specialized a
lgorithm design by experts. However, most constructive NCO methods cannot solve
problems with large-scale instance sizes, which significantly diminishes their u
sefulness for real-world applications. In this work, we propose a novel Light En
coder and Heavy Decoder (LEHD) model with a strong generalization ability to add
ress this critical issue. The LEHD model can learn to dynamically capture the re
lationships between all available nodes of varying sizes, which is beneficial fo
r model generalization to problems of various scales. Moreover, we develop a dat
a-efficient training scheme and a flexible solution construction mechanism for t
he proposed LEHD model. By training on small-scale problem instances, the LEHD m
odel can generate nearly optimal solutions for the Travelling Salesman Problem (
TSP) and the Capacitated Vehicle Routing Problem (CVRP) with up to 1000 nodes, a
nd also generalizes well to solve real-world TSPLib and CVRPLib problems. These

results confirm our proposed LEHD model can significantly improve the state-of-the-art performance for constructive NCO.
****
Topological Obstructions and How to Avoid Them
Babak Esmaeili, Robin Walters, Heiko Zimmermann, Jan-Willem van de Meent
Incorporating geometric inductive biases into models can aid interpretability and generalization, but encoding to a specific geometric structure can be challenging due to the imposed topological constraints. In this paper, we theoretically and empirically characterize obstructions to training encoders with geometric latent spaces. We show that local optima can arise due to singularities (e.g. self-intersection) or due to an incorrect degree or winding number. We then discuss how normalizing flows can potentially circumvent these obstructions by defining multimodal variational distributions. Inspired by this observation, we propose a new flow-based model that maps data points to multimodal distributions over geometric spaces and empirically evaluate our model on 2 domains. We observe improved stability during training and a higher chance of converging to a homeomorphic encoder.
****
The Double-Edged Sword of Implicit Bias: Generalization vs. Robustness in ReLU Networks
Spencer Frei, Gal Vardi, Peter Bartlett, Nati Srebro
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
PromptRestorer: A Prompting Image Restoration Method with Degradation Perception
Cong Wang, Jinshan Pan, Wei Wang, Jiangxin Dong, Mengzhu Wang, Yakun Ju, Junyang Chen
We show that raw degradation features can effectively guide deep restoration models, providing accurate degradation priors to facilitate better restoration. While networks that do not consider them for restoration forget gradually degradation during the learning process, model capacity is severely hindered. To address this, we propose a Prompting image Restorer, termed as PromptRestorer. Specifically, PromptRestorer contains two branches: a restoration branch and a prompting branch. The former is used to restore images, while the latter perceives degradation priors to prompt the restoration branch with reliable perceived content to guide the restoration process for better recovery. To better perceive the degradation which is extracted by a pre-trained model from given degradation observations, we propose a prompting degradation perception modulator, which adequately considers the characters of the self-attention mechanism and pixel-wise modulation, to better perceive the degradation priors from global and local perspectives. To control the propagation of the perceived content for the restoration branch, we propose gated degradation perception propagation, enabling the restoration branch to adaptively learn more useful features for better recovery. Extensive experimental results show that our PromptRestorer achieves state-of-the-art results on 4 image restoration tasks, including image deraining, deblurring, dehazing, and desnowing.
****
Beyond MLE: Convex Learning for Text Generation
Chenze Shao, Zhengrui Ma, Min Zhang, Yang Feng
Maximum likelihood estimation (MLE) is a statistical method used to estimate the parameters of a probability distribution that best explain the observed data. In the context of text generation, MLE is often used to train generative language models, which can then be used to generate new text. However, we argue that MLE is not always necessary and optimal, especially for closed-ended text generation tasks like machine translation. In these tasks, the goal of model is to generate the most appropriate response, which does not necessarily require it to estimate the entire data distribution with MLE. To this end, we propose a novel class of training objectives based on convex functions, which enables text generation

models to focus on highly probable outputs without having to estimate the entire data distribution. We investigate the theoretical properties of the optimal predicted distribution when applying convex functions to the loss, demonstrating that convex functions can sharpen the optimal distribution, thereby enabling the model to better capture outputs with high probabilities. Experiments on various text generation tasks and models show the effectiveness of our approach. It enables autoregressive models to bridge the gap between greedy and beam search, and facilitates the learning of non-autoregressive models with a maximum improvement of 9+ BLEU points. Moreover, our approach also exhibits significant impact on large language models (LLMs), substantially enhancing their generative capability on various tasks. Source code is available at \url{https://github.com/ictnlp/Convex-Learning}.
****
Bandit Task Assignment with Unknown Processing Time
Shinji Ito, Daisuke Hatano, Hanna Sumita, Kei Takemura, Takuro Fukunaga, Naonori Kakimura, Ken-Ichi Kawarabayashi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Multimodal C4: An Open, Billion-scale Corpus of Images Interleaved with Text
Wanrong Zhu, Jack Hessel, Anas Awadalla, Samir Yitzhak Gadre, Jesse Dodge, Alex Fang, Youngjae Yu, Ludwig Schmidt, William Yang Wang, Yejin Choi
In-context vision and language models like Flamingo support arbitrarily interleaved sequences of images and text as input.This format not only enables few-shot learning via interleaving independent supervised (image, text) examples, but also, more complex prompts involving interaction between images, e.g., ``What do image A and image B have in common?''To support this interface, pretraining occurs over web corpora that similarly contain interleaved images+text.To date, however, large-scale data of this form have not been publicly available.We release Multimodal C4, an augmentation of the popular text-only C4 corpus with images interleaved.We use a linear assignment algorithm to place images into longer bodies of text using CLIP features, a process that we show outperforms alternatives.Multimodal C4 spans everyday topics like cooking, travel, technology, etc. A manual inspection of a random sample of documents shows that a vast majority (88\%) of images are topically relevant, and that linear assignment frequently selects individual sentences specifically well-aligned with each image (80\%). After filtering NSFW images, ads, etc., the resulting corpus consists of 101.2M documents with 571M images interleaved in 43B English tokens.
****
Towards Self-Interpretable Graph-Level Anomaly Detection
Yixin Liu, Kaize Ding, Qinghua Lu, Fuyi Li, Leo Yu Zhang, Shirui Pan
Graph-level anomaly detection (GLAD) aims to identify graphs that exhibit notable dissimilarity compared to the majority in a collection. However, current works primarily focus on evaluating graph-level abnormality while failing to provide meaningful explanations for the predictions, which largely limits their reliability and application scope. In this paper, we investigate a new challenging problem, explainable GLAD, where the learning objective is to predict the abnormality of each graph sample with corresponding explanations, i.e., the vital subgraph that leads to the predictions. To address this challenging problem, we propose a Self-Interpretable Graph aNomaly dETection model (SIGNET for short) that detects anomalous graphs as well as generates informative explanations simultaneously. Specifically, we first introduce the multi-view subgraph information bottleneck (MSIB) framework, serving as the design basis of our self-interpretable GLAD approach. This way SIGNET is able to not only measure the abnormality of each graph based on cross-view mutual information but also provide informative graph rationales by extracting bottleneck subgraphs from the input graph and its dual hypergraph in a self-supervised way. Extensive experiments on 16 datasets demonstrate the anomaly detection capability and self-interpretability of SIGNET.

****
AMAG: Additive, Multiplicative and Adaptive Graph Neural Network For Forecasting Neuron Activity
Jingyuan Li, Leo Scholl, Trung Le, Pavithra Rajeswaran, Amy Orsborn, Eli Shlizerman
Latent Variable Models (LVMs) propose to model the dynamics of neural populations by capturing low-dimensional structures that represent features involved in neural activity. Recent LVMs are based on deep learning methodology where a deep neural network is trained to reconstruct the same neural activity given as input and as a result to build the latent representation. Without taking past or future activity into account such a task is non-causal. In contrast, the task of forecasting neural activity based on given input extends the reconstruction task. LVMs that are trained on such a task could potentially capture temporal causality constraints within its latent representation. Forecasting has received less attention than reconstruction due to recording challenges such as limited neural measurements and trials. In this work, we address modeling neural population dynamics via the forecasting task and improve forecasting performance by including a prior, which consists of pairwise neural unit interaction as a multivariate dynamic system. Our proposed model---Additive, Multiplicative, and Adaptive Graph Neural Network (AMAG)---leverages additive and multiplicative message-passing operations analogous to the interactions in neuronal systems and adaptively learns the interaction among neural units to forecast their future activity. We demonstrate the advantage of AMAG compared to non-GNN based methods on synthetic data and multiple modalities of neural recordings (field potentials from penetrating electrodes or surface-level micro-electrocorticography) from four rhesus macaques. Our results show the ability of AMAG to recover ground truth spatial interactions and yield estimation for future dynamics of the neural population.
****
PackQViT: Faster Sub-8-bit Vision Transformers via Full and Packed Quantization on the Mobile
Peiyan Dong, LEI LU, Chao Wu, Cheng Lyu, Geng Yuan, Hao Tang, Yanzhi Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Extending the Design Space of Graph Neural Networks by Rethinking Folklore Weisfeiler-Lehman
Jiarui Feng, Lecheng Kong, Hao Liu, Dacheng Tao, Fuhai Li, Muhan Zhang, Yixin Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Off-Policy Evaluation for Human Feedback
Qitong Gao, Ge Gao, Juncheng Dong, Vahid Tarokh, Min Chi, Miroslav Pajic
Off-policy evaluation (OPE) is important for closing the gap between offline training and evaluation of reinforcement learning (RL), by estimating performance and/or rank of target (evaluation) policies using offline trajectories only. It can improve the safety and efficiency of data collection and policy testing procedures in situations where online deployments are expensive, such as healthcare. However, existing OPE methods fall short in estimating human feedback (HF) signals, as HF may be conditioned over multiple underlying factors and are only sparsely available; as opposed to the agent-defined environmental rewards (used in policy optimization), which are usually determined over parametric functions or distributions. Consequently, the nature of HF signals makes extrapolating accurate OPE estimations to be challenging. To resolve this, we introduce an OPE for HF (OPEHF) framework that revives existing OPE methods in order to accurately evaluate the HF signals. Specifically, we develop an immediate human reward (IHR) rec

onstruction approach, regularized by environmental knowledge distilled in a latent space that captures the underlying dynamics of state transitions as well as issuing HF signals. Our approach has been tested over two real-world experiments, adaptive in-vivo neurostimulation and intelligent tutoring, and a simulation environment (visual Q&A). Results show that our approach significantly improves the performance toward estimating HF signals accurately, compared to directly applying (variants of) existing OPE methods.
****
Contrastive Lift: 3D Object Instance Segmentation by Slow-Fast Contrastive Fusion
Yash Bhalgat, Iro Laina, João F. Henriques, Andrea Vedaldi, Andrew Zisserman
Instance segmentation in 3D is a challenging task due to the lack of large-scale annotated datasets. In this paper, we show that this task can be addressed effectively by leveraging instead 2D pre-trained models for instance segmentation. We propose a novel approach to lift 2D segments to 3D and fuse them by means of a neural field representation, which encourages multi-view consistency across frames. The core of our approach is a slow-fast clustering objective function, which is scalable and well-suited for scenes with a large number of objects. Unlike previous approaches, our method does not require an upper bound on the number of objects or object tracking across frames. To demonstrate the scalability of the slow-fast clustering, we create a new semi-realistic dataset called the Messy Rooms dataset, which features scenes with up to 500 objects per scene. Our approach outperforms the state-of-the-art on challenging scenes from the ScanNet, Hypersim, and Replica datasets, as well as on our newly created Messy Rooms dataset, demonstrating the effectiveness and scalability of our slow-fast clustering method.
****
GALOPA: Graph Transport Learning with Optimal Plan Alignment
Yejiang Wang, Yuhai Zhao, Daniel Zhengkui Wang, Ling Li
Self-supervised learning on graph aims to learn graph representations in an unsupervised manner. While graph contrastive learning (GCL - relying on graph augmentation for creating perturbation views of anchor graphs and maximizing/minimizing similarity for positive/negative pairs) is a popular self-supervised method, it faces challenges in finding label-invariant augmented graphs and determining the exact extent of similarity between sample pairs to be achieved. In this work, we propose an alternative self-supervised solution that (i) goes beyond the label invariance assumption without distinguishing between positive/negative samples, (ii) can calibrate the encoder for preserving not only the structural information inside the graph, but the matching information between different graphs, (iii) learns isometric embeddings that preserve the distance between graphs, a by-product of our objective. Motivated by optimal transport theory, this scheme relays on an observation that the optimal transport plans between node representations at the output space, which measure the matching probability between two distributions, should be consistent to the plans between the corresponding graphs at the input space. The experimental findings include: (i) The plan alignment strategy significantly outperforms the counterpart using the transport distance; (ii) The proposed model shows superior performance using only node attributes as calibration signals, without relying on edge information; (iii) Our model maintains robust results even under high perturbation rates; (iv) Extensive experiments on various benchmarks validate the effectiveness of the proposed method.
****
Adaptive Topological Feature via Persistent Homology: Filtration Learning for Point Clouds
Naoki Nishikawa, Yuichi Ike, Kenji Yamanishi
Machine learning for point clouds has been attracting much attention, with many applications in various fields, such as shape recognition and material science. For enhancing the accuracy of such machine learning methods, it is often effective to incorporate global topological features, which are typically extracted by persistent homology. In the calculation of persistent homology for a point cloud, we choose a filtration for the point cloud, an increasing sequence of spaces.

Since the performance of machine learning methods combined with persistent homol
ogy is highly affected by the choice of a filtration, we need to tune it dependi
ng on data and tasks. In this paper, we propose a framework that learns a filtra
tion adaptively with the use of neural networks. In order to make the resulting
persistent homology isometry-invariant, we develop a neural network architecture
with such invariance. Additionally, we show a theoretical result on a finite-di
mensional approximation of filtration functions, which justifies the proposed ne
twork architecture. Experimental results demonstrated the efficacy of our framew
ork in several classification tasks.
****
Accurate Interpolation for Scattered Data through Hierarchical Residual Refineme
nt
Shizhe Ding, Boyang Xia, Dongbo Bu
Accurate interpolation algorithms are highly desired in various theoretical and
engineering scenarios. Unlike the traditional numerical algorithms that have exa
ct zero-residual constraints on observed points, the neural network-based interp
olation methods exhibit non-zero residuals at these points. These residuals, whi
ch provide observations of an underlying residual function, can guide predicting
interpolation functions, but have not been exploited by the existing approaches
. To fill this gap, we propose Hierarchical INTerpolation Network (HINT), which
utilizes the residuals on observed points to guide target function estimation in
a hierarchical fashion. HINT consists of several sequentially arranged lightwei
ght interpolation blocks. The first interpolation block estimates the main compo
nent of the target function, while subsequent blocks predict the residual compon
ents using observed points residuals of the preceding blocks. The main component
and residual components are accumulated to form the final interpolation results
. Furthermore, under the assumption that finer residual prediction requires a mo
re focused attention range on observed points, we utilize hierarchical local con
straints in correlation modeling between observed and target points. Extensive e
xperiments demonstrate that HINT outperforms existing interpolation algorithms s
ignificantly in terms of interpolation accuracy across a wide variety of dataset
s, which underscores its potential for practical scenarios.
****
Learning Universal Policies via Text-Guided Video Generation
Yilun Du, Sherry Yang, Bo Dai, Hanjun Dai, Ofir Nachum, Josh Tenenbaum, Dale Sch
uurmans, Pieter Abbeel
A goal of artificial intelligence is to construct an agent that can solve a wide
variety of tasks. Recent progress in text-guided image synthesis has yielded mo
dels with an impressive ability to generate complex novel images, exhibiting com
binatorial generalization across domains. Motivated by this success, we investig
ate whether such tools can be used to construct more general-purpose agents. Spe
cifically, we cast the sequential decision making problem as a text-conditioned
video generation problem, where, given a text-encoded specification of a desired
goal, a planner synthesizes a set of future frames depicting its planned action
s in the future, after which control actions are extracted from the generated vi
deo. By leveraging text as the underlying goal specification, we are able to nat
urally and combinatorially generalize to novel goals. The proposed policy-as-vid
eo formulation can further represent environments with different state and actio
n spaces in a unified space of images, which, for example, enables learning and
generalization across a variety of robot manipulation tasks. Finally, by leverag
ing pretrained language embeddings and widely available videos from the internet
, the approach enables knowledge transfer through predicting highly realistic vi
deo plans for real robots.
****
Necessary and Sufficient Conditions for Optimal Decision Trees using Dynamic Pro
gramming
Jacobus van der Linden, Mathijs de Weerdt, Emir Demirovi■
Global optimization of decision trees has shown to be promising in terms of accu
racy, size, and consequently human comprehensibility. However, many of the metho
ds used rely on general-purpose solvers for which scalability remains an issue.D

ynamic programming methods have been shown to scale much better because they exp loit the tree structure by solving subtrees as independent subproblems. However, this only works when an objective can be optimized separately for subtrees.We e xplore this relationship in detail and show the necessary and sufficient conditi ons for such separability and generalize previous dynamic programming approaches into a framework that can optimize any combination of separable objectives and constraints.Experiments on five application domains show the general applicabili ty of this framework, while outperforming the scalability of general-purpose sol vers by a large margin.
****
Polyhedron Attention Module: Learning Adaptive-order Interactions
Tan Zhu, Fei Dou, Xinyu Wang, Jin Lu, Jinbo Bi
Learning feature interactions can be the key for multivariate predictive modelin g. ReLU-activated neural networks create piecewise linear prediction models, and other nonlinear activation functions lead to models with only high-order featur e interactions. Recent methods incorporate candidate polynomial terms of fixed o rders into deep learning, which is subject to the issue of combinatorial explosi on, or learn the orders that are difficult to adapt to different regions of the feature space. We propose a Polyhedron Attention Module (PAM) to create piecewis e polynomial models where the input space is split into polyhedrons which define the different pieces and on each piece the hyperplanes that define the polyhedr on boundary multiply to form the interactive terms, resulting in interactions of adaptive order to each piece. PAM is interpretable to identify important intera ctions in predicting a target. Theoretic analysis shows that PAM has stronger ex pression capability than ReLU-activated networks. Extensive experimental results demonstrate the superior classification performance of PAM on massive datasets of the click-through rate prediction and PAM can learn meaningful interaction ef fects in a medical problem.
****
Faster Query Times for Fully Dynamic k-Center Clustering with Outliers
Leyla Biabani, Annika Hennes, Morteza Monemizadeh, Melanie Schmidt
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Natural Language Instruction-following with Task-related Language Development an d Translation
Jing-Cheng Pang, Xin-Yu Yang, Si-Hang Yang, Xiong-Hui Chen, Yang Yu
Natural language-conditioned reinforcement learning (RL) enables agents to follo w human instructions. Previous approaches generally implemented language-conditi oned RL by providing the policy with human instructions in natural language (NL) and training the policy to follow instructions. In this is outside-in approach, the policy must comprehend the NL and manage the task simultaneously. However, the unbounded NL examples often bring much extra complexity for solving concrete RL tasks, which can distract policy learning from completing the task. To ease the learning burden of the policy, we investigate an inside-out scheme for natur al language-conditioned RL by developing a task language (TL) that is task-relat ed and easily understood by the policy, thus reducing the policy learning burden . Besides, we employ a translator to translate natural language into the TL, whi ch is used in RL to achieve efficient policy training. We implement this scheme as TALAR (TAsk Language with predicAte Representation) that learns multiple pred icates to model object relationships as the TL. Experiments indicate that TALAR not only better comprehends NL instructions but also leads to a better instructi on-following policy that significantly improves the success rate over baselines and adapts to unseen expressions of NL instruction. Besides, the TL is also an e ffective sub-task abstraction compatible with hierarchical RL.
****
Convergence of Actor-Critic with Multi-Layer Neural Networks
Haoxing Tian, Alex Olshevsky, Yannis Paschalidis

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Percentile Criterion Optimization in Offline Reinforcement Learning
Cyrus Cousins, Elita Lobo, Marek Petrik, Yair Zick
In reinforcement learning, robust policies for high-stakes decision-making problems with limited data are usually computed by optimizing the percentile criterion. The percentile criterion is optimized by constructing an uncertainty set that contains the true model with high probability and optimizing the policy for the worst model in the set. Since the percentile criterion is non-convex, constructing these sets itself is challenging. Existing works use Bayesian credible regions as uncertainty sets, but they are often unnecessarily large and result in learning overly conservative policies. To overcome these shortcomings, we propose a novel Value-at-Risk based dynamic programming algorithm to optimize the percentile criterion without explicitly constructing any uncertainty sets. Our theoretical and empirical results show that our algorithm implicitly constructs much smaller uncertainty sets and learns less-conservative robust policies.
****
TextDiffuser: Diffusion Models as Text Painters
Jingye Chen, Yupan Huang, Tengchao Lv, Lei Cui, Qifeng Chen, Furu Wei
Diffusion models have gained increasing attention for their impressive generation abilities but currently struggle with rendering accurate and coherent text. To address this issue, we introduce TextDiffuser, focusing on generating images with visually appealing text that is coherent with backgrounds. TextDiffuser consists of two stages: first, a Transformer model generates the layout of keywords extracted from text prompts, and then diffusion models generate images conditioned on the text prompt and the generated layout. Additionally, we contribute the first large-scale text images dataset with OCR annotations, MARIO-10M, containing 10 million image-text pairs with text recognition, detection, and character-level segmentation annotations. We further collect the MARIO-Eval benchmark to serve as a comprehensive tool for evaluating text rendering quality. Through experiments and user studies, we demonstrate that TextDiffuser is flexible and controllable to create high-quality text images using text prompts alone or together with text template images, and conduct text inpainting to reconstruct incomplete images with text. We will make the code, model and dataset publicly available.
****
Object-centric Learning with Cyclic Walks between Parts and Whole
Ziyu Wang, Mike Zheng Shou, Mengmi Zhang
Learning object-centric representations from complex natural environments enables both humans and machines with reasoning abilities from low-level perceptual features. To capture compositional entities of the scene, we proposed cyclic walks between perceptual features extracted from vision transformers and object entities. First, a slot-attention module interfaces with these perceptual features and produces a finite set of slot representations. These slots can bind to any object entities in the scene via inter-slot competitions for attention. Next, we establish entity-feature correspondence with cyclic walks along high transition probability based on the pairwise similarity between perceptual features (aka "parts") and slot-binded object representations (aka "whole"). The whole is greater than its parts and the parts constitute the whole. The part-whole interactions form cycle consistencies, as supervisory signals, to train the slot-attention module. Our rigorous experiments on \textit{seven} image datasets in \textit{three} \textit{unsupervised} tasks demonstrate that the networks trained with our cyclic walks can disentangle foregrounds and backgrounds, discover objects, and segment semantic objects in complex scenes. In contrast to object-centric models attached with a decoder for the pixel-level or feature-level reconstructions, our cyclic walks provide strong learning signals, avoiding computation overheads and enhancing memory efficiency. Our source code and data are available at: \href{https://github.com/ZhangLab-DeepNeuroCogLab/Parts-Whole-Object-Centric-Learning/}{

link}.
****
Experiment Planning with Function Approximation
Aldo Pacchiano, Jonathan Lee, Emma Brunskill
We study the problem of experiment planning with function approximation in contextual bandit problems. In settings where there is a significant overhead to deploying adaptive algorithms---for example, when the execution of the data collection policies is required to be distributed, or a human in the loop is needed to implement these policies---producing in advance a set of policies for data collection is paramount. We study the setting where a large dataset of contexts but not rewards is available and may be used by the learner to design an effective data collection strategy. Although when rewards are linear this problem has been well studied, results are still missing for more complex reward models. In this work we propose two experiment planning strategies compatible with function approximation. The first is an eluder planning and sampling procedure that can recover optimality guarantees depending on the eluder dimension of the reward function class. For the second, we show that a uniform sampler achieves competitive optimality rates in the setting where the number of actions is small. We finalize our results introducing a statistical gap fleshing out the fundamental differences between planning and adaptive learning and provide results for planning with model selection.
****
White-Box Transformers via Sparse Rate Reduction
Yaodong Yu, Sam Buchanan, Druv Pai, Tianzhe Chu, Ziyang Wu, Shengbang Tong, Benjamin Haeffele, Yi Ma
In this paper, we contend that the objective of representation learning is to compress and transform the distribution of the data, say sets of tokens, towards a mixture of low-dimensional Gaussian distributions supported on incoherent subspaces. The quality of the final representation can be measured by a unified objective function called sparse rate reduction. From this perspective, popular deep networks such as transformers can be naturally viewed as realizing iterative schemes to optimize this objective incrementally. Particularly, we show that the standard transformer block can be derived from alternating optimization on complementary parts of this objective: the multi-head self-attention operator can be viewed as a gradient descent step to compress the token sets by minimizing their lossy coding rate, and the subsequent multi-layer perceptron can be viewed as attempting to sparsify the representation of the tokens. This leads to a family of white-box transformer-like deep network architectures which are mathematically fully interpretable. Despite their simplicity, experiments show that these networks indeed learn to optimize the designed objective: they compress and sparsify representations of large-scale real-world vision datasets such as ImageNet, and achieve performance very close to thoroughly engineered transformers such as ViT. Code is at https://github.com/Ma-Lab-Berkeley/CRATE.
****
Task-Robust Pre-Training for Worst-Case Downstream Adaptation
Jianghui Wang, Yang Chen, Xingyu Xie, Cong Fang, Zhouchen Lin
Pre-training has achieved remarkable success when transferred to downstream tasks. In machine learning, we care about not only the good performance of a model but also its behavior under reasonable shifts of condition. The same philosophy holds when pre-training a foundation model. However, the foundation model may not uniformly behave well for a series of related downstream tasks. This happens, for example, when conducting mask recovery regression where the recovery ability or the training instances diverge like pattern features are extracted dominantly on pre-training, but semantic features are also required on a downstream task. This paper considers pre-training a model that guarantees a uniformly good performance over the downstream tasks. We call this goal as downstream-task robustness.Our method first separates the upstream task into several representative ones and applies a simple minimax loss for pre-training. We then design an efficient algorithm to solve the minimax lossand prove its convergence in the convex setting. In the experiments, we show both on large-scale natural language processing

and computer vision datasets our method increases the metrics on worse-case downstream tasks. Additionally, some theoretical explanations for why our loss is beneficial are provided. Specifically, we show fewer samples are inherently required for the most challenging downstream task in some cases.
****

Inconsistency, Instability, and Generalization Gap of Deep Neural Network Training

Rie Johnson, Tong Zhang

As deep neural networks are highly expressive, it is important to find solutions with small generalization gap (the difference between the performance on the training data and unseen data). Focusing on the stochastic nature of training, we first present a theoretical analysis in which the bound of generalization gap depends on what we call inconsistency and instability of model outputs, which can be estimated on unlabeled data. Our empirical study based on this analysis shows that instability and inconsistency are strongly predictive of generalization gap in various settings. In particular, our finding indicates that inconsistency is a more reliable indicator of generalization gap than the sharpness of the loss landscape. Furthermore, we show that algorithmic reduction of inconsistency leads to superior performance. The results also provide a theoretical basis for existing methods such as co-distillation and ensemble.
****

Neural approximation of Wasserstein distance via a universal architecture for symmetric and factorwise group invariant functions

Samantha Chen, Yusu Wang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Revisiting the Evaluation of Image Synthesis with GANs

mengping yang, Ceyuan Yang, Yichi Zhang, Qingyan Bai, Yujun Shen, Bo Dai

A good metric, which promises a reliable comparison between solutions, is essential for any well-defined task. Unlike most vision tasks that have per-sample ground-truth, image synthesis tasks target generating unseen data and hence are usually evaluated through a distributional distance between one set of real samples and another set of generated samples. This study presents an empirical investigation into the evaluation of synthesis performance, with generative adversarial networks (GANs) as a representative of generative models. In particular, we make in-depth analyses of various factors, including how to represent a data point in the representation space, how to calculate a fair distance using selected samples, and how many instances to use from each set. Extensive experiments conducted on multiple datasets and settings reveal several important findings. Firstly, a group of models that include both CNN-based and ViT-based architectures serve as reliable and robust feature extractors for measurement evaluation. Secondly, Centered Kernel Alignment (CKA) provides a better comparison across various extractors and hierarchical layers in one model. Finally, CKA is more sample-efficient and enjoys better agreement with human judgment in characterizing the similarity between two internal data correlations. These findings contribute to the development of a new measurement system, which enables a consistent and reliable re-evaluation of current state-of-the-art generative models.
****

What Do Deep Saliency Models Learn about Visual Attention?

Shi Chen, Ming Jiang, Qi Zhao

In recent years, deep saliency models have made significant progress in predicting human visual attention. However, the mechanisms behind their success remain largely unexplained due to the opaque nature of deep neural networks. In this paper, we present a novel analytic framework that sheds light on the implicit features learned by saliency models and provides principled interpretation and quantification of their contributions to saliency prediction. Our approach decomposes these implicit features into interpretable bases that are explicitly aligned wit

h semantic attributes and reformulates saliency prediction as a weighted combina tion of probability maps connecting the bases and saliency. By applying our fram ework, we conduct extensive analyses from various perspectives, including the po sitive and negative weights of semantics, the impact of training data and archit ectural designs, the progressive influences of fine-tuning, and common error pat terns of state-of-the-art deep saliency models. Additionally, we demonstrate the effectiveness of our framework by exploring visual attention characteristics in various application scenarios, such as the atypical attention of people with au tism spectrum disorder, attention to emotion-eliciting stimuli, and attention ev olution over time. Our code is publicly available at \url{https://github.com/szz expoi/saliency_analysis}.
****
Three Iterations of (d − 1)-WL Test Distinguish Non Isometric Clouds of d-dimens ional Points
Valentino Delle Rose, Alexander Kozachinskiy, Cristobal Rojas, Mircea Petrache, Pablo Barceló
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Puzzlefusion: Unleashing the Power of Diffusion Models for Spatial Puzzle Solvin g
Sepidehsadat (Sepid) Hossieni, Mohammad Amin Shabani, Saghar Irandoust, Yasutaka Furukawa
This paper presents an end-to-end neural architecture based on Diffusion Models for spatial puzzle solving, particularly jigsaw puzzle and room arrangement task s.In the latter task, for instance, the proposed system ``PuzzleFusion'' takes a set of room layouts as polygonal curves in the top-down view and aligns the roo m layout pieces by estimating their 2D translations and rotations, akin to solvi ng the jigsaw puzzle of room layouts. A surprising discovery of the paper is tha t the simple use of a Diffusion Model effectively solves these challenging spati al puzzle tasks as a conditional generation process. To enable learning of an en d-to-end neural system, the paper introduces new datasets with ground-truth arra ngements: 1) 2D Voronoi Jigsaw Dataset, a synthetic one where pieces are generat ed by voronoi diagram of 2D pointset; and 2) MagicPlan Dataset, a real one from a production pipeline by MagicPlan, where pieces are room layouts constructed by augmented reality App by real-estate consumers.The qualitative and quantitative evaluations demonstrate that the proposed approach outperforms the competing me thods by significant margins in all three spatial puzzle tasks. We have provided code and data in https://sepidsh.github.io/puzzlefusion.
****
Visual Instruction Inversion: Image Editing via Image Prompting
Thao Nguyen, Yuheng Li, Utkarsh Ojha, Yong Jae Lee
Text-conditioned image editing has emerged as a powerful tool for editing images .However, in many situations, language can be ambiguous and ineffective in descr ibing specific image edits.When faced with such challenges, visual prompts can b e a more informative and intuitive way to convey ideas.We present a method for i mage editing via visual prompting.Given pairs of example that represent the "bef ore" and "after" images of an edit, our goal is to learn a text-based editing di rection that can be used to perform the same edit on new images.We leverage the rich, pretrained editing capabilities of text-to-image diffusion models by inver ting visual prompts into editing instructions.Our results show that with just on e example pair, we can achieve competitive results compared to state-of-the-art text-conditioned image editing frameworks.
****
Algorithm Selection for Deep Active Learning with Imbalanced Datasets
Jifan Zhang, Shuai Shao, Saurabh Verma, Robert Nowak
Label efficiency has become an increasingly important objective in deep learning applications. Active learning aims to reduce the number of labeled examples nee

ded to train deep networks, but the empirical performance of active learning algorithms can vary dramatically across datasets and applications. It is difficult to know in advance which active learning strategy will perform well or best in a given application. To address this, we propose the first adaptive algorithm selection strategy for deep active learning. For any unlabeled dataset, our (meta) algorithm TAILOR (Thompson ActIve Learning algORithm selection) iteratively and adaptively chooses among a set of candidate active learning algorithms. TAILOR uses novel reward functions aimed at gathering class-balanced examples. Extensive experiments in multi-class and multi-label applications demonstrate TAILOR's effectiveness in achieving accuracy comparable or better than that of the best of the candidate algorithms. Our implementation of TAILOR is open-sourced at https://github.com/jifanz/TAILOR.
****
Federated Compositional Deep AUC Maximization
Xinwen Zhang, Yihan Zhang, Tianbao Yang, Richard Souvenir, Hongchang Gao
Federated learning has attracted increasing attention due to the promise of balancing privacy and large-scale learning; numerous approaches have been proposed. However, most existing approaches focus on problems with balanced data, and prediction performance is far from satisfactory for many real-world applications where the number of samples in different classes is highly imbalanced. To address this challenging problem, we developed a novel federated learning method for imbalanced data by directly optimizing the area under curve (AUC) score. In particular, we formulate the AUC maximization problem as a federated compositional minimax optimization problem, develop a local stochastic compositional gradient descent ascent with momentum algorithm, and provide bounds on the computational and communication complexities of our algorithm. To the best of our knowledge, this is the first work to achieve such favorable theoretical results. Finally, extensive experimental results confirm the efficacy of our method.
****
On Learning Latent Models with Multi-Instance Weak Supervision
Kaifu Wang, Efthymia Tsamoura, Dan Roth
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Degraded Polygons Raise Fundamental Questions of Neural Network Perception
Leonard Tang, Dan Ley
It is well-known that modern computer vision systems often exhibit behaviors misaligned with those of humans: from adversarial attacks to image corruptions, deeplearning vision models suffer in a variety of settings that humans capably handle. Inlight of these phenomena, here we introduce another, orthogonal perspective studying the human-machine vision gap. We revisit the task of recovering images underdegradation, first introduced over 30 years ago in the Recognition-by-Componentstheory of human vision. Specifically, we study the performance and behavior ofneural networks on the seemingly simple task of classifying regular polygons atvarying orders of degradation along their perimeters. To this end, we implement theAutomated Shape Recoverability Testfor rapidly generating large-scale datasetsof perimeter-degraded regular polygons, modernizing the historically manual creation of image recoverability experiments. We then investigate the capacity ofneural networks to recognize and recover such degraded shapes when initialized with different priors. Ultimately, we find that neural networks' behavior on thissimple task conflicts with human behavior, raising a fundamental question of therobustness and learning capabilities of modern computer vision models
****
Dynamics of Finite Width Kernel and Prediction Fluctuations in Mean Field Neural Networks
Blake Bordelon, Cengiz Pehlevan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
TART: A plug-and-play Transformer module for task-agnostic reasoning
Kush Bhatia, Avanika Narayan, Christopher M. De Sa, Christopher Ré
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Navigating the Pitfalls of Active Learning Evaluation: A Systematic Framework for Meaningful Performance Assessment
Carsten Lüth, Till Bungert, Lukas Klein, Paul Jaeger
Active Learning (AL) aims to reduce the labeling burden by interactively selecting the most informative samples from a pool of unlabeled data. While there has been extensive research on improving AL query methods in recent years, some studies have questioned the effectiveness of AL compared to emerging paradigms such as semi-supervised (Semi-SL) and self-supervised learning (Self-SL), or a simple optimization of classifier configurations. Thus, today’s AL literature presents an inconsistent and contradictory landscape, leaving practitioners uncertain about whether and how to use AL in their tasks. In this work, we make the case that this inconsistency arises from a lack of systematic and realistic evaluation of AL methods. Specifically, we identify five key pitfalls in the current literature that reflect the delicate considerations required for AL evaluation. Further, we present an evaluation framework that overcomes these pitfalls and thus enables meaningful statements about the performance of AL methods. To demonstrate the relevance of our protocol, we present a large-scale empirical study and benchmark for image classification spanning various data sets, query methods, AL settings, and training paradigms. Our findings clarify the inconsistent picture in the literature and enable us to give hands-on recommendations for practitioners. The benchmark is hosted at https://github.com/IML-DKFZ/realistic-al.
****
Semantic HELM: A Human-Readable Memory for Reinforcement Learning
Fabian Paischer, Thomas Adler, Markus Hofmarcher, Sepp Hochreiter
Reinforcement learning agents deployed in the real world often have to cope with partially observable environments. Therefore, most agents employ memory mechanisms to approximate the state of the environment. Recently, there have been impressive success stories in mastering partially observable environments, mostly in the realm of computer games like Dota 2, StarCraft II, or MineCraft. However, existing methods lack interpretability in the sense that it is not comprehensible for humans what the agent stores in its memory.In this regard, we propose a novel memory mechanism that represents past events in human language.Our method uses CLIP to associate visual inputs with language tokens. Then we feed these tokens to a pretrained language model that serves the agent as memory and provides it with a coherent and human-readable representation of the past.We train our memory mechanism on a set of partially observable environments and find that it excels on tasks that require a memory component, while mostly attaining performance on-par with strong baselines on tasks that do not. On a challenging continuous recognition task, where memorizing the past is crucial, our memory mechanism converges two orders of magnitude faster than prior methods.Since our memory mechanism is human-readable, we can peek at an agent's memory and check whether crucial pieces of information have been stored.This significantly enhances troubleshooting and paves the way toward more interpretable agents.
****
Empowering Convolutional Neural Nets with MetaSin Activation
Farnood Salehi, Tunç Aydin, André Gaillard, Guglielmo Camporese, Yuxuan Wang
ReLU networks have remained the default choice for models in the area of image prediction despite their well-established spectral bias towards learning low frequencies faster, and consequently their difficulty of reproducing high frequency visual details. As an alternative, sin networks showed promising results in lear

ning implicit representations of visual data. However training these networks in practically relevant settings proved to be difficult, requiring careful initialization, dealing with issues due to inconsistent gradients, and a degeneracy in local minima. In this work, we instead propose replacing a baseline network's existing activations with a novel ensemble function with trainable parameters. The proposed MetaSin activation can be trained reliably without requiring intricate initialization schemes, and results in consistently lower test loss compared to alternatives. We demonstrate our method in the areas of Monte-Carlo denoising and image resampling where we set new state-of-the-art through a knowledge distillation based training procedure. We present ablations on hyper-parameter settings, comparisons with alternative activation function formulations, and discuss the use of our method in other domains, such as image classification.
****
When Does Confidence-Based Cascade Deferral Suffice?
Wittawat Jitkrittum, Neha Gupta, Aditya K. Menon, Harikrishna Narasimhan, Ankit Rawat, Sanjiv Kumar
Cascades are a classical strategy to enable inference cost to vary adaptively across samples, wherein a sequence of classifiers are invoked in turn. A deferral rule determines whether to invoke the next classifier in the sequence, or to terminate prediction. One simple deferral rule employs the confidence of the current classifier, e.g., based on the maximum predicted softmax probability. Despite being oblivious to the structure of the cascade --- e.g., not modelling the errors of downstream models --- such confidence-based deferral often works remarkably well in practice. In this paper, we seek to better understand the conditions under which confidence-based deferral may fail, and when alternate deferral strategies can perform better. We first present a theoretical characterisation of the optimal deferral rule, which precisely characterises settings under which confidence-based deferral may suffer. We then study post-hoc deferral mechanisms, and demonstrate they can significantly improve upon confidence-based deferral in settings where (i) downstream models are specialists that only work well on a subset of inputs, (ii) samples are subject to label noise, and (iii) there is distribution shift between the train and test set.
****
DeWave: Discrete Encoding of EEG Waves for EEG to Text Translation
Yiqun Duan, Charles Chau, Zhen Wang, Yu-Kai Wang, Chin-teng Lin
The translation of brain dynamics into natural language is pivotal for brain-computer interfaces (BCIs), a field that has seen substantial growth in recent years. With the swift advancement of large language models, such as ChatGPT, the need to bridge the gap between the brain and languages becomes increasingly pressing. Current methods, however, require eye-tracking fixations or event markers to segment brain dynamics into word-level features, which can restrict the practical application of these systems. These event markers may not be readily available or could be challenging to acquire during real-time inference, and the sequence of eye fixations may not align with the order of spoken words. To tackle these issues, we introduce a novel framework, DeWave, that integrates discrete encoding sequences into open-vocabulary EEG-to-text translation tasks. DeWave uses a quantized variational encoder to derive discrete codex encoding and align it with pre-trained language models. This discrete codex representation brings forth two advantages: 1) it alleviates the order mismatch between eye fixations and spoken words by introducing text-EEG contrastive alignment training, and 2) it minimizes the interference caused by individual differences in EEG waves through an invariant discrete codex. Our model surpasses the previous baseline (40.1 and 31.7) by 3.06% and 6.34\%, respectively, achieving 41.35 BLEU-1 and 33.71 Rouge-F on the ZuCo Dataset. Furthermore, this work is the first to facilitate the translation of entire EEG signal periods without the need for word-level order markers (e.g., eye fixations), scoring 20.5 BLEU-1 and 29.5 Rouge-1 on the ZuCo Dataset, respectively.
****
SpatialRank: Urban Event Ranking with NDCG Optimization on Spatiotemporal Data
BANG AN, Xun Zhou, YONGJIAN ZHONG, Tianbao Yang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
An Information-Theoretic Evaluation of Generative Models in Learning Multi-modal Distributions
Mohammad Jalali, Cheuk Ting Li, Farzan Farnia
The evaluation of generative models has received significant attention in the machine learning community. When applied to a multi-modal distribution which is common among image datasets, an intuitive evaluation criterion is the number of modes captured by the generative model. While several scores have been proposed to evaluate the quality and diversity of a model's generated data, the correspondence between existing scores and the number of modes in the distribution is unclear. In this work, we propose an information-theoretic diversity evaluation method for multi-modal underlying distributions. We utilize the R\'enyi Kernel Entropy (RKE) as an evaluation score based on quantum information theory to measure the number of modes in generated samples. To interpret the proposed evaluation method, we show that the RKE score can output the number of modes of a mixture of sub-Gaussian components. We also prove estimation error bounds for estimating the RKE score from limited data, suggesting a fast convergence of the empirical RKE score to the score for the underlying data distribution. Utilizing the RKE score, we conduct an extensive evaluation of state-of-the-art generative models over standard image datasets. The numerical results indicate that while the recent algorithms for training generative models manage to improve the mode-based diversity over the earlier architectures, they remain incapable of capturing the full diversity of real data. Our empirical results provide a ranking of widely-used generative models based on the RKE score of their generated samples.
****
A Cross-Moment Approach for Causal Effect Estimation
Yaroslav Kivva, Saber Salehkaleybar, Negar Kiyavash
We consider the problem of estimating the causal effect of a treatment on an outcome in linear structural causal models (SCM) with latent confounders when we have access to a single proxy variable.Several methods (such as difference-in-difference (DiD) estimator or negative outcome control) have been proposed in this setting in the literature. However, these approaches require either restrictive assumptions on the data generating model or having access to at least two proxy variables.We propose a method to estimate the causal effect using cross moments between the treatment, the outcome, and the proxy variable. In particular, we show that the causal effect can be identified with simple arithmetic operations on the cross moments if the latent confounder in linear SCM is non-Gaussian.In this setting, DiD estimator provides an unbiased estimate only in the special case where the latent confounder has exactly the same direct causal effects on the outcomes in the pre-treatment and post-treatment phases. This translates to the common trend assumption in DiD, which we effectively relax.Additionally, we provide an impossibility result that shows the causal effect cannot be identified if the observational distribution over the treatment, the outcome, and the proxy is jointly Gaussian. Our experiments on both synthetic and real-world datasets showcase the effectivenessof the proposed approach in estimating the causal effect.
****
Combining Behaviors with the Successor Features Keyboard
Wilka Carvalho Carvalho, Andre Saraiva, Angelos Filos, Andrew Lampinen, Loic Matthey, Richard L Lewis, Honglak Lee, Satinder Singh, Danilo Jimenez Rezende, Daniel Zoran
The Option Keyboard (OK) was recently proposed as a method for transferring behavioral knowledge across tasks. OK transfers knowledge by adaptively combining subsets of known behaviors using Successor Features (SFs) and Generalized Policy Improvement (GPI).However, it relies on hand-designed state-features and task encodings which are cumbersome to design for every new environment.In this work, we propose the "Successor Features Keyboard" (SFK), which enables transfer with di

scovered state-features and task encodings.To enable discovery, we propose the "Categorical Successor Feature Approximator" (CSFA), a novel learning algorithm for estimating SFs while jointly discovering state-features and task encodings.With SFK and CSFA, we achieve the first demonstration of transfer with SFs in a challenging 3D environment where all the necessary representations are discovered. We first compare CSFA against other methods for approximating SFs and show that only CSFA discovers representations compatible with SF&GPI at this scale.We then compare SFK against transfer learning baselines and show that it transfers most quickly to long-horizon tasks.
****
Rethinking Semi-Supervised Medical Image Segmentation: A Variance-Reduction Perspective
Chenyu You, Weicheng Dai, Yifei Min, Fenglin Liu, David Clifton, S. Kevin Zhou, Lawrence Staib, James Duncan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Boosting Verification of Deep Reinforcement Learning via Piece-Wise Linear Decision Neural Networks
Jiaxu Tian, Dapeng Zhi, Si Liu, Peixin Wang, Cheng Chen, Min Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Star-Shaped Denoising Diffusion Probabilistic Models
Andrey Okhotin, Dmitry Molchanov, Arkhipkin Vladimir, Grigory Bartosh, Viktor Ohanesian, Aibek Alanov, Dmitry P. Vetrov
Denoising Diffusion Probabilistic Models (DDPMs) provide the foundation for the recent breakthroughs in generative modeling.Their Markovian structure makes it difficult to define DDPMs with distributions other than Gaussian or discrete.In this paper, we introduce Star-Shaped DDPM (SS-DDPM).Its star-shaped diffusion process allows us to bypass the need to define the transition probabilities or compute posteriors.We establish duality between star-shaped and specific Markovian diffusions for the exponential family of distributions and derive efficient algorithms for training and sampling from SS-DDPMs.In the case of Gaussian distributions, SS-DDPM is equivalent to DDPM.However, SS-DDPMs provide a simple recipe for designing diffusion models with distributions such as Beta, von Mises–Fisher, Dirichlet, Wishart and others, which can be especially useful when data lies on a constrained manifold.We evaluate the model in different settings and find it competitive even on image data, where Beta SS-DDPM achieves results comparable to a Gaussian DDPM.Our implementation is available at https://github.com/andrey-okhotin/star-shaped
****
An Adaptive Algorithm for Learning with Unknown Distribution Drift
Alessio Mazzetto, Eli Upfal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
QLoRA: Efficient Finetuning of Quantized LLMs
Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, Luke Zettlemoyer
We present QLoRA, an efficient finetuning approach that reduces memory usage enough to finetune a 65B parameter model on a single 48GB GPU while preserving full 16-bit finetuning task performance. QLoRA backpropagates gradients through a frozen, 4-bit quantized pretrained language model into Low Rank Adapters~(LoRA). Our best model family, which we name Guanaco, outperforms all previous openly rel

eased models on the Vicuna benchmark, reaching 99.3% of the performance level of ChatGPT while only requiring 24 hours of finetuning on a single GPU. QLoRA introduces a number of innovations to save memory without sacrificing performance: (a) 4-bit NormalFloat (NF4), a new data type that is information-theoretically optimal for normally distributed weights (b) Double Quantization to reduce the average memory footprint by quantizing the quantization constants, and (c) Paged Optimziers to manage memory spikes. We use QLoRA to finetune more than 1,000 models, providing a detailed analysis of instruction following and chatbot performance across 8 instruction datasets, multiple model types (LLaMA, T5), and model scales that would be infeasible to run with regular finetuning (e.g. 33B and 65B parameter models). Our results show that QLoRA finetuning on a small, high-quality dataset leads to state-of-the-art results, even when using smaller models than the previous SoTA. We provide a detailed analysis of chatbot performance based on both human and GPT-4 evaluations, showing that GPT-4 evaluations are a cheap and reasonable alternative to human evaluation. Furthermore, we find that current chatbot benchmarks are not trustworthy to accurately evaluate the performance levels of chatbots. A lemon-picked analysis demonstrates where Guanaco fails compared to ChatGPT. We release all of our models and code, including CUDA kernels for 4-bit training.
****
EMMA-X: An EM-like Multilingual Pre-training Algorithm for Cross-lingual Representation Learning
Ping Guo, Xiangpeng Wei, Yue Hu, Baosong Yang, Dayiheng Liu, Fei Huang, jun xie
Expressing universal semantics common to all languages is helpful to understand the meanings of complex and culture-specific sentences. The research theme underlying this scenario focuses on learning universal representations across languages with the usage of massive parallel corpora. However, due to the sparsity and scarcity of parallel data, there is still a big challenge in learning authentic ``universals'' for any two languages. In this paper, we propose Emma-X: an EM-like Multilingual pre-training Algorithm, to learn Cross-lingual universals with the aid of excessive multilingual non-parallel data. Emma-X unifies the cross-lingual representation learning task and an extra semantic relation prediction task within an EM framework. Both the extra semantic classifier and the cross-lingual sentence encoder approximate the semantic relation of two sentences, and supervise each other until convergence. To evaluate Emma-X, we conduct experiments on xrete, a newly introduced benchmark containing 12 widely studied cross-lingual tasks that fully depend on sentence-level representations. Results reveal that Emma-X achieves state-of-the-art performance. Further geometric analysis of the built representation space with three requirements demonstrates the superiority of Emma-X over advanced models.
****
Tester-Learners for Halfspaces: Universal Algorithms
Aravind Gollakota, Adam Klivans, Konstantinos Stavropoulos, Arsen Vasilyan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Revisit the Power of Vanilla Knowledge Distillation: from Small Scale to Large Scale
Zhiwei Hao, Jianyuan Guo, Kai Han, Han Hu, Chang Xu, Yunhe Wang
The tremendous success of large models trained on extensive datasets demonstrates that scale is a key ingredient in achieving superior results. Therefore, the reflection on the rationality of designing knowledge distillation (KD) approaches for limited-capacity architectures solely based on small-scale datasets is now deemed imperative. In this paper, we identify the small data pitfall that presents in previous KD methods, which results in the underestimation of the power of vanilla KD framework on large-scale datasets such as ImageNet-1K. Specifically, we show that employing stronger data augmentation techniques and using larger datasets can directly decrease the gap between vanilla KD and other meticulously d

esigned KD variants. This highlights the necessity of designing and evaluating KD approaches in the context of practical scenarios, casting off the limitations of small-scale datasets. Our investigation of the vanilla KD and its variants in more complex schemes, including stronger training strategies and different model capacities, demonstrates that vanilla KD is elegantly simple but astonishingly effective in large-scale scenarios. Without bells and whistles, we obtain state-of-the-art ResNet-50, ViT-S, and ConvNeXtV2-T models for ImageNet, which achieve 83.1%, 84.3%, and 85.0% top-1 accuracy, respectively. PyTorch code and checkpoints can be found at https://github.com/Hao840/vanillaKD.
****

Humans in Kitchens: A Dataset for Multi-Person Human Motion Forecasting with Scene Context
Julian Tanke, Oh-Hun Kwon, Felix B Mueller, Andreas Doering, Jürgen Gall
Forecasting human motion of multiple persons is very challenging. It requires to model the interactions between humans and the interactions with objects and the environment. For example, a person might want to make a coffee, but if the coffee machine is already occupied the person will haveto wait. These complex relations between scene geometry and persons ariseconstantly in our daily lives, and models that wish to accurately forecasthuman behavior will have to take them into consideration. To facilitate research in this direction, we propose Humans in Kitchens, alarge-scale multi-person human motion dataset with annotated 3D human poses, scene geometry and activities per person and frame.Our dataset consists of over 7.3h recorded data of up to 16 persons at the same time in four kitchen scenes, with more than 4M annotated human poses, represented by a parametric 3D body model. In addition, dynamic scene geometry and objects like chair or cupboard are annotated per frame. As first benchmarks, we propose two protocols for short-term and long-term human motion forecasting.
****

LD2: Scalable Heterophilous Graph Neural Network with Decoupled Embeddings
Ningyi Liao, Siqiang Luo, Xiang Li, Jieming Shi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

On Single-Index Models beyond Gaussian Data
Aaron Zweig, Loucas PILLAUD-VIVIEN, Joan Bruna
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Privacy Assessment on Reconstructed Images: Are Existing Evaluation Metrics Faithful to Human Perception?
Xiaoxiao Sun, Nidham Gazagnadou, Vivek Sharma, Lingjuan Lyu, Hongdong Li, Liang Zheng
Hand-crafted image quality metrics, such as PSNR and SSIM, are commonly used to evaluate model privacy risk under reconstruction attacks. Under these metrics, reconstructed images that are determined to resemble the original one generally indicate more privacy leakage. Images determined as overall dissimilar, on the other hand, indicate higher robustness against attack. However, there is no guarantee that these metrics well reflect human opinions, which offers trustworthy judgement for model privacy leakage. In this paper, we comprehensively study the faithfulness of these hand-crafted metrics to human perception of privacy information from the reconstructed images. On 5 datasets ranging from natural images, faces, to fine-grained classes, we use 4 existing attack methods to reconstruct images from many different classification models and, for each reconstructed image, we ask multiple human annotators to assess whether this image is recognizable. Our studies reveal that the hand-crafted metrics only have a weak correlation with the human evaluation of privacy leakage and that even these metrics themsel

ves often contradict each other. These observations suggest risks of current met rics in the community. To address this potential risk, we propose a learning-ba sed measure called SemSim to evaluate the Semantic Similarity between the origin al and reconstructed images. SemSim is trained with a standard triplet loss, usi ng an original image as an anchor, one of its recognizable reconstructed images as a positive sample, and an unrecognizable one as a negative. By training on hu man annotations, SemSim exhibits a greater reflection of privacy leakage on the semantic level. We show that SemSim has a significantly higher correlation with human judgment compared with existing metrics. Moreover, this strong correlation generalizes to unseen datasets, models and attack methods. We envision this wor k as a milestone for image quality evaluation closer to the human level. The pro ject webpage can be accessed at https://sites.google.com/view/semsim.
****
Graph Denoising Diffusion for Inverse Protein Folding
Kai Yi, Bingxin Zhou, Yiqing Shen, Pietro Lió, Yuguang Wang
Inverse protein folding is challenging due to its inherent one-to-many mapping c haracteristic, where numerous possible amino acid sequences can fold into a sing le, identical protein backbone. This task involves not only identifying viable s equences but also representing the sheer diversity of potential solutions. Howev er, existing discriminative models, such as transformer-based auto-regressive mo dels, struggle to encapsulate the diverse range of plausible solutions. In contr ast, diffusion probabilistic models, as an emerging genre of generative approach es, offer the potential to generate a diverse set of sequence candidates for det ermined protein backbones. We propose a novel graph denoising diffusion model fo r inverse protein folding, where a given protein backbone guides the diffusion p rocess on the corresponding amino acid residue types. The model infers the joint distribution of amino acids conditioned on the nodes' physiochemical properties and local environment. Moreover, we utilize amino acid replacement matrices for the diffusion forward process, encoding the biologically-meaningful prior knowl edge of amino acids from their spatial and sequential neighbors as well as thems elves, which reduces the sampling space of the generative process. Our model ach ieves state-of-the-art performance over a set of popular baseline methods in seq uence recovery and exhibits great potential in generating diverse protein sequen ces for a determined protein backbone structure.
****
AttrSeg: Open-Vocabulary Semantic Segmentation via Attribute Decomposition-Aggre gation
Chaofan Ma, Yang Yuhuan, Chen Ju, Fei Zhang, Ya Zhang, Yanfeng Wang
Open-vocabulary semantic segmentation is a challenging task that requires segmen ting novel object categories at inference time. Recent works explore vision-lang uage pre-training to handle this task, but suffer from unrealistic assumptions i n practical scenarios, i.e., low-quality textual category names.For example, thi s paradigm assumes that new textual categories will be accurately and completely provided, and exist in lexicons during pre-training.However, exceptions often h appen when meet with ambiguity for brief or incomplete names, new words that are not present in the pre-trained lexicons, and difficult-to-describe categories f or users.To address these issues, this work proposes a novel attribute decomposi tion-aggregation framework, AttrSeg, inspired by human cognition in understandin g new concepts. Specifically, in the decomposition stage, we decouple class name s into diverse attribute descriptions to complement semantic contexts from multi ple perspectives.Two attribute construction strategies are designed: using large language models for common categories, and involving manually labelling for hum an-invented categories. In the aggregation stage, we group diverse attributes in to an integrated global description, to form a discriminative classifier that di stinguishes the target object from others. One hierarchical aggregation architec ture is further proposed to achieve multi-level aggregation, leveraging the meti culously designed clustering module.The final result is obtained by computing th e similarity between aggregated attributes and images embedding.To evaluate the effectiveness, we annotate three datasets with attribute descriptions, and condu ct extensive experiments and ablation studies. The results show the superior per

formance of attribute decomposition-aggregation.We refer readers to the latest arXiv version at https://arxiv.org/abs/2309.00096.
****
Stable and low-precision training for large-scale vision-language models
Mitchell Wortsman, Tim Dettmers, Luke Zettlemoyer, Ari Morcos, Ali Farhadi, Ludwig Schmidt
We introduce new methods for 1) accelerating and 2) stabilizing training for large language-vision models. 1) For acceleration, we introduce SwitchBack, a linear layer for int8 quantized training which provides a speed-up of 13-25% while matching the performance of bfloat16 training within 0.1 percentage points for the 1B parameter CLIP ViT-Huge---the largest int8 training to date. Our main focus is int8 as GPU support for float8 is rare, though we also analyze float8 training through simulation. While SwitchBack proves effective for float8, we show that standard techniques are also successful if the network is trained and initialized so that large feature magnitudes are discouraged, which we accomplish via layer-scale initialized with zeros. 2) For stability, we analyze loss spikes and find they consistently occur 1-8 iterations after the squared gradients become under-estimated by their AdamW second moment estimator. As a result, we recommend an AdamW-Adafactor hybrid which avoids loss spikes when training a CLIP ViT-Huge model and outperforms gradient clipping at the scales we test.
****
Recommender Systems with Generative Retrieval
Shashank Rajput, Nikhil Mehta, Anima Singh, Raghunandan Hulikal Keshavan, Trung Vu, Lukasz Heldt, Lichan Hong, Yi Tay, Vinh Tran, Jonah Samost, Maciej Kula, Ed Chi, Maheswaran Sathiamoorthy
Modern recommender systems perform large-scale retrieval by embedding queries and item candidates in the same unified space, followed by approximate nearest neighbor search to select top candidates given a query embedding. In this paper, we propose a novel generative retrieval approach, where the retrieval model autoregressively decodes the identifiers of the target candidates. To that end, we create semantically meaningful tuple of codewords to serve as a Semantic ID for each item. Given Semantic IDs for items in a user session, a Transformer-based sequence-to-sequence model is trained to predict the Semantic ID of the next item that the user will interact with. We show that recommender systems trained with the proposed paradigm significantly outperform the current SOTA models on various datasets. In addition, we show that incorporating Semantic IDs into the sequence-to-sequence model enhances its ability to generalize, as evidenced by the improved retrieval performance observed for items with no prior interaction history.
****
Active Vision Reinforcement Learning under Limited Visual Observability
Jinghuan Shang, Michael S Ryoo
In this work, we investigate Active Vision Reinforcement Learning (ActiveVision-RL), where an embodied agent simultaneously learns action policy for the task while also controlling its visual observations in partially observable environments. We denote the former as motor policy and the latter as sensory policy. For example, humans solve real world tasks by hand manipulation (motor policy) together with eye movements (sensory policy). ActiveVision-RL poses challenges on coordinating two policies given their mutual influence. We propose SUGARL, Sensorimotor Understanding Guided Active Reinforcement Learning, a framework that models motor and sensory policies separately, but jointly learns them using with an intrinsic sensorimotor reward. This learnable reward is assigned by sensorimotor reward module, incentivizes the sensory policy to select observations that are optimal to infer its own motor action, inspired by the sensorimotor stage of humans. Through a series of experiments, we show the effectiveness of our method across a range of observability conditions and its adaptability to existed RL algorithms. The sensory policies learned through our method are observed to exhibit effective active vision strategies.
****
Optimization of Inter-group criteria for clustering with minimum size constraints

Eduardo Laber, Lucas Murtinho
Internal measures that are used to assess the quality of a clustering usually take into account intra-group and/or inter-group criteria.There are many papers in the literature that propose algorithms with provable approximation guarantees for optimizing the former. However, the optimization of inter-group criteria is much less understood.Here, we contribute to the state-of-the-art of this literature by devising algorithms with provable guarantees for the maximization of two natural inter-group criteria, namely the minimum spacing and the minimum spanning tree spacing. The former is the minimum distance between points in different groups while the latter captures separability through the cost of the minimum spanning tree that connects all groups. We obtain results for both the unrestricted case, in which no constraint on the clusters is imposed, and for the constrained case where each group is required to have a minimum number of points. Our constraint is motivated by the fact that the popular Single-Linkage, which optimizes both criteria in the unrestricted case, produces clustering with many tiny groups.To complement our work, we present an empirical study with 10 real datasets that provides evidence that our methods work very well in practical settings.
****
CycleNet: Rethinking Cycle Consistency in Text-Guided Diffusion for Image Manipulation
Sihan Xu, Ziqiao Ma, Yidong Huang, Honglak Lee, Joyce Chai
Diffusion models (DMs) have enabled breakthroughs in image synthesis tasks but lack an intuitive interface for consistent image-to-image (I2I) translation. Various methods have been explored to address this issue, including mask-based methods, attention-based methods, and image-conditioning. However, it remains a critical challenge to enable unpaired I2I translation with pre-trained DMs while maintaining satisfying consistency. This paper introduces Cyclenet, a novel but simple method that incorporates cycle consistency into DMs to regularize image manipulation. We validate Cyclenet on unpaired I2I tasks of different granularities. Besides the scene and object level translation, we additionally contribute a multi-domain I2I translation dataset to study the physical state changes of objects. Our empirical studies show that Cyclenet is superior in translation consistency and quality, and can generate high-quality images for out-of-domain distributions with a simple change of the textual prompt. Cyclenet is a practical framework, which is robust even with very limited training data (around 2k) and requires minimal computational resources (1 GPU) to train. Project homepage: https://cyclenetweb.github.io/
****
Bandit Social Learning under Myopic Behavior
Kiarash Banihashem, MohammadTaghi Hajiaghayi, Suho Shin, Aleksandrs Slivkins
We study social learning dynamics motivated by reviews on online platforms. Theagents collectively follow a simple multi-armed bandit protocol, but each agentacts myopically, without regards to exploration. We allow a wide range of myopicbehaviors that are consistent with (parameterized) confidence intervals for the arms'expected rewards. We derive stark exploration failures for any such behavior, andprovide matching positive results. As a special case, we obtain the first generalresults on failure of the greedy algorithm in bandits, thus providing a theoreticalfoundation for why bandit algorithms should explore.
****
Gradient Flossing: Improving Gradient Descent through Dynamic Control of Jacobians
Rainer Engelken
Training recurrent neural networks (RNNs) remains a challenge due to the instability of gradients across long time horizons, which can lead to exploding and vanishing gradients. Recent research has linked these problems to the values of Lyapunov exponents for the forward-dynamics, which describe the growth or shrinkage of infinitesimal perturbations. Here, we propose gradient flossing, a novel approach to tackling gradient instability by pushing Lyapunov exponents of the forward dynamics toward zero during learning. We achieve this by regularizing Lyapunov exponents through backpropagation using differentiable linear algebra. This e

nables us to "floss" the gradients, stabilizing them and thus improving network training. We show that gradient flossing controls not only the gradient norm but also the condition number of the long-term Jacobian, facilitating multidimensional error feedback propagation. We find that applying gradient flossing before training enhances both the success rate and convergence speed for tasks involving long time horizons.For challenging tasks, we show that gradient flossing during training can further increase the time horizon that can be bridged by backpropagation through time. Moreover, we demonstrate the effectiveness of our approach on various RNN architectures and tasks of variable temporal complexity. Additionally, we provide a simple implementation of our gradient flossing algorithm that can be used in practice. Our results indicate that gradient flossing via regularizing Lyapunov exponents can significantly enhance the effectiveness of RNN training and mitigate the exploding and vanishing gradients problem.
****
How to Data in Datathons
Carlos Mougan, Richard Plant, Clare Teng, Marya Bazzi, Alvaro Cabrejas Egea, Ryan Chan, David Salvador Jasin, Martin Stoffel, Kirstie Whitaker, JULES MANSER
The rise of datathons, also known as data or data science hackathons, has provided a platform to collaborate, learn, and innovate quickly. Despite their significant potential benefits, organizations often struggle to effectively work with data due to a lack of clear guidelines and best practices for potential issues that might arise. Drawing on our own experiences and insights from organizing +80 datathon challenges with +60 partnership organizations since 2016, we provide a guide that serves as a resource for organizers to navigate the data-related complexities of datathons. We apply our proposed framework to 10 case studies.
****
Convolution Monge Mapping Normalization for learning on sleep data
Théo Gnassounou, Rémi Flamary, Alexandre Gramfort
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Online learning of long-range dependencies
Nicolas Zucchet, Robert Meier, Simon Schug, Asier Mujika, Joao Sacramento
Online learning holds the promise of enabling efficient long-term credit assignment in recurrent neural networks. However, current algorithms fall short of offline backpropagation by either not being scalable or failing to learn long-range dependencies. Here we present a high-performance online learning algorithm that merely doubles the memory and computational requirements of a single inference pass. We achieve this by leveraging independent recurrent modules in multi-layer networks, an architectural motif that has recently been shown to be particularly powerful. Experiments on synthetic memory problems and on the challenging long-range arena benchmark suite reveal that our algorithm performs competitively, establishing a new standard for what can be achieved through online learning. This ability to learn long-range dependencies offers a new perspective on learning in the brain and opens a promising avenue in neuromorphic computing.
****
Fast Rank-1 Lattice Targeted Sampling for Black-box Optimization
Yueming LYU
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DreamHuman: Animatable 3D Avatars from Text
Nikos Kolotouros, Thiemo Alldieck, Andrei Zanfir, Eduard Bazavan, Mihai Fieraru, Cristian Sminchisescu
We present \emph{DreamHuman}, a method to generate realistic animatable 3D human avatar models entirely from textual descriptions. Recent text-to-3D methods hav

e made considerable strides in generation, but are still lacking in important as pects. Control and often spatial resolution remain limited, existing methods pro duce fixed rather than 3D human models that can be placed in different poses (i. e. re-posable or animatable), and anthropometric consistency for complex structu res like people remains a challenge. \emph{DreamHuman} connects large text-to-im age synthesis models, neural radiance fields, and statistical human body models in a novel optimization framework. This makes it possible to generate dynamic 3D human avatars with high-quality textures and learnt per-instance rigid and non rigid geometric deformations. We demonstrate that our method is capable to gener ate a wide variety of animatable, realistic 3D human models from text. These hav e diverse appearance, clothing, skin tones and body shapes, and outperform both generic text-to-3D approaches and previous text-based 3D avatar generators in vi sual fidelity.
****
Rank-1 Matrix Completion with Gradient Descent and Small Random Initialization
Daesung Kim, Hye Won Chung
The nonconvex formulation of the matrix completion problem has received signific ant attention in recent years due to its affordable complexity compared to the c onvex formulation. Gradient Descent (GD) is a simple yet efficient baseline algo rithm for solving nonconvex optimization problems. The success of GD has been wi tnessed in many different problems in both theory and practice when it is combin ed with random initialization. However, previous works on matrix completion requ ire either careful initialization or regularizers to prove the convergence of GD . In this paper, we study the rank-1 symmetric matrix completion and prove that GD converges to the ground truth when small random initialization is used. We sh ow that in a logarithmic number of iterations, the trajectory enters the region where local convergence occurs. We provide an upper bound on the initialization size that is sufficient to guarantee the convergence, and show that a larger ini tialization can be used as more samples are available. We observe that the impli cit regularization effect of GD plays a critical role in the analysis, and for t he entire trajectory, it prevents each entry from becoming much larger than the others.
****
No-Regret Learning in Dynamic Competition with Reference Effects Under Logit Dem and
Mengzi Amy Guo, Donghao Ying, Javad Lavaei, Zuo-Jun Shen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
VoxDet: Voxel Learning for Novel Instance Detection
Bowen Li, Jiashun Wang, Yaoyu Hu, Chen Wang, Sebastian Scherer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Evaluating and Inducing Personality in Pre-trained Language Models
Guangyuan Jiang, Manjie Xu, Song-Chun Zhu, Wenjuan Han, Chi Zhang, Yixin Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
On Measuring Fairness in Generative Models
Christopher Teo, Milad Abdollahzadeh, Ngai-Man (Man) Cheung
Recently, there has been increased interest in fair generative models. In this w ork,we conduct, for the first time, an in-depth study on fairness measurement, a critical component in gauging progress on fair generative models. We make threec

ontributions. First, we conduct a study that reveals that the existing fairnessm easurement framework has considerable measurement errors, even when highlyaccura te sensitive attribute (SA) classifiers are used. These findings cast doubtson p reviously reported fairness improvements. Second, to address this issue,we propo se CLassifier Error-Aware Measurement (CLEAM), a new frameworkwhich uses a stati stical model to account for inaccuracies in SA classifiers. Ourproposed CLEAM re duces measurement errors significantly, e.g., 4.98%→0.62%for StyleGAN2 w.r.t. Ge nder. Additionally, CLEAM achieves this with minimaladditional overhead. Third, we utilize CLEAM to measure fairness in importanttext-to-image generator and GAN s, revealing considerable biases in these modelsthat raise concerns about their applications. Code and more resources: https://sutd-visual-computing-group.githu b.io/CLEAM/.
****
IMPRESS: Evaluating the Resilience of Imperceptible Perturbations Against Unauth orized Data Usage in Diffusion-Based Generative AI
Bochuan Cao, Changjiang Li, Ting Wang, Jinyuan Jia, Bo Li, Jinghui Chen
Diffusion-based image generation models, such as Stable Diffusion or DALL·E 2, are able to learn from given images and generate high-quality samples following the guidance from prompts. For instance, they can be used to create artistic ima ges that mimic the style of an artist based on his/her original artworks or to m aliciously edit the original images for fake content. However, such ability also brings serious ethical issues without proper authorization from the owner of th e original images. In response, several attempts have been made to protect the o riginal images from such unauthorized data usage by adding imperceptible perturb ations, which are designed to mislead the diffusion model and make it unable to properly generate new samples. In this work, we introduce a perturbation purific ation platform, named IMPRESS, to evaluate the effectiveness of imperceptible pe rturbations as a protective measure.IMPRESS is based on the key observation that imperceptible perturbations could lead to a perceptible inconsistency between t he original image and the diffusion-reconstructed image, which can be used to de vise a new optimization strategy for purifying the image, which may weaken the p rotection of the original image from unauthorized data usage (e.g., style mimick ing, malicious editing).The proposed IMPRESS platform offers a comprehensive eva luation of several contemporary protection methods, and can be used as an evalua tion platform for future protection methods.
****
Robust Contrastive Language-Image Pretraining against Data Poisoning and Backdoo r Attacks
Wenhan Yang, Jingdong Gao, Baharan Mirzasoleiman
Contrastive vision-language representation learning has achieved state-of-the-ar t performance for zero-shot classification, by learning from millions of image-c aption pairs crawled from the internet. However, the massive data that powers la rge multimodal models such as CLIP, makes them extremely vulnerable to various t ypes of targeted data poisoning and backdoor attacks. Despite this vulnerability , robust contrastive vision-language pre-training against such attacks has remai ned unaddressed. In this work, we propose RoCLIP, the first effective method for robust pre-training multimodal vision-language models against targeted data poi soning and backdoor attacks. RoCLIP effectively breaks the association between p oisoned image-caption pairs by considering a relatively large and varying pool o f random captions, and matching every image with the text that is most similar t o it in the pool instead of its own caption, every few epochs.It also leverages image and text augmentations to further strengthen the defense and improve the p erformance of the model. Our extensive experiments show that RoCLIP renders stat e-of-the-art targeted data poisoning and backdoor attacks ineffective during pre -training CLIP models. In particular, RoCLIP decreases the success rate for targ eted data poisoning attacks from 93.75% to 12.5% and that of backdoor attacks do wn to 0%, while improving the model's linear probe performance by 10% and mainta ins a similar zero shot performance compared to CLIP. By increasing the frequenc y of matching, RoCLIP is able to defend strong attacks, which add up to 1% poiso ned examples to the data, and successfully maintain a low attack success rate of

12.5%, while trading off the performance on some tasks.
****
Block-Coordinate Methods and Restarting for Solving Extensive-Form Games
Darshan Chakrabarti, Jelena Diakonikolas, Christian Kroer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ReContrast: Domain-Specific Anomaly Detection via Contrastive Reconstruction
Jia Guo, shuai lu, Lize Jia, Weihang Zhang, Huiqi Li
Most advanced unsupervised anomaly detection (UAD) methods rely on modeling feature representations of frozen encoder networks pre-trained on large-scale datasets, e.g. ImageNet. However, the features extracted from the encoders that are borrowed from natural image domains coincide little with the features required in the target UAD domain, such as industrial inspection and medical imaging. In this paper, we propose a novel epistemic UAD method, namely ReContrast, which optimizes the entire network to reduce biases towards the pre-trained image domain and orients the network in the target domain. We start with a feature reconstruction approach that detects anomalies from errors. Essentially, the elements of contrastive learning are elegantly embedded in feature reconstruction to prevent the network from training instability, pattern collapse, and identical shortcut, while simultaneously optimizing both the encoder and decoder on the target domain. To demonstrate our transfer ability on various image domains, we conduct extensive experiments across two popular industrial defect detection benchmarks and three medical image UAD tasks, which shows our superiority over current state-of-the-art methods.
****
Transferable Adversarial Robustness for Categorical Data via Universal Robust Embeddings
Klim Kireev, Maksym Andriushchenko, Carmela Troncoso, Nicolas Flammarion
Research on adversarial robustness is primarily focused on image and text data. Yet, many scenarios in which lack of robustness can result in serious risks, such as fraud detection, medical diagnosis, or recommender systems often do not rely on images or text but instead on tabular data. Adversarial robustness in tabular data poses two serious challenges. First, tabular datasets often contain categorical features, and therefore cannot be tackled directly with existing optimization procedures. Second, in the tabular domain, algorithms that are not based on deep networks are widely used and offer great performance, but algorithms to enhance robustness are tailored to neural networks (e.g. adversarial training).In this paper, we tackle both challenges. We present a method that allows us to train adversarially robust deep networks for tabular data and to transfer this robustness to other classifiers via universal robust embeddings tailored to categorical data. These embeddings, created using a bilevel alternating minimization framework, can be transferred to boosted trees or random forests making them robust without the need for adversarial training while preserving their high accuracy on tabular data. We show that our methods outperform existing techniques within a practical threat model suitable for tabular data.
****
Drift doesn't Matter: Dynamic Decomposition with Diffusion Reconstruction for Unstable Multivariate Time Series Anomaly Detection
Chengsen Wang, Zirui Zhuang, Qi Qi, Jingyu Wang, Xingyu Wang, Haifeng Sun, Jianxin Liao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
MetaBox: A Benchmark Platform for Meta-Black-Box Optimization with Reinforcement Learning

Zeyuan Ma, Hongshu Guo, Jiacheng Chen, Zhenrui Li, Guojun Peng, Yue-Jiao Gong, Yining Ma, Zhiguang Cao

Recently, Meta-Black-Box Optimization with Reinforcement Learning (MetaBBO-RL) has showcased the power of leveraging RL at the meta-level to mitigate manual fine-tuning of low-level black-box optimizers. However, this field is hindered by the lack of a unified benchmark. To fill this gap, we introduce MetaBox, the first benchmark platform expressly tailored for developing and evaluating MetaBBO-RL methods. MetaBox offers a flexible algorithmic template that allows users to effortlessly implement their unique designs within the platform. Moreover, it provides a broad spectrum of over 300 problem instances, collected from synthetic to realistic scenarios, and an extensive library of 19 baseline methods, including both traditional black-box optimizers and recent MetaBBO-RL methods. Besides, MetaBox introduces three standardized performance metrics, enabling a more thorough assessment of the methods. In a bid to illustrate the utility of MetaBox for facilitating rigorous evaluation and in-depth analysis, we carry out a wide-ranging benchmarking study on existing MetaBBO-RL methods. Our MetaBox is open-source and accessible at: https://github.com/GMC-DRL/MetaBox.

****

Improving Adversarial Robustness via Information Bottleneck Distillation

Huafeng Kuang, Hong Liu, Yongjian Wu, Shin'ichi Satoh, Rongrong Ji

Previous studies have shown that optimizing the information bottleneck can significantly improve the robustness of deep neural networks. Our study closely examines the information bottleneck principle and proposes an Information Bottleneck Distillation approach. This specially designed, robust distillation technique utilizes prior knowledge obtained from a robust pre-trained model to boost information bottlenecks. Specifically, we propose two distillation strategies that align with the two optimization processes of the information bottleneck. Firstly, we use a robust soft-label distillation method to increase the mutual information between latent features and output prediction. Secondly, we introduce an adaptive feature distillation method that automatically transfers relevant knowledge from the teacher model to the student model, thereby reducing the mutual information between the input and latent features. We conduct extensive experiments to evaluate our approach's robustness against state-of-the-art adversarial attackers such as PGD-attack and AutoAttack. Our experimental results demonstrate the effectiveness of our approach in significantly improving adversarial robustness. Our code is available at https://github.com/SkyKuang/IBD.

****

Reading Relevant Feature from Global Representation Memory for Visual Object Tracking

Xinyu Zhou, Pinxue Guo, Lingyi Hong, Jinglun Li, Wei Zhang, Weifeng Ge, Wenqiang Zhang

Reference features from a template or historical frames are crucial for visual object tracking. Prior works utilize all features from a fixed template or memory for visual object tracking. However, due to the dynamic nature of videos, the required reference historical information for different search regions at different time steps is also inconsistent. Therefore, using all features in the template and memory can lead to redundancy and impair tracking performance. To alleviate this issue, we propose a novel tracking paradigm, consisting of a relevance attention mechanism and a global representation memory, which can adaptively assist the search region in selecting the most relevant historical information from reference features. Specifically, the proposed relevance attention mechanism in this work differs from previous approaches in that it can dynamically choose and build the optimal global representation memory for the current frame by accessing cross-frame information globally. Moreover, it can flexibly read the relevant historical information from the constructed memory to reduce redundancy and counteract the negative effects of harmful information. Extensive experiments validate the effectiveness of the proposed method, achieving competitive performance on five challenging datasets with 71 FPS.

****

Provable Guarantees for Nonlinear Feature Learning in Three-Layer Neural Network

s
Eshaan Nichani, Alex Damian, Jason D. Lee
One of the central questions in the theory of deep learning is to understand how neural networks learn hierarchical features. The ability of deep networks to extract salient features is crucial to both their outstanding generalization ability and the modern deep learning paradigm of pretraining and finetuneing. However, this feature learning process remains poorly understood from a theoretical perspective, with existing analyses largely restricted to two-layer networks. In this work we show that three-layer neural networks have provably richer feature learning capabilities than two-layer networks. We analyze the features learned by a three-layer network trained with layer-wise gradient descent, and present a general purpose theorem which upper bounds the sample complexity and width needed to achieve low test error when the target has specific hierarchical structure. We instantiate our framework in specific statistical learning settings -- single-index models and functions of quadratic features -- and show that in the latter setting three-layer networks obtain a sample complexity improvement over all existing guarantees for two-layer networks. Crucially, this sample complexity improvement relies on the ability of three-layer networks to efficiently learn nonlinear features. We then establish a concrete optimization-based depth separation by constructing a function which is efficiently learnable via gradient descent on a three-layer network, yet cannot be learned efficiently by a two-layer network. Our work makes progress towards understanding the provable benefit of three-layer neural networks over two-layer networks in the feature learning regime.
****
Lockdown: Backdoor Defense for Federated Learning with Isolated Subspace Training
Tiansheng Huang, Sihao Hu, Ka-Ho Chow, Fatih Ilhan, Selim Tekin, Ling Liu
Federated learning (FL) is vulnerable to backdoor attacks due to its distributed computing nature. Existing defense solution usually requires larger amount of computation in either the training or testing phase, which limits their practicality in the resource-constrain scenarios. A more practical defense, i.e., neural network (NN) pruning based defense has been proposed in centralized backdoor setting. However, our empirical study shows that traditional pruning-based solution suffers \textit{poison-coupling} effect in FL, which significantly degrades the defense performance.This paper presents Lockdown, an isolated subspace training method to mitigate the poison-coupling effect. Lockdown follows three key procedures. First, it modifies the training protocol by isolating the training subspaces for different clients. Second, it utilizes randomness in initializing isolated subspacess, and performs subspace pruning and subspace recovery to segregate the subspaces between malicious and benign clients. Third, it introduces quorum consensus to cure the global model by purging malicious/dummy parameters. Empirical results show that Lockdown achieves \textit{superior} and \textit{consistent} defense performance compared to existing representative approaches against backdoor attacks. Another value-added property of Lockdown is the communication-efficiency and model complexity reduction, which are both critical for resource-constrain FL scenario. Our code is available at \url{https://github.com/git-disl/Lockdown}.
****
Robust Lipschitz Bandits to Adversarial Corruptions
Yue Kang, Cho-Jui Hsieh, Thomas Chun Man Lee
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Predicting Global Label Relationship Matrix for Graph Neural Networks under Heterophily
Langzhang Liang, Xiangjing Hu, Zenglin Xu, Zixing Song, Irwin King
Graph Neural Networks (GNNs) have been shown to achieve remarkable performance on node classification tasks by exploiting both graph structures and node feature

s. The majority of existing GNNs rely on the implicit homophily assumption. Rece
nt studies have demonstrated that GNNs may struggle to model heterophilous graph
s where nodes with different labels are more likely connected. To address this i
ssue, we propose a generic GNN applicable to both homophilous and heterophilous
graphs, namely Low-Rank Graph Neural Network (LRGNN). Our analysis demonstrates
that a signed graph's global label relationship matrix has a low rank. This insi
ght inspires us to predict the label relationship matrix by solving a robust low
-rank matrix approximation problem, as prior research has proven that low-rank a
pproximation could achieve perfect recovery under certain conditions. The experi
mental results reveal that the solution bears a strong resemblance to the label
relationship matrix, presenting two advantages for graph modeling: a block diago
nal structure and varying distributions of within-class and between-class entrie
s.
****
Into the Single Cell Multiverse: an End-to-End Dataset for Procedural Knowledge
Extraction in Biomedical Texts
Ruth Dannenfelser, Jeffrey Zhong, Ran Zhang, Vicky Yao
Many of the most commonly explored natural language processing (NLP) information
 extraction tasks can be thought of as evaluations of declarative knowledge, or
fact-based information extraction. Procedural knowledge extraction, i.e., breaki
ng down a described process into a series of steps, has received much less atten
tion, perhaps in part due to the lack of structured datasets that capture the kn
owledge extraction process from end-to-end. To address this unmet need, we prese
nt FlaMBé (Flow annotations for Multiverse Biological entities), a collection of
 expert-curated datasets across a series of complementary tasks that capture pro
cedural knowledge in biomedical texts. This dataset is inspired by the observati
on that one ubiquitous source of procedural knowledge that is described as unstr
uctured text is within academic papers describing their methodology. The workflo
ws annotated in FlaMBé are from texts in the burgeoning field of single cell res
earch, a research area that has become notorious for the number of software tool
s and complexity of workflows used. Additionally, FlaMBé provides, to our knowle
dge, the largest manually curated named entity recognition (NER) and disambiguat
ion (NED) datasets for tissue/cell type, a fundamental biological entity that is
 critical for knowledge extraction in the biomedical research domain. Beyond pro
viding a valuable dataset to enable further development of NLP models for proced
ural knowledge extraction, automating the process of workflow mining also has im
portant implications for advancing reproducibility in biomedical research.
****
RRHF: Rank Responses to Align Language Models with Human Feedback
Hongyi Yuan, Zheng Yuan, Chuanqi Tan, Wei Wang, Songfang Huang, Fei Huang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Sparsity-Preserving Differentially Private Training of Large Embedding Models
Badih Ghazi, Yangsibo Huang, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Amer
Sinha, Chiyuan Zhang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Bayesian target optimisation for high-precision holographic optogenetics
Marcus Triplett, Marta Gajowa, Hillel Adesnik, Liam Paninski
Two-photon optogenetics has transformed our ability to probe the structure and f
unction of neural circuits. However, achieving precise optogenetic control of ne
ural ensemble activity has remained fundamentally constrained by the problem of
off-target stimulation (OTS): the inadvertent activation of nearby non-target ne
urons due to imperfect confinement of light onto target neurons. Here we propose

a novel computational approach to this problem called Bayesian target optimisation. Our approach uses nonparametric Bayesian inference to model neural responses to optogenetic stimulation, and then optimises the laser powers and optical target locations needed to achieve a desired activity pattern with minimal OTS. We validate our approach in simulations and using data from in vitro experiments, showing that Bayesian target optimisation considerably reduces OTS across all conditions we test. Together, these results establish our ability to overcome OTS, enabling optogenetic stimulation with substantially improved precision.
****
From ViT Features to Training-free Video Object Segmentation via Streaming-data Mixture Models
Roy Uziel, Or Dinari, Oren Freifeld
In the task of semi-supervised video object segmentation, the input is the binary mask of an object in the first frame, and the desired output consists of the corresponding masks of that object in the subsequent frames. Existing leading solutions have two main drawbacks: 1) an expensive and typically-supervised training on videos; 2) a large memory footprint during inference. Here we present a training-free solution, with a low-memory footprint, that yields state-of-the-art results. The proposed method combines pre-trained deep learning-based features (trained on still images) with more classical methods for streaming-data clustering. Designed to adapt to temporal concept drifts and generalize to diverse video content without relying on annotated images or videos, the method eliminates the need for additional training or fine-tuning, ensuring fast inference and immediate applicability to new videos. Concretely, we represent an object via a dynamic ensemble of temporally- and spatially-coherent mixtures over a representation built from pre-trained ViT features and positional embeddings. A convolutional conditional random field further improves spatial coherence and helps reject outliers. We demonstrate the efficacy of the method on key benchmarks: the DAVIS-2017 and YouTube-VOS 2018 validation datasets. Moreover, by the virtue of the low-memory footprint of the compact cluster-based representation, the method scales gracefully to high-resolution ViT features. Our code is available at https://github.com/BGU-CS-VIL/Training-Free-VOS
****
How hard are computer vision datasets? Calibrating dataset difficulty to viewing time
David Mayo, Jesse Cummings, Xinyu Lin, Dan Gutfreund, Boris Katz, Andrei Barbu
Humans outperform object recognizers despite the fact that models perform well on current datasets, including those explicitly designed to challenge machines with debiased images or distribution shift. This problem persists, in part, because we have no guidance on the absolute difficulty of an image or dataset making it hard to objectively assess progress toward human-level performance, to cover the range of human abilities, and to increase the challenge posed by a dataset. We develop a dataset difficulty metric MVT, Minimum Viewing Time, that addresses these three problems. Subjects view an image that flashes on screen and then classify the object in the image. Images that require brief flashes to recognize are easy, those which require seconds of viewing are hard. We compute the ImageNet and ObjectNet image difficulty distribution, which we find significantly undersamples hard images. Nearly 90% of current benchmark performance is derived from images that are easy for humans. Rather than hoping that we will make harder datasets, we can for the first time objectively guide dataset difficulty during development. We can also subset recognition performance as a function of difficulty: model performance drops precipitously while human performance remains stable. Difficulty provides a new lens through which to view model performance, one which uncovers new scaling laws: vision-language models stand out as being the most robust and human-like while all other techniques scale poorly. We release tools to automatically compute MVT, along with image sets which are tagged by difficulty. Objective image difficulty has practical applications – one can measure how hard a test set is before deploying a real-world system – and scientific applications such as discovering the neural correlates of image difficulty and enabling new object recognition techniques that eliminate the benchmark-vs- real-world p

erformance gap.
****
What Knowledge Gets Distilled in Knowledge Distillation?
Utkarsh Ojha, Yuheng Li, Anirudh Sundara Rajan, Yingyu Liang, Yong Jae Lee
Knowledge distillation aims to transfer useful information from a teacher network to a student network, with the primary goal of improving the student's performance for the task at hand. Over the years, there has a been a deluge of novel techniques and use cases of knowledge distillation. Yet, despite the various improvements, there seems to be a glaring gap in the community's fundamental understanding of the process. Specifically, what is the knowledge that gets distilled in knowledge distillation? In other words, in what ways does the student become similar to the teacher? Does it start to localize objects in the same way? Does it get fooled by the same adversarial samples? Does its data invariance properties become similar? Our work presents a comprehensive study to try to answer these questions. We show that existing methods can indeed indirectly distill these properties beyond improving task performance. We further study why knowledge distillation might work this way, and show that our findings have practical implications as well.
****
Kernelized Cumulants: Beyond Kernel Mean Embeddings
Patric Bonnier, Harald Oberhauser, Zoltan Szabo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Contrastive Training of Complex-Valued Autoencoders for Object Discovery
Aleksandar Stani■, Anand Gopalakrishnan, Kazuki Irie, Jürgen Schmidhuber
Current state-of-the-art object-centric models use slots and attention-based routing for binding. However, this class of models has several conceptual limitations: the number of slots is hardwired; all slots have equal capacity; training has high computational cost; there are no object-level relational factors within slots. Synchrony-based models in principle can address these limitations by using complex-valued activations which store binding information in their phase components. However, working examples of such synchrony-based models have been developed only very recently, and are still limited to toy grayscale datasets and simultaneous storage of less than three objects in practice. Here we introduce architectural modifications and a novel contrastive learning method that greatly improve the state-of-the-art synchrony-based model. For the first time, we obtain a class of synchrony-based models capable of discovering objects in an unsupervised manner in multi-object color datasets and simultaneously representing more than three objects.
****
Non-adversarial training of Neural SDEs with signature kernel scores
Zacharia Issa, Blanka Horvath, Maud Lemercier, Cristopher Salvi
Neural SDEs are continuous-time generative models for sequential data. State-of-the-art performance for irregular time series generation has been previously obtained by training these models adversarially as GANs. However, as typical for GAN architectures, training is notoriously unstable, often suffers from mode collapse, and requires specialised techniques such as weight clipping and gradient penalty to mitigate these issues. In this paper, we introduce a novel class of scoring rules on pathspace based on signature kernels and use them as objective for training Neural SDEs non-adversarially. By showing strict properness of such kernel scores and consistency of the corresponding estimators, we provide existence and uniqueness guarantees for the minimiser. With this formulation, evaluating the generator-discriminator pair amounts to solving a system of linear path-dependent PDEs which allows for memory-efficient adjoint-based backpropagation. Moreover, because the proposed kernel scores are well-defined for paths with values in infinite dimensional spaces of functions, our framework can be easily extended to generate spatiotemporal data. Our procedure significantly outperforms alte

rnative ways of training Neural SDEs on a variety of tasks including the simulation of rough volatility models, the conditional probabilistic forecasts of real-world forex pairs where the conditioning variable is an observed past trajectory, and the mesh-free generation of limit order book dynamics.
****

Uni-ControlNet: All-in-One Control to Text-to-Image Diffusion Models
Shihao Zhao, Dongdong Chen, Yen-Chun Chen, Jianmin Bao, Shaozhe Hao, Lu Yuan, Kwan-Yee K. Wong
Text-to-Image diffusion models have made tremendous progress over the past two years, enabling the generation of highly realistic images based on open-domain text descriptions. However, despite their success, text descriptions often struggle to adequately convey detailed controls, even when composed of long and complex texts. Moreover, recent studies have also shown that these models face challenges in understanding such complex texts and generating the corresponding images. Therefore, there is a growing need to enable more control modes beyond text description. In this paper, we introduce Uni-ControlNet, a unified framework that allows for the simultaneous utilization of different local controls (e.g., edge maps, depth map, segmentation masks) and global controls (e.g., CLIP image embeddings) in a flexible and composable manner within one single model. Unlike existing methods, Uni-ControlNet only requires the fine-tuning of two additional adapters upon frozen pre-trained text-to-image diffusion models, eliminating the huge cost of training from scratch. Moreover, thanks to some dedicated adapter designs, Uni-ControlNet only necessitates a constant number (i.e., 2) of adapters, regardless of the number of local or global controls used. This not only reduces the fine-tuning costs and model size, making it more suitable for real-world deployment, but also facilitate composability of different conditions. Through both quantitative and qualitative comparisons, Uni-ControlNet demonstrates its superiority over existing methods in terms of controllability, generation quality and composability. Code is available at https://github.com/ShihaoZhaoZSH/Uni-ControlNet.
****

Loss Decoupling for Task-Agnostic Continual Learning
Yan-Shuo Liang, Wu-Jun Li
Continual learning requires the model to learn multiple tasks in a sequential order. To perform continual learning, the model must possess the abilities to maintain performance on old tasks (stability) and adapt itself to learn new tasks (plasticity). Task-agnostic problem in continual learning is a challenging problem, in which task identities are not available in the inference stage and hence the model must learn to distinguish all the classes in all the tasks. In task-agnostic problem, the model needs to learn two new objectives for learning a new task, including distinguishing new classes from old classes and distinguishing between different new classes. For task-agnostic problem, replay-based methods are commonly used. These methods update the model with both saved old samples and new samples for continual learning. Most existing replay-based methods mix the two objectives in task-agnostic problem together, inhibiting the models from achieving a good trade-off between stability and plasticity. In this paper, we propose a simple yet effective method, called loss decoupling (LODE), for task-agnostic continual learning. LODE separates the two objectives for the new task by decoupling the loss of the new task. As a result, LODE can assign different weights for different objectives, which provides a way to obtain a better trade-off between stability and plasticity than those methods with coupled loss. Experiments show that LODE can outperform existing state-of-the-art replay-based methods on multiple continual learning datasets.
****

Federated Learning via Meta-Variational Dropout
Insu Jeon, Minui Hong, Junhyeog Yun, Gunhee Kim
Federated Learning (FL) aims to train a global inference model from remotely distributed clients, gaining popularity due to its benefit of improving data privacy. However, traditional FL often faces challenges in practical applications, including model overfitting and divergent local models due to limited and non-IID d

ata among clients. To address these issues, we introduce a novel Bayesian meta-l earning approach called meta-variational dropout (MetaVD). MetaVD learns to pred ict client-dependent dropout rates via a shared hypernetwork, enabling effective model personalization of FL algorithms in limited non-IID data settings. We als o emphasize the posterior adaptation view of meta-learning and the posterior agg regation view of Bayesian FL via the conditional dropout posterior. We conducted extensive experiments on various sparse and non-IID FL datasets. MetaVD demonst rated excellent classification accuracy and uncertainty calibration performance, especially for out-of-distribution (OOD) clients. MetaVD compresses the local m odel parameters needed for each client, mitigating model overfitting and reducin g communication costs. Code is available at https://github.com/insujeon/MetaVD.
****
Horospherical Decision Boundaries for Large Margin Classification in Hyperbolic Space
Xiran Fan, Chun-Hao Yang, Baba Vemuri
Hyperbolic spaces have been quite popular in the recent past for representing hi erarchically organized data. Further, several classification algorithms for data in these spaces have been proposed in the literature. These algorithms mainly u se either hyperplanes or geodesics for decision boundaries in a large margin cla ssifiers setting leading to a non-convex optimization problem. In this paper, we propose a novel large margin classifier based on horospherical decision boundar ies that leads to a geodesically convex optimization problem that can be optimiz ed using any Riemannian gradient descent technique guaranteeing a globally optim al solution. We present several experiments depicting the competitive performanc e of our classifier in comparison to SOTA.
****
MIM4DD: Mutual Information Maximization for Dataset Distillation
Yuzhang Shang, Zhihang Yuan, Yan Yan
Dataset distillation (DD) aims to synthesize a small dataset whose test performa nce is comparable to a full dataset using the same model. State-of-the-art (SoTA) methods optimize synthetic datasets primarily by matching heuristic indicators extracted from two networks: one from real data and one from synthetic data (se e Fig.1, Left), such as gradients and training trajectories. DD is essentially a compression problem that emphasizes on maximizing the preservation of informati on contained in the data. We argue that well-defined metrics which measure the a mount of shared information between variables in information theory are necessar y for success measurement, but are never considered by previous works. Thus, we introduce mutual information (MI) as the metric to quantify the shared informati on between the synthetic and the real datasets, and devise MIM4DD numerically ma ximizing the MI via a newly designed optimizable objective within a contrastive learning framework to update the synthetic dataset. Specifically, we designate t he samples in different datasets who share the same labels as positive pairs, an d vice versa negative pairs. Then we respectively pull and push those samples in positive and negative pairs into contrastive space via minimizing NCE loss. As a result, the targeted MI can be transformed into a lower bound represented by f eature maps of samples, which is numerically feasible. Experiment results show t hat MIM4DD can be implemented as an add-on module to existing SoTA DD methods.
****
Hypernetwork-based Meta-Learning for Low-Rank Physics-Informed Neural Networks
Woojin Cho, Kookjin Lee, Donsub Rim, Noseong Park
In various engineering and applied science applications, repetitive numerical si mulations of partial differential equations (PDEs) for varying input parameters are often required (e.g., aircraft shape optimization over many design parameter s) and solvers are required to perform rapid execution. In this study, we sugges t a path that potentially opens up a possibility for physics-informed neural net works (PINNs), emerging deep-learning-based solvers, to be considered as one suc h solver. Although PINNs have pioneered a proper integration of deep-learning an d scientific computing, they require repetitive time-consuming training of neura l networks, which is not suitable for many-query scenarios. To address this issu e, we propose a lightweight low-rank PINNs containing only hundreds of model par

ameters and an associated hypernetwork-based meta-learning algorithm, which allows efficient approximation of solutions of PDEs for varying ranges of PDE input parameters. Moreover, we show that the proposed method is effective in overcoming a challenging issue, known as "failure modes" of PINNs.
****

Solving a Class of Non-Convex Minimax Optimization in Federated Learning
Xidong Wu, Jianhui Sun, Zhengmian Hu, Aidong Zhang, Heng Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

End-to-End Meta-Bayesian Optimisation with Transformer Neural Processes
Alexandre Maraval, Matthieu Zimmer, Antoine Grosnit, Haitham Bou Ammar
Meta-Bayesian optimisation (meta-BO) aims to improve the sample efficiency of Bayesian optimisation by leveraging data from related tasks. While previous methods successfully meta-learn either a surrogate model or an acquisition function independently, joint training of both components remains an open challenge. This paper proposes the first end-to-end differentiable meta-BO framework that generalises neural processes to learn acquisition functions via transformer architectures. We enable this end-to-end framework with reinforcement learning (RL) to tackle the lack of labelled acquisition data. Early on, we notice that training transformer-based neural processes from scratch with RL is challenging due to insufficient supervision, especially when rewards are sparse. We formalise this claim with a combinatorial analysis showing that the widely used notion of regret as a reward signal exhibits a logarithmic sparsity pattern in trajectory lengths. To tackle this problem, we augment the RL objective with an auxiliary task that guides part of the architecture to learn a valid probabilistic model as an inductive bias. We demonstrate that our method achieves state-of-the-art regret results against various baselines in experiments on standard hyperparameter optimisation tasks and also outperforms others in the real-world problems of mixed-integer programming tuning, antibody design, and logic synthesis for electronic design automation.
****

Contextual Bandits and Imitation Learning with Preference-Based Active Queries
Ayush Sekhari, Karthik Sridharan, Wen Sun, Runzhe Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Laplacian Canonization: A Minimalist Approach to Sign and Basis Invariant Spectral Embedding
George Ma, Yifei Wang, Yisen Wang
Spectral embedding is a powerful graph embedding technique that has received a lot of attention recently due to its effectiveness on Graph Transformers. However, from a theoretical perspective, the universal expressive power of spectral embedding comes at the price of losing two important invariance properties of graphs, sign and basis invariance, which also limits its effectiveness on graph data. To remedy this issue, many previous methods developed costly approaches to learn new invariants and suffer from high computation complexity. In this work, we explore a minimal approach that resolves the ambiguity issues by directly finding canonical directions for the eigenvectors, named Laplacian Canonization (LC). As a pure pre-processing method, LC is light-weighted and can be applied to any existing GNNs. We provide a thorough investigation, from theory to algorithm, on this approach, and discover an efficient algorithm named Maximal Axis Projection (MAP) that works for both sign and basis invariance and successfully canonizes more than 90\% of all eigenvectors. Experiments on real-world benchmark datasets like ZINC, MOLTOX21, and MOLPCBA show that MAP consistently outperforms existing methods while bringing minimal computation overhead. Code is available at http

s://github.com/PKU-ML/LaplacianCanonization.
****
Localized Symbolic Knowledge Distillation for Visual Commonsense Models
Jae Sung Park, Jack Hessel, Khyathi Chandu, Paul Pu Liang, Ximing Lu, Peter West, Youngjae Yu, Qiuyuan Huang, Jianfeng Gao, Ali Farhadi, Yejin Choi
Instruction following vision-language (VL) models offer a flexibleinterface that supports a broad range of multimodal tasks in a zero-shot fashion.However, interfaces that operate on full images do not directly enable the user to“point to" and access specific regions within images. This capability is importantnot only to support reference-grounded VL benchmarks, but also, for practicalapplications that require precise within-image reasoning. We build LocalizedVisual Commonsense model which allows users to specify (multiple) regions-as-input. We train our model by sampling localized commonsense knowledgefrom a large language model (LLM): specifically, we prompt a LLM to collectcommonsense knowledge given a global literal image description and a localliteral region description automatically generated by a set of VL models. Thispipeline is scalable and fully automatic, as no aligned or human-authored imageand text pairs are required. With a separately trained critic model that selectshigh quality examples, we find that training on the localized commonsense corpusexpanded solely from images can successfully distill existing VL models to supporta reference-as-input interface. Empirical results and human evaluations in zero-shotsettings demonstrate that our distillation method results in more precise VL modelsof reasoning compared to a baseline of passing a generated referring expression.
****
SmooSeg: Smoothness Prior for Unsupervised Semantic Segmentation
Mengcheng Lan, Xinjiang Wang, Yiping Ke, Jiaxing Xu, Litong Feng, Wayne Zhang
Unsupervised semantic segmentation is a challenging task that segments images into semantic groups without manual annotation. Prior works have primarily focused on leveraging prior knowledge of semantic consistency or priori concepts from self-supervised learning methods, which often overlook the coherence property of image segments. In this paper, we demonstrate that the smoothness prior, asserting that close features in a metric space share the same semantics, can significantly simplify segmentation by casting unsupervised semantic segmentation as an energy minimization problem. Under this paradigm, we propose a novel approach called SmooSeg that harnesses self-supervised learning methods to model the closeness relationships among observations as smoothness signals. To effectively discover coherent semantic segments, we introduce a novel smoothness loss that promotes piecewise smoothness within segments while preserving discontinuities across different segments. Additionally, to further enhance segmentation quality, we design an asymmetric teacher-student style predictor that generates smoothly updated pseudo labels, facilitating an optimal fit between observations and labeling outputs. Thanks to the rich supervision cues of the smoothness prior, our SmooSeg significantly outperforms STEGO in terms of pixel accuracy on three datasets: COCOStuff (+14.9\%), Cityscapes (+13.0\%), and Potsdam-3 (+5.7\%).
****
Fast Trainable Projection for Robust Fine-tuning
Junjiao Tian, Yen-Cheng Liu, James S Smith, Zsolt Kira
Robust fine-tuning aims to achieve competitive in-distribution (ID) performance while maintaining the out-of-distribution (OOD) robustness of a pre-trained model when transferring it to a downstream task. Recently, projected gradient descent has been successfully used in robust fine-tuning by constraining the deviation from the initialization of the fine-tuned model explicitly through projection. However, algorithmically, two limitations prevent this method from being adopted more widely, scalability and efficiency. In this paper, we propose a new projection-based fine-tuning algorithm, Fast Trainable Projection (FTP) for computationally efficient learning of per-layer projection constraints, resulting in an average 35% speedup on our benchmarks compared to prior works. FTP can be combined with existing optimizers such as AdamW, and be used in a plug-and-play fashion. Finally, we show that FTP is a special instance of hyper-optimizers that tune the hyper-parameters of optimizers in a learnable manner through nested differen

tiation. Empirically, we show superior robustness on OOD datasets, including dom ain shifts and natural corruptions, across four different vision tasks with five different pre-trained models. Additionally, we demonstrate that FTP is broadly applicable and beneficial to other learning scenarios such as low-label and cont inual learning settings thanks to its easy adaptability. The code will be availa ble at https://github.com/GT-RIPL/FTP.git.
****
Counterfactual-Augmented Importance Sampling for Semi-Offline Policy Evaluation
Shengpu Tang, Jenna Wiens
In applying reinforcement learning (RL) to high-stakes domains, quantitative and qualitative evaluation using observational data can help practitioners understa nd the generalization performance of new policies. However, this type of off-pol icy evaluation (OPE) is inherently limited since offline data may not reflect th e distribution shifts resulting from the application of new policies. On the oth er hand, online evaluation by collecting rollouts according to the new policy is often infeasible, as deploying new policies in these domains can be unsafe. In this work, we propose a semi-offline evaluation framework as an intermediate ste p between offline and online evaluation, where human users provide annotations o f unobserved counterfactual trajectories. While tempting to simply augment exist ing data with such annotations, we show that this naive approach can lead to bia sed results. Instead, we design a new family of OPE estimators based on importan ce sampling (IS) and a novel weighting scheme that incorporate counterfactual an notations without introducing additional bias. We analyze the theoretical proper ties of our approach, showing its potential to reduce both bias and variance com pared to standard IS estimators. Our analyses reveal important practical conside rations for handling biased, noisy, or missing annotations. In a series of proof -of-concept experiments involving bandits and a healthcare-inspired simulator, w e demonstrate that our approach outperforms purely offline IS estimators and is robust to imperfect annotations. Our framework, combined with principled human-c entered design of annotation solicitation, can enable the application of RL in h igh-stakes domains.
****
Are GATs Out of Balance?
Nimrah Mustafa, Aleksandar Bojchevski, Rebekka Burkholz
While the expressive power and computational capabilities of graph neural networ ks (GNNs) have been theoretically studied, their optimization and learning dynam ics, in general, remain largely unexplored. Our study undertakes the Graph Atten tion Network (GAT), a popular GNN architecture in which a node's neighborhood ag gregation is weighted by parameterized attention coefficients. We derive a conse rvation law of GAT gradient flow dynamics, which explains why a high portion of parameters in GATs with standard initialization struggle to change during traini ng. This effect is amplified in deeper GATs, which perform significantly worse t han their shallow counterparts. To alleviate this problem, we devise an initiali zation scheme that balances the GAT network. Our approach i) allows more effecti ve propagation of gradients and in turn enables trainability of deeper networks, and ii) attains a considerable speedup in training and convergence time in comp arison to the standard initialization. Our main theorem serves as a stepping sto ne to studying the learning dynamics of positive homogeneous models with attenti on mechanisms.
****
SMPLer-X: Scaling Up Expressive Human Pose and Shape Estimation
Zhongang Cai, Wanqi Yin, Ailing Zeng, CHEN WEI, Qingping SUN, Wang Yanjun, Hui E n Pang, Haiyi Mei, Mingyuan Zhang, Lei Zhang, Chen Change Loy, Lei Yang, Ziwei L iu
Expressive human pose and shape estimation (EHPS) unifies body, hands, and face motion capture with numerous applications. Despite encouraging progress, current state-of-the-art methods still depend largely on a confined set of training dat asets. In this work, we investigate scaling up EHPS towards the first generalist foundation model (dubbed SMPLer-X), with up to ViT-Huge as the backbone and tra ining with up to 4.5M instances from diverse data sources. With big data and the

large model, SMPLer-X exhibits strong performance across diverse test benchmarks and excellent transferability to even unseen environments. 1) For the data scaling, we perform a systematic investigation on 32 EHPS datasets, including a wide range of scenarios that a model trained on any single dataset cannot handle. More importantly, capitalizing on insights obtained from the extensive benchmarking process, we optimize our training scheme and select datasets that lead to a significant leap in EHPS capabilities. 2) For the model scaling, we take advantage of vision transformers to study the scaling law of model sizes in EHPS. Moreover, our finetuning strategy turn SMPLer-X into specialist models, allowing them to achieve further performance boosts. Notably, our foundation model SMPLer-X consistently delivers state-of-the-art results on seven benchmarks such as AGORA (107.2 mm NMVE), UBody (57.4 mm PVE), EgoBody (63.6 mm PVE), and EHF (62.3 mm PVE without finetuning).
****

Fast Asymptotically Optimal Algorithms for Non-Parametric Stochastic Bandits
Dorian Baudry, Fabien Pesquerel, Rémy Degenne, Odalric-Ambrym Maillard
We consider the problem of regret minimization in non-parametric stochastic bandits. When the rewards are known to be bounded from above, there exists asymptotically optimal algorithms, with asymptotic regret depending on an infimum of Kullback-Leibler divergences (KL). These algorithms are computationally expensive and require storing all past rewards, thus simpler but non-optimal algorithms are often used instead. We introduce several methods to approximate the infimum KL which reduce drastically the computational and memory costs of existing optimal algorithms, while keeping their regret guaranties. We apply our findings to design new variants of the MED and IMED algorithms, and demonstrate their interest with extensive numerical simulations.
****

SHAP-IQ: Unified Approximation of any-order Shapley Interactions
Fabian Fumagalli, Maximilian Muschalik, Patrick Kolpaczki, Eyke Hüllermeier, Barbara Hammer
Predominately in explainable artificial intelligence (XAI) research, the Shapley value (SV) is applied to determine feature attributions for any black box model. Shapley interaction indices extend the SV to define any-order feature interactions. Defining a unique Shapley interaction index is an open research question and, so far, three definitions have been proposed, which differ by their choice of axioms. Moreover, each definition requires a specific approximation technique. Here, we propose SHAPley Interaction Quantification (SHAP-IQ), an efficient sampling-based approximator to compute Shapley interactions for arbitrary cardinal interaction indices (CII), i.e. interaction indices that satisfy the linearity, symmetry and dummy axiom. SHAP-IQ is based on a novel representation and, in contrast to existing methods, we provide theoretical guarantees for its approximation quality, as well as estimates for the variance of the point estimates. For the special case of SV, our approach reveals a novel representation of the SV and corresponds to Unbiased KernelSHAP with a greatly simplified calculation. We illustrate the computational efficiency and effectiveness by explaining language, image classification and high-dimensional synthetic models.
****

Towards Last-layer Retraining for Group Robustness with Fewer Annotations
Tyler LaBonte, Vidya Muthukumar, Abhishek Kumar
Empirical risk minimization (ERM) of neural networks is prone to over-reliance on spurious correlations and poor generalization on minority groups. The recent deep feature reweighting (DFR) technique achieves state-of-the-art group robustness via simple last-layer retraining, but it requires held-out group and class annotations to construct a group-balanced reweighting dataset. In this work, we examine this impractical requirement and find that last-layer retraining can be surprisingly effective with no group annotations (other than for model selection) and only a handful of class annotations. We first show that last-layer retraining can greatly improve worst-group accuracy even when the reweighting dataset has only a small proportion of worst-group data. This implies a "free lunch" where holding out a subset of training data to retrain the last layer can substantiall

y outperform ERM on the entire dataset with no additional data, annotations, or computation for training. To further improve group robustness, we introduce a lightweight method called selective last-layer finetuning (SELF), which constructs the reweighting dataset using misclassifications or disagreements. Our experiments present the first evidence that model disagreement upsamples worst-group data, enabling SELF to nearly match DFR on four well-established benchmarks across vision and language tasks with no group annotations and less than 3% of the held-out class annotations.
****
Analysis of Variance of Multiple Causal Networks
Zhongli Jiang, Dabao Zhang
Constructing a directed cyclic graph (DCG) is challenged by both algorithmic difficulty and computational burden. Comparing multiple DCGs is even more difficult, compounded by the need to identify dynamic causalities across graphs. We propose to unify multiple DCGs with a single structural model and develop a limited-information-based method to simultaneously construct multiple networks and infer their disparities, which can be visualized by appropriate correspondence analysis. The algorithm provides DCGs with robust non-asymptotic theoretical properties. It is designed with two sequential stages, each of which involves parallel computation tasks that are scalable to the network complexity. Taking advantage of high-performance clusters, our method makes it possible to evaluate the statistical significance of DCGs using the bootstrap method. We demonstrated the effectiveness of our method by applying it to synthetic and real datasets.
****
Revisiting the Minimalist Approach to Offline Reinforcement Learning
Denis Tarasov, Vladislav Kurenkov, Alexander Nikulin, Sergey Kolesnikov
Recent years have witnessed significant advancements in offline reinforcement learning (RL), resulting in the development of numerous algorithms with varying degrees of complexity. While these algorithms have led to noteworthy improvements, many incorporate seemingly minor design choices that impact their effectiveness beyond core algorithmic advances. However, the effect of these design choices on established baselines remains understudied. In this work, we aim to bridge this gap by conducting a retrospective analysis of recent works in offline RL and propose ReBRAC, a minimalistic algorithm that integrates such design elements built on top of the TD3+BC method. We evaluate ReBRAC on 51 datasets with both proprioceptive and visual state spaces using D4RL and V-D4RL benchmarks, demonstrating its state-of-the-art performance among ensemble-free methods in both offline and offline-to-online settings. To further illustrate the efficacy of these design choices, we perform a large-scale ablation study and hyperparameter sensitivity analysis on the scale of thousands of experiments.
****
Complementary Benefits of Contrastive Learning and Self-Training Under Distribution Shift
Saurabh Garg, Amrith Setlur, Zachary Lipton, Sivaraman Balakrishnan, Virginia Smith, Aditi Raghunathan
Self-training and contrastive learning have emerged as leading techniques for incorporating unlabeled data, both under distribution shift (unsupervised domain adaptation) and when it is absent (semi-supervised learning). However, despite the popularity and compatibility of these techniques, their efficacy in combination remains surprisingly unexplored. In this paper, we first undertake a systematic empirical investigation of this combination, finding (i) that in domain adaptation settings, self-training and contrastive learning offer significant complementary gains; and (ii) that in semi-supervised learning settings, surprisingly, the benefits are not synergistic. Across eight distribution shift datasets (e.g., BREEDs, WILDS), we demonstrate that the combined method obtains 3--8\% higher accuracy than either approach independently. Finally, we theoretically analyze these techniques in a simplified model of distribution shift demonstrating scenarios under which the features produced by contrastive learning can yield a good initialization for self-training to further amplify gains and achieve optimal performance, even when either method alone would fail.

****
Low Tensor Rank Learning of Neural Dynamics
Arthur Pellegrino, N Alex Cayco Gajic, Angus Chadwick
Learning relies on coordinated synaptic changes in recurrently connected populat
ions of neurons. Therefore, understanding the collective evolution of synaptic c
onnectivity over learning is a key challenge in neuroscience and machine learnin
g. In particular, recent work has shown that the weight matrices of task-trained
 RNNs are typically low rank, but how this low rank structure unfolds over learn
ing is unknown. To address this, we investigate the rank of the 3-tensor formed
by the weight matrices throughout learning. By fitting RNNs of varying rank to l
arge-scale neural recordings during a motor learning task, we find that the infe
rred weights are low-tensor-rank and therefore evolve over a fixed low-dimension
al subspace throughout the entire course of learning. We next validate the obser
vation of low-tensor-rank learning on an RNN trained to solve the same task. Fin
ally, we present a set of mathematical results bounding the matrix and tensor ra
nks of gradient descent learning dynamics which show that low-tensor-rank weight
s emerge naturally in RNNs trained to solve low-dimensional tasks. Taken togethe
r, our findings provide insight on the evolution of population connectivity over
 learning in both biological and artificial neural networks, and enable reverse
engineering of learning-induced changes in recurrent dynamics from large-scale n
eural recordings.
****
MonoUNI: A Unified Vehicle and Infrastructure-side Monocular 3D Object Detection
 Network with Sufficient Depth Clues
Jia Jinrang, Zhenjia Li, Yifeng Shi
Monocular 3D detection of vehicle and infrastructure sides are two important top
ics in autonomous driving. Due to diverse sensor installations and focal lengths
, researchers are faced with the challenge of constructing algorithms for the tw
o topics based on different prior knowledge. In this paper, by taking into accou
nt the diversity of pitch angles and focal lengths, we propose a unified optimiz
ation target named normalized depth, which realizes the unification of 3D detect
ion problems for the two sides. Furthermore, to enhance the accuracy of monocula
r 3D detection, 3D normalized cube depth of obstacle is developed to promote the
 learning of depth information. We posit that the richness of depth clues is a p
ivotal factor impacting the detection performance on both the vehicle and infras
tructure sides. A richer set of depth clues facilitates the model to learn bette
r spatial knowledge, and the 3D normalized cube depth offers sufficient depth cl
ues. Extensive experiments demonstrate the effectiveness of our approach. Withou
t introducing any extra information, our method, named MonoUNI, achieves state-o
f-the-art performance on five widely used monocular 3D detection benchmarks, inc
luding Rope3D and DAIR-V2X-I for the infrastructure side, KITTI and Waymo for th
e vehicle side, and nuScenes for the cross-dataset evaluation.
****
Active Reasoning in an Open-World Environment
Manjie Xu, Guangyuan Jiang, Wei Liang, Chi Zhang, Yixin Zhu
Recent advances in vision-language learning have achieved notable success on com
plete-information question-answering datasets through the integration of extensi
ve world knowledge. Yet, most models operate passively, responding to questions
based on pre-stored knowledge. In stark contrast, humans possess the ability to
actively explore, accumulate, and reason using both newfound and existing inform
ation to tackle incomplete-information questions. In response to this gap, we in
troduce Conan, an interactive open-world environment devised for the assessment
of active reasoning. Conan facilitates active exploration and promotes multi-rou
nd abductive inference, reminiscent of rich, open-world settings like Minecraft.
 Diverging from previous works that lean primarily on single-round deduction via
 instruction following, Conan compels agents to actively interact with their sur
roundings, amalgamating new evidence with prior knowledge to elucidate events fr
om incomplete observations. Our analysis on \bench underscores the shortcomings
of contemporary state-of-the-art models in active exploration and understanding
complex scenarios. Additionally, we explore Abduction from Deduction, where agen

ts harness Bayesian rules to recast the challenge of abduction as a deductive process. Through Conan, we aim to galvanize advancements in active reasoning and set the stage for the next generation of artificial intelligence agents adept at dynamically engaging in environments.
****
2Direction: Theoretically Faster Distributed Training with Bidirectional Communication Compression
Alexander Tyurin, Peter Richtarik
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Tree of Thoughts: Deliberate Problem Solving with Large Language Models
Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Tom Griffiths, Yuan Cao, Karthik Narasimhan
Language models are increasingly being deployed for general problem solving across a wide range of tasks, but are still confined to token-level, left-to-right decision-making processes during inference. This means they can fall short in tasks that require exploration, strategic lookahead, or where initial decisions play a pivotal role. To surmount these challenges, we introduce a new framework for language model inference, Tree of Thoughts (ToT), which generalizes over the popular Chain of Thought approach to prompting language models, and enables exploration over coherent units of text (thoughts) that serve as intermediate steps toward problem solving. ToT allows LMs to perform deliberate decision making by considering multiple different reasoning paths and self-evaluating choices to decide the next course of action, as well as looking ahead or backtracking when necessary to make global choices.Our experiments show that ToT significantly enhances language models’ problem-solving abilities on three novel tasks requiring non-trivial planning or search: Game of 24, Creative Writing, and Mini Crosswords. For instance, in Game of 24, while GPT-4 with chain-of-thought prompting only solved 4\% of tasks, our method achieved a success rate of 74\%. Code repo with all prompts: https://github.com/princeton-nlp/tree-of-thought-llm.
****
What functions can Graph Neural Networks compute on random graphs? The role of Positional Encoding
Nicolas Keriven, Samuel Vaiter
We aim to deepen the theoretical understanding of Graph Neural Networks (GNNs) on large graphs, with a focus on their expressive power.Existing analyses relate this notion to the graph isomorphism problem, which is mostly relevant for graphs of small sizes, or studied graph classification or regression tasks, while prediction tasks on \emph{nodes} are far more relevant on large graphs. Recently, several works showed that, on very general random graphs models, GNNs converge to certains functions as the number of nodes grows.In this paper, we provide a more complete and intuitive description of the function space generated by equivariant GNNs for node-tasks, through general notions of convergence that encompass several previous examples. We emphasize the role of input node features, and study the impact of \emph{node Positional Encodings} (PEs), a recent line of work that has been shown to yield state-of-the-art results in practice. Through the study of several examples of PEs on large random graphs, we extend previously known universality results to significantly more general models. Our theoretical results hint at some normalization tricks, which is shown numerically to have a positive impact on GNN generalization on synthetic and real data. Our proofs contain new concentration inequalities of independent interest.
****
High-dimensional Asymptotics of Denoising Autoencoders
Hugo Cui, Lenka Zdeborová
We address the problem of denoising data from a Gaussian mixture using a two-layer non-linear autoencoder with tied weights and a skip connection. We consider the high-dimensional limit where the number of training samples and the input dim

ension jointly tend to infinity while the number of hidden units remains bounded . We provide closed-form expressions for the denoising mean-squared test error. Building on this result, we quantitatively characterize the advantage of the considered architecture over the autoencoder without the skip connection that relates closely to principal component analysis. We further show that our results capture accurately the learning curves on a range of real datasets.
****
Learning to Reason and Memorize with Self-Notes
Jack Lanchantin, Shubham Toshniwal, Jason Weston, arthur szlam, Sainbayar Sukhbaatar
Large language models have been shown to struggle with multi-step reasoning, and do not retain previous reasoning steps for future use. We propose a simple method for solving both of these problems by allowing the model to take Self-Notes. Unlike recent chain-of-thought or scratchpad approaches, the model can deviate from the input context at any time to explicitly think and write down its thoughts. This allows the model to perform reasoning on the fly as it reads the context and even integrate previous reasoning steps, thus enhancing its memory with useful information and enabling multi-step reasoning. Experiments across a wide variety of tasks demonstrate that our method can outperform chain-of-thought and scratchpad methods by taking Self-Notes that interleave the input text.
****
What can a Single Attention Layer Learn? A Study Through the Random Features Lens
Hengyu Fu, Tianyu Guo, Yu Bai, Song Mei
Attention layers---which map a sequence of inputs to a sequence of outputs---are core building blocks of the Transformer architecture which has achieved significant breakthroughs in modern artificial intelligence. This paper presents a rigorous theoretical study on the learning and generalization of a single multi-head attention layer, with a sequence of key vectors and a separate query vector as input. We consider the random feature setting where the attention layer has a large number of heads, with randomly sampled frozen query and key matrices, and trainable value matrices. We show that such a random-feature attention layer can express a broad class of target functions that are permutation invariant to the key vectors. We further provide quantitative excess risk bounds for learning these target functions from finite samples, using random feature attention with finitely many heads.Our results feature several implications unique to the attention structure compared with existing random features theory for neural networks, such as (1) Advantages in the sample complexity over standard two-layer random-feature networks; (2) Concrete and natural classes of functions that can be learned efficiently by a random-feature attention layer; and (3) The effect of the sampling distribution of the query-key weight matrix (the product of the query and key matrix), where Gaussian random weights with a non-zero mean result in better sample complexities over the zero-mean counterpart for learning certain natural target functions. Experiments on simulated data corroborate our theoretical findings and further illustrate the interplay between the sample size and the complexity of the target function.
****
Let the Flows Tell: Solving Graph Combinatorial Problems with GFlowNets
Dinghuai Zhang, Hanjun Dai, Nikolay Malkin, Aaron C. Courville, Yoshua Bengio, Ling Pan
Combinatorial optimization (CO) problems are often NP-hard and thus out of reach for exact algorithms, making them a tempting domain to apply machine learning methods. The highly structured constraints in these problems can hinder either optimization or sampling directly in the solution space.On the other hand, GFlowNets have recently emerged as a powerful machinery to efficiently sample from composite unnormalized densities sequentially and have the potential to amortize such solution-searching processes in CO, as well as generate diverse solution candidates.In this paper, we design Markov decision processes (MDPs) for different combinatorial problems and propose to train conditional GFlowNets to sample from the solution space. Efficient training techniques are also developed to benefit l

ong-range credit assignment.Through extensive experiments on a variety of different CO tasks with synthetic and realistic data, we demonstrate that GFlowNet policies can efficiently find high-quality solutions.Our implementation is open-sourced at https://github.com/zdhNarsil/GFlowNet-CombOpt.
****

Debiasing Scores and Prompts of 2D Diffusion for View-consistent Text-to-3D Generation
Susung Hong, Donghoon Ahn, Seungryong Kim
Existing score-distilling text-to-3D generation techniques, despite their considerable promise, often encounter the view inconsistency problem. One of the most notable issues is the Janus problem, where the most canonical view of an object (\textit{e.g}., face or head) appears in other views. In this work, we explore existing frameworks for score-distilling text-to-3D generation and identify the main causes of the view inconsistency problem---the embedded bias of 2D diffusion models. Based on these findings, we propose two approaches to debias the score-distillation frameworks for view-consistent text-to-3D generation. Our first approach, called score debiasing, involves cutting off the score estimated by 2D diffusion models and gradually increasing the truncation value throughout the optimization process. Our second approach, called prompt debiasing, identifies conflicting words between user prompts and view prompts using a language model, and adjusts the discrepancy between view prompts and the viewing direction of an object. Our experimental results show that our methods improve the realism of the generated 3D objects by significantly reducing artifacts and achieve a good trade-off between faithfulness to the 2D diffusion models and 3D consistency with little overhead. Our project page is available at~\url{https://susunghong.github.io/Debiased-Score-Distillation-Sampling/}.
****

On the Learnability of Multilabel Ranking
Vinod Raman, UNIQUE SUBEDI, Ambuj Tewari
Multilabel ranking is a central task in machine learning. However, the most fundamental question of learnability in a multilabel ranking setting with relevance-score feedback remains unanswered. In this work, we characterize the learnability of multilabel ranking problems in both batch and online settings for a large family of ranking losses. Along the way, we give two equivalence classes of ranking losses based on learnability that capture most losses used in practice.
****

MAG-GNN: Reinforcement Learning Boosted Graph Neural Network
Lecheng Kong, Jiarui Feng, Hao Liu, Dacheng Tao, Yixin Chen, Muhan Zhang
While Graph Neural Networks (GNNs) recently became powerful tools in graph learning tasks, considerable efforts have been spent on improving GNNs' structural encoding ability. A particular line of work proposed subgraph GNNs that use subgraph information to improve GNNs' expressivity and achieved great success. However, such effectivity sacrifices the efficiency of GNNs by enumerating all possible subgraphs. In this paper, we analyze the necessity of complete subgraph enumeration and show that a model can achieve a comparable level of expressivity by considering a small subset of the subgraphs. We then formulate the identification of the optimal subset as a combinatorial optimization problem and propose Magnetic Graph Neural Network (MAG-GNN), a reinforcement learning (RL) boosted GNN, to solve the problem. Starting with a candidate subgraph set, MAG-GNN employs an RL agent to iteratively update the subgraphs to locate the most expressive set for prediction. This reduces the exponential complexity of subgraph enumeration to the constant complexity of a subgraph search algorithm while keeping good expressivity. We conduct extensive experiments on many datasets, showing that MAG-GNN achieves competitive performance to state-of-the-art methods and even outperforms many subgraph GNNs. We also demonstrate that MAG-GNN effectively reduces the running time of subgraph GNNs.
****

RanPAC: Random Projections and Pre-trained Models for Continual Learning
Mark D. McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, Anton van den Hengel

Continual learning (CL) aims to incrementally learn different tasks (such as classification) in a non-stationary data stream without forgetting old ones. Most CL works focus on tackling catastrophic forgetting under a learning-from-scratch paradigm. However, with the increasing prominence of foundation models, pre-trained models equipped with informative representations have become available for various downstream requirements. Several CL methods based on pre-trained models have been explored, either utilizing pre-extracted features directly (which makes bridging distribution gaps challenging) or incorporating adaptors (which may be subject to forgetting). In this paper, we propose a concise and effective approach for CL with pre-trained models. Given that forgetting occurs during parameter updating, we contemplate an alternative approach that exploits training-free random projectors and class-prototype accumulation, which thus bypasses the issue. Specifically, we inject a frozen Random Projection layer with nonlinear activation between the pre-trained model's feature representations and output head, which captures interactions between features with expanded dimensionality, providing enhanced linear separability for class-prototype-based CL. We also demonstrate the importance of decorrelating the class-prototypes to reduce the distribution disparity when using pre-trained representations. These techniques prove to be effective and circumvent the problem of forgetting for both class- and domain-incremental continual learning. Compared to previous methods applied to pre-trained ViT-B/16 models, we reduce final error rates by between 20% and 62% on seven class-incremental benchmark datasets, despite not using any rehearsal memory. We conclude that the full potential of pre-trained models for simple, effective, and fast continual learning has not hitherto been fully tapped. Code is available at https://github.com/RanPAC/RanPAC.
****
FGPrompt: Fine-grained Goal Prompting for Image-goal Navigation
Xinyu Sun, Peihao Chen, Jugang Fan, Jian Chen, Thomas Li, Mingkui Tan
Learning to navigate to an image-specified goal is an important but challenging task for autonomous systems like household robots. The agent is required to well understand and reason the location of the navigation goal from a picture shot in the goal position. Existing methods try to solve this problem by learning a navigation policy, which captures semantic features of the goal image and observation image independently and lastly fuses them for predicting a sequence of navigation actions. However, these methods suffer from two major limitations. 1) They may miss detailed information in the goal image, and thus fail to reason the goal location. 2) More critically, it is hard to focus on the goal-relevant regions in the observation image, because they attempt to understand observation without goal conditioning. In this paper, we aim to overcome these limitations by designing a Fine-grained Goal Prompting (\sexyname) method for image-goal navigation. In particular, we leverage fine-grained and high-resolution feature maps in the goal image as prompts to perform conditioned embedding, which preserves detailed information in the goal image and guides the observation encoder to pay attention to goal-relevant regions. Compared with existing methods on the image-goal navigation benchmark, our method brings significant performance improvement on 3 benchmark datasets (\textit{i.e.,} Gibson, MP3D, and HM3D). Especially on Gibson, we surpass the state-of-the-art success rate by 8\% with only 1/50 model size.
****
Inner-Outer Aware Reconstruction Model for Monocular 3D Scene Reconstruction
Yu-Kun Qiu, Guo-Hao Xu, Wei-Shi Zheng
Monocular 3D scene reconstruction aims to reconstruct the 3D structure of scenes based on posed images. Recent volumetric-based methods directly predict the truncated signed distance function (TSDF) volume and have achieved promising results. The memory cost of volumetric-based methods will grow cubically as the volume size increases, so a coarse-to-fine strategy is necessary for saving memory. Specifically, the coarse-to-fine strategy distinguishes surface voxels from non-surface voxels, and only potential surface voxels are considered in the succeeding procedure. However, the non-surface voxels have various features, and in particular, the voxels on the inner side of the surface are quite different from those

on the outer side since there exists an intrinsic gap between them. Therefore, grouping inner-surface and outer-surface voxels into the same class will force the classifier to spend its capacity to bridge the gap. By contrast, it is relatively easy for the classifier to distinguish inner-surface and outer-surface voxels due to the intrinsic gap. Inspired by this, we propose the inner-outer aware reconstruction (IOAR) model. IOAR explores a new coarse-to-fine strategy to classify outer-surface, inner-surface and surface voxels. In addition, IOAR separates occupancy branches from TSDF branches to avoid mutual interference between them. Since our model can better classify the surface, outer-surface and inner-surface voxels, it can predict more precise meshes than existing methods. Experiment results on ScanNet, ICL-NUIM and TUM-RGBD datasets demonstrate the effectiveness and generalization of our model. The code is available at https://github.com/YorkQiu/InnerOuterAwareReconstruction.
****

Sheaf Hypergraph Networks

Iulia Duta, Giulia Cassarà, Fabrizio Silvestri, Pietro Lió

Higher-order relations are widespread in nature, with numerous phenomena involving complex interactions that extend beyond simple pairwise connections. As a result, advancements in higher-order processing can accelerate the growth of various fields requiring structured data. Current approaches typically represent these interactions using hypergraphs.We enhance this representation by introducing cellular sheaves for hypergraphs, a mathematical construction that adds extra structure to the conventional hypergraph while maintaining their local, higher-order connectivity. Drawing inspiration from existing Laplacians in the literature, we develop two unique formulations of sheaf hypergraph Laplacians: linear and non-linear. Our theoretical analysis demonstrates that incorporating sheaves into the hypergraph Laplacian provides a more expressive inductive bias than standard hypergraph diffusion, creating a powerful instrument for effectively modelling complex data structures.We employ these sheaf hypergraph Laplacians to design two categories of models: Sheaf Hypergraph Neural Networks and Sheaf Hypergraph Convolutional Networks. These models generalize classical Hypergraph Networks often found in the literature. Through extensive experimentation, we show that this generalization significantly improves performance, achieving top results on multiple benchmark datasets for hypergraph node classification.
****

f-Policy Gradients: A General Framework for Goal-Conditioned RL using f-Divergences

Siddhant Agarwal, Ishan Durugkar, Peter Stone, Amy Zhang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Counterfactually Fair Representation

Zhiqun Zuo, Mahdi Khalili, Xueru Zhang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Truncating Trajectories in Monte Carlo Policy Evaluation: an Adaptive Approach

Riccardo Poiani, Nicole Nobili, Alberto Maria Metelli, Marcello Restelli

Policy evaluation via Monte Carlo (MC) simulation is at the core of many MC Reinforcement Learning (RL) algorithms (e.g., policy gradient methods). In this context, the designer of the learning system specifies an interaction budget that the agent usually spends by collecting trajectories of fixed length within a simulator. However, is this data collection strategy the best option? To answer this question, in this paper, we consider as quality index the variance of an unbiased policy return estimator that uses trajectories of different lengths, i.e., truncated. We first derive a closed-form expression of this variance that clearly s

hows the sub-optimality of the fixed-length trajectory schedule. Furthermore, it suggests that adaptive data collection strategies that spend the available budget sequentially might be able to allocate a larger portion of transitions in timesteps in which more accurate sampling is required to reduce the variance of the final estimate. Building on these findings, we present an adaptive algorithm called Robust and Iterative Data collection strategy Optimization (RIDO). The main intuition behind RIDO is to split the available interaction budget into mini-batches. At each round, the agent determines the most convenient schedule of trajectories that minimizes an empirical and robust estimate of the estimator's variance. After discussing the theoretical properties of our method, we conclude by assessing its performance across multiple domains. Our results show that RIDO can adapt its trajectory schedule toward timesteps where more sampling is required to increase the quality of the final estimation.
****
DP-Mix: Mixup-based Data Augmentation for Differentially Private Learning
Wenxuan Bao, Francesco Pittaluga, Vijay Kumar B G, Vincent Bindschaedler
Data augmentation techniques, such as image transformations and combinations, are highly effective at improving the generalization of computer vision models, especially when training data is limited. However, such techniques are fundamentally incompatible with differentially private learning approaches, due to the latter’s built-in assumption that each training image’s contribution to the learned model is bounded. In this paper, we investigate why naive applications of multi-sample data augmentation techniques, such as mixup, fail to achieve good performance and propose two novel data augmentation techniques specifically designed for the constraints of differentially private learning. Our first technique, DP-MixSelf, achieves SoTA classification performance across a range of datasets and settings by performing mixup on self-augmented data. Our second technique, DP-MixDiff, further improves performance by incorporating synthetic data from a pre-trained diffusion model into the mixup process. We open-source the code at https://github.com/wenxuan-Bao/DP-Mix.
****
Point Cloud Completion with Pretrained Text-to-Image Diffusion Models
Yoni Kasten, Ohad Rahamim, Gal Chechik
Point cloud data collected in real-world applications are often incomplete. This is because they are observed from partial viewpoints, which capture only a specific perspective or angle, or due to occlusion and low resolution. Existing completion approaches rely on datasets of specific predefined objects to guide the completion of incomplete, and possibly noisy, point clouds. However, these approaches perform poorly with Out-Of-Distribution (OOD) objects, which are either absent from the dataset or poorly represented. In recent years, the field of text-guided image generation has made significant progress, leading to major breakthroughs in text guided shape generation. We describe an approach called SDS-Complete that uses a pre-trained text-to-image diffusion model and leverages the text semantic of a given incomplete point cloud of an object, to obtain a complete surface representation. SDS-Complete can complete a variety of objects at test time optimization without the need for an expensive collection of 3D information. We evaluate SDS-Complete on incomplete scanned objects, captured by real-world depth sensors and LiDAR scanners, and demonstrate that is effective in handling objects which are typically absent from common datasets.
****
Eliciting User Preferences for Personalized Multi-Objective Decision Making through Comparative Feedback
Han Shao, Lee Cohen, Avrim Blum, Yishay Mansour, Aadirupa Saha, Matthew Walter
In this work, we propose a multi-objective decision making framework that accommodates different user preferences over objectives, where preferences are learned via policy comparisons. Our model consists of a known Markov decision process with a vector-valued reward function, with each user having an unknown preference vector that expresses the relative importance of each objective. The goal is to efficiently compute a near-optimal policy for a given user. We consider two user feedback models. We first address the case where a user is provided with two p

olicies and returns their preferred policy as feedback. We then move to a different user feedback model, where a user is instead provided with two small weighted sets of representative trajectories and selects the preferred one. In both cases, we suggest an algorithm that finds a nearly optimal policy for the user using a number of comparison queries that scales quasilinearly in the number of objectives.
****

Deep Recurrent Optimal Stopping
Niranjan Damera Venkata, Chiranjib Bhattacharyya
Deep neural networks (DNNs) have recently emerged as a powerful paradigm for solving Markovian optimal stopping problems. However, a ready extension of DNN-based methods to non-Markovian settings requires significant state and parameter space expansion, manifesting the curse of dimensionality. Further, efficient state-space transformations permitting Markovian approximations, such as those afforded by recurrent neural networks (RNNs), are either structurally infeasible or are confounded by the curse of non-Markovianity. Considering these issues, we introduce, for the first time, an optimal stopping policy gradient algorithm (OSPG) that can leverage RNNs effectively in non-Markovian settings by implicitly optimizing value functions without recursion, mitigating the curse of non-Markovianity. The OSPG algorithm is derived from an inference procedure on a novel Bayesian network representation of discrete-time non-Markovian optimal stopping trajectories and, as a consequence, yields an offline policy gradient algorithm that eliminates expensive Monte Carlo policy rollouts.
****

A Partially-Supervised Reinforcement Learning Framework for Visual Active Search
Anindya Sarkar, Nathan Jacobs, Yevgeniy Vorobeychik
Visual active search (VAS) has been proposed as a modeling framework in which visual cues are used to guide exploration, with the goal of identifying regions of interest in a large geospatial area. Its potential applications include identifying hot spots of rare wildlife poaching activity, search-and-rescue scenarios, identifying illegal trafficking of weapons, drugs, or people, and many others. State of the art approaches to VAS include applications of deep reinforcement learning (DRL), which yield end-to-end search policies, and traditional active search, which combines predictions with custom algorithmic approaches. While the DRL framework has been shown to greatly outperform traditional active search in such domains, its end-to-end nature does not make full use of supervised information attained either during training, or during actual search, a significant limitation if search tasks differ significantly from those in the training distribution. We propose an approach that combines the strength of both DRL and conventional active search approaches by decomposing the search policy into a prediction module, which produces a geospatial distribution of regions of interest based on task embedding and search history, and a search module, which takes the predictions and search history as input and outputs the search distribution. In addition, we develop a novel meta-learning approach for jointly learning the resulting combined policy that can make effective use of supervised information obtained both at training and decision time. Our extensive experiments demonstrate that the proposed representation and meta-learning frameworks significantly outperform state of the art in visual active search on several problem domains.
****

Koopa: Learning Non-stationary Time Series Dynamics with Koopman Predictors
Yong Liu, Chenyu Li, Jianmin Wang, Mingsheng Long
Real-world time series are characterized by intrinsic non-stationarity that poses a principal challenge for deep forecasting models. While previous models suffer from complicated series variations induced by changing temporal distribution, we tackle non-stationary time series with modern Koopman theory that fundamentally considers the underlying time-variant dynamics. Inspired by Koopman theory of portraying complex dynamical systems, we disentangle time-variant and time-invariant components from intricate non-stationary series by Fourier Filter and design Koopman Predictor to advance respective dynamics forward. Technically, we propose Koopa as a novel Koopman forecaster composed of stackable blocks that learn

hierarchical dynamics. Koopa seeks measurement functions for Koopman embedding and utilizes Koopman operators as linear portraits of implicit transition. To cope with time-variant dynamics that exhibits strong locality, Koopa calculates context-aware operators in the temporal neighborhood and is able to utilize incoming ground truth to scale up forecast horizon. Besides, by integrating Koopman Predictors into deep residual structure, we ravel out the binding reconstruction loss in previous Koopman forecasters and achieve end-to-end forecasting objective optimization. Compared with the state-of-the-art model, Koopa achieves competitive performance while saving 77.3% training time and 76.0% memory.
****
Bridging Discrete and Backpropagation: Straight-Through and Beyond
Liyuan Liu, Chengyu Dong, Xiaodong Liu, Bin Yu, Jianfeng Gao
Backpropagation, the cornerstone of deep learning, is limited to computing gradients for continuous variables. This limitation poses challenges for problems involving discrete latent variables. To address this issue, we propose a novel approach to approximate the gradient of parameters involved in generating discrete latent variables. First, we examine the widely used Straight-Through (ST) heuristic and demonstrate that it works as a first-order approximation of the gradient. Guided by our findings, we propose ReinMax, which achieves second-order accuracy by integrating Heun’s method, a second-order numerical method for solving ODEs. ReinMax does not require Hessian or other second-order derivatives, thus having negligible computation overheads. Extensive experimental results on various tasks demonstrate the superiority of ReinMax over the state of the art.
****
Distributed Inference and Fine-tuning of Large Language Models Over The Internet
Alexander Borzunov, Max Ryabinin, Artem Chumachenko, Dmitry Baranchuk, Tim Dettmers, Younes Belkada, Pavel Samygin, Colin A. Raffel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Contrast, Attend and Diffuse to Decode High-Resolution Images from Brain Activities
Jingyuan Sun, Mingxiao Li, Zijiao Chen, Yunhao Zhang, Shaonan Wang, Marie-Francine Moens
Decoding visual stimuli from neural responses recorded by functional Magnetic Resonance Imaging (fMRI) presents an intriguing intersection between cognitive neuroscience and machine learning, promising advancements in understanding human visual perception. However, the task is challenging due to the noisy nature of fMRI signals and the intricate pattern of brain visual representations. To mitigate these challenges, we introduce a two-phase fMRI representation learning framework. The first phase pre-trains an fMRI feature learner with a proposed Double-contrastive Mask Auto-encoder to learn denoised representations. The second phase tunes the feature learner to attend to neural activation patterns most informative for visual reconstruction with guidance from an image auto-encoder. The optimized fMRI feature learner then conditions a latent diffusion model to reconstruct image stimuli from brain activities. Experimental results demonstrate our model's superiority in generating high-resolution and semantically accurate images, substantially exceeding previous state-of-the-art methods by 39.34% in the 50-way-top-1 semantic classification accuracy. The code implementations is available at https://github.com/soinx0629/visdecneurips/.
****
Diffusion with Forward Models: Solving Stochastic Inverse Problems Without Direct Supervision
Ayush Tewari, Tianwei Yin, George Cazenavette, Semon Rezchikov, Josh Tenenbaum, Fredo Durand, Bill Freeman, Vincent Sitzmann
Denoising diffusion models are a powerful type of generative models used to capture complex distributions of real-world signals. However, their applicability is limited to scenarios where training samples are readily available, which is not

always the case in real-world applications. For example, in inverse graphics, the goal is to generate samples from a distribution of 3D scenes that align with a given image, but ground-truth 3D scenes are unavailable and only 2D images are accessible. To address this limitation, we propose a novel class of denoising diffusion probabilistic models that learn to sample from distributions of signals that are never directly observed. Instead, these signals are measured indirectly through a known differentiable forward model, which produces partial observations of the unknown signal. Our approach involves integrating the forward model directly into the denoising process. A key contribution of our work is the integration of a differentiable forward model into the denoising process. This integration effectively connects the generative modeling of observations with the generative modeling of the underlying signals, allowing for end-to-end training of a conditional generative model over signals. During inference, our approach enables sampling from the distribution of underlying signals that are consistent with a given partial observation. We demonstrate the effectiveness of our method on three challenging computer vision tasks. For instance, in the context of inverse graphics, our model enables direct sampling from the distribution of 3D scenes that align with a single 2D input image.
****
Computational Guarantees for Doubly Entropic Wasserstein Barycenters
Tomas Vaskevicius, Lénaïc Chizat
We study the computation of doubly regularized Wasserstein barycenters, a recently introduced family of entropic barycenters governed by inner and outer regularization strengths. Previous research has demonstrated that various regularization parameter choices unify several notions of entropy-penalized barycenters while also revealing new ones, including a special case of debiased barycenters. In this paper, we propose and analyze an algorithm for computing doubly regularized Wasserstein barycenters. Our procedure builds on damped Sinkhorn iterations followed by exact maximization/minimization steps and guarantees convergence for any choice of regularization parameters. An inexact variant of our algorithm, implementable using approximate Monte Carlo sampling, offers the first non-asymptotic convergence guarantees for approximating Wasserstein barycenters between discrete point clouds in the free-support/grid-free setting.
****
WITRAN: Water-wave Information Transmission and Recurrent Acceleration Network for Long-range Time Series Forecasting
Yuxin Jia, Youfang Lin, Xinyan Hao, Yan Lin, Shengnan Guo, Huaiyu Wan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Spatio-Angular Convolutions for Super-resolution in Diffusion MRI
Matthew Lyon, Paul Armitage, Mauricio A Álvarez
Diffusion MRI (dMRI) is a widely used imaging modality, but requires long scanning times to acquire high resolution datasets. By leveraging the unique geometry present within this domain, we present a novel approach to dMRI angular super-resolution that extends upon the parametric continuous convolution (PCConv) framework. We introduce several additions to the operation including a Fourier feature mapping, 'global' co-ordinates, and domain specific context. Using this framework, we build a fully parametric continuous convolution network (PCCNN) and compare against existing models. We demonstrate the PCCNN performs competitively while using significantly fewer parameters. Moreover, we show that this formulation generalises well to clinically relevant downstream analyses such as fixel-based analysis, and neurite orientation dispersion and density imaging.
****
Disentangled Counterfactual Learning for Physical Audiovisual Commonsense Reasoning
Changsheng Lv, Shuai Zhang, Yapeng Tian, Mengshi Qi, Huadong Ma
In this paper, we propose a Disentangled Counterfactual Learning (DCL) approach

for physical audiovisual commonsense reasoning. The task aims to infer objects' physics commonsense based on both video and audio input, with the main challenge is how to imitate the reasoning ability of humans. Most of the current methods fail to take full advantage of different characteristics in multi-modal data, and lacking causal reasoning ability in models impedes the progress of implicit physical knowledge inferring. To address these issues, our proposed DCL method decouples videos into static (time-invariant) and dynamic (time-varying) factors in the latent space by the disentangled sequential encoder, which adopts a variational autoencoder (VAE) to maximize the mutual information with a contrastive loss function. Furthermore, we introduce a counterfactual learning module to augment the model's reasoning ability by modeling physical knowledge relationships among different objects under counterfactual intervention. Our proposed method is a plug-and-play module that can be incorporated into any baseline. In experiments, we show that our proposed method improves baseline methods and achieves state-of-the-art performance. Our source code is available at https://github.com/Andy20178/DCL.
****
Protein Design with Guided Discrete Diffusion
Nate Gruver, Samuel Stanton, Nathan Frey, Tim G. J. Rudner, Isidro Hotzel, Julien Lafrance-Vanasse, Arvind Rajpal, Kyunghyun Cho, Andrew G. Wilson
A popular approach to protein design is to combine a generative model with a discriminative model for conditional sampling. The generative model samples plausible sequences while the discriminative model guides a search for sequences with high fitness. Given its broad success in conditional sampling, classifier-guided diffusion modeling is a promising foundation for protein design, leading many to develop guided diffusion models for structure with inverse folding to recover sequences. In this work, we propose diffusioN Optimized Sampling (NOS), a guidance method for discrete diffusion models that follows gradients in the hidden states of the denoising network. NOS makes it possible to perform design directly in sequence space, circumventing significant limitations of structure-based methods, including scarce data and challenging inverse design. Moreover, we use NOS to generalize LaMBO, a Bayesian optimization procedure for sequence design that facilitates multiple objectives and edit-based constraints. The resulting method, LaMBO-2, enables discrete diffusions and stronger performance with limited edits through a novel application of saliency maps. We apply LaMBO-2 to a real-world protein design task, optimizing antibodies for higher expression yield and binding affinity to several therapeutic targets under locality and developability constraints, attaining a 99\% expression rate and 40\% binding rate in exploratory in vitro experiments.
****
Adaptive whitening with fast gain modulation and slow synaptic plasticity
Lyndon Duong, Eero Simoncelli, Dmitri Chklovskii, David Lipshutz
Neurons in early sensory areas rapidly adapt to changing sensory statistics, both by normalizing the variance of their individual responses and by reducing correlations between their responses. Together, these transformations may be viewed as an adaptive form of statistical whitening. Existing mechanistic models of adaptive whitening exclusively use either synaptic plasticity or gain modulation as the biological substrate for adaptation; however, on their own, each of these models has significant limitations. In this work, we unify these approaches in a normative multi-timescale mechanistic model that adaptively whitens its responses with complementary computational roles for synaptic plasticity and gain modulation. Gains are modified on a fast timescale to adapt to the current statistical context, whereas synapses are modified on a slow timescale to match structural properties of the input statistics that are invariant across contexts. Our model is derived from a novel multi-timescale whitening objective that factorizes the inverse whitening matrix into basis vectors, which correspond to synaptic weights, and a diagonal matrix, which corresponds to neuronal gains. We test our model on synthetic and natural datasets and find that the synapses learn optimal configurations over long timescales that enable adaptive whitening on short timescales using gain modulation.

****
Tanh Works Better with Asymmetry
Dongjin Kim, Woojeong Kim, Suhyun Kim
Batch Normalization is commonly located in front of activation functions, as proposed by the original paper. Swapping the order, i.e., using Batch Normalization after activation functions, has also been attempted, but its performance is generally not much different from the conventional order when ReLU or a similar activation function is used. However, in the case of bounded activation functions like Tanh, we discovered that the swapped order achieves considerably better performance than the conventional order on various benchmarks and architectures. This paper reports this remarkable phenomenon and closely examines what contributes to this performance improvement. By looking at the output distributions of individual activation functions, not the whole layers, we found that many of them are asymmetrically saturated. The experiments designed to induce a different degree of asymmetric saturation support the hypothesis that asymmetric saturation helps improve performance. In addition, Batch Normalization after bounded activation functions relocates the asymmetrically saturated output of activation functions near zero, enabling the swapped model to have high sparsity, further improving performance. Extensive experiments with Tanh, LeCun Tanh, and Softsign show that the swapped models achieve improved performance with a high degree of asymmetric saturation. Finally, based on this investigation, we test a Tanh function shifted to be asymmetric. This shifted Tanh function that is manipulated to have consistent asymmetry shows even higher accuracy than the original Tanh used in the swapped order, confirming the asymmetry's importance. The code is available at https://github.com/hipros/tanhworksbetterwithasymmetry.
****
Constraint-Conditioned Policy Optimization for Versatile Safe Reinforcement Learning
Yihang Yao, ZUXIN LIU, Zhepeng Cen, Jiacheng Zhu, Wenhao Yu, Tingnan Zhang, DING ZHAO
Safe reinforcement learning (RL) focuses on training reward-maximizing agents subject to pre-defined safety constraints. Yet, learning versatile safe policies that can adapt to varying safety constraint requirements during deployment without retraining remains a largely unexplored and challenging area. In this work, we formulate the versatile safe RL problem and consider two primary requirements: training efficiency and zero-shot adaptation capability. To address them, we introduce the Conditioned Constrained Policy Optimization (CCPO) framework, consisting of two key modules: (1) Versatile Value Estimation (VVE) for approximating value functions under unseen threshold conditions, and (2) Conditioned Variational Inference (CVI) for encoding arbitrary constraint thresholds during policy optimization. Our extensive experiments demonstrate that CCPO outperforms the baselines in terms of safety and task performance while preserving zero-shot adaptation capabilities to different constraint thresholds data-efficiently. This makes our approach suitable for real-world dynamic applications.
****
Prompt Pre-Training with Twenty-Thousand Classes for Open-Vocabulary Visual Recognition
Shuhuai Ren, Aston Zhang, Yi Zhu, Shuai Zhang, Shuai Zheng, Mu Li, Alexander J. Smola, Xu Sun
This work proposes POMP, a prompt pre-training method for vision-language models. Being memory and computation efficient, POMP enables the learned prompt to condense semantic information for a rich set of visual concepts with over twenty-thousand classes. Once pre-trained, the prompt with a strong transferable ability can be directly plugged into a variety of visual recognition tasks including image classification, semantic segmentation, and object detection, to boost recognition performances in a zero-shot manner. Empirical evaluation shows that POMP achieves state-of-the-art performances on 21 datasets, e.g., 67.0% average accuracy on 10 classification datasets (+3.1% compared to CoOp) and 84.4 hIoU on open-vocabulary Pascal VOC segmentation (+6.9 compared to ZSSeg).
****

Composing Parameter-Efficient Modules with Arithmetic Operation
Jinghan Zhang, shiqi chen, Junteng Liu, Junxian He
As an efficient alternative to conventional full fine-tuning, parameter-efficien
t fine-tuning (PEFT) is becoming the prevailing method to adapt pretrained langu
age models. In PEFT, a lightweight module is learned on each dataset while the u
nderlying pretrained language model remains unchanged, resulting in multiple com
pact modules representing diverse skills when applied to various domains and tas
ks. In this paper, we propose to compose these parameter-efficient modules throu
gh linear arithmetic operations in the weight space, thereby integrating differe
nt module capabilities. Specifically, we first define an addition and negation o
perator for the module, and then further compose these two basic operators to pe
rform flexible arithmetic. Our approach requires no additional training and enab
les highly flexible module composition. We apply different arithmetic operation
s to compose the parameter-efficient modules for (1) distribution generalization
, (2) multi-tasking, (3) detoxifying, and (4) domain transfer. Additionally, we
extend our approach to detoxify Alpaca-LoRA, the latest instruction-tuned large
language model based on LLaMA. Empirical results demonstrate that our approach p
roduces new and effective parameter-efficient modules that significantly outperf
orm existing ones across all settings.
****
UltraRE: Enhancing RecEraser for Recommendation Unlearning via Error Decompositi
on
Yuyuan Li, Chaochao Chen, Yizhao Zhang, Weiming Liu, Lingjuan Lyu, Xiaolin Zheng
, Dan Meng, Jun Wang
With growing concerns regarding privacy in machine learning models, regulations
have committed to granting individuals the right to be forgotten while mandating
 companies to develop non-discriminatory machine learning systems, thereby fueli
ng the study of the machine unlearning problem. Our attention is directed toward
 a practical unlearning scenario, i.e., recommendation unlearning. As the state-
of-the-art framework, i.e., RecEraser, naturally achieves full unlearning comple
teness, our objective is to enhance it in terms of model utility and unlearning
efficiency. In this paper, we rethink RecEraser from an ensemble-based perspecti
ve and focus on its three potential losses, i.e., redundancy, relevance, and com
bination. Under the theoretical guidance of the above three losses, we propose a
 new framework named UltraRE, which simplifies and powers RecEraser for recommen
dation tasks. Specifically, for redundancy loss, we incorporate transport weight
s in the clustering algorithm to optimize the equilibrium between collaboration
and balance while enhancing efficiency; for relevance loss, we ensure that sub-m
odels reach convergence on their respective group data; for combination loss, we
 simplify the combination estimator without compromising its efficacy. Extensive
 experiments on three real-world datasets demonstrate the effectiveness of Ultra
RE.
****
WCLD: Curated Large Dataset of Criminal Cases from Wisconsin Circuit Courts
Elliott Ash, Naman Goel, Nianyun Li, Claudia Marangon, Peiyao Sun
Machine learning based decision-support tools in criminal justice systems are su
bjects of intense discussions and academic research. There are important open qu
estions about the utility and fairness of such tools. Academic researchers often
 rely on a few small datasets that are not sufficient to empirically study vario
us real-world aspects of these questions. In this paper, we contribute WCLD, a c
urated large dataset of 1.5 million criminal cases from circuit courts in the U.
S. state of Wisconsin. We used reliable public data from 1970 to 2020 to curate
attributes like prior criminal counts and recidivism outcomes. The dataset conta
ins large number of samples from five racial groups, in addition to information
like sex and age (at judgment and first offense). Other attributes in this datas
et include neighborhood characteristics obtained from census data, detailed type
s of offense, charge severity, case decisions, sentence lengths, year of filing
etc. We also provide pseudo-identifiers for judge, county and zipcode. The datas
et will not only enable researchers to more rigorously study algorithmic fairnes
s in the context of criminal justice, but also relate algorithmic challenges wit

h various systemic issues. We also discuss in detail the process of constructing the dataset and provide a datasheet. The WCLD dataset is available at https://clezdata.github.io/wcld/.
****

Weitzman's Rule for Pandora's Box with Correlations
Evangelia Gergatsouli, Christos Tzamos
Pandora’s Box is a central problem in decision making under uncertainty that can model various real life scenarios. In this problem we are given n boxes, each with a fixed opening cost, and an unknown value drawn from a known distribution, only revealed if we pay the opening cost. Our goal is to find a strategy for opening boxes to minimize the sum of the value selected and the opening cost paid.In this work we revisit Pandora’s Box when the value distributions are correlated, first studied in [CGT+20]. We show that the optimal algorithm for the independent case, given by Weitzman’s rule, directly works for the correlated case. In fact, our algorithm results in significantly improved approximation guarantees compared to the previous work, while also being substantially simpler. We also show how to implement the rule given only sample access to the correlated distribution of values. Specifically, we find that a number of samples that is polynomial in the number of boxes is sufficient for the algorithm to work.
****

Compositional Sculpting of Iterative Generative Processes
Timur Garipov, Sebastiaan De Peuter, Ge Yang, Vikas Garg, Samuel Kaski, Tommi Jaakkola
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Face Reconstruction from Facial Templates by Learning Latent Space of a Generator Network
Hatef Otroshi Shahreza, Sébastien Marcel
In this paper, we focus on the template inversion attack against face recognition systems and propose a new method to reconstruct face images from facial templates. Within a generative adversarial network (GAN)-based framework, we learn a mapping from facial templates to the intermediate latent space of a pre-trained face generation network, from which we can generate high-resolution realistic reconstructed face images. We show that our proposed method can be applied in whitebox and blackbox attacks against face recognition systems. Furthermore, we evaluate the transferability of our attack when the adversary uses the reconstructed face image to impersonate the underlying subject in an attack against another face recognition system. Considering the adversary's knowledge and the target face recognition system, we define five different attacks and evaluate the vulnerability of state-of-the-art face recognition systems. Our experiments show that our proposed method achieves high success attack rates in whitebox and blackbox scenarios. Furthermore, the reconstructed face images are transferable and can be used to enter target face recognition systems with a different feature extractor model. We also explore important areas in the reconstructed face images that can fool the target face recognition system.
****

Triangulation Residual Loss for Data-efficient 3D Pose Estimation
Jiachen Zhao, Tao Yu, Liang An, Yipeng Huang, Fang Deng, Qionghai Dai
This paper presents Triangulation Residual loss (TR loss) for multiview 3D pose estimation in a data-efficient manner. Existing 3D supervised models usually require large-scale 3D annotated datasets, but the amount of existing data is still insufficient to train supervised models to achieve ideal performance, especially for animal pose estimation. To employ unlabeled multiview data for training, previous epipolar-based consistency provides a self-supervised loss that considers only the local consistency in pairwise views, resulting in limited performance and heavy calculations. In contrast, TR loss enables self-supervision with global multiview geometric consistency. Starting from initial 2D keypoint estimates,

the TR loss can fine-tune the corresponding 2D detector without 3D supervision by simply minimizing the smallest singular value of the triangulation matrix in an end-to-end fashion. Our method achieves the state-of-the-art 25.8mm MPJPE and competitive 28.7mm MPJPE with only 5\% 2D labeled training data on the Human3.6M dataset. Experiments on animals such as mice demonstrate our TR loss's data-efficient training ability.
****
A Long N-step Surrogate Stage Reward for Deep Reinforcement Learning
Junmin Zhong, Ruofan Wu, Jennie Si
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CODA: Generalizing to Open and Unseen Domains with Compaction and Disambiguation
Chaoqi Chen, Luyao Tang, Yue Huang, Xiaoguang Han, Yizhou Yu
The generalization capability of machine learning systems degenerates notably when the test distribution drifts from the training distribution. Recently, Domain Generalization (DG) has been gaining momentum in enabling machine learning models to generalize to unseen domains. However, most DG methods assume that training and test data share an identical label space, ignoring the potential unseen categories in many real-world applications. In this paper, we delve into a more general but difficult problem termed Open Test-Time DG (OTDG), where both domain shift and open class may occur on the unseen test data. We propose Compaction and Disambiguation (CODA), a novel two-stage framework for learning compact representations and adapting to open classes in the wild. To meaningfully regularize the model's decision boundary, CODA introduces virtual unknown classes and optimizes a new training objective to insert unknowns into the latent space by compacting the embedding space of source known classes. To adapt target samples to the source model, we then disambiguate the decision boundaries between known and unknown classes with a test-time training objective, mitigating the adaptivity gap and catastrophic forgetting challenges. Experiments reveal that CODA can significantly outperform the previous best method on standard DG datasets and harmonize the classification accuracy between known and unknown classes.
****
Scale-Space Hypernetworks for Efficient Biomedical Image Analysis
Jose Javier Gonzalez Ortiz, John Guttag, Adrian Dalca
Convolutional Neural Networks (CNNs) are the predominant model used for a variety of medical image analysis tasks. At inference time, these models are computationally intensive, especially with volumetric data.In principle, it is possible to trade accuracy for computational efficiency by manipulating the rescaling factor in the downsample and upsample layers of CNN architectures.However, properly exploring the accuracy-efficiency trade-off is prohibitively expensive with existing models.To address this, we introduce Scale-Space HyperNetworks (SSHN), a method that learns a spectrum of CNNs with varying internal rescaling factors.A single SSHN characterizes an entire Pareto accuracy-efficiency curve of models that match, and occasionally surpass, the outcomes of training many separate networks with fixed rescaling factors.We demonstrate the proposed approach in several medical image analysis applications, comparing SSHN against strategies with both fixed and dynamic rescaling factors.We find that SSHN consistently provides a better accuracy-efficiency trade-off at a fraction of the training cost. Trained SSHNs enable the user to quickly choose a rescaling factor that appropriately balances accuracy and computational efficiency for their particular needs at inference.
****
Follow-ups Also Matter: Improving Contextual Bandits via Post-serving Contexts
Chaoqi Wang, Ziyu Ye, Zhe Feng, Ashwinkumar Badanidiyuru Varadaraja, Haifeng Xu
Standard contextual bandit problem assumes that all the relevant contexts are observed before the algorithm chooses an arm. This modeling paradigm, while useful, often falls short when dealing with problems in which additional valuable cont

exts can be observed after arm selection. For example, content recommendation pl atforms like Youtube, Instagram, Tiktok receive much additional features about a user's reward after the user clicks a content (e.g., how long the user stayed, what is the user's watch speed, etc.). To improve online learning efficiency in these applications, we study a novel contextual bandit problem with post-servi ng contexts and design a new algorithm, poLinUCB, that achieves tight regret un der standard assumptions. Core to our technical proof is a robustified and gener alized version of the well-known Elliptical Potential Lemma (EPL), which can acc ommodate noise in data. Such robustification is necessary for tackling our prob lem, though we believe it could also be of general interest.Extensive empirical tests on both synthetic and real-world datasets demonstrate the significant ben efit of utilitzing post-serving contexts as well as the superior performance of our algorithm over the state-of-the-art approaches.
****
Offline Minimax Soft-Q-learning Under Realizability and Partial Coverage
Masatoshi Uehara, Nathan Kallus, Jason D. Lee, Wen Sun
We consider offline reinforcement learning (RL) where we only have only access t o offline data. In contrast to numerous offline RL algorithms that necessitate t he uniform coverage of the offline data over state and action space, we propose value-based algorithms with PAC guarantees under partial coverage, specifically, coverage of offline data against a single policy, and realizability of soft Q-f unction (a.k.a., entropy-regularized Q-function) and another function, which is defined as a solution to a saddle point of certain minimax optimization problem) . Furthermore, we show the analogous result for Q-functions instead of soft Q-fu nctions. To attain these guarantees, we use novel algorithms with minimax loss f unctions to accurately estimate soft Q-functions and Q-functions with -convergen ce guarantees measured on the offline data. We introduce these loss functions by casting the estimation problems into nonlinear convex optimization problems and taking the Lagrange functions.
****
Restless Bandits with Average Reward: Breaking the Uniform Global Attractor Assu mption
Yige Hong, Qiaomin Xie, Yudong Chen, Weina Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Smooth Flipping Probability for Differential Private Sign Random Projection Meth ods
Ping Li, Xiaoyun Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Idempotent Learned Image Compression with Right-Inverse
Yanghao Li, Tongda Xu, Yan Wang, Jingjing Liu, Ya-Qin Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
A Simple Yet Effective Strategy to Robustify the Meta Learning Paradigm
Qi Wang, Yiqin Lv, yanghe feng, Zheng Xie, Jincai Huang
Meta learning is a promising paradigm to enable skill transfer across tasks.Most previous methods employ the empirical risk minimization principle in optimizati on.However, the resulting worst fast adaptation to a subset of tasks can be cata strophic in risk-sensitive scenarios.To robustify fast adaptation, this paper op timizes meta learning pipelines from a distributionally robust perspective and m

eta trains models with the measure of tail task risk.We take the two-stage strategy as heuristics to solve the robust meta learning problem, controlling the worst fast adaptation cases at a certain probabilistic level. Experimental results show that our simple method can improve the robustness of meta learning to task distributions and reduce the conditional expectation of the worst fast adaptation risk.
****
Stabilized Neural Differential Equations for Learning Dynamics with Explicit Constraints
Alistair White, Niki Kilbertus, Maximilian Gelbrecht, Niklas Boers
Many successful methods to learn dynamical systems from data have recently been introduced. However, ensuring that the inferred dynamics preserve known constraints, such as conservation laws or restrictions on the allowed system states, remains challenging. We propose stabilized neural differential equations (SNDEs), a method to enforce arbitrary manifold constraints for neural differential equations. Our approach is based on a stabilization term that, when added to the original dynamics, renders the constraint manifold provably asymptotically stable. Due to its simplicity, our method is compatible with all common neural differential equation (NDE) models and broadly applicable. In extensive empirical evaluations, we demonstrate that SNDEs outperform existing methods while broadening the types of constraints that can be incorporated into NDE training.
****
On the Importance of Exploration for Generalization in Reinforcement Learning
Yiding Jiang, J. Zico Kolter, Roberta Raileanu
Existing approaches for improving generalization in deep reinforcement learning (RL) have mostly focused on representation learning, neglecting RL-specific aspects such as exploration. We hypothesize that the agent's exploration strategy plays a key role in its ability to generalize to new environments.Through a series of experiments in a tabular contextual MDP, we show that exploration is helpful not only for efficiently finding the optimal policy for the training environments but also for acquiring knowledge that helps decision making in unseen environments. Based on these observations, we propose EDE: Exploration via Distributional Ensemble, a method that encourages the exploration of states with high epistemic uncertainty through an ensemble of Q-value distributions. The proposed algorithm is the first value-based approach to achieve strong performance on both Procgen and Crafter, two benchmarks for generalization in RL with high-dimensional observations. The open-sourced implementation can be found at https://github.com/facebookresearch/ede.
****
Uniform Convergence with Square-Root Lipschitz Loss
Lijia Zhou, Zhen Dai, Frederic Koehler, Nati Srebro
We establish generic uniform convergence guarantees for Gaussian data in terms of the Radamacher complexity of the hypothesis class and the Lipschitz constant of the square root of the scalar loss function. We show how these guarantees substantially generalize previous results based on smoothness (Lipschitz constant of the derivative), and allow us to handle the broader class of square-root-Lipschtz losses, which includes also non-smooth loss functions appropriate for studying phase retrieval and ReLU regression, as well as rederive and better understand “optimistic rate” and interpolation learning guarantees.
****
A Fractional Graph Laplacian Approach to Oversmoothing
Sohir Maskey, Raffaele Paolino, Aras Bacho, Gitta Kutyniok
Graph neural networks (GNNs) have shown state-of-the-art performances in various applications. However, GNNs often struggle to capture long-range dependencies in graphs due to oversmoothing. In this paper, we generalize the concept of oversmoothing from undirected to directed graphs. To this aim, we extend the notion of Dirichlet energy by considering a directed symmetrically normalized Laplacian. As vanilla graph convolutional networks are prone to oversmooth, we adopt a neural graph ODE framework. Specifically, we propose fractional graph Laplacian neural ODEs, which describe non-local dynamics. We prove that our approach allows p

ropagating information between distant nodes while maintaining a low probabilit
y of long-distance jumps. Moreover, we show that our method is more flexible wit
h respect to the convergence of the graph’s Dirichlet energy, thereby mitigating
 oversmoothing. We conduct extensive experiments on synthetic and real-world gra
phs, both directed and undirected, demonstrating our method’s versatility acros
s diverse graph homophily levels. Ourcode is available at https://github.com/RP
aolino/fLode
****
On the Convergence and Sample Complexity Analysis of Deep Q-Networks with $\epsi
lon$-Greedy Exploration
Shuai Zhang, Hongkang Li, Meng Wang, Miao Liu, Pin-Yu Chen, Songtao Lu, Sijia Li
u, Keerthiram Murugesan, Subhajit Chaudhury
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Joint Data-Task Generation for Auxiliary Learning
Hong Chen, Xin Wang, Yuwei Zhou, Yijian Qin, Chaoyu Guan, Wenwu Zhu
Current auxiliary learning methods mainly adopt the methodology of reweighing lo
sses for the manually collected auxiliary data and tasks. However, these methods
 heavily rely on domain knowledge during data collection, which may be hardly av
ailable in reality. Therefore, current methods will become less effective and ev
en do harm to the primary task when unhelpful auxiliary data and tasks are emplo
yed. To tackle the problem, we propose a joint data-task generation framework fo
r auxiliary learning (DTG-AuxL), which can bring benefits to the primary task by
 generating the new auxiliary data and task in a joint manner. The proposed DTG-
AuxL framework contains a joint generator and a bi-level optimization strategy.
Specifically, the joint generator contains a feature generator and a label gener
ator, which are designed to be applicable and expressive for various auxiliary
learning scenarios. The bi-level optimization strategy optimizes the joint gener
ator and the task learning model, where the joint generator is effectively optim
ized in the upper level via the implicit gradient from the primary loss and the
explicit gradient of our proposed instance regularization, while the task learni
ng model is optimized in the lower level by the generated data and task. Extensi
ve experiments show that our proposed DTG-AuxL framework consistently outperform
s existing methods in various auxiliary learning scenarios, particularly when th
e manually collected auxiliary data and tasks are unhelpful.
****
On Proper Learnability between Average- and Worst-case Robustness
Vinod Raman, UNIQUE SUBEDI, Ambuj Tewari
Recently, Montasser at al. (2019) showed that finite VC dimension is not suffici
ent for proper adversarially robust PAC learning. In light of this hardness, the
re is a growing effort to study what type of relaxations to the adversarially ro
bust PAC learning setup can enable proper learnability. In this work, we initiat
e the study of proper learning under relaxations of the worst-case robust loss.
We give a family of robust loss relaxations under which VC classes are properly
PAC learnable with sample complexity close to what one would require in the stan
dard PAC learning setup. On the other hand, we show that for an existing and nat
ural relaxation of the worst-case robust loss, finite VC dimension is not suffic
ient for proper learning. Lastly, we give new generalization guarantees for the
adversarially robust empirical risk minimizer.
****
Distributional Policy Evaluation: a Maximum Entropy approach to Representation L
earning
Riccardo Zamboni, Alberto Maria Metelli, Marcello Restelli
The Maximum Entropy (Max-Ent) framework has been effectively employed in a varie
ty of Reinforcement Learning (RL) tasks. In this paper, we first propose a novel
 Max-Ent framework for policy evaluation in a distributional RL setting, named D
istributional Maximum Entropy Policy Evaluation (D-Max-Ent PE). We derive a gene

ralization-error bound that depends on the complexity of the representation employed, showing that this framework can explicitly take into account the features used to represent the state space while evaluating a policy. Then, we exploit these favorable properties to drive the representation learning of the state space in a Structural Risk Minimization fashion. We employ state-aggregation functions as feature functions and we specialize the D-Max-Ent approach into an algorithm, named D-Max-Ent Progressive Factorization, which constructs a progressively finer-grained representation of the state space by balancing the trade-off between preserving information (bias) and reducing the effective number of states, i.e., the complexity of the representation space (variance). Finally, we report the results of some illustrative numerical simulations, showing that the proposed algorithm matches the expected theoretical behavior and highlighting the relationship between aggregations and sample regimes.
****

Thin and deep Gaussian processes
Daniel Augusto de Souza, Alexander Nikitin, ST John, Magnus Ross, Mauricio A Álvarez, Marc Deisenroth, João Paulo Gomes, Diego Mesquita, César Lincoln Mattos
Gaussian processes (GPs) can provide a principled approach to uncertainty quantification with easy-to-interpret kernel hyperparameters, such as the lengthscale, which controls the correlation distance of function values.However, selecting an appropriate kernel can be challenging.Deep GPs avoid manual kernel engineering by successively parameterizing kernels with GP layers, allowing them to learn low-dimensional embeddings of the inputs that explain the output data.Following the architecture of deep neural networks, the most common deep GPs warp the input space layer-by-layer but lose all the interpretability of shallow GPs. An alternative construction is to successively parameterize the lengthscale of a kernel, improving the interpretability but ultimately giving away the notion of learning lower-dimensional embeddings. Unfortunately, both methods are susceptible to particular pathologies which may hinder fitting and limit their interpretability. This work proposes a novel synthesis of both previous approaches: {Thin and Deep GP} (TDGP). Each TDGP layer defines locally linear transformations of the original input data maintaining the concept of latent embeddings while also retaining the interpretation of lengthscales of a kernel. Moreover, unlike the prior solutions, TDGP induces non-pathological manifolds that admit learning lower-dimensional representations.We show with theoretical and experimental results that i) TDGP is, unlike previous models, tailored to specifically discover lower-dimensional manifolds in the input data, ii) TDGP behaves well when increasing the number of layers, and iii) TDGP performs well in standard benchmark datasets.
****

Human-like Few-Shot Learning via Bayesian Reasoning over Natural Language
Kevin Ellis
A core tension in models of concept learning is that the model must carefully balance the tractability of inference against the expressivity of the hypothesis class. Humans, however, can efficiently learn a broad range of concepts. We introduce a model of inductive learning that seeks to be human-like in that sense.It implements a Bayesian reasoning process where a language model first proposes candidate hypotheses expressed in natural language, which are then re-weighed by a prior and a likelihood.By estimating the prior from human data, we can predict human judgments on learning problems involving numbers and sets, spanning concepts that are generative, discriminative, propositional, and higher-order.
****

CSLP-AE: A Contrastive Split-Latent Permutation Autoencoder Framework for Zero-Shot Electroencephalography Signal Conversion
Anders Nørskov, Alexander Neergaard Zahid, Morten Mørup
Electroencephalography (EEG) is a prominent non-invasive neuroimaging technique providing insights into brain function. Unfortunately, EEG data exhibit a high degree of noise and variability across subjects hampering generalizable signal extraction. Therefore, a key aim in EEG analysis is to extract the underlying neural activation (content) as well as to account for the individual subject variability (style). We hypothesize that the ability to convert EEG signals between tas

ks and subjects requires the extraction of latent representations accounting for
 content and style. Inspired by recent advancements in voice conversion technolo
gies, we propose a novel contrastive split-latent permutation autoencoder (CSLP-
AE) framework that directly optimizes for EEG conversion. Importantly, the laten
t representations are guided using contrastive learning to promote the latent sp
lits to explicitly represent subject (style) and task (content). We contrast CSL
P-AE to conventional supervised, unsupervised (AE), and self-supervised (contras
tive learning) training and find that the proposed approach provides favorable g
eneralizable characterizations of subject and task. Importantly, the procedure a
lso enables zero-shot conversion between unseen subjects. While the present work
 only considers conversion of EEG, the proposed CSLP-AE provides a general frame
work for signal conversion and extraction of content (task activation) and style
 (subject variability) components of general interest for the modeling and analy
sis of biological signals.
****
Delegated Classification
Eden Saig, Inbal Talgam-Cohen, Nir Rosenfeld
When machine learning is outsourced to a rational agent, conflicts of interest m
ight arise and severely impact predictive performance. In this work, we propose
a theoretical framework for incentive-aware delegation of machine learning tasks
. We model delegation as a principal-agent game, in which accurate learning can
be incentivized by the principal using performance-based contracts. Adapting the
 economic theory of contract design to this setting, we define budget-optimal co
ntracts and prove they take a simple threshold form under reasonable assumptions
. In the binary-action case, the optimality of such contracts is shown to be equ
ivalent to the classic Neyman-Pearson lemma, establishing a formal connection be
tween contract design and statistical hypothesis testing. Empirically, we demons
trate that budget-optimal contracts can be constructed using small-scale data, l
everaging recent advances in the study of learning curves and scaling laws. Perf
ormance and economic outcomes are evaluated using synthetic and real-world class
ification tasks.
****
PTQD: Accurate Post-Training Quantization for Diffusion Models
Yefei He, Luping Liu, Jing Liu, Weijia Wu, Hong Zhou, Bohan Zhuang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Reward Finetuning for Faster and More Accurate Unsupervised Object Discovery
Katie Luo, Zhenzhen Liu, Xiangyu Chen, Yurong You, Sagie Benaim, Cheng Perng Pho
o, Mark Campbell, Wen Sun, Bharath Hariharan, Kilian Q. Weinberger
Recent advances in machine learning have shown that Reinforcement Learning from
Human Feedback (RLHF) can improve machine learning models and align them with hu
man preferences. Although very successful for Large Language Models (LLMs), thes
e advancements have not had a comparable impact in research for autonomous vehic
les—where alignment with human expectations can be imperative. In this paper, we
 propose to adapt similar RL-based methods to unsupervised object discovery, i.e
. learning to detect objects from LiDAR points without any training labels. Inst
ead of labels, we use simple heuristics to mimic human feedback. More explicitly
, we combine multiple heuristics into a simple reward function that positively c
orrelates its score with bounding box accuracy, i.e., boxes containing objects a
re scored higher than those without. We start from the detector’s own prediction
s to explore the space and reinforce boxes with high rewards through gradient up
dates. Empirically, we demonstrate that our approach is not only more accurate,
but also orders of magnitudes faster to train compared to prior works on object
discovery. Code is available at https://github.com/katieluo88/DRIFT.
****
Doubly Constrained Fair Clustering
John Dickerson, Seyed Esmaeili, Jamie H. Morgenstern, Claire Jie Zhang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ResShift: Efficient Diffusion Model for Image Super-resolution by Residual Shifting
Zongsheng Yue, Jianyi Wang, Chen Change Loy
Diffusion-based image super-resolution (SR) methods are mainly limited by the low inference speed due to the requirements of hundreds or even thousands of sampling steps. Existing acceleration sampling techniques inevitably sacrifice performance to some extent, leading to over-blurry SR results. To address this issue, we propose a novel and efficient diffusion model for SR that significantly reduces the number of diffusion steps, thereby eliminating the need for post-acceleration during inference and its associated performance deterioration. Our method constructs a Markov chain that transfers between the high-resolution image and the low-resolution image by shifting the residual between them, substantially improving the transition efficiency. Additionally, an elaborate noise schedule is developed to flexibly control the shifting speed and the noise strength during the diffusion process. Extensive experiments demonstrate that the proposed method obtains superior or at least comparable performance to current state-of-the-art methods on both synthetic and real-world datasets, \textit{\textbf{even only with 20 sampling steps}}. Our code and model will be made publicly.
****
WalkLM: A Uniform Language Model Fine-tuning Framework for Attributed Graph Embedding
Yanchao Tan, Zihao Zhou, Hang Lv, Weiming Liu, Carl Yang
Graphs are widely used to model interconnected entities and improve downstream predictions in various real-world applications. However, real-world graphs nowadays are often associated with complex attributes on multiple types of nodes and even links that are hard to model uniformly, while the widely used graph neural networks (GNNs) often require sufficient training toward specific downstream predictions to achieve strong performance. In this work, we take a fundamentally different approach than GNNs, to simultaneously achieve deep joint modeling of complex attributes and flexible structures of real-world graphs and obtain unsupervised generic graph representations that are not limited to specific downstream predictions. Our framework, built on a natural integration of language models (LMs) and random walks (RWs), is straightforward, powerful and data-efficient. Specifically, we first perform attributed RWs on the graph and design an automated program to compose roughly meaningful textual sequences directly from the attributed RWs; then we fine-tune an LM using the RW-based textual sequences and extract embedding vectors from the LM, which encapsulates both attribute semantics and graph structures. In our experiments, we evaluate the learned node embeddings towards different downstream prediction tasks on multiple real-world attributed graph datasets and observe significant improvements over a comprehensive set of state-of-the-art unsupervised node embedding methods. We believe this work opens a door for more sophisticated technical designs and empirical evaluations toward the leverage of LMs for the modeling of real-world graphs.
****
Generalizing Nonlinear ICA Beyond Structural Sparsity
Yujia Zheng, Kun Zhang
Nonlinear independent component analysis (ICA) aims to uncover the true latent sources from their observable nonlinear mixtures. Despite its significance, the identifiability of nonlinear ICA is known to be impossible without additional assumptions. Recent advances have proposed conditions on the connective structure from sources to observed variables, known as Structural Sparsity, to achieve identifiability in an unsupervised manner. However, the sparsity constraint may not hold universally for all sources in practice. Furthermore, the assumptions of bijectivity of the mixing process and independence among all sources, which arise from the setting of ICA, may also be violated in many real-world scenarios. To a

ddress these limitations and generalize nonlinear ICA, we propose a set of new i dentifiability results in the general settings of undercompleteness, partial spa rsity and source dependence, and flexible grouping structures. Specifically, we prove identifiability when there are more observed variables than sources (under complete), and when certain sparsity and/or source independence assumptions are not met for some changing sources. Moreover, we show that even in cases with fle xible grouping structures (e.g., part of the sources can be divided into irreduc ible independent groups with various sizes), appropriate identifiability results can also be established. Theoretical claims are supported empirically on both s ynthetic and real-world datasets.
****
Towards Characterizing the First-order Query Complexity of Learning (Approximate) Nash Equilibria in Zero-sum Matrix Games
Hedi Hadiji, Sarah Sachs, Tim van Erven, Wouter M. Koolen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Fast Conditional Mixing of MCMC Algorithms for Non-log-concave Distributions
Xiang Cheng, Bohan Wang, Jingzhao Zhang, Yusong Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
How to Select Which Active Learning Strategy is Best Suited for Your Specific Pr oblem and Budget
Guy Hacohen, Daphna Weinshall
In the domain of Active Learning (AL), a learner actively selects which unlabele d examples to seek labels from an oracle, while operating within predefined budg et constraints. Importantly, it has been recently shown that distinct query stra tegies are better suited for different conditions and budgetary constraints. In practice, the determination of the most appropriate AL strategy for a given situ ation remains an open problem. To tackle this challenge, we propose a practical derivative-based method that dynamically identifies the best strategy for a give n budget. Intuitive motivation for our approach is provided by the theoretical a nalysis of a simplified scenario. We then introduce a method to dynamically sele ct an AL strategy, which takes into account the unique characteristics of the pr oblem and the available budget. Empirical results showcase the effectiveness of our approach across diverse budgets and computer vision tasks.
****
Aligning Synthetic Medical Images with Clinical Knowledge using Human Feedback
Shenghuan Sun, Greg Goldgof, Atul Butte, Ahmed M. Alaa
Generative models capable of precisely capturing nuanced clinical features in me dical images hold great promise for facilitating clinical data sharing, enhancin g rare disease datasets, and efficiently synthesizing (annotated) medical images at scale. Despite their potential, assessing the quality of synthetic medical i mages remains a challenge. While modern generative models can synthesize visuall y-realistic medical images, the clinical plausibility of these images may be cal led into question. Domain-agnostic scores, such as FID score, precision, and rec all, cannot incorporate clinical knowledge and are, therefore, not suitable for assessing clinical sensibility. Additionally, there are numerous unpredictable w ays in which generative models may fail to synthesize clinically plausible image s, making it challenging to anticipate potential failures and design automated s cores for their detection. To address these challenges, this paper introduces a pathologist-in-the-loop framework for generating clinically-plausible synthetic medical images. Our framework comprises three steps: (1) pretraining a condition al diffusion model to generate medical images conditioned on a clinical concept, (2) expert pathologist evaluation of the generated images to assess whether the

y satisfy clinical desiderata, and (3) training a reward model that predicts hum
an feedback on new samples, which we use to incorporate expert knowledge into th
e finetuning objective of the diffusion model. Our results show that human feedb
ack significantly improves the quality of synthetic images in terms of fidelity,
 diversity, utility in downstream applications, and plausibility as evaluated by
 experts. We also demonstrate that human feedback can teach the model new clinic
al concepts not annotated in the original training data. Our results demonstrate
 the value of incorporating human feedback in clinical applications where genera
tive models may struggle to capture extensive domain knowledge from raw data alo
ne.
****
Interpretable Graph Networks Formulate Universal Algebra Conjectures
Francesco Giannini, Stefano Fioravanti, Oguzhan Keskin, Alisia Lupidi, Lucie Cha
rlotte Magister, Pietro Lió, Pietro Barbiero
The rise of Artificial Intelligence (AI) recently empowered researchers to inves
tigate hard mathematical problems which eluded traditional approaches for decade
s. Yet, the use of AI in Universal Algebra (UA)---one of the fields laying the f
oundations of modern mathematics---is still completely unexplored. This work pro
poses the first use of AI to investigate UA's conjectures with an equivalent equ
ational and topological characterization. While topological representations woul
d enable the analysis of such properties using graph neural networks, the limite
d transparency and brittle explainability of these models hinder their straightf
orward use to empirically validate existing conjectures or to formulate new ones
. To bridge these gaps, we propose a general algorithm generating AI-ready datas
ets based on UA's conjectures, and introduce a novel neural layer to build fully
 interpretable graph networks. The results of our experiments demonstrate that i
nterpretable graph networks: (i) enhance interpretability without sacrificing ta
sk accuracy, (ii) strongly generalize when predicting universal algebra's proper
ties, (iii) generate simple explanations that empirically validate existing conj
ectures, and (iv) identify subgraphs suggesting the formulation of novel conject
ures.
****
GraphAdapter: Tuning Vision-Language Models With Dual Knowledge Graph
Xin Li, Dongze Lian, Zhihe Lu, Jiawang Bai, Zhibo Chen, Xinchao Wang
Adapter-style efficient transfer learning (ETL) has shown excellent performance
in the tuning of vision-language models (VLMs) under the low-data regime, where
only a few additional parameters are introduced to excavate the task-specific kn
owledge based on the general and powerful representation of VLMs. However, most
adapter-style works face two limitations: (i) modeling task-specific knowledge w
ith a single modality only; and (ii) overlooking the exploitation of the inter-c
lass relationships in downstream tasks, thereby leading to sub-optimal solutions
. To mitigate that, we propose an effective adapter-style tuning strategy, dubbe
d GraphAdapter, which performs the textual adapter by explicitly modeling the du
al-modality structure knowledge (i.e., the correlation of different semantics/cl
asses in textual and visual modalities) with a dual knowledge graph. In particul
ar, the dual knowledge graph is established with two sub-graphs, i.e., a textual
 knowledge sub-graph, and a visual knowledge sub-graph, where the nodes and edge
s represent the semantics/classes and their correlations in two modalities, resp
ectively. This enables the textual feature of each prompt to leverage the task-s
pecific structure knowledge from both textual and visual modalities, yielding a
more effective classifier for downstream tasks. Extensive experimental results o
n 11 benchmark datasets reveal that our GraphAdapter significantly outperforms t
he previous adapter-based methods.
****
FaceComposer: A Unified Model for Versatile Facial Content Creation
Jiayu Wang, Kang Zhao, Yifeng Ma, Shiwei Zhang, Yingya Zhang, Yujun Shen, Deli Z
hao, Jingren Zhou
This work presents FaceComposer, a unified generative model that accomplishes a
variety of facial content creation tasks, including text-conditioned face synthe
sis, text-guided face editing, face animation etc. Based on the latent diffusion

 framework, FaceComposer follows the paradigm of compositional generation and em
ploys diverse face-specific conditions, e.g., Identity Feature and Projected Nor
malized Coordinate Code, to release the model creativity at all possible. To sup
port text control and animation, we clean up some existing face image datasets a
nd collect around 500 hours of talking-face videos, forming a high-quality large
-scale multi-modal face database. A temporal self-attention module is incorporat
ed into the U-Net structure, which allows learning the denoising process on the
mixture of images and videos. Extensive experiments suggest that our approach no
t only achieves comparable or even better performance than state-of-the-arts on
each single task, but also facilitates some combined tasks with one-time forward
, demonstrating its potential in serving as a foundation generative model in fac
e domain. We further develop an interface such that users can enjoy our one-step
 service to create, edit, and animate their own characters. Code, dataset, model
, and interface will be made publicly available.
****
A Unified Solution for Privacy and Communication Efficiency in Vertical Federate
d Learning
Ganyu Wang, Bin Gu, Qingsong Zhang, Xiang Li, Boyu Wang, Charles X. Ling
Vertical Federated Learning (VFL) is a collaborative machine learning paradigm t
hat enables multiple participants to jointly train a model on their private data
 without sharing it.To make VFL practical, privacy security and communication ef
ficiency should both be satisfied. Recent research has shown that Zero-Order Opt
imization (ZOO) in VFL can effectively conceal the internal information of the m
odel without adding costly privacy protective add-ons, making it a promising app
roach for privacy and efficiency.However, there are still two key problems that
have yet to be resolved. First, the convergence rate of ZOO-based VFL is signifi
cantly slower compared to gradient-based VFL, resulting in low efficiency in mod
el training and more communication round, which hinders its application on large
 neural networks. Second, although ZOO-based VFL has demonstrated resistance to
state-of-the-art (SOTA) attacks, its privacy guarantee lacks a theoretical expla
nation.To address these challenges, we propose a novel cascaded hybrid optimizat
ion approach that employs a zeroth-order (ZO) gradient on the most critical outp
ut layer of the clients, with other parts utilizing the first-order (FO) gradien
t. This approach preserves the privacy protection of ZOO while significantly enh
ancing convergence.Moreover, we theoretically prove that applying ZOO to the VFL
 is equivalent to adding Gaussian Mechanism to the gradient information, which o
ffers an implicit differential privacy guarantee. Experimental results demonstra
te that our proposed framework achieves similar utility as the Gaussian mechanis
m under the same privacy budget, while also having significantly lower communica
tion costs compared with SOTA communication-efficient VFL frameworks.
****
Optimization and Bayes: A Trade-off for Overparameterized Neural Networks
Zhengmian Hu, Heng Huang
This paper proposes a novel algorithm, Transformative Bayesian Learning (TansBL)
, which bridges the gap between empirical risk minimization (ERM) and Bayesian l
earning for neural networks. We compare ERM, which uses gradient descent to opti
mize, and Bayesian learning with importance sampling for their generalization an
d computational complexity. We derive the first algorithm-dependent PAC-Bayesian
 generalization bound for infinitely wide networks based on an exact KL divergen
ce between the trained posterior distribution obtained by infinitesimal step siz
e gradient descent and a Gaussian prior. Moreover, we show how to transform grad
ient-based optimization into importance sampling by incorporating a weight. Whil
e Bayesian learning has better generalization, it suffers from low sampling effi
ciency. Optimization methods, on the other hand, have good sampling efficiency b
ut poor generalization. Our proposed algorithm TansBL enables a trade-off betwee
n generalization and sampling efficiency.
****
Understanding Social Reasoning in Language Models with Language Models
Kanishk Gandhi, Jan-Philipp Fraenken, Tobias Gerstenberg, Noah Goodman
As Large Language Models (LLMs) become increasingly integrated into our everyday

lives, understanding their ability to comprehend human mental states becomes critical for ensuring effective interactions. However, despite the recent attempts to assess the Theory-of-Mind (ToM) reasoning capabilities of LLMs, the degree to which these models can align with human ToM remains a nuanced topic of exploration. This is primarily due to two distinct challenges: (1) the presence of inconsistent results from previous evaluations, and (2) concerns surrounding the validity of existing evaluation methodologies. To address these challenges, we present a novel framework for procedurally generating evaluations with LLMs by populating causal templates. Using our framework, we create a new social reasoning benchmark (BigToM) for LLMs which consists of 25 controls and 5,000 model-written evaluations. We find that human participants rate the quality of our benchmark higher than previous crowd-sourced evaluations and comparable to expert-written evaluations. Using BigToM, we evaluate the social reasoning capabilities of a variety of LLMs and compare model performances with human performance. Our results suggest that GPT4 has ToM capabilities that mirror human inference patterns, though less reliable, while other LLMs struggle.
****

Reproducibility in Multiple Instance Learning: A Case For Algorithmic Unit Tests

Edward Raff, James Holt

Multiple Instance Learning (MIL) is a sub-domain of classification problems with positive and negative labels and a "bag" of inputs, where the label is positive if and only if a positive element is contained within the bag, and otherwise is negative. Training in this context requires associating the bag-wide label to instance-level information, and implicitly contains a causal assumption and asymmetry to the task (i.e., you can't swap the labels without changing the semantics). MIL problems occur in healthcare (one malignant cell indicates cancer), cyber security (one malicious executable makes an infected computer), and many other tasks. In this work, we examine five of the most prominent deep-MIL models and find that none of them respects the standard MIL assumption. They are able to learn anti-correlated instances, i.e., defaulting to "positive" labels until seeing a negative counter-example, which should not be possible for a correct MIL model. We suspect that enhancements and other works derived from these models will share the same issue. In any context in which these models are being used, this creates the potential for learning incorrect models, which creates risk of operational failure. We identify and demonstrate this problem via a proposed ``algorithmic unit test'', where we create synthetic datasets that can be solved by a MIL respecting model, and which clearly reveal learning that violates MIL assumptions. The five evaluated methods each fail one or more of these tests. This provides a model-agnostic way to identify violations of modeling assumptions, which we hope will be useful for future development and evaluation of MIL models.
****

Meta-learning families of plasticity rules in recurrent spiking networks using simulation-based inference

Basile Confavreux, Poornima Ramesh, Pedro J. Goncalves, Jakob H Macke, Tim Vogels

There is substantial experimental evidence that learning and memory-related behaviours rely on local synaptic changes, but the search for distinct plasticity rules has been driven by human intuition, with limited success for multiple, co-active plasticity rules in biological networks. More recently, automated meta-learning approaches have been used in simplified settings, such as rate networks and small feed-forward spiking networks. Here, we develop a simulation-based inference (SBI) method for sequentially filtering plasticity rules through an increasingly fine mesh of constraints that can be modified on-the-fly. This method, filter SBI, allows us to infer entire families of complex and co-active plasticity rules in spiking networks. We first consider flexibly parameterized doublet (Hebbian) rules, and find that the set of inferred rules contains solutions that extend and refine -and also reject- predictions from mean-field theory. Next, we expand the search space of plasticity rules by modelling them as multi-layer perceptrons that combine several plasticity-relevant factors, such as weight, voltage, triplets and co-dependency. Out of the millions of possible rules, we identify

thousands of unique rule combinations that satisfy biological constraints like p
lausible activity and weight dynamics. The resulting rules can be used as a star
ting point for further investigations into specific network computations, and al
ready suggest refinements and predictions for classical experimental approaches
on plasticity. This flexible approach for principled exploration of complex plas
ticity rules in large recurrent spiking networks presents the most advanced sear
ch tool to date for enabling robust predictions and deep insights into the plast
icity mechanisms underlying brain function.
****
Joint Training of Deep Ensembles Fails Due to Learner Collusion
Alan Jeffares, Tennison Liu, Jonathan Crabbé, Mihaela van der Schaar
Ensembles of machine learning models have been well established as a powerful me
thod of improving performance over a single model. Traditionally, ensembling alg
orithms train their base learners independently or sequentially with the goal of
 optimizing their joint performance. In the case of deep ensembles of neural net
works, we are provided with the opportunity to directly optimize the true object
ive: the joint performance of the ensemble as a whole. Surprisingly, however, di
rectly minimizing the loss of the ensemble appears to rarely be applied in pract
ice. Instead, most previous research trains individual models independently with
 ensembling performed post hoc. In this work, we show that this is for good reas
on - joint optimization of ensemble loss results in degenerate behavior. We appr
oach this problem by decomposing the ensemble objective into the strength of the
 base learners and the diversity between them. We discover that joint optimizati
on results in a phenomenon in which base learners collude to artificially inflat
e their apparent diversity. This pseudo-diversity fails to generalize beyond the
 training data, causing a larger generalization gap. We proceed to comprehensive
ly demonstrate the practical implications of this effect on a range of standard
machine learning tasks and architectures by smoothly interpolating between indep
endent training and joint optimization.
****
Flexible Attention-Based Multi-Policy Fusion for Efficient Deep Reinforcement Le
arning
Zih-Yun Chiu, Yi-Lin Tuan, William Yang Wang, Michael Yip
Reinforcement learning (RL) agents have long sought to approach the efficiency o
f human learning. Humans are great observers who can learn by aggregating extern
al knowledge from various sources, including observations from others' policies
of attempting a task. Prior studies in RL have incorporated external knowledge p
olicies to help agents improve sample efficiency. However, it remains non-trivia
l to perform arbitrary combinations and replacements of those policies, an essen
tial feature for generalization and transferability. In this work, we present Kn
owledge-Grounded RL (KGRL), an RL paradigm fusing multiple knowledge policies an
d aiming for human-like efficiency and flexibility. We propose a new actor archi
tecture for KGRL, Knowledge-Inclusive Attention Network (KIAN), which allows fre
e knowledge rearrangement due to embedding-based attentive action prediction. KI
AN also addresses entropy imbalance, a problem arising in maximum entropy KGRL t
hat hinders an agent from efficiently exploring the environment, through a new d
esign of policy distributions. The experimental results demonstrate that KIAN ou
tperforms alternative methods incorporating external knowledge policies and achi
eves efficient and flexible learning. Our implementation is available at https:/
/github.com/Pascalson/KGRL.git .
****
Balanced Training for Sparse GANs
Yite Wang, Jing Wu, NAIRA HOVAKIMYAN, Ruoyu Sun
Over the past few years, there has been growing interest in developing larger an
d deeper neural networks, including deep generative models like generative adver
sarial networks (GANs). However, GANs typically come with high computational com
plexity, leading researchers to explore methods for reducing the training and i
nference costs. One such approach gaining popularity in supervised learning is
dynamic sparse training (DST), which maintains good performance while enjoying e
xcellent training efficiency. Despite its potential benefits, applying DST to G

ANs presents challenges due to the adversarial nature of the training process. I
n this paper, we propose a novel metric called the balance ratio (BR) to study t
he balance between the sparse generator and discriminator. We also introduce a n
ew method called balanced dynamic sparse training (ADAPT), which seeks to contro
l the BR during GAN training to achieve a good trade-off between performance and
 computational cost. Our proposed method shows promising results on multiple dat
asets, demonstrating its effectiveness.
****
Policy Optimization for Continuous Reinforcement Learning
HANYANG ZHAO, Wenpin Tang, David Yao
We study reinforcement learning (RL) in the setting of continuous time and space
, for an infinite horizon with a discounted objective and the underlying dynamic
s driven by a stochastic differential equation. Built upon recent advances in th
e continuous approach to RL, we develop a notion of occupation time (specificall
y for a discounted objective), and show how it can be effectively used to deriv
e performance difference and local approximation formulas. We further extend the
se results to illustrate their applications in the PG (policy gradient) and TRPO
/PPO (trust region policy optimization/ proximal policy optimization) methods,
which have been familiar and powerful tools in the discrete RL setting but under
-developed in continuous RL. Through numerical experiments, we demonstrate the e
ffectiveness and advantages of our approach.
****
PrimDiffusion: Volumetric Primitives Diffusion for 3D Human Generation
Zhaoxi Chen, Fangzhou Hong, Haiyi Mei, Guangcong Wang, Lei Yang, Ziwei Liu
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
A Closer Look at the Robustness of Contrastive Language-Image Pre-Training (CLIP
)
Weijie Tu, Weijian Deng, Tom Gedeon
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Model Spider: Learning to Rank Pre-Trained Models Efficiently
Yi-Kai Zhang, Ting-Ji Huang, Yao-Xiang Ding, De-Chuan Zhan, Han-Jia Ye
Figuring out which Pre-Trained Model (PTM) from a model zoo fits the target task
 is essential to take advantage of plentiful model resources. With the availabil
ity of numerous heterogeneous PTMs from diverse fields, efficiently selecting th
e most suitable one is challenging due to the time-consuming costs of carrying o
ut forward or backward passes over all PTMs. In this paper, we propose Model Spi
der, which tokenizes both PTMs and tasks by summarizing their characteristics in
to vectors to enable efficient PTM selection. By leveraging the approximated per
formance of PTMs on a separate set of training tasks, Model Spider learns to con
struct representation and measure the fitness score between a model-task pair vi
a their representation. The ability to rank relevant PTMs higher than others gen
eralizes to new tasks. With the top-ranked PTM candidates, we further learn to e
nrich task repr. with their PTM-specific semantics to re-rank the PTMs for bette
r selection. Model Spider balances efficiency and selection ability, making PTM
selection like a spider preying on a web. Model Spider exhibits promising perfor
mance across diverse model zoos, including visual models and Large Language Mode
ls (LLMs). Code is available at https://github.com/zhangyikaii/Model-Spider.
****
Investigating how ReLU-networks encode symmetries
Georg Bökman, Fredrik Kahl
Many data symmetries can be described in terms of group equivariance and the mos
t common way of encoding group equivariances in neural networks is by building l

inear layers that are group equivariant.In this work we investigate whether equivariance of a network implies that all layers are equivariant.On the theoretical side we find cases where equivariance implies layerwise equivariance, but alsod emonstrate that this is not the case generally.Nevertheless, we conjecture that CNNs that are trained to be equivariant will exhibit layerwise equivariance and explain how this conjecture is a weaker version of the recent permutation conjecture by Entezari et al.\ [2022].We perform quantitative experiments with VGG-nets on CIFAR10 and qualitative experiments with ResNets on ImageNet to illustrate and support our theoretical findings. These experiments are not only of interest for understanding how group equivariance is encoded in ReLU-networks, but they also give a new perspective on Entezari et al.'s permutation conjecture as we find that itis typically easier to merge a network with a group-transformed version of itself than merging two different networks.
****
Optimal and Fair Encouragement Policy Evaluation and Learning
Angela Zhou
In consequential domains, it is often impossible to compel individuals to take treatment, so that optimal policy rules are merely suggestions in the presence of human non-adherence to treatment recommendations. In these same domains, there may be heterogeneity both in who responds in taking-up treatment, and heterogeneity in treatment efficacy. For example, in social services, a persistent puzzle is the gap in take-up of beneficial services among those who may benefit from them the most. When in addition the decision-maker has distributional preferences over both access and average outcomes, the optimal decision rule changes. We study identification, doubly-robust estimation, and robust estimation under potential violations of positivity. We consider fairness constraints such as demographic parity in treatment take-up, and other constraints, via constrained optimization. Our framework can be extended to handle algorithmic recommendations under an often-reasonable covariate-conditional exclusion restriction, using our robustness checks for lack of positivity in the recommendation. We develop a two-stage, online learning-based algorithm for solving over parametrized policy classes under general constraints to obtain variance-sensitive regret bounds. We assess improved recommendation rules in a stylized case study of optimizing recommendation of supervised release in the PSA-DMF pretrial risk-assessment tool while reducing surveillance disparities.
****
NICE: NoIse-modulated Consistency rEgularization for Data-Efficient GANs
Yao Ni, Piotr Koniusz
Generative Adversarial Networks (GANs) are powerful tools for image synthesis. However, they require access to vast amounts of training data, which is often costly and prohibitive. Limited data affects GANs, leading to discriminator overfitting and training instability. In this paper, we present a novel approach called NoIse-modulated Consistency rEgularization (NICE) to overcome these challenges. To this end, we introduce an adaptive multiplicative noise into the discriminator to modulate its latent features. We demonstrate the effectiveness of such a modulation in preventing discriminator overfitting by adaptively reducing the Rademacher complexity of the discriminator. However, this modulation leads to an unintended consequence of increased gradient norm, which can undermine the stability of GAN training. To mitigate this undesirable effect, we impose a constraint on the discriminator, ensuring its consistency for the same inputs under different noise modulations. The constraint effectively penalizes the first and second-order gradients of latent features, enhancing GAN stability. Experimental evidence aligns with our theoretical analysis, demonstrating the reduction of generalization error and gradient penalization of NICE. This substantiates the efficacy of NICE in reducing discriminator overfitting and improving stability of GAN training. NICE achieves state-of-the-art results on CIFAR-10, CIFAR-100, ImageNet and FFHQ datasets when trained with limited data, as well as in low-shot generation tasks.
****
Not All Out-of-Distribution Data Are Harmful to Open-Set Active Learning

Yang Yang, Yuxuan Zhang, XIN SONG, Yi Xu
Active learning (AL) methods have been proven to be an effective way to reduce the labeling effort by intelligently selecting valuable instances for annotation. Despite their great success with in-distribution (ID) scenarios, AL methods suffer from performance degradation in many real-world applications because out-of-distribution (OOD) instances are always inevitably contained in unlabeled data, which may lead to inefficient sampling. Therefore, several attempts have been explored open-set AL by strategically selecting pure ID instances while filtering OOD instances. However, concentrating solely on selecting pseudo-ID instances may cause the training constraint of the ID classifier and OOD detector. To address this issue, we propose a simple yet effective sampling scheme, Progressive Active Learning (PAL), which employs a progressive sampling mechanism to leverage the active selection of valuable OOD instances. The proposed PAL measures unlabeled instances by synergistically evaluating instances' informativeness and representativeness, and thus it can balance the pseudo-ID and pseudo-OOD instances in each round to enhance both the capacity of the ID classifier and the OOD detector. %Meanwhile, PAL measures unlabeled instances by synergistically evaluating instances' informativeness and representativeness, which can more effectively estimate the values of instances. Extensive experiments on various open-set AL scenarios demonstrate the effectiveness of the proposed PAL, compared with the state-of-the-art methods. The code is available at \url{https://github.com/njustkmg/PAL}.
****
Improving the Privacy and Practicality of Objective Perturbation for Differentially Private Linear Learners
Rachel Redberg, Antti Koskela, Yu-Xiang Wang
In the arena of privacy-preserving machine learning, differentially private stochastic gradient descent (DP-SGD) has outstripped the objective perturbation mechanism in popularity and interest. Though unrivaled in versatility, DP-SGD requires a non-trivial privacy overhead (for privately tuning the model’s hyperparameters) and a computational complexity which might be extravagant for simple models such as linear and logistic regression. This paper revamps the objective perturbation mechanism with tighter privacy analyses and new computational tools that boost it to perform competitively with DP-SGD on unconstrained convex generalized linear problems.
****
Implicit Differentiable Outlier Detection Enable Robust Deep Multimodal Analysis
Zhu Wang, Sourav Medya, Sathya Ravi
Deep network models are often purely inductive during both training and inference on unseen data. When these models are used for prediction, but they may fail to capture important semantic information and implicit dependencies within datasets. Recent advancements have shown that combining multiple modalities in large-scale vision and language settings can improve understanding and generalization performance. However, as the model size increases, fine-tuning and deployment become computationally expensive, even for a small number of downstream tasks. Moreover, it is still unclear how domain or prior modal knowledge can be specified in a backpropagation friendly manner, especially in large-scale and noisy settings. To address these challenges, we propose a simplified alternative of combining features from pretrained deep networks and freely available semantic explicit knowledge. In order to remove irrelevant explicit knowledge that does not correspond well to the images, we introduce an implicit Differentiable Out-of-Distribution (OOD) detection layer. This layer addresses outlier detection by solving for fixed points of a differentiable function and using the last iterate of fixed point solver to backpropagate. In practice, we apply our model on several vision and language downstream tasks including visual question answering, visual reasoning, and image-text retrieval on different datasets. Our experiments show that it is possible to design models that perform similarly to state-of-the-art results but with significantly fewer samples and less training time. Our models and code are available here: https://github.com/ellenzhuwang/implicit_vkood
****

DSR: Dynamical Surface Representation as Implicit Neural Networks for Protein
Daiwen Sun, He Huang, Yao Li, Xinqi Gong, Qiwei Ye
We propose a novel neural network-based approach to modeling protein dynamics using an implicit representation of a protein's surface in 3D and time. Our method utilizes the zero-level set of signed distance functions (SDFs) to represent protein surfaces, enabling temporally and spatially continuous representations of protein dynamics. Our experimental results demonstrate that our model accurately captures protein dynamic trajectories and can interpolate and extrapolate in 3D and time. Importantly, this is the first study to introduce this method and successfully model large-scale protein dynamics. This approach offers a promising alternative to current methods, overcoming the limitations of first-principles-based and deep learning methods, and provides a more scalable and efficient approach to modeling protein dynamics. Additionally, our surface representation approach simplifies calculations and allows identifying movement trends and amplitudes of protein domains, making it a useful tool for protein dynamics research. Codes are available at https://github.com/Sundw-818/DSR, and we have a project webpage that shows some video results, https://sundw-818.github.io/DSR/.
****
A Theory of Transfer-Based Black-Box Attacks: Explanation and Implications
Yanbo Chen, Weiwei Liu
Transfer-based attacks are a practical method of black-box adversarial attacks, in which the attacker aims to craft adversarial examples from a source (surrogate) model that is transferable to the target model. A wide range of empirical works has tried to explain the transferability of adversarial examples from different angles. However, these works only provide ad hoc explanations without quantitative analyses. The theory behind transfer-based attacks remains a mystery.This paper studies transfer-based attacks under a unified theoretical framework. We propose an explanatory model, called the manifold attack model, that formalizes popular beliefs and explains the existing empirical results. Our model explains why adversarial examples are transferable even when the source model is inaccurate. Moreover, our model implies that the existence of transferable adversarial examples depends on the “curvature” of the data manifold, which quantitatively explains why the success rates of transfer-based attacks are hard to improve. We also discuss the expressive power and the possible extensions of our model in general applications.
****
Explaining V1 Properties with a Biologically Constrained Deep Learning Architecture
Galen Pogoncheff, Jacob Granley, Michael Beyeler
Convolutional neural networks (CNNs) have recently emerged as promising models of the ventral visual stream, despite their lack of biological specificity.While current state-of-the-art models of the primary visual cortex (V1) have surfaced from training with adversarial examples and extensively augmented data, these models are still unable to explain key neural properties observed in V1 that arise from biological circuitry.To address this gap, we systematically incorporated neuroscience-derived architectural components into CNNs to identify a set of mechanisms and architectures that more comprehensively explain V1 activity.Upon enhancing task-driven CNNs with architectural components that simulate center-surround antagonism, local receptive fields, tuned normalization, and cortical magnification, we uncover models with latent representations that yield state-of-the-art explanation of V1 neural activity and tuning properties.Moreover, analyses of the learned parameters of these components and stimuli that maximally activate neurons of the evaluated networks provide support for their role in explaining neural properties of V1.Our results highlight an important advancement in the field of NeuroAI, as we systematically establish a set of architectural components that contribute to unprecedented explanation of V1.The neuroscience insights that could be gleaned from increasingly accurate in-silico models of the brain have the potential to greatly advance the fields of both neuroscience and artificial intelligence.
****

Revisiting Adversarial Training for ImageNet: Architectures, Training and Genera lization across Threat Models
Naman Deep Singh, Francesco Croce, Matthias Hein
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
URL: A Representation Learning Benchmark for Transferable Uncertainty Estimates
Michael Kirchhof, Bálint Mucsányi, Seong Joon Oh, Dr. Enkelejda Kasneci
Representation learning has significantly driven the field to develop pretrained models that can act as a valuable starting point when transferring to new datas ets. With the rising demand for reliable machine learning and uncertainty quanti fication, there is a need for pretrained models that not only provide embeddings but also transferable uncertainty estimates. To guide the development of such m odels, we propose the Uncertainty-aware Representation Learning (URL) benchmark. Besides the transferability of the representations, it also measures the zero-s hot transferability of the uncertainty estimate using a novel metric. We apply U RL to evaluate ten uncertainty quantifiers that are pretrained on ImageNet and t ransferred to eight downstream datasets. We find that approaches that focus on t he uncertainty of the representation itself or estimate the prediction risk dire ctly outperform those that are based on the probabilities of upstream classes. Y et, achieving transferable uncertainty quantification remains an open challenge. Our findings indicate that it is not necessarily in conflict with traditional r epresentation learning goals. Code is available at https://github.com/mkirchhof/ url.
****
FineMoGen: Fine-Grained Spatio-Temporal Motion Generation and Editing
Mingyuan Zhang, Huirong Li, Zhongang Cai, Jiawei Ren, Lei Yang, Ziwei Liu
Text-driven motion generation has achieved substantial progress with the emergen ce of diffusion models. However, existing methods still struggle to generate com plex motion sequences that correspond to fine-grained descriptions, depicting de tailed and accurate spatio-temporal actions.This lack of fine controllability li mits the usage of motion generation to a larger audience. To tackle these challe nges, we present FineMoGen, a diffusion-based motion generation and editing fram ework that can synthesize fine-grained motions, with spatial-temporal compositio n to the user instructions. Specifically, FineMoGen builds upon diffusion model with a novel transformer architecture dubbed Spatio-Temporal Mixture Attention S AMI. SAMI optimizes the generation of the global attention template from two per spectives: 1) explicitly modeling the constraints of spatio-temporal composition ; and 2) utilizing sparsely-activated mixture-of-experts to adaptively extract f ine-grained features. To facilitate a large-scale study on this new fine-grained motion generation task, we contribute the HuMMan-MoGen dataset, which consists of 2,968 videos and 102,336 fine-grained spatio-temporal descriptions. Extensive experiments validate that FineMoGen exhibits superior motion generation qualit y over state-of-the-art methods. Notably, FineMoGen further enables zero-shot mo tion editing capabilities with the aid of modern large language models (LLM), wh ich faithfully manipulates motion sequences with fine-grained instructions.
****
StEik: Stabilizing the Optimization of Neural Signed Distance Functions and Fine r Shape Representation
Huizong Yang, Yuxin Sun, Ganesh Sundaramoorthi, Anthony Yezzi
We present new insights and a novel paradigm for learning implicit neural repres entations (INR) of shapes. In particular, we shed light on the popular eikonal l oss used for imposing a signed distance function constraint in INR. We show anal ytically that as the representation power of the network increases, the optimiza tion approaches a partial differential equation (PDE) in the continuum limit tha t is unstable. We show that this instability can manifest in existing network op timization, leading to irregularities in the reconstructed surface and/or conver gence to sub-optimal local minima, and thus fails to capture fine geometric and

topological structure. We show analytically how other terms added to the loss, currently used in the literature for other purposes, can actually eliminate these instabilities. However, such terms can over-regularize the surface, preventing the representation of fine shape detail. Based on a similar PDE theory for the continuum limit, we introduce a new regularization term that still counteracts the eikonal instability but without over-regularizing. Furthermore, since stability is now guaranteed in the continuum limit, this stabilization also allows for considering new network structures that are able to represent finer shape detail. We introduce such a structure based on quadratic layers. Experiments on multiple benchmark data sets show that our new regularization and network are able to capture more precise shape details and more accurate topology than existing state-of-the-art.
****
Voicebox: Text-Guided Multilingual Universal Speech Generation at Scale
Matthew Le, Apoorv Vyas, Bowen Shi, Brian Karrer, Leda Sari, Rashel Moritz, Mary Williamson, Vimal Manohar, Yossi Adi, Jay Mahadeokar, Wei-Ning Hsu
Large-scale generative models such as GPT and DALL-E have revolutionized the research community. These models not only generate high fidelity outputs, but are also generalists which can solve tasks not explicitly taught. In contrast, speech generative models are still primitive in terms of scale and task generalization. In this paper, we present Voicebox, the most versatile text-guided generative model for speech at scale. Voicebox is a non-autoregressive flow-matching model trained to infill speech, given audio context and text, trained on over 50K hours of speech that are not filtered or enhanced. Similar to GPT, Voicebox can perform many different tasks through in-context learning, but is more flexible as it can also condition on future context. Voicebox can be used for mono or cross-lingual zero-shot text-to-speech synthesis, noise removal, content editing, style conversion, and diverse sample generation. In particular, Voicebox outperforms the state-of-the-art zero-shot TTS model VALL-E on both intelligibility (5.9\% vs 1.9\% word error rates) and audio similarity (0.580 vs 0.681) while being up to 20 times faster. Audio samples can be found in \url{https://voicebox.metademolab.com}.
****
Optimizing Solution-Samplers for Combinatorial Problems: The Landscape of Policy-Gradient Method
Constantine Caramanis, Dimitris Fotakis, Alkis Kalavasis, Vasilis Kontonis, Christos Tzamos
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SoTTA: Robust Test-Time Adaptation on Noisy Data Streams
Taesik Gong, Yewon Kim, Taeckyung Lee, Sorn Chottananurak, Sung-Ju Lee
Test-time adaptation (TTA) aims to address distributional shifts between training and testing data using only unlabeled test data streams for continual model adaptation. However, most TTA methods assume benign test streams, while test samples could be unexpectedly diverse in the wild. For instance, an unseen object or noise could appear in autonomous driving. This leads to a new threat to existing TTA algorithms; we found that prior TTA algorithms suffer from those noisy test samples as they blindly adapt to incoming samples. To address this problem, we present Screening-out Test-Time Adaptation (SoTTA), a novel TTA algorithm that is robust to noisy samples. The key enabler of SoTTA is two-fold: (i) input-wise robustness via high-confidence uniform-class sampling that effectively filters out the impact of noisy samples and (ii) parameter-wise robustness via entropy-sharpness minimization that improves the robustness of model parameters against large gradients from noisy samples. Our evaluation with standard TTA benchmarks with various noisy scenarios shows that our method outperforms state-of-the-art TTA methods under the presence of noisy samples and achieves comparable accuracy to those methods without noisy samples. The source code is available at https://g

ithub.com/taeckyung/SoTTA.
****
FouriDown: Factoring Down-Sampling into Shuffling and Superposing
Qi Zhu, man zhou, Jie Huang, Naishan Zheng, Hongzhi Gao, Chongyi Li, Yuan Xu, Feng Zhao
Spatial down-sampling techniques, such as strided convolution, Gaussian, and Nearest down-sampling, are essential in deep neural networks. In this study, we revisit the working mechanism of the spatial down-sampling family and analyze the biased effects caused by the static weighting strategy employed in previous approaches. To overcome this limitation, we propose a novel down-sampling paradigm in the Fourier domain, abbreviated as FouriDown, which unifies existing down-sampling techniques. Drawing inspiration from the signal sampling theorem, we parameterize the non-parameter static weighting down-sampling operator as a learnable and context-adaptive operator within a unified Fourier function. Specifically, we organize the corresponding frequency positions of the 2D plane in a physically-closed manner within a single channel dimension. We then perform point-wise channel shuffling based on an indicator that determines whether a channel's signal frequency bin is susceptible to aliasing, ensuring the consistency of the weighting parameter learning. FouriDown, as a generic operator, comprises four key components: 2D discrete Fourier transform, context shuffling rules, Fourier weighting-adaptively superposing rules, and 2D inverse Fourier transform. These components can be easily integrated into existing image restoration networks. To demonstrate the efficacy of FouriDown, we conduct extensive experiments on image de-blurring and low-light image enhancement. The results consistently show that FouriDown can provide significant performance improvements. We will make the code publicly available to facilitate further exploration and application of FouriDown.
****
Participatory Personalization in Classification
Hailey Joren, Chirag Nagpal, Katherine A. Heller, Berk Ustun
Machine learning models are often personalized based on information that is protected, sensitive, self-reported, or costly to acquire. These models use information about people, but do not facilitate nor inform their consent. Individuals cannot opt out of reporting information that a model needs to personalize their predictions nor tell if they benefit from personalization in the first place. We introduce a new family of prediction models, called participatory systems, that let individuals opt into personalization at prediction time. We present a model-agnostic algorithm to learn participatory systems for supervised learning tasks where models are personalized with categorical group attributes. We conduct a comprehensive empirical study of participatory systems in clinical prediction tasks, comparing them to common approaches for personalization and imputation. Our results show that participatory systems can facilitate and inform consent in a way that improves performance and privacy across all groups who report personal data.
****
A Neural Collapse Perspective on Feature Evolution in Graph Neural Networks
Vignesh Kothapalli, Tom Tirer, Joan Bruna
Graph neural networks (GNNs) have become increasingly popular for classification tasks on graph-structured data. Yet, the interplay between graph topology and feature evolution in GNNs is not well understood. In this paper, we focus on node-wise classification, illustrated with community detection on stochastic block model graphs, and explore the feature evolution through the lens of the "Neural Collapse" (NC) phenomenon. When training instance-wise deep classifiers (e.g. for image classification) beyond the zero training error point, NC demonstrates a reduction in the deepest features' within-class variability and an increased alignment of their class means to certain symmetric structures. We start with an empirical study that shows that a decrease in within-class variability is also prevalent in the node-wise classification setting, however, not to the extent observed in the instance-wise case. Then, we theoretically study this distinction. Specifically, we show that even an "optimistic" mathematical model requires that the graphs obey a strict structural condition in order to possess a minimizer with

exact collapse. Furthermore, by studying the gradient dynamics of this model, we provide reasoning for the partial collapse observed empirically. Finally, we present a study on the evolution of within- and between-class feature variability across layers of a well-trained GNN and contrast the behavior with spectral methods.
****
ResoNet: Noise-Trained Physics-Informed MRI Off-Resonance Correction
Alfredo De Goyeneche Macaya, Shreya Ramachandran, Ke Wang, Ekin Karasan, Joseph Y. Cheng, Stella X. Yu, Michael Lustig
Magnetic Resonance Imaging (MRI) is a powerful medical imaging modality that offers diagnostic information without harmful ionizing radiation. Unlike optical imaging, MRI sequentially samples the spatial Fourier domain (k-space) of the image. Measurements are collected in multiple shots, or readouts, and in each shot, data along a smooth trajectory is sampled.Conventional MRI data acquisition relies on sampling k-space row-by-row in short intervals, which is slow and inefficient. More efficient, non-Cartesian sampling trajectories (e.g., Spirals) use longer data readout intervals, but are more susceptible to magnetic field inhomogeneities, leading to off-resonance artifacts. Spiral trajectories cause off-resonance blurring in the image, and the mathematics of this blurring resembles that of optical blurring, where magnetic field variation corresponds to depth and readout duration to aperture size. Off-resonance blurring is a system issue with a physics-based, accurate forward model. We present a physics-informed deep learning framework for off-resonance correction in MRI, which is trained exclusively on synthetic, noise-like data with representative marginal statistics. Our approach allows for fat/water separation and is compatible with parallel imaging acceleration. Through end-to-end training using synthetic randomized data (i.e., noise-like images, coil sensitivities, field maps), we train the network to reverse off-resonance effects across diverse anatomies and contrasts without retraining. We demonstrate the effectiveness of our approach through results on phantom and in-vivo data. This work has the potential to facilitate the clinical adoption of non-Cartesian sampling trajectories, enabling efficient, rapid, and motion-robust MRI scans. Code is publicly available at: https://github.com/mikgroup/ResoNet.
****
Eliminating Domain Bias for Federated Learning in Representation Space
Jianqing Zhang, Yang Hua, Jian Cao, Hao Wang, Tao Song, Zhengui XUE, Ruhui Ma, Haibing Guan
Recently, federated learning (FL) is popular for its privacy-preserving and collaborative learning abilities. However, under statistically heterogeneous scenarios, we observe that biased data domains on clients cause a representation bias phenomenon and further degenerate generic representations during local training, i.e., the representation degeneration phenomenon. To address these issues, we propose a general framework Domain Bias Eliminator (DBE) for FL. Our theoretical analysis reveals that DBE can promote bi-directional knowledge transfer between server and client, as it reduces the domain discrepancy between server and client in representation space. Besides, extensive experiments on four datasets show that DBE can greatly improve existing FL methods in both generalization and personalization abilities. The DBE-equipped FL method can outperform ten state-of-the-art personalized FL methods by a large margin. Our code is public at https://github.com/TsingZ0/DBE.
****
Pretraining task diversity and the emergence of non-Bayesian in-context learning for regression
Allan Raventós, Mansheej Paul, Feng Chen, Surya Ganguli
Pretrained transformers exhibit the remarkable ability of in-context learning (ICL): they can learn tasks from just a few examples provided in the prompt without updating any weights. This raises a foundational question: can ICL solve fundamentally new tasks that are very different from those seen during pretraining? To probe this question, we examine ICL’s performance on linear regression while varying the diversity of tasks in the pretraining dataset. We empirically demonst

rate a task diversity threshold for the emergence of ICL. Below this threshold, the pretrained transformer cannot solve unseen regression tasks, instead behaving like a Bayesian estimator with the non-diverse pretraining task distribution as the prior. Beyond this threshold, the transformer significantly outperforms this estimator; its behavior aligns with that of ridge regression, corresponding to a Gaussian prior over all tasks, including those not seen during pretraining. Thus, when pretrained on data with task diversity greater than the threshold, transformers can optimally solve fundamentally new tasks in-context. Importantly, this capability hinges on it deviating from the Bayes optimal estimator with the pretraining distribution as the prior. This study also explores the effect of regularization, model capacity and task structure and underscores, in a concrete example, the critical role of task diversity, alongside data and model scale, in the emergence of ICL.
****

Two-Stage Predict+Optimize for MILPs with Unknown Parameters in Constraints
Xinyi Hu, Jasper Lee, Jimmy Lee
Consider the setting of constrained optimization, with some parameters unknown at solving time and requiring prediction from relevant features. Predict+Optimize is a recent framework for end-to-end training supervised learning models for such predictions, incorporating information about the optimization problem in the training process in order to yield better predictions in terms of the quality of the predicted solution under the true parameters. Almost all prior works have focused on the special case where the unknowns appear only in the optimization objective and not the constraints. Hu et al. proposed the first adaptation of Predict+Optimize to handle unknowns appearing in constraints, but the framework has somewhat ad-hoc elements, and they provided a training algorithm only for covering and packing linear programs. In this work, we give a new simpler and more powerful framework called Two-Stage Predict+Optimize, which we believe should be the canonical framework for the Predict+Optimize setting. We also give a training algorithm usable for all mixed integer linear programs, vastly generalizing the applicability of the framework. Experimental results demonstrate the superior prediction performance of our training framework over all classical and state-of-the-art methods.
****

Adaptive Normalization for Non-stationary Time Series Forecasting: A Temporal Slice Perspective
Zhiding Liu, Mingyue Cheng, Zhi Li, Zhenya Huang, Qi Liu, Yanhu Xie, Enhong Chen
Deep learning models have progressively advanced time series forecasting due to their powerful capacity in capturing sequence dependence. Nevertheless, it is still challenging to make accurate predictions due to the existence of non-stationarity in real-world data, denoting the data distribution rapidly changes over time. To mitigate such a dilemma, several efforts have been conducted by reducing the non-stationarity with normalization operation. However, these methods typically overlook the distribution discrepancy between the input series and the horizon series, and assume that all time points within the same instance share the same statistical properties, which is too ideal and may lead to suboptimal relative improvements. To this end, we propose a novel slice-level adaptive normalization, referred to \textbf{SAN}, which is a novel scheme for empowering time series forecasting with more flexible normalization and denormalization. SAN includes two crucial designs. First, SAN tries to eliminate the non-stationarity of time series in units of a local temporal slice (i.e., sub-series) rather than a global instance. Second, SAN employs a slight network module to independently model the evolving trends of statistical properties of raw time series. Consequently, SAN could serve as a general model-agnostic plugin and better alleviate the impact of the non-stationary nature of time series data. We instantiate the proposed SAN on four widely used forecasting models and test their prediction results on benchmark datasets to evaluate its effectiveness. Also, we report some insightful findings to deeply analyze and understand our proposed SAN. We make our codes publicly available.
****

Distance-Restricted Folklore Weisfeiler-Leman GNNs with Provable Cycle Counting Power
Junru Zhou, Jiarui Feng, Xiyuan Wang, Muhan Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Computing a human-like reaction time metric from stable recurrent vision models
Lore Goetschalckx, Lakshmi Narasimhan Govindarajan, Alekh Karkada Ashok, Aarit Ahuja, David Sheinberg, Thomas Serre
The meteoric rise in the adoption of deep neural networks as computational models of vision has inspired efforts to ``align" these models with humans. One dimension of interest for alignment includes behavioral choices, but moving beyond characterizing choice patterns to capturing temporal aspects of visual decision-making has been challenging. Here, we sketch a general-purpose methodology to construct computational accounts of reaction times from a stimulus-computable, task-optimized model. Specifically, we introduce a novel metric leveraging insights from subjective logic theory summarizing evidence accumulation in recurrent vision models. We demonstrate that our metric aligns with patterns of human reaction times for stimulus manipulations across four disparate visual decision-making tasks spanning perceptual grouping, mental simulation, and scene categorization. This work paves the way for exploring the temporal alignment of model and human visual strategies in the context of various other cognitive tasks toward generating testable hypotheses for neuroscience. Links to the code and data can be found on the project page: https://serre-lab.github.io/rnnrtssite/.
****
ReHLine: Regularized Composite ReLU-ReHU Loss Minimization with Linear Computation and Linear Convergence
Ben Dai, Yixuan Qiu
Empirical risk minimization (ERM) is a crucial framework that offers a general approach to handling a broad range of machine learning tasks. In this paper, we propose a novel algorithm, called ReHLine, for minimizing a set of regularized ERMs with convex piecewise linear-quadratic loss functions and optional linear constraints. The proposed algorithm can effectively handle diverse combinations of loss functions, regularization, and constraints, making it particularly well-suited for complex domain-specific problems. Examples of such problems include FairSVM, elastic net regularized quantile regression, Huber minimization, etc. In addition, ReHLine enjoys a provable linear convergence rate and exhibits a per-iteration computational complexity that scales linearly with the sample size. The algorithm is implemented with both Python and R interfaces, and its performance is benchmarked on various tasks and datasets. Our experimental results demonstrate that ReHLine significantly surpasses generic optimization solvers in terms of computational efficiency on large-scale datasets. Moreover, it also outperforms specialized solvers such as Liblinear in SVMs, hqreg in Huber minimization, and Lightning (SAGA, SAG, SDCA, SVRG) in smoothed SVMs, exhibiting exceptional flexibility and efficiency. The source code, project page, accompanying software, and the Python/R interface can be accessed through the link: https://github.com/softmin/ReHLine.
****
Improved Frequency Estimation Algorithms with and without Predictions
Anders Aamand, Justin Chen, Huy Nguyen, Sandeep Silwal, Ali Vakilian
Estimating frequencies of elements appearing in a data stream is a key task in large-scale data analysis. Popular sketching approaches to this problem (e.g., CountMin and CountSketch) come with worst-case guarantees that probabilistically bound the error of the estimated frequencies for any possible input. The work of Hsu et al.~(2019) introduced the idea of using machine learning to tailor sketching algorithms to the specific data distribution they are being run on. In particular, their learning-augmented frequency estimation algorithm uses a learned heavy-hitter oracle which predicts which elements will appear many times in the st

ream. We give a novel algorithm, which in some parameter regimes, already theoretically outperforms the learning based algorithm of Hsu et al. without the use of any predictions. Augmenting our algorithm with heavy-hitter predictions further reduces the error and improves upon the state of the art. Empirically, our algorithms achieve superior performance in all experiments compared to prior approaches.
****
BCDiff: Bidirectional Consistent Diffusion for Instantaneous Trajectory Prediction
Rongqing Li, Changsheng Li, Dongchun Ren, Guangyi Chen, Ye Yuan, Guoren Wang
The objective of pedestrian trajectory prediction is to estimate the future paths of pedestrians by leveraging historical observations, which plays a vital role in ensuring the safety of self-driving vehicles and navigation robots. Previous works usually rely on a sufficient amount of observation time to accurately predict future trajectories. However, there are many real-world situations where the model lacks sufficient time to observe, such as when pedestrians abruptly emerge from blind spots, resulting in inaccurate predictions and even safety risks. Therefore, it is necessary to perform trajectory prediction based on instantaneous observations, which has rarely been studied before. In this paper, we propose a Bi-directional Consistent Diffusion framework tailored for instantaneous trajectory prediction, named BCDiff. At its heart, we develop two coupled diffusion models by designing a mutual guidance mechanism which can bidirectionally and consistently generate unobserved historical trajectories and future trajectories step-by-step, to utilize the complementary information between them. Specifically, at each step, the predicted unobserved historical trajectories and limited observed trajectories guide one diffusion model to generate future trajectories, while the predicted future trajectories and observed trajectories guide the other diffusion model to predict unobserved historical trajectories. Given the presence of relatively high noise in the generated trajectories during the initial steps, we introduce a gating mechanism to learn the weights between the predicted trajectories and the limited observed trajectories for automatically balancing their contributions. By means of this iterative and mutually guided generation process, both the future and unobserved historical trajectories undergo continuous refinement, ultimately leading to accurate predictions. Essentially, BCDiff is an encoder-free framework that can be compatible with existing trajectory prediction models in principle. Experiments show that our proposed BCDiff significantly improves the accuracy of instantaneous trajectory prediction on the ETH/UCY and Stanford Drone datasets, compared to related approaches.
****
Leave No Stone Unturned: Mine Extra Knowledge for Imbalanced Facial Expression Recognition
Yuhang Zhang, Yaqi Li, lixiong Qin, Xuannan Liu, Weihong Deng
Facial expression data is characterized by a significant imbalance, with most collected data showing happy or neutral expressions and fewer instances of fear or disgust. This imbalance poses challenges to facial expression recognition (FER) models, hindering their ability to fully understand various human emotional states. Existing FER methods typically report overall accuracy on highly imbalanced test sets but exhibit low performance in terms of the mean accuracy across all expression classes. In this paper, our aim is to address the imbalanced FER problem. Existing methods primarily focus on learning knowledge of minor classes solely from minor-class samples. However, we propose a novel approach to extract extra knowledge related to the minor classes from both major and minor class samples. Our motivation stems from the belief that FER resembles a distribution learning task, wherein a sample may contain information about multiple classes. For instance, a sample from the major class surprise might also contain useful features of the minor class fear. Inspired by that, we propose a novel method that leverages re-balanced attention maps to regularize the model, enabling it to extract transformation invariant information about the minor classes from all training samples. Additionally, we introduce re-balanced smooth labels to regulate the cross-entropy loss, guiding the model to pay more attention to the minor classes

by utilizing the extra information regarding the label distribution of the imbalanced training data. Extensive experiments on different datasets and backbones show that the two proposed modules work together to regularize the model and achieve state-of-the-art performance under the imbalanced FER task. Code is available at https://github.com/zyh-uaiaaaa.

****

ARTree: A Deep Autoregressive Model for Phylogenetic Inference

Tianyu Xie, Cheng Zhang

Designing flexible probabilistic models over tree topologies is important for developing efficient phylogenetic inference methods. To do that, previous works often leverage the similarity of tree topologies via hand-engineered heuristic features which would require domain expertise and may suffer from limited approximation capability. In this paper, we propose a deep autoregressive model for phylogenetic inference based on graph neural networks (GNNs), called ARTree. By decomposing a tree topology into a sequence of leaf node addition operations and modeling the involved conditional distributions based on learnable topological features via GNNs, ARTree can provide a rich family of distributions over tree topologies that have simple sampling algorithms, without using heuristic features. We demonstrate the effectiveness and efficiency of our method on a benchmark of challenging real data tree topology density estimation and variational Bayesian phylogenetic inference problems.

****

A One-Size-Fits-All Approach to Improving Randomness in Paper Assignment

Yixuan Xu, Steven Jecmen, Zimeng Song, Fei Fang

The assignment of papers to reviewers is a crucial part of the peer review processes of large publication venues, where organizers (e.g., conference program chairs) rely on algorithms to perform automated paper assignment. As such, a major challenge for the organizers of these processes is to specify paper assignment algorithms that find appropriate assignments with respect to various desiderata. Although the main objective when choosing a good paper assignment is to maximize the expertise of each reviewer for their assigned papers, several other considerations make introducing randomization into the paper assignment desirable: robustness to malicious behavior, the ability to evaluate alternative paper assignments, reviewer diversity, and reviewer anonymity. However, it is unclear in what way one should randomize the paper assignment in order to best satisfy all of these considerations simultaneously. In this work, we present a practical, one-size-fits-all method for randomized paper assignment intended to perform well across different motivations for randomness. We show theoretically and experimentally that our method outperforms currently-deployed methods for randomized paper assignment on several intuitive randomness metrics, demonstrating that the randomized assignments produced by our method are general-purpose.

****

Loss Dynamics of Temporal Difference Reinforcement Learning

Blake Bordelon, Paul Masset, Henry Kuo, Cengiz Pehlevan

Reinforcement learning has been successful across several applications in which agents have to learn to act in environments with sparse feedback. However, despite this empirical success there is still a lack of theoretical understanding of how the parameters of reinforcement learning models and the features used to represent states interact to control the dynamics of learning. In this work, we use concepts from statistical physics, to study the typical case learning curves for temporal difference learning of a value function with linear function approximators. Our theory is derived under a Gaussian equivalence hypothesis where averages over the random trajectories are replaced with temporally correlated Gaussian feature averages and we validate our assumptions on small scale Markov Decision Processes. We find that the stochastic semi-gradient noise due to subsampling the space of possible episodes leads to significant plateaus in the value error, unlike in traditional gradient descent dynamics. We study how learning dynamics and plateaus depend on feature structure, learning rate, discount factor, and reward function. We then analyze how strategies like learning rate annealing and reward shaping can favorably alter learning dynamics and plateaus. To conclude,

our work introduces new tools to open a new direction towards developing a theory of learning dynamics in reinforcement learning.
****
REASONER: An Explainable Recommendation Dataset with Comprehensive Labeling Ground Truths
Xu Chen, Jingsen Zhang, Lei Wang, Quanyu Dai, Zhenhua Dong, Ruiming Tang, Rui Zhang, Li Chen, Xin Zhao, Ji-Rong Wen
Explainable recommendation has attracted much attention from the industry and academic communities. It has shown great potential to improve the recommendation persuasiveness, informativeness and user satisfaction. In the past few years, while a lot of promising explainable recommender models have been proposed, the datasets used to evaluate them still suffer from several limitations, for example, the explanation ground truths are not labeled by the real users, the explanations are mostly single-modal and around only one aspect. To bridge these gaps, in this paper, we build a new explainable recommendation dataset, which, to our knowledge, is the first contribution that provides a large amount of real user labeled multi-modal and multi-aspect explaination ground truths. In specific, we firstly develop a video recommendation platform, where a series of questions around the recommendation explainability are carefully designed. Then, we recruit about 3000 high-quality labelers with different backgrounds to use the system, and collect their behaviors and feedback to our questions. In this paper, we detail the construction process of our dataset and also provide extensive analysis on its characteristics. In addition, we develop a library, where ten well-known explainable recommender models are implemented in a unified framework. Based on this library, we build several benchmarks for different explainable recommendation tasks. At last, we present many new opportunities brought by our dataset, which are expected to promote the field of explainable recommendation. Our dataset, library and the related documents have been released at https://reasoner2023.github.io/.
****
Fast Model DeBias with Machine Unlearning
Ruizhe Chen, Jianfei Yang, Huimin Xiong, Jianhong Bai, Tianxiang Hu, Jin Hao, YANG FENG, Joey Tianyi Zhou, Jian Wu, Zuozhu Liu
Recent discoveries have revealed that deep neural networks might behave in a biased manner in many real-world scenarios. For instance, deep networks trained on a large-scale face recognition dataset CelebA tend to predict blonde hair for females and black hair for males. Such biases not only jeopardize the robustness of models but also perpetuate and amplify social biases, which is especially concerning for automated decision-making processes in healthcare, recruitment, etc., as they could exacerbate unfair economic and social inequalities among different groups. Existing debiasing methods suffer from high costs in bias labeling or model re-training, while also exhibiting a deficiency in terms of elucidating the origins of biases within the model. To this respect, we propose a fast model debiasing method (FMD) which offers an efficient approach to identify, evaluate and remove biases inherent in trained models. The FMD identifies biased attributes through an explicit counterfactual concept and quantifies the influence of data samples with influence functions. Moreover, we design a machine unlearning-based strategy to efficiently and effectively remove the bias in a trained model with a small counterfactual dataset. Experiments on the Colored MNIST, CelebA, and Adult Income datasets demonstrate that our method achieves superior or competing classification accuracies compared with state-of-the-art retraining-based methods while attaining significantly fewer biases and requiring much less debiasing cost. Notably, our method requires only a small external dataset and updating a minimal amount of model parameters, without the requirement of access to training data that may be too large or unavailable in practice.
****
Coherent Soft Imitation Learning
Joe Watson, Sandy Huang, Nicolas Heess
Imitation learning methods seek to learn from an expert either through behavioral cloning (BC) for the policy or inverse reinforcement learning (IRL) for the re

ward.Such methods enable agents to learn complex tasks from humans that are diff
icult to capture with hand-designed reward functions.Choosing between BC or IRL
for imitation depends on the quality and state-action coverage of the demonstrat
ions, as well as additional access to the Markov decision process. Hybrid strate
gies that combine BC and IRL are rare, as initial policy optimization against in
accurate rewards diminishes the benefit of pretraining the policy with BC.Our wo
rk derives an imitation method that captures the strengths of both BC and IRL.In
 the entropy-regularized (`soft') reinforcement learning setting, we show that t
he behavioral-cloned policy can be used as both a shaped reward and a critic hyp
othesis space by inverting the regularized policy update. This coherency facilit
ates fine-tuning cloned policies using the reward estimate and additional intera
ctions with the environment.This approach conveniently achieves imitation learni
ng through initial behavioral cloning and subsequent refinement via RL with onli
ne or offline data sources.The simplicity of the approach enables graceful scali
ng to high-dimensional and vision-based tasks, with stable learning and minimal
hyperparameter tuning, in contrast to adversarial approaches.For the open-source
 implementation and simulation results, see https://joemwatson.github.io/csil/.
****
Exact Generalization Guarantees for (Regularized) Wasserstein Distributionally R
obust Models
Waïss Azizian, Franck Iutzeler, Jérôme Malick
Wasserstein distributionally robust estimators have emerged as powerful models f
or prediction and decision-making under uncertainty. These estimators provide at
tractive generalization guarantees: the robust objective obtained from the train
ing distribution is an exact upper bound on the true risk with high probability.
 However, existing guarantees either suffer from the curse of dimensionality, ar
e restricted to specific settings, or lead to spurious error terms. In this pape
r, we show that these generalization guarantees actually hold on general classes
 of models, do not suffer from the curse of dimensionality, and can even cover d
istribution shifts at testing. We also prove that these results carry over to th
e newly-introduced regularized versions of Wasserstein distributionally robust p
roblems.
****
Supply-Side Equilibria in Recommender Systems
Meena Jagadeesan, Nikhil Garg, Jacob Steinhardt
Algorithmic recommender systems such as Spotify and Netflix affect not only cons
umer behavior but also producer incentives. Producers seek to create content tha
t will be shown by the recommendation algorithm, which can impact both the diver
sity and quality of their content. In this work, we investigate the resulting su
pply-side equilibria in personalized content recommender systems. We model the d
ecisions of producers as choosing multi-dimensional content vectors and users as
 having heterogenous preferences, which contrasts with classical low-dimensional
 models. Multi-dimensionality and heterogeneity creates the potential for specia
lization, where different producers create different types of content at equilib
rium. Using a duality argument, we derive necessary and sufficient conditions fo
r whether specialization occurs. Then, we characterize the distribution of conte
nt at equilibrium in concrete settings with two populations of users. Lastly, we
 show that specialization can enable producers to achieve positive profit at equ
ilibrium, which means that specialization can reduce the competitiveness of the
marketplace. At a conceptual level, our analysis of supply-side competition take
s a step towards elucidating how personalized recommendations shape the marketpl
ace of digital goods.
****
Human-Aligned Calibration for AI-Assisted Decision Making
Nina Corvelo Benz, Manuel Rodriguez
Whenever a binary classifier is used to provide decision support, it typically p
rovides both a label prediction and a confidence value. Then, the decision maker
 is supposed to use the confidence value to calibrate how much to trust the pred
iction. In this context, it has been often argued that the confidence value shou
ld correspond to a well calibrated estimate of the probability that the predicte

d label matches the ground truth label. However, multiple lines of empirical evidence suggest that decision makers have difficulties at developing a good sense on when to trust a prediction using these confidence values. In this paper, our goal is first to understand why and then investigate how to construct more useful confidence values. We first argue that, for a broad class of utility functions, there exists data distributions for which a rational decision maker is, in general, unlikely to discover the optimal decision policy using the above confidence values—an optimal decision maker would need to sometimes place more (less) trust on predictions with lower (higher) confidence values. However, we then show that, if the confidence values satisfy a natural alignment property with respect to the decision maker’s confidence on her own predictions, there always exists an optimal decision policy under which the level of trust the decision maker would need to place on predictions is monotone on the confidence values, facilitating its discoverability. Further, we show that multicalibration with respect to the decision maker’s confidence on her own prediction is a sufficient condition for alignment. Experiments on a real AI-assisted decision making scenario where a classifier provides decision support to human decision makers validate our theoretical results and suggest that alignment may lead to better decisions.
****
Transformer as a hippocampal memory consolidation model based on NMDAR-inspired nonlinearity
Dong Kyum Kim, Jea Kwon, Meeyoung Cha, C. Lee
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Gaussian Differential Privacy on Riemannian Manifolds
Yangdi Jiang, Xiaotian Chang, Yi Liu, Lei Ding, Linglong Kong, Bei Jiang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Agnostically Learning Single-Index Models using Omnipredictors
Aravind Gollakota, Parikshit Gopalan, Adam Klivans, Konstantinos Stavropoulos
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On skip connections and normalisation layers in deep optimisation
Lachlan MacDonald, Jack Valmadre, Hemanth Saratchandran, Simon Lucey
We introduce a general theoretical framework, designed for the study of gradient optimisation of deep neural networks, that encompasses ubiquitous architecture choices including batch normalisation, weight normalisation and skip connections. Our framework determines the curvature and regularity properties of multilayer loss landscapes in terms of their constituent layers, thereby elucidating the roles played by normalisation layers and skip connections in globalising these properties. We then demonstrate the utility of this framework in two respects. First, we give the only proof of which we are aware that a class of deep neural networks can be trained using gradient descent to global optima even when such optima only exist at infinity, as is the case for the cross-entropy cost. Second, we identify a novel causal mechanism by which skip connections accelerate training, which we verify predictively with ResNets on MNIST, CIFAR10, CIFAR100 and ImageNet.
****
Efficient Low-rank Backpropagation for Vision Transformer Adaptation
Yuedong Yang, Hung-Yueh Chiang, Guihong Li, Diana Marculescu, Radu Marculescu
The increasing scale of vision transformers (ViT) has made the efficient fine-tu

ning of these large models for specific needs a significant challenge in various applications. This issue originates from the computationally demanding matrix multiplications required during the backpropagation process through linear layers in ViT.In this paper, we tackle this problem by proposing a new Low-rank BackPropagation via Walsh-Hadamard Transformation (LBP-WHT) method. Intuitively, LBP-WHT projects the gradient into a low-rank space and carries out backpropagation. This approach substantially reduces the computation needed for adapting ViT, as matrix multiplication in the low-rank space is far less resource-intensive. We conduct extensive experiments with different models (ViT, hybrid convolution-ViT model) on multiple datasets to demonstrate the effectiveness of our method. For instance, when adapting an EfficientFormer-L1 model on CIFAR100, our LBP-WHT achieves 10.4\% higher accuracy than the state-of-the-art baseline, while requiring 9 MFLOPs less computation.As the first work to accelerate ViT adaptation with low-rank backpropagation, our LBP-WHT method is complementary to many prior efforts and can be combined with them for better performance.
****
AdaVAE: Bayesian Structural Adaptation for Variational Autoencoders
Paribesh Regmi, Rui Li
The neural network structures of generative models and their corresponding inference models paired in variational autoencoders (VAEs) play a critical role in the models' generative performance. However, powerful VAE network structures are hand-crafted and fixed prior to training, resulting in a one-size-fits-all approach that requires heavy computation to tune for given data. Moreover, existing VAE regularization methods largely overlook the importance of network structures and fail to prevent overfitting in deep VAE models with cascades of hidden layers. To address these issues, we propose a Bayesian inference framework that automatically adapts VAE network structures to data and prevent overfitting as they grow deeper. We model the number of hidden layers with a beta process to infer the most plausible encoding/decoding network depths warranted by data and perform layer-wise dropout regularization with a conjugate Bernoulli process. We develop a scalable estimator that performs joint inference on both VAE network structures and latent variables. Our experiments show that the inference framework effectively prevents overfitting in both shallow and deep VAE models, yielding state-of-the-art performance. We demonstrate that our framework is compatible with different types of VAE backbone networks and can be applied to various VAE variants, further improving their performance.
****
Safety Verification of Decision-Tree Policies in Continuous Time
Christian Schilling, Anna Lukina, Emir Demirovi■, Kim Larsen
Decision trees have gained popularity as interpretable surrogate models for learning-based control policies. However, providing safety guarantees for systems controlled by decision trees is an open challenge. We show that the problem is undecidable even for systems with the simplest dynamics, and PSPACE-complete for finite-horizon properties. The latter can be verified for discrete-time systems via bounded model checking. However, for continuous-time systems, such an approach requires discretization, thereby weakening the guarantees for the original system. This paper presents the first algorithm to directly verify decision-tree controlled system in continuous time. The key aspect of our method is exploiting the decision-tree structure to propagate a set-based approximation through the decision nodes. We demonstrate the effectiveness of our approach by verifying safety of several decision trees distilled to imitate neural-network policies for nonlinear systems.
****
Quasi-Monte Carlo Graph Random Features
Isaac Reid, Krzysztof M Choromanski, Adrian Weller
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Functional Renyi Differential Privacy for Generative Modeling
Dihong Jiang, Sun Sun, Yaoliang Yu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
FedL2P: Federated Learning to Personalize
Royson Lee, Minyoung Kim, Da Li, Xinchi Qiu, Timothy Hospedales, Ferenc Huszar, Nicholas Lane
Federated learning (FL) research has made progress in developing algorithms for distributed learning of global models, as well as algorithms for local personalization of those common models to the specifics of each client’s local data distribution. However, different FL problems may require different personalization strategies, and it may not even be possible to define an effective one-size-fits-all personalization strategy for all clients: Depending on how similar each client’s optimal predictor is to that of the global model, different personalization strategies may be preferred. In this paper, we consider the federated meta-learning problem of learning personalization strategies. Specifically, we consider meta-nets that induce the batch-norm and learning rate parameters for each client given local data statistics. By learning these meta-nets through FL, we allow the whole FL network to collaborate in learning a customized personalization strategy for each client. Empirical results show that this framework improves on a range of standard hand-crafted personalization baselines in both label and feature shift situations.
****
Learning Reliable Logical Rules with SATNet
Zhaoyu Li, Jinpei Guo, Yuhe Jiang, Xujie Si
Bridging logical reasoning and deep learning is crucial for advanced AI systems. In this work, we present a new framework that addresses this goal by generating interpretable and verifiable logical rules through differentiable learning, without relying on pre-specified logical structures. Our approach builds upon SATNet, a differentiable MaxSAT solver that learns the underlying rules from input-output examples. Despite its efficacy, the learned weights in SATNet are not straightforwardly interpretable, failing to produce human-readable rules. To address this, we propose a novel specification method called ``maximum equality'', which enables the interchangeability between the learned weights of SATNet and a set of propositional logical rules in weighted MaxSAT form. With the decoded weighted MaxSAT formula, we further introduce several effective verification techniques to validate it against the ground truth rules. Experiments on stream transformations and Sudoku problems show that our decoded rules are highly reliable: using exact solvers on them could achieve 100% accuracy, whereas the original SATNet fails to give correct solutions in many cases. Furthermore, we formally verify that our decoded logical rules are functionally equivalent to the ground truth ones.
****
Demo2Code: From Summarizing Demonstrations to Synthesizing Code via Extended Chain-of-Thought
Yuki Wang, Gonzalo Gonzalez-Pumariega, Yash Sharma, Sanjiban Choudhury
Language instructions and demonstrations are two natural ways for users to teach robots personalized tasks. Recent progress in Large Language Models (LLMs) has shown impressive performance in translating language instructions into code for robotic tasks. However, translating demonstrations into task code continues to be a challenge due to the length and complexity of both demonstrations and code, making learning a direct mapping intractable. This paper presents Demo2Code, a novel framework that generates robot task code from demonstrations via an extended chain-of-thought and defines a common latent specification to connect the two. Our framework employs a robust two-stage process: (1) a recursive summarization technique that condenses demonstrations into concise specifications, and (2) a code synthesis approach that expands each function recursively from the generate

d specifications. We conduct extensive evaluation on various robot task benchmarks, including a novel game benchmark Robotouille, designed to simulate diverse cooking tasks in a kitchen environment.
****

Easy Learning from Label Proportions
Róbert Busa-Fekete, Heejin Choi, Travis Dick, Claudio Gentile, Andres Munoz Medina
We consider the problem of Learning from Label Proportions (LLP), a weakly supervised classification setup where instances are grouped into i.i.d. “bags”, and only the frequency of class labels at each bag is available. Albeit, the objective of the learner is to achieve low task loss at an individual instance level. Here we propose EASYLLP, a flexible and simple-to-implement debiasing approach based on aggregate labels, which operates on arbitrary loss functions. Our technique allows us to accurately estimate the expected loss of an arbitrary model at an individual level. We elucidate the differences between our method and standard methods based on label proportion matching, in terms of applicability and optimality conditions. We showcase the flexibility of our approach compared to alternatives by applying our method to popular learning frameworks, like Empirical Risk Minimization (ERM) and Stochastic Gradient Descent (SGD) with provable guarantees on instance level performance. Finally, we validate our theoretical results on multiple datasets, empirically illustrating the conditions under which our algorithm is expected to perform better or worse than previous LLP approaches
****

Jigsaw: Learning to Assemble Multiple Fractured Objects
Jiaxin Lu, Yifan Sun, Qixing Huang
Automated assembly of 3D fractures is essential in orthopedics, archaeology, and our daily life. This paper presents Jigsaw, a novel framework for assembling physically broken 3D objects from multiple pieces. Our approach leverages hierarchical features of global and local geometry to match and align the fracture surfaces. Our framework consists of four components: (1) front-end point feature extractor with attention layers, (2) surface segmentation to separate fracture and original parts, (3) multi-parts matching to find correspondences among fracture surface points, and (4) robust global alignment to recover the global poses of the pieces. We show how to jointly learn segmentation and matching and seamlessly integrate feature matching and rigidity constraints. We evaluate Jigsaw on the Breaking Bad dataset and achieve superior performance compared to state-of-the-art methods. Our method also generalizes well to diverse fracture modes, objects, and unseen instances. To the best of our knowledge, this is the first learning-based method designed specifically for 3D fracture assembly over multiple pieces. Our code is available at https://jiaxin-lu.github.io/Jigsaw/.
****

Persuading Farsighted Receivers in MDPs: the Power of Honesty
Martino Bernasconi, Matteo Castiglioni, Alberto Marchesi, Mirco Mutti
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

When are ensembles really effective?
Ryan Theisen, Hyunsuk Kim, Yaoqing Yang, Liam Hodgkinson, Michael W. Mahoney
Ensembling has a long history in statistical data analysis, with many impactful applications. However, in many modern machine learning settings, the benefits of ensembling are less ubiquitous and less obvious. We study, both theoretically and empirically, the fundamental question of when ensembling yields significant performance improvements in classification tasks. Theoretically, we prove new results relating the \emph{ensemble improvement rate} (a measure of how much ensembling decreases the error rate versus a single model, on a relative scale) to the \emph{disagreement-error ratio}. We show that ensembling improves performance significantly whenever the disagreement rate is large relative to the average error rate; and that, conversely, one classifier is often enough whenever the disag

reement rate is low relative to the average error rate. On the way to proving these results, we derive, under a mild condition called \emph{competence}, improved upper and lower bounds on the average test error rate of the majority vote classifier.To complement this theory, we study ensembling empirically in a variety of settings, verifying the predictions made by our theory, and identifying practical scenarios where ensembling does and does not result in large performance improvements. Perhaps most notably, we demonstrate a distinct difference in behavior between interpolating models (popular in current practice) and non-interpolating models (such as tree-based methods, where ensembling is popular), demonstrating that ensembling helps considerably more in the latter case than in the former.
****
A Unified Approach to Domain Incremental Learning with Memory: Theory and Algorithm
Haizhou Shi, Hao Wang
Domain incremental learning aims to adapt to a sequence of domains with access to only a small subset of data (i.e., memory) from previous domains. Various methods have been proposed for this problem, but it is still unclear how they are related and when practitioners should choose one method over another. In response, we propose a unified framework, dubbed Unified Domain Incremental Learning (UDIL), for domain incremental learning with memory. Our UDIL unifies various existing methods, and our theoretical analysis shows that UDIL always achieves a tighter generalization error bound compared to these methods. The key insight is that different existing methods correspond to our bound with different fixed coefficients; based on insights from this unification, our UDIL allows adaptive coefficients during training, thereby always achieving the tightest bound. Empirical results show that our UDIL outperforms the state-of-the-art domain incremental learning methods on both synthetic and real-world datasets. Code will be available at https://github.com/Wang-ML-Lab/unified-continual-learning.
****
Few-Shot Class-Incremental Learning via Training-Free Prototype Calibration
Qi-Wei Wang, Da-Wei Zhou, Yi-Kai Zhang, De-Chuan Zhan, Han-Jia Ye
Real-world scenarios are usually accompanied by continuously appearing classes with scare labeled samples, which require the machine learning model to incrementally learn new classes and maintain the knowledge of base classes. In this Few-Shot Class-Incremental Learning (FSCIL) scenario, existing methods either introduce extra learnable components or rely on a frozen feature extractor to mitigate catastrophic forgetting and overfitting problems. However, we find a tendency for existing methods to misclassify the samples of new classes into base classes, which leads to the poor performance of new classes. In other words, the strong discriminability of base classes distracts the classification of new classes. To figure out this intriguing phenomenon, we observe that although the feature extractor is only trained on base classes, it can surprisingly represent the semantic similarity between the base and unseen new classes. Building upon these analyses, we propose a simple yet effective Training-frEE calibratioN (TEEN) strategy to enhance the discriminability of new classes by fusing the new prototypes (i.e., mean features of a class) with weighted base prototypes. In addition to standard benchmarks in FSCIL, TEEN demonstrates remarkable performance and consistent improvements over baseline methods in the few-shot learning scenario. Code is available at: https://github.com/wangkiw/TEEN
****
RADAR: Robust AI-Text Detection via Adversarial Learning
Xiaomeng Hu, Pin-Yu Chen, Tsung-Yi Ho
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Alleviating the Semantic Gap for Generalized fMRI-to-Image Reconstruction
Tao Fang, Qian Zheng, Gang Pan

Although existing fMRI-to-image reconstruction methods could predict high-quality images, they do not explicitly consider the semantic gap between training and testing data, resulting in reconstruction with unstable and uncertain semantics. This paper addresses the problem of generalized fMRI-to-image reconstruction by explicitly alleviates the semantic gap. Specifically, we leverage the pre-trained CLIP model to map the training data to a compact feature representation, which essentially extends the sparse semantics of training data to dense ones, thus alleviating the semantic gap of the instances nearby known concepts (i.e., inside the training super-classes). Inspired by the robust low-level representation in fMRI data, which could help alleviate the semantic gap for instances that far from the known concepts (i.e., outside the training super-classes), we leverage structural information as a general cue to guide image reconstruction. Further, we quantify the semantic uncertainty based on probability density estimation and achieve Generalized fMRI-to-image reconstruction by adaptively integrating Expanded Semantics and Structural information (GESS) within a diffusion process. Experimental results demonstrate that the proposed GESS model outperforms state-of-the-art methods, and we propose a generalized scenario split strategy to evaluate the advantage of GESS in closing the semantic gap.
****
Softmax Output Approximation for Activation Memory-Efficient Training of Attention-based Networks
Changhyeon Lee, Seulki Lee
In this paper, we propose to approximate the softmax output, which is the key product of the attention mechanism, to reduce its activation memory usage when training attention-based networks (aka Transformers). During the forward pass of the network, the proposed softmax output approximation method stores only a small fraction of the entire softmax output required for back-propagation and evicts the rest of the softmax output from memory. Then, during the backward pass, the evicted softmax activation output is approximated to compose the gradient to perform back-propagation for model training. Considering most attention-based models heavily rely on the softmax-based attention module that usually takes one of the biggest portions of the network, approximating the softmax activation output can be a simple yet effective way to decrease the training memory requirement of many attention-based networks. The experiment with various attention-based models and relevant tasks, i.e., machine translation, text classification, and sentiment analysis, shows that it curtails the activation memory usage of the softmax-based attention module by up to 84% (6.2× less memory) in model training while achieving comparable or better performance, e.g., up to 5.4% higher classification accuracy.
****
Data Portraits: Recording Foundation Model Training Data
Marc Marone, Benjamin Van Durme
Foundation models are trained on increasingly immense and opaque datasets. Even while these models are now key in AI system building, it can be difficult to answer the straightforward question: has the model already encountered a given example during training? We therefore propose a widespread adoption of Data Portraits: artifacts that record training data and allow for downstream inspection. First we outline the properties of such an artifact and discuss how existing solutions can be used to increase transparency. We then propose and implement a solution based on data sketching, stressing fast and space efficient querying. Using our tools, we document a popular language modeling corpus (The Pile) and a recently released code modeling dataset (The Stack). We show that our solution enables answering questions about test set leakage and model plagiarism. Our tool is lightweight and fast, costing only 3% of the dataset size in overhead. We release a live interface of our tools at https://dataportraits.org/ and call on dataset and model creators to release Data Portraits as a complement to current documentation practices.
****
Memory Efficient Optimizers with 4-bit States
Bingrui Li, Jianfei Chen, Jun Zhu

Optimizer states are a major source of memory consumption for training neural networks, limiting the maximum trainable model within given memory budget. Compressing the optimizer states from 32-bit floating points to lower bitwidth is promising to reduce the training memory footprint, while the current lowest achievable bitwidth is 8-bit. In this work, we push optimizer states bitwidth down to 4-bit through a detailed empirical analysis of first and second moments. Specifically, we find that moments have complicated outlier patterns, that current block-wise quantization cannot accurately approximate. We use a smaller block size and propose to utilize both row-wise and column-wise information for better quantization. We further identify a zero point problem of quantizing the second moment, and solve this problem with a linear quantizer that excludes the zero point. Our 4-bit optimizers are evaluated on a wide variety of benchmarks including natural language understanding, machine translation, image classification, and instruction tuning. On all the tasks our optimizers can achieve comparable accuracy with their full-precision counterparts, while enjoying better memory efficiency.

****

Replicable Reinforcement Learning

Eric Eaton, Marcel Hussing, Michael Kearns, Jessica Sorrell

The replicability crisis in the social, behavioral, and data sciences has led to the formulation of algorithm frameworks for replicability --- i.e., a requirement that an algorithm produce identical outputs (with high probability) when run on two different samples from the same underlying distribution. While still in its infancy, provably replicable algorithms have been developed for many fundamental tasks in machine learning and statistics, including statistical query learning, the heavy hitters problem, and distribution testing. In this work we initiate the study of replicable reinforcement learning, providing a provably replicable algorithm for parallel value iteration, and a provably replicable version of R-Max in the episodic setting. These are the first formal replicability results for control problems, which present different challenges for replication than batch learning settings.

****

Make Pre-trained Model Reversible: From Parameter to Memory Efficient Fine-Tuning

Baohao Liao, Shaomu Tan, Christof Monz

Parameter-efficient fine-tuning (PEFT) of pre-trained language models (PLMs) has emerged as a highly successful approach, with training only a small number of parameters without sacrificing performance and becoming the de-facto learning paradigm with the increasing size of PLMs. However, existing PEFT methods are not memory-efficient, because they still require caching most of the intermediate activations for the gradient calculation, akin to fine-tuning. One effective way to reduce the activation memory is to apply a reversible model, so the intermediate activations are not necessary to be cached and can be recomputed. Nevertheless, modifying a PLM to its reversible variant is not straightforward, since the reversible model has a distinct architecture from the currently released PLMs. In this paper, we first investigate what is a key factor for the success of existing PEFT methods, and realize that it's essential to preserve the PLM's starting point when initializing a PEFT method. With this finding, we propose memory-efficient fine-tuning (MEFT) that inserts adapters into a PLM, preserving the PLM's starting point and making it reversible without additional pre-training. We evaluate MEFT on the GLUE benchmark and five question-answering tasks with various backbones, BERT, RoBERTa, BART and OPT. MEFT significantly reduces the activation memory up to 84% of full fine-tuning with a negligible amount of trainable parameters. Moreover, MEFT achieves the same score on GLUE and a comparable score on the question-answering tasks as full fine-tuning. A similar finding is also observed for the image classification task.

****

Design from Policies: Conservative Test-Time Adaptation for Offline Policy Optimization

Jinxin Liu, Hongyin Zhang, Zifeng Zhuang, Yachen Kang, Donglin Wang, Bin Wang

In this work, we decouple the iterative bi-level offline RL (value estimation an

d policy extraction) from the offline training phase, forming a non-iterative bi-level paradigm and avoiding the iterative error propagation over two levels. Specifically, this non-iterative paradigm allows us to conduct inner-level optimization (value estimation) in training, while performing outer-level optimization (policy extraction) in testing. Naturally, such a paradigm raises three core questions that are not fully answered by prior non-iterative offline RL counterparts like reward-conditioned policy: (q1) What information should we transfer from the inner-level to the outer-level? (q2) What should we pay attention to when exploiting the transferred information for safe/confident outer-level optimization? (q3) What are the benefits of concurrently conducting outer-level optimization during testing? Motivated by model-based optimization (MBO), we propose DROP (design from policies), which fully answers the above questions. Specifically, in the inner-level, DROP decomposes offline data into multiple subsets, and learns an MBO score model (a1). To keep safe exploitation to the score model in the outer-level, we explicitly learn a behavior embedding and introduce a conservative regularization (a2). During testing, we show that DROP permits deployment adaptation, enabling an adaptive inference across states (a3). Empirically, we evaluate DROP on various tasks, showing that DROP gains comparable or better performance compared to prior methods.
****
Lightweight Vision Transformer with Bidirectional Interaction
Qihang Fan, Huaibo Huang, Xiaoqiang Zhou, Ran He
Recent advancements in vision backbones have significantly improved their performance by simultaneously modeling images’ local and global contexts. However, the bidirectional interaction between these two contexts has not been well explored and exploited, which is important in the human visual system. This paper proposes a Fully Adaptive Self-Attention (FASA) mechanism for vision transformer to model the local and global information as well as the bidirectional interaction between them in context-aware ways. Specifically, FASA employs self-modulated convolutions to adaptively extract local representation while utilizing self-attention in down-sampled space to extract global representation. Subsequently, it conducts a bidirectional adaptation process between local and global representation to model their interaction. In addition, we introduce a fine-grained downsampling strategy to enhance the down-sampled self-attention mechanism for finer-grained global perception capability. Based on FASA, we develop a family of lightweight vision backbones, Fully Adaptive Transformer (FAT) family. Extensive experiments on multiple vision tasks demonstrate that FAT achieves impressive performance. Notably, FAT accomplishes a 77.6% accuracy on ImageNet-1K using only 4.5M parameters and 0.7G FLOPs, which surpasses the most advanced ConvNets and Transformers with similar model size and computational costs. Moreover, our model exhibits faster speed on modern GPU compared to other models.
****
Fused Gromov-Wasserstein Graph Mixup for Graph-level Classifications
Xinyu Ma, Xu Chu, Yasha Wang, Yang Lin, Junfeng Zhao, Liantao Ma, Wenwu Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Modeling Dynamics over Meshes with Gauge Equivariant Nonlinear Message Passing
Jung Yeon Park, Lawson Wong, Robin Walters
Data over non-Euclidean manifolds, often discretized as surface meshes, naturally arise in computer graphics and biological and physical systems. In particular, solutions to partial differential equations (PDEs) over manifolds depend critically on the underlying geometry. While graph neural networks have been successfully applied to PDEs, they do not incorporate surface geometry and do not consider local gauge symmetries of the manifold. Alternatively, recent works on gauge equivariant convolutional and attentional architectures on meshes leverage the underlying geometry but underperform in modeling surface PDEs with complex nonlinear dynamics. To address these issues, we introduce a new gauge equivariant archi

tecture using nonlinear message passing. Our novel architecture achieves higher performance than either convolutional or attentional networks on domains with highly complex and nonlinear dynamics. However, similar to the non-mesh case, design trade-offs favor convolutional, attentional, or message passing networks for different tasks; we investigate in which circumstances our message passing method provides the most benefit.

****

ASIF: Coupled Data Turns Unimodal Models to Multimodal without Training

Antonio Norelli, Marco Fumero, Valentino Maiorca, Luca Moschella, Emanuele Rodolà, Francesco Locatello

CLIP proved that aligning visual and language spaces is key to solving many vision tasks without explicit training, but required to train image and text encoders from scratch on a huge dataset. LiT improved this by only training the text encoder and using a pre-trained vision network. In this paper, we show that a common space can be created without any training at all, using single-domain encoders (trained with or without supervision) and a much smaller amount of image-text pairs. Furthermore, our model has unique properties. Most notably, deploying a new version with updated training samples can be done in a matter of seconds. Additionally, the representations in the common space are easily interpretable as every dimension corresponds to the similarity of the input to a unique entry in the multimodal dataset. Experiments on standard zero-shot visual benchmarks demonstrate the typical transfer ability of image-text models. Overall, our method represents a simple yet surprisingly strong baseline for foundation multi-modal models, raising important questions on their data efficiency and on the role of retrieval in machine learning.

****

A Metadata-Driven Approach to Understand Graph Neural Networks

Ting Wei Li, Qiaozhu Mei, Jiaqi Ma

Graph Neural Networks (GNNs) have achieved remarkable success in various applications, but their performance can be sensitive to specific data properties of the graph datasets they operate on. Current literature on understanding the limitations of GNNs has primarily employed a \emph{model-driven} approach that leverage heuristics and domain knowledge from network science or graph theory to model the GNN behaviors, which is time-consuming and highly subjective. In this work, we propose a \emph{metadata-driven} approach to analyze the sensitivity of GNNs to graph data properties, motivated by the increasing availability of graph learning benchmarks. We perform a multivariate sparse regression analysis on the metadata derived from benchmarking GNN performance across diverse datasets, yielding a set of salient data properties. To validate the effectiveness of our data-driven approach, we focus on one identified data property, the degree distribution, and investigate how this property influences GNN performance through theoretical analysis and controlled experiments. Our theoretical findings reveal that datasets with more balanced degree distribution exhibit better linear separability of node representations, thus leading to better GNN performance. We also conduct controlled experiments using synthetic datasets with varying degree distributions, and the results align well with our theoretical findings. Collectively, both the theoretical analysis and controlled experiments verify that the proposed metadata-driven approach is effective in identifying critical data properties for GNNs.

****

Multimodal Deep Learning Model Unveils Behavioral Dynamics of V1 Activity in Freely Moving Mice

Aiwen Xu, Yuchen Hou, Cristopher Niell, Michael Beyeler

Despite their immense success as a model of macaque visual cortex, deep convolutional neural networks (CNNs) have struggled to predict activity in visual cortex of the mouse, which is thought to be strongly dependent on the animal’s behavioral state. Furthermore, most computational models focus on predicting neural responses to static images presented under head fixation, which are dramatically different from the dynamic, continuous visual stimuli that arise during movement in the real world. Consequently, it is still unknown how natural visual input and

different behavioral variables may integrate over time to generate responses in primary visual cortex (V1). To address this, we introduce a multimodal recurrent neural network that integrates gaze-contingent visual input with behavioral and temporal dynamics to explain V1 activity in freely moving mice. We show that the model achieves state-of-the-art predictions of V1 activity during free exploration and demonstrate the importance of each component in an extensive ablation study. Analyzing our model using maximally activating stimuli and saliency maps, we reveal new insights into cortical function, including the prevalence of mixed selectivity for behavioral variables in mouse V1. In summary, our model offers a comprehensive deep-learning framework for exploring the computational principles underlying V1 neurons in freely-moving animals engaged in natural behavior.
****
Goal-conditioned Offline Planning from Curious Exploration
Marco Bagatella, Georg Martius
Curiosity has established itself as a powerful exploration strategy in deep reinforcement learning. Notably, leveraging expected future novelty as intrinsic motivation has been shown to efficiently generate exploratory trajectories, as well as a robust dynamics model. We consider the challenge of extracting goal-conditioned behavior from the products of such unsupervised exploration techniques, without any additional environment interaction. We find that conventional goal-conditioned reinforcement learning approaches for extracting a value function and policy fall short in this difficult offline setting. By analyzing the geometry of optimal goal-conditioned value functions, we relate this issue to a specific class of estimation artifacts in learned values. In order to mitigate their occurrence, we propose to combine model-based planning over learned value landscapes with a graph-based value aggregation scheme. We show how this combination can correct both local and global artifacts, obtaining significant improvements in zero-shot goal-reaching performance across diverse simulated environments.
****
Rethinking the Role of Token Retrieval in Multi-Vector Retrieval
Jinhyuk Lee, Zhuyun Dai, Sai Meher Karthik Duddu, Tao Lei, Iftekhar Naim, Ming-Wei Chang, Vincent Zhao
Multi-vector retrieval models such as ColBERT [Khattab et al., 2020] allow token-level interactions between queries and documents, and hence achieve state of the art on many information retrieval benchmarks. However, their non-linear scoring function cannot be scaled to millions of documents, necessitating a three-stage process for inference: retrieving initial candidates via token retrieval, accessing all token vectors, and scoring the initial candidate documents. The non-linear scoring function is applied over all token vectors of each candidate document, making the inference process complicated and slow. In this paper, we aim to simplify the multi-vector retrieval by rethinking the role of token retrieval. We present XTR, ConteXtualized Token Retriever, which introduces a simple, yet novel, objective function that encourages the model to retrieve the most important document tokens first. The improvement to token retrieval allows XTR to rank candidates only using the retrieved tokens rather than all tokens in the document, and enables a newly designed scoring stage that is two-to-three orders of magnitude cheaper than that of ColBERT. On the popular BEIR benchmark, XTR advances the state-of-the-art by 2.8 nDCG@10 without any distillation. Detailed analysis confirms our decision to revisit the token retrieval stage, as XTR demonstrates much better recall of the token retrieval stage compared to ColBERT.
****
Optimal Exploration for Model-Based RL in Nonlinear Systems
Andrew Wagenmaker, Guanya Shi, Kevin G. Jamieson
Learning to control unknown nonlinear dynamical systems is a fundamental problem in reinforcement learning and control theory. A commonly applied approach is to first explore the environment (exploration), learn an accurate model of it (system identification), and then compute an optimal controller with the minimum cost on this estimated system (policy optimization). While existing work has shown that it is possible to learn a uniformly good model of the system (Mania et al., 2020), in practice, if we aim to learn a good controller with a low cost on the

actual system, certain system parameters may be significantly more critical than others, and we therefore ought to focus our exploration on learning such parameters.In this work, we consider the setting of nonlinear dynamical systems and seek to formally quantify, in such settings, (a) which parameters are most relevant to learning a good controller, and (b) how we can best explore so as to minimize uncertainty in such parameters. Inspired by recent work in linear systems (Wagenmaker et al., 2021), we show that minimizing the controller loss in nonlinear systems translates to estimating the system parameters in a particular, task-dependent metric. Motivated by this, we develop an algorithm able to efficiently explore the system to reduce uncertainty in this metric, and prove a lower bound showing that our approach learns a controller at a near-instance-optimal rate. Our algorithm relies on a general reduction from policy optimization to optimal experiment design in arbitrary systems, and may be of independent interest. We conclude with experiments demonstrating the effectiveness of our method in realistic nonlinear robotic systems.
****
ELDEN: Exploration via Local Dependencies
Zizhao Wang, Jiaheng Hu, Peter Stone, Roberto Martín-Martín
Tasks with large state space and sparse rewards present a longstanding challenge to reinforcement learning. In these tasks, an agent needs to explore the state space efficiently until it finds a reward. To deal with this problem, the community has proposed to augment the reward function with intrinsic reward, a bonus signal that encourages the agent to visit interesting states. In this work, we propose a new way of defining interesting states for environments with factored state spaces and complex chained dependencies, where an agent's actions may change the value of one entity that, in order, may affect the value of another entity. Our insight is that, in these environments, interesting states for exploration are states where the agent is uncertain whether (as opposed to how) entities such as the agent or objects have some influence on each other. We present ELDEN, Exploration via Local DepENdencies, a novel intrinsic reward that encourages the discovery of new interactions between entities. ELDEN utilizes a novel scheme --- the partial derivative of the learned dynamics to model the local dependencies between entities accurately and computationally efficiently. The uncertainty of the predicted dependencies is then used as an intrinsic reward to encourage exploration toward new interactions. We evaluate the performance of ELDEN on four different domains with complex dependencies, ranging from 2D grid worlds to 3D robotic tasks. In all domains, ELDEN correctly identifies local dependencies and learns successful policies, significantly outperforming previous state-of-the-art exploration methods.
****
Maximization of Average Precision for Deep Learning with Adversarial Ranking Robustness
Gang Li, Wei Tong, Tianbao Yang
This paper seeks to address a gap in optimizing Average Precision (AP) while ensuring adversarial robustness, an area that has not been extensively explored to the best of our knowledge. AP maximization for deep learning has widespread applications, particularly when there is a significant imbalance between positive and negative examples. Although numerous studies have been conducted on adversarial training, they primarily focus on robustness concerning accuracy, ensuring that the average accuracy on adversarially perturbed examples is well maintained. However, this type of adversarial robustness is insufficient for many applications, as minor perturbations on a single example can significantly impact AP while not greatly influencing the accuracy of the prediction system. To tackle this issue, we introduce a novel formulation that combines an AP surrogate loss with a regularization term representing adversarial ranking robustness, which maintains the consistency between ranking of clean data and that of perturbed data. We then devise an efficient stochastic optimization algorithm to optimize the resulting objective. Our empirical studies, which compare our method to current leading adversarial training baselines and other robust AP maximization strategies, demonstrate the effectiveness of the proposed approach. Notably, our methods outpe

rform a state-of-the-art method (TRADES) by more than 4\% in terms of robust AP
against PGD attacks while achieving 7\% higher AP on clean data simultaneously o
n CIFAR10 and CIFAR100.The code is available at: https://github.com/GangLii/Adve
rsarial-AP
****
Act As You Wish: Fine-Grained Control of Motion Diffusion Model with Hierarchica
l Semantic Graphs
Peng Jin, Yang Wu, Yanbo Fan, Zhongqian Sun, Wei Yang, Li Yuan
Most text-driven human motion generation methods employ sequential modeling appr
oaches, e.g., transformer, to extract sentence-level text representations automa
tically and implicitly for human motion synthesis. However, these compact text r
epresentations may overemphasize the action names at the expense of other import
ant properties and lack fine-grained details to guide the synthesis of subtly di
stinct motion. In this paper, we propose hierarchical semantic graphs for fine-g
rained control over motion generation. Specifically, we disentangle motion descr
iptions into hierarchical semantic graphs including three levels of motions, act
ions, and specifics. Such global-to-local structures facilitate a comprehensive
understanding of motion description and fine-grained control of motion generatio
n. Correspondingly, to leverage the coarse-to-fine topology of hierarchical sema
ntic graphs, we decompose the text-to-motion diffusion process into three semant
ic levels, which correspond to capturing the overall motion, local actions, and
action specifics. Extensive experiments on two benchmark human motion datasets,
including HumanML3D and KIT, with superior performances, justify the efficacy of
 our method. More encouragingly, by modifying the edge weights of hierarchical s
emantic graphs, our method can continuously refine the generated motion, which m
ay have a far-reaching impact on the community. Code and pre-trained weights are
 available at https://github.com/jpthu17/GraphMotion.
****
DynaDojo: An Extensible Platform for Benchmarking Scaling in Dynamical System Id
entification
Logan M Bhamidipaty, Tommy Bruzzese, Caryn Tran, Rami Ratl Mrad, Maxinder S. Kan
wal
Modeling complex dynamical systems poses significant challenges, with traditiona
l methods struggling to work on a variety of systems and scale to high-dimension
al dynamics. In response, we present DynaDojo, a novel benchmarking platform des
igned for data-driven dynamical system identification. DynaDojo provides diagnos
tics on three ways an algorithm’s performance scales: across the number of train
ing samples, the complexity of a dynamical system, and a target error to achieve
. Furthermore, DynaDojo enables studying out-of-distribution generalization (by
providing unique test conditions for each system) and active learning (by suppor
ting closed-loop control). Through its user-friendly and easily extensible API,
DynaDojo accommodates a wide range of user-defined \texttt{Algorithms}, \texttt{
Systems}, and \texttt{Challenges} (evaluation metrics). The platform also priori
tizes resource-efficient training with parallel processing strategies for runnin
g on a cluster. To showcase its utility, in DynaDojo 0.9, we include implementat
ions of 7 baseline algorithms and 20 dynamical systems, along with several demos
 exhibiting insights researchers can glean using our platform. This work aspires
 to make DynaDojo a unifying benchmarking platform for system identification, pa
ralleling the role of OpenAI’s Gym in reinforcement learning.
****
Online Constrained Meta-Learning: Provable Guarantees for Generalization
Siyuan Xu, Minghui Zhu
Meta-learning has attracted attention due to its strong ability to learn experie
nces from known tasks, which can speed up and enhance the learning process for n
ew tasks. However, most existing meta-learning approaches only can learn from ta
sks without any constraint. This paper proposes an online constrained meta-learn
ing framework, which continuously learns meta-knowledge from sequential learning
 tasks, and the learning tasks are subject to hard constraints. Beyond existing
meta-learning analyses, we provide the upper bounds of optimality gaps and const
raint violations produced by the proposed framework, which considers the dynamic

regret of online learning, as well as the generalization ability of the task-specific models. Moreover, we provide a practical algorithm for the framework, and validate its superior effectiveness through experiments conducted on meta-imitation learning and few-shot image classification.
****
Convergence of mean-field Langevin dynamics: time-space discretization, stochastic gradient, and variance reduction
Taiji Suzuki, Denny Wu, Atsushi Nitanda
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Public Opinion Field Effect Fusion in Representation Learning for Trending Topics Diffusion
Junliang Li, Yang Yajun, Qinghua Hu, Xin Wang, Hong Gao
Trending topic diffusion and prediction analysis is an important problem and has been well studied in social networks. Representation learning is an effective way to extract node embeddings, which can help for topic propagation analysis by completing downstream tasks such as link prediction and node classification. In real world, there are often several trending topics or opinion leaders in public opinion space at the same time and they can be regarded as different centers of public opinion. A public opinion field will be formed surrounding every center. These public opinion fields compete for public's attention and it will potentially affect the development of public opinion. However, the existing methods do not consider public opinion field effect for trending topics diffusion. In this paper, we introduce three well-known observations about public opinion field effect in media and communication studies, and propose a novel and effective heterogeneous representation learning framework to incorporate public opinion field effect and social circle influence effect. To the best of our knowledge, our work is the first to consider these effects in representation learning for trending topic diffusion. Extensive experiments on real-world datasets validate the superiority of our model.
****
Distributional Pareto-Optimal Multi-Objective Reinforcement Learning
Xin-Qiang Cai, Pushi Zhang, Li Zhao, Jiang Bian, Masashi Sugiyama, Ashley Llorens
Multi-objective reinforcement learning (MORL) has been proposed to learn control policies over multiple competing objectives with each possible preference over returns. However, current MORL algorithms fail to account for distributional preferences over the multi-variate returns, which are particularly important in real-world scenarios such as autonomous driving. To address this issue, we extend the concept of Pareto-optimality in MORL into distributional Pareto-optimality, which captures the optimality of return distributions, rather than the expectations. Our proposed method, called Distributional Pareto-Optimal Multi-Objective Reinforcement Learning~(DPMORL), is capable of learning distributional Pareto-optimal policies that balance multiple objectives while considering the return uncertainty. We evaluated our method on several benchmark problems and demonstrated its effectiveness in discovering distributional Pareto-optimal policies and satisfying diverse distributional preferences compared to existing MORL methods.
****
Large Language Models Are Latent Variable Models: Explaining and Finding Good Demonstrations for In-Context Learning
Xinyi Wang, Wanrong Zhu, Michael Saxon, Mark Steyvers, William Yang Wang
In recent years, pre-trained large language models (LLMs) have demonstrated remarkable efficiency in achieving an inference-time few-shot learning capability known as in-context learning. However, existing literature has highlighted the sensitivity of this capability to the selection of few-shot demonstrations. Current understandings of the underlying mechanisms by which this capability arises from regular language model pretraining objectives remain disconnected from the rea

l-world LLMs. This study aims to examine the in-context learning phenomenon thro ugh a Bayesian lens, viewing real-world LLMs as latent variable models. On this premise, we propose an algorithm to select optimal demonstrations from a set of annotated data with a small LM, and then directly generalize the selected demons trations to larger LMs. We demonstrate significant improvement over baselines, a veraged over eight GPT models on eight real-world text classification datasets. We also demonstrate the real-world usefulness of our algorithm on GSM8K, a math word problem dataset. Our empirical findings support our hypothesis that LLMs im plicitly infer a latent variable containing task information.
****
Physics-Driven ML-Based Modelling for Correcting Inverse Estimation
ruiyuan kang, Tingting Mu, Panagiotis Liatsis, Dimitrios Kyritsis
When deploying machine learning estimators in science and engineering (SAE) dom ains, it is critical to avoid failed estimations that can have disastrous conse quences, e.g., in aero engine design. This work focuses on detecting and correct ing failed state estimations before adopting them in SAE inverse problems, by utilizing simulations and performance metrics guided by physical laws. We sugge st to flag a machine learning estimation when its physical model error exceeds a feasible threshold, and propose a novel approach, GEESE, to correct it through optimization, aiming at delivering both low error and high efficiency. The key designs of GEESE include (1) a hybrid surrogate error model to provide fast er ror estimations to reduce simulation cost and to enable gradient based backprop agation of error feedback, and (2) two generative models to approximate the prob ability distributions of the candidate states for simulating the exploitation a nd exploration behaviours. All three models are constructed as neural networks. GEESE is tested on three real-world SAE inverse problems and compared to a numbe r of state-of-the-art optimization/search approaches. Results show that it fails the least number of times in terms of finding a feasible state correction, and requires physical evaluations less frequently in general.
****
One-Pass Distribution Sketch for Measuring Data Heterogeneity in Federated Learn ing
Zichang Liu, Zhaozhuo Xu, Benjamin Coleman, Anshumali Shrivastava
Federated learning (FL) is a machine learning paradigm where multiple client dev ices train models collaboratively without data exchange. Data heterogeneity prob lem is naturally inherited in FL since data in different clients follow diverse distributions. To mitigate the negative influence of data heterogeneity, we nee d to start by measuring it across clients. However, the efficient measurement be tween distributions is a challenging problem, especially in high dimensionality. In this paper, we propose a one-pass distribution sketch to represent the clien t data distribution. Our sketching algorithm only requires a single pass of the client data, which is efficient in terms of time and memory. Moreover, we show i n both theory and practice that the distance between two distribution sketches r epresents the divergence between their corresponding distributions. Furthermore, we demonstrate with extensive experiments that our distribution sketch improves the client selection in the FL training. We also showcase that our distribution sketch is an efficient solution to the cold start problem in FL for new clients with unlabeled data.
****
Kernel-Based Tests for Likelihood-Free Hypothesis Testing
Patrik Robert Gerber, Tianze Jiang, Yury Polyanskiy, Rui Sun
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Deep Equilibrium Based Neural Operators for Steady-State PDEs
Tanya Marwah, Ashwini Pokle, J. Zico Kolter, Zachary Lipton, Jianfeng Lu, Andrej Risteski
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
An Efficient and Robust Framework for Approximate Nearest Neighbor Search with Attribute Constraint
Mengzhao Wang, Lingwei Lv, Xiaoliang Xu, Yuxiang Wang, Qiang Yue, Jiongkang Ni
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Parameter-efficient Tuning of Large-scale Multimodal Foundation Model
Haixin Wang, Xinlong Yang, Jianlong Chang, Dian Jin, Jinan Sun, Shikun Zhang, Xiao Luo, Qi Tian
Driven by the progress of large-scale pre-training, parameter-efficient transfer learning has gained immense popularity across different subfields of Artificial Intelligence. The core is to adapt the model to downstream tasks with only a small set of parameters. Recently, researchers have leveraged such proven techniques in multimodal tasks and achieve promising results. However, two critical issues remain unresolved: how to further reduce the complexity with lightweight design and how to boost alignment between modalities under extremely low parameters. In this paper, we propose A graceFUl pRompt framewOrk for cRoss-modal trAnsfer (AURORA) to overcome these challenges. Considering the redundancy in existing architectures, we first utilize the mode approximation to generate 0.1M trainable parameters to implement the multimodal parameter-efficient tuning, which explores the low intrinsic dimension with only 0.04% parameters of the pre-trained model. Then, for better modality alignment, we propose the Informative Context Enhancement and Gated Query Transformation module under extremely few parameters scenes. A thorough evaluation on six cross-modal benchmarks shows that it not only outperforms the state-of-the-art but even outperforms the full fine-tuning approach. Our code is available at: https://github.com/WillDreamer/Aurora.
****
Balance, Imbalance, and Rebalance: Understanding Robust Overfitting from a Minimax Game Perspective
Yifei Wang, Liangchen Li, Jiansheng Yang, Zhouchen Lin, Yisen Wang
Adversarial Training (AT) has become arguably the state-of-the-art algorithm for extracting robust features. However, researchers recently notice that AT suffers from severe robust overfitting problems, particularly after learning rate (LR) decay. In this paper, we explain this phenomenon by viewing adversarial training as a dynamic minimax game between the model trainer and the attacker. Specifically, we analyze how LR decay breaks the balance between the minimax game by empowering the trainer with a stronger memorization ability, and show such imbalance induces robust overfitting as a result of memorizing non-robust features. We validate this understanding with extensive experiments, and provide a holistic view of robust overfitting from the dynamics of both the two game players. This understanding further inspires us to alleviate robust overfitting by rebalancing the two players by either regularizing the trainer's capacity or improving the attack strength. Experiments show that the proposed ReBalanced Adversarial Training (ReBAT) can attain good robustness and does not suffer from robust overfitting even after very long training. Code is available at https://github.com/PKU-ML/ReBAT.
****
Should I Stop or Should I Go: Early Stopping with Heterogeneous Populations
Hammaad Adam, Fan Yin, Huibin Hu, Neil Tenenholtz, Lorin Crawford, Lester Mackey, Allison Koenecke
Randomized experiments often need to be stopped prematurely due to the treatment having an unintended harmful effect. Existing methods that determine when to stop an experiment early are typically applied to the data in aggregate and do not account for treatment effect heterogeneity. In this paper, we study the early s

topping of experiments for harm on heterogeneous populations. We first establish that current methods often fail to stop experiments when the treatment harms a minority group of participants. We then use causal machine learning to develop CLASH, the first broadly-applicable method for heterogeneous early stopping. We demonstrate CLASH's performance on simulated and real data and show that it yields effective early stopping for both clinical trials and A/B tests.
****
Adaptive Privacy Composition for Accuracy-first Mechanisms
Ryan M. Rogers, Gennady Samorodnitsk, Steven Z. Wu, Aaditya Ramdas
Although there has been work to develop ex-post private mechanisms from Ligett et al. '17 and Whitehouse et al '22 that seeks to provide privacy guarantees subject to a target level of accuracy, there was not a way to use them in conjunction with differentially private mechanisms. Furthermore, there has yet to be work in developing a theory for how these ex-post privacy mechanisms compose, so that we can track the accumulated privacy over several mechanisms. We develop privacy filters that allow an analyst to adaptively switch between differentially private mechanisms and ex-post private mechanisms subject to an overall privacy loss guarantee. We show that using a particular ex-post private mechanism --- noise reduction mechanisms --- can substantially outperform baseline approaches that use existing privacy loss composition bounds. We use the common task of returning as many counts as possible subject to a relative error guarantee and an overall privacy budget as a motivating example.
****
CaMP: Causal Multi-policy Planning for Interactive Navigation in Multi-room Scenes
Xiaohan Wang, Yuehu Liu, Xinhang Song, Beibei Wang, Shuqiang Jiang
Visual navigation has been widely studied under the assumption that there may be several clear routes to reach the goal. However, in more practical scenarios such as a house with several messy rooms, there may not. Interactive Navigation (InterNav) considers agents navigating to their goals more effectively with object interactions, posing new challenges of learning interaction dynamics and extra action space. Previous works learn single vision-to-action policy with the guidance of designed representations. However, the causality between actions and outcomes is prone to be confounded when the attributes of obstacles are diverse and hard to measure. Learning policy for long-term action planning in complex scenes also leads to extensive inefficient exploration. In this paper, we introduce a causal diagram of InterNav clarifying the confounding bias caused by obstacles. To address the problem, we propose a multi-policy model that enables the exploration of counterfactual interactions as well as reduces unnecessary exploration. We develop a large-scale dataset containing 600k task episodes in 12k multi-room scenes based on the ProcTHOR simulator and showcase the effectiveness of our method with the evaluations on our dataset.
****
DiffSketcher: Text Guided Vector Sketch Synthesis through Latent Diffusion Models
XiMing Xing, Chuang Wang, Haitao Zhou, Jing Zhang, Qian Yu, Dong Xu
Even though trained mainly on images, we discover that pretrained diffusion models show impressive power in guiding sketch synthesis. In this paper, we present DiffSketcher, an innovative algorithm that creates \textit{vectorized} free-hand sketches using natural language input. DiffSketcher is developed based on a pre-trained text-to-image diffusion model. It performs the task by directly optimizing a set of Bézier curves with an extended version of the score distillation sampling (SDS) loss, which allows us to use a raster-level diffusion model as a prior for optimizing a parametric vectorized sketch generator. Furthermore, we explore attention maps embedded in the diffusion model for effective stroke initialization to speed up the generation process. The generated sketches demonstrate multiple levels of abstraction while maintaining recognizability, underlying structure, and essential visual details of the subject drawn. Our experiments show that DiffSketcher achieves greater quality than prior work. The code and demo of DiffSketcher can be found at https://ximinng.github.io/DiffSketcher-project/.

****
Mix-of-Show: Decentralized Low-Rank Adaptation for Multi-Concept Customization of Diffusion Models
Yuchao Gu, Xintao Wang, Jay Zhangjie Wu, Yujun Shi, Yunpeng Chen, Zihan Fan, WUYOU XIAO, Rui Zhao, Shuning Chang, Weijia Wu, Yixiao Ge, Ying Shan, Mike Zheng Shou
Public large-scale text-to-image diffusion models, such as Stable Diffusion, have gained significant attention from the community. These models can be easily customized for new concepts using low-rank adaptations (LoRAs). However, the utilization of multiple-concept LoRAs to jointly support multiple customized concepts presents a challenge. We refer to this scenario as decentralized multi-concept customization, which involves single-client concept tuning and center-node concept fusion. In this paper, we propose a new framework called Mix-of-Show that addresses the challenges of decentralized multi-concept customization, including concept conflicts resulting from existing single-client LoRA tuning and identity loss during model fusion. Mix-of-Show adopts an embedding-decomposed LoRA (ED-LoRA) for single-client tuning and gradient fusion for the center node to preserve the in-domain essence of single concepts and support theoretically limitless concept fusion. Additionally, we introduce regionally controllable sampling, which extends spatially controllable sampling (e.g., ControlNet and T2I-Adapter) to address attribute binding and missing object problems in multi-concept sampling. Extensive experiments demonstrate that Mix-of-Show is capable of composing multiple customized concepts with high fidelity, including characters, objects, and scenes.
****
ImageReward: Learning and Evaluating Human Preferences for Text-to-Image Generation
Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, Yuxiao Dong
We present a comprehensive solution to learn and improve text-to-image models from human preference feedback.To begin with, we build ImageReward---the first general-purpose text-to-image human preference reward model---to effectively encode human preferences.Its training is based on our systematic annotation pipeline including rating and ranking, which collects 137k expert comparisons to date.In human evaluation, ImageReward outperforms existing scoring models and metrics, making it a promising automatic metric for evaluating text-to-image synthesis.On top of it, we propose Reward Feedback Learning (ReFL), a direct tuning algorithm to optimize diffusion models against a scorer.Both automatic and human evaluation support ReFL's advantages over compared methods.All code and datasets are provided at \url{https://github.com/THUDM/ImageReward}.
****
To Stay or Not to Stay in the Pre-train Basin: Insights on Ensembling in Transfer Learning
Ildus Sadrtdinov, Dmitrii Pozdeev, Dmitry P. Vetrov, Ekaterina Lobacheva
Transfer learning and ensembling are two popular techniques for improving the performance and robustness of neural networks. Due to the high cost of pre-training, ensembles of models fine-tuned from a single pre-trained checkpoint are often used in practice. Such models end up in the same basin of the loss landscape, which we call the pre-train basin, and thus have limited diversity. In this work, we show that ensembles trained from a single pre-trained checkpoint may be improved by better exploring the pre-train basin, however, leaving the basin results in losing the benefits of transfer learning and in degradation of the ensemble quality. Based on the analysis of existing exploration methods, we propose a more effective modification of the Snapshot Ensembles (SSE) for transfer learning setup, StarSSE, which results in stronger ensembles and uniform model soups.
****
Statistical Limits of Adaptive Linear Models: Low-Dimensional Estimation and Inference
Licong Lin, Mufang Ying, Suvrojit Ghosh, Koulik Khamaru, Cun-Hui Zhang
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Future-Dependent Value-Based Off-Policy Evaluation in POMDPs
Masatoshi Uehara, Haruka Kiyohara, Andrew Bennett, Victor Chernozhukov, Nan Jiang, Nathan Kallus, Chengchun Shi, Wen Sun
We study off-policy evaluation (OPE) for partially observable MDPs (POMDPs) with general function approximation. Existing methods such as sequential importance sampling estimators and fitted-Q evaluation suffer from the curse of horizon in POMDPs. To circumvent this problem, we develop a novel model-free OPE method by introducing future-dependent value functions that take future proxies as inputs. Future-dependent value functions play similar roles as classical value functions in fully-observable MDPs. We derive a new off-policy Bellman equation for future-dependent value functions as conditional moment equations that use history proxies as instrumental variables. We further propose a minimax learning method to learn future-dependent value functions using the new Bellman equation. We obtain the PAC result, which implies our OPE estimator is close to the true policy value as long as futures and histories contain sufficient information about latent states, and the Bellman completeness. Our code is available at https://github.com/aiueola/neurips2023-future-dependent-ope
****

Visual Explanations of Image-Text Representations via Multi-Modal Information Bottleneck Attribution
Ying Wang, Tim G. J. Rudner, Andrew G. Wilson
Vision-language pretrained models have seen remarkable success, but their application to safety-critical settings is limited by their lack of interpretability. To improve the interpretability of vision-language models such as CLIP, we propose a multi-modal information bottleneck (M2IB) approach that learns latent representations that compress irrelevant information while preserving relevant visual and textual features. We demonstrate how M2IB can be applied to attribution analysis of vision-language pretrained models, increasing attribution accuracy and improving the interpretability of such models when applied to safety-critical domains such as healthcare. Crucially, unlike commonly used unimodal attribution methods, M2IB does not require ground truth labels, making it possible to audit representations of vision-language pretrained models when multiple modalities but no ground-truth data is available. Using CLIP as an example, we demonstrate the effectiveness of M2IB attribution and show that it outperforms gradient-based, perturbation-based, and attention-based attribution methods both qualitatively and quantitatively.
****

PlanE: Representation Learning over Planar Graphs
Radoslav Dimitrov, Zeyang Zhao, Ralph Abboud, Ismail Ceylan
Graph neural networks are prominent models for representation learning over graphs, where the idea is to iteratively compute representations of nodes of an input graph through a series of transformations in such a way that the learned graph function is isomorphism-invariant on graphs, which makes the learned representations graph invariants. On the other hand, it is well-known that graph invariants learned by these class of models are incomplete: there are pairs of non-isomorphic graphs which cannot be distinguished by standard graph neural networks. This is unsurprising given the computational difficulty of graph isomorphism testing on general graphs, but the situation begs to differ for special graph classes, for which efficient graph isomorphism testing algorithms are known, such as planar graphs. The goal of this work is to design architectures for efficiently learning complete invariants of planar graphs. Inspired by the classical planar graph isomorphism algorithm of Hopcroft and Tarjan, we propose PlanE as a framework for planar representation learning. PlanE includes architectures which can learn complete invariants over planar graphs while remaining practically scalable. We empirically validate the strong performance of the resulting model architectures on well-known planar graph benchmarks, achieving multiple state-of-the-art re

sults.
****
Optimal Regret Is Achievable with Bounded Approximate Inference Error: An Enhanced Bayesian Upper Confidence Bound Framework
Ziyi Huang, Henry Lam, Amirhossein Meisami, Haofeng Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Any-to-Any Generation via Composable Diffusion
Zineng Tang, Ziyi Yang, Chenguang Zhu, Michael Zeng, Mohit Bansal
We present Composable Diffusion (CoDi), a novel generative model capable of generating any combination of output modalities, such as language, image, video, or audio, from any combination of input modalities. Unlike existing generative AI systems, CoDi can generate multiple modalities in parallel and its input is not limited to a subset of modalities like text or image. Despite the absence of training datasets for many combinations of modalities, we propose to align modalities in both the input and output space. This allows CoDi to freely condition on any input combination and generate any group of modalities, even if they are not present in the training data. CoDi employs a novel composable generation strategy which involves building a shared multimodal space by bridging alignment in the diffusion process, enabling the synchronized generation of intertwined modalities, such as temporally aligned video and audio. Highly customizable and flexible, CoDi achieves strong joint-modality generation quality, and outperforms or is on par with the unimodal state-of-the-art for single-modality synthesis.
****
Rank-DETR for High Quality Object Detection
Yifan Pu, Weicong Liang, Yiduo Hao, YUHUI YUAN, Yukang Yang, Chao Zhang, Han Hu, Gao Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Towards Efficient Pre-Trained Language Model via Feature Correlation Distillation
Kun Huang, Xin Guo, Meng Wang
Knowledge Distillation (KD) has emerged as a promising approach for compressing large Pre-trained Language Models (PLMs). The performance of KD relies on how to effectively formulate and transfer the knowledge from the teacher model to the student model. Prior arts mainly focus on directly aligning output features from the transformer block, which may impose overly strict constraints on the student model's learning process and complicate the training process by introducing extra parameters and computational cost. Moreover, our analysis indicates that the different relations within self-attention, as adopted in other works, involves more computation complexities and can easily be constrained by the number of heads, potentially leading to suboptimal solutions. To address these issues, we propose a novel approach that builds relationships directly from output features. Specifically, we introduce token-level and sequence-level relations concurrently to fully exploit the knowledge from the teacher model. Furthermore, we propose a correlation-based distillation loss to alleviate the exact match properties inherent in traditional KL divergence or MSE loss functions. Our method, dubbed FCD, presents a simple yet effective method to compress various architectures (BERT, RoBERTa, and GPT) and model sizes (base-size and large-size). Extensive experimental results demonstrate that our distilled, smaller language models significantly surpass existing KD methods across various NLP tasks.
****
Simple and Asymmetric Graph Contrastive Learning without Augmentations
Teng Xiao, Huaisheng Zhu, Zhengyu Chen, Suhang Wang

Graph Contrastive Learning (GCL) has shown superior performance in representation learning in graph-structured data. Despite their success, most existing GCL methods rely on prefabricated graph augmentation and homophily assumptions. Thus, they fail to generalize well to heterophilic graphs where connected nodes may have different class labels and dissimilar features. In this paper, we study the problem of conducting contrastive learning on homophilic and heterophilic graphs. We find that we can achieve promising performance simply by considering an asymmetric view of the neighboring nodes. The resulting simple algorithm, Asymmetric Contrastive Learning for Graphs (GraphACL), is easy to implement and does not rely on graph augmentations and homophily assumptions. We provide theoretical and empirical evidence that GraphACL can capture one-hop local neighborhood information and two-hop monophily similarity, which are both important for modeling heterophilic graphs. Experimental results show that the simple GraphACL significantly outperforms state-of-the-art graph contrastive learning and self-supervised learning methods on homophilic and heterophilic graphs. The code of GraphACL is available at https://github.com/tengxiao1/GraphACL.
****
Large-Scale Distributed Learning via Private On-Device LSH
Tahseen Rabbani, Marco Bornstein, Furong Huang
Locality-sensitive hashing (LSH) based frameworks have been used efficiently to select weight vectors in a dense hidden layer with high cosine similarity to an input, enabling dynamic pruning. While this type of scheme has been shown to improve computational training efficiency, existing algorithms require repeated randomized projection of the full layer weight, which is impractical for computational- and memory-constrained devices. In a distributed setting, deferring LSH analysis to a centralized host is (i) slow if the device cluster is large and (ii) requires access to input data which is forbidden in a federated context. Using a new family of hash functions, we develop the first private, personalized, and memory-efficient on-device LSH framework.Our framework enables privacy and personalization by allowing each device to generate hash tables, without the help of a central host, using device-specific hashing hyper-parameters (e.g., number of hash tables or hash length).Hash tables are generated with a compressed set of the full weights, and can be serially generated and discarded if the process is memory-intensive.This allows devices to avoid maintaining (i) the fully-sized model and (ii) large amounts of hash tables in local memory for LSH analysis. We prove several statistical and sensitivity properties of our hash functions, and experimentally demonstrate that our framework is competitive in training large scale recommender networks compared to other LSH frameworks which assume unrestricted on-device capacity.
****
Facilitating Graph Neural Networks with Random Walk on Simplicial Complexes
Cai Zhou, Xiyuan Wang, Muhan Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Koopman Kernel Regression
Petar Bevanda, Max Beier, Armin Lederer, Stefan Sosnowski, Eyke Hüllermeier, Sandra Hirche
Many machine learning approaches for decision making, such as reinforcement learning, rely on simulators or predictive models to forecast the time-evolution of quantities of interest, e.g., the state of an agent or the reward of a policy. Forecasts of such complex phenomena are commonly described by highly nonlinear dynamical systems, making their use in optimization-based decision-making challenging.Koopman operator theory offers a beneficial paradigm for addressing this problem by characterizing forecasts via linear time-invariant (LTI) ODEs, turning multi-step forecasts into sparse matrix multiplication.Though there exists a variety of learning approaches, they usually lack crucial learning-theoretic guarantees, making the behavior of the obtained models with increasing data and dimensi

onality unclear.We address the aforementioned by deriving a universal Koopman-invariant reproducing kernel Hilbert space (RKHS) that solely spans transformations into LTI dynamical systems. The resulting Koopman Kernel Regression (KKR) framework enables the use of statistical learning tools from function approximation for novel convergence results and generalization error bounds under weaker assumptions than existing work. Our experiments demonstrate superior forecasting performance compared to Koopman operator and sequential data predictors in RKHS.
****

Diffusion Self-Guidance for Controllable Image Generation
Dave Epstein, Allan Jabri, Ben Poole, Alexei Efros, Aleksander Holynski
Large-scale generative models are capable of producing high-quality images from detailed prompts. However, many aspects of an image are difficult or impossible to convey through text. We introduce self-guidance, a method that provides precise control over properties of the generated image by guiding the internal representations of diffusion models. We demonstrate that the size, location, and appearance of objects can be extracted from these representations, and show how to use them to steer the sampling process. Self-guidance operates similarly to standard classifier guidance, but uses signals present in the pretrained model itself, requiring no additional models or training. We demonstrate the flexibility and effectiveness of self-guided generation through a wide range of challenging image manipulations, such as modifying the position or size of a single object (keeping the rest of the image unchanged), merging the appearance of objects in one image with the layout of another, composing objects from multiple images into one, and more. We also propose a new method for reconstruction using self-guidance, which allows extending our approach to editing real images.
****

Neural (Tangent Kernel) Collapse
Mariia Seleznova, Dana Weitzner, Raja Giryes, Gitta Kutyniok, Hung-Hsu Chou
This work bridges two important concepts: the Neural Tangent Kernel (NTK), which captures the evolution of deep neural networks (DNNs) during training, and the Neural Collapse (NC) phenomenon, which refers to the emergence of symmetry and structure in the last-layer features of well-trained classification DNNs. We adopt the natural assumption that the empirical NTK develops a block structure aligned with the class labels, i.e., samples within the same class have stronger correlations than samples from different classes. Under this assumption, we derive the dynamics of DNNs trained with mean squared (MSE) loss and break them into interpretable phases. Moreover, we identify an invariant that captures the essence of the dynamics, and use it to prove the emergence of NC in DNNs with block-structured NTK. We provide large-scale numerical experiments on three common DNN architectures and three benchmark datasets to support our theory.
****

Fast Optimal Locally Private Mean Estimation via Random Projections
Hilal Asi, Vitaly Feldman, Jelani Nelson, Huy Nguyen, Kunal Talwar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Expert load matters: operating networks at high accuracy and low manual effort
Sara Sangalli, Ertunc Erdil, Ender Konukoglu
In human-AI collaboration systems for critical applications, in order to ensure minimal error, users should set an operating point based on model confidence to determine when the decision should be delegated to human experts. Samples for which model confidence is lower than the operating point would be manually analysed by experts to avoid mistakes.Such systems can become truly useful only if they consider two aspects: models should be confident only for samples for which they are accurate, and the number of samples delegated to experts should be minimized.The latter aspect is especially crucial for applications where available expert time is limited and expensive, such as healthcare. The trade-off between the model accuracy and the number of samples delegated to experts can be represented

by a curve that is similar to an ROC curve, which we refer to as confidence operating characteristic (COC) curve. In this paper, we argue that deep neural networks should be trained by taking into account both accuracy and expert load and, to that end, propose a new complementary loss function for classification that maximizes the area under this COC curve.This promotes simultaneously the increase in network accuracy and the reduction in number of samples delegated to humans.We perform experiments on multiple computer vision and medical image datasets for classification.Our results demonstrate that the proposed loss improves classification accuracy and delegates less number of decisions to experts, achieves better out-of-distribution samples detection and on par calibration performance compared to existing loss functions.
****
PRODIGY: Enabling In-context Learning Over Graphs
Qian Huang, Hongyu Ren, Peng Chen, Gregor Kržmanc, Daniel Zeng, Percy S. Liang, Jure Leskovec
In-context learning is the ability of a pretrained model to adapt to novel and diverse downstream tasks by conditioning on prompt examples, without optimizing any parameters. While large language models have demonstrated this ability, how in-context learning could be performed over graphs is unexplored. In this paper, we develop \textbf{Pr}etraining \textbf{O}ver \textbf{D}iverse \textbf{I}n-Context \textbf{G}raph S\textbf{y}stems (PRODIGY), the first pretraining framework that enables in-context learning over graphs. The key idea of our framework is to formulate in-context learning over graphs with a novel \emph{prompt graph} representation, which connects prompt examples and queries. We then propose a graph neural network architecture over the prompt graph and a corresponding family of in-context pretraining objectives. With PRODIGY, the pretrained model can directly perform novel downstream classification tasks on unseen graphs via in-context learning. We provide empirical evidence of the effectiveness of our framework by showcasing its strong in-context learning performance on tasks involving citation networks and knowledge graphs. Our approach outperforms the in-context learning accuracy of contrastive pretraining baselines with hard-coded adaptation by 18\% on average across all setups. Moreover, it also outperforms standard finetuning with limited data by 33\% on average with in-context learning.
****
Towards Automated Circuit Discovery for Mechanistic Interpretability
Arthur Conmy, Augustine Mavor-Parker, Aengus Lynch, Stefan Heimersheim, Adrià Garriga-Alonso
Through considerable effort and intuition, several recent works have reverse-engineered nontrivial behaviors oftransformer models. This paper systematizes the mechanistic interpretability process they followed. First, researcherschoose a metric and dataset that elicit the desired model behavior. Then, they apply activation patching to find whichabstract neural network units are involved in the behavior. By varying the dataset, metric, and units underinvestigation, researchers can understand the functionality of each component.We automate one of the process' steps: finding the connections between the abstract neural network units that form a circuit. We propose several algorithms and reproduce previous interpretability results to validate them. Forexample, the ACDC algorithm rediscovered 5/5 of the component types in a circuit in GPT-2 Small that computes theGreater-Than operation. ACDC selected 68 of the 32,000 edges in GPT-2 Small, all of which were manually found byprevious work. Our code is available at https://github.com/ArthurConmy/Automatic-Circuit-Discovery
****
Find What You Want: Learning Demand-conditioned Object Attribute Space for Demand-driven Navigation
Hongcheng Wang, Andy Guan Hong Chen, Xiaoqi Li, Mingdong Wu, Hao Dong
The task of Visual Object Navigation (VON) involves an agent's ability to locate a particular object within a given scene. To successfully accomplish the VON task, two essential conditions must be fulfiled: 1) the user knows the name of the desired object; and 2) the user-specified object actually is present within the scene. To meet these conditions, a simulator can incorporate predefined object

names and positions into the metadata of the scene. However, in real-world scenarios, it is often challenging to ensure that these conditions are always met. Humans in an unfamiliar environment may not know which objects are present in the scene, or they may mistakenly specify an object that is not actually present. Nevertheless, despite these challenges, humans may still have a demand for an object, which could potentially be fulfilled by other objects present within the scene in an equivalent manner. Hence, this paper proposes Demand-driven Navigation (DDN), which leverages the user's demand as the task instruction and prompts the agent to find an object which matches the specified demand. DDN aims to relax the stringent conditions of VON by focusing on fulfilling the user's demand rather than relying solely on specified object names. This paper proposes a method of acquiring textual attribute features of objects by extracting common sense knowledge from a large language model (LLM). These textual attribute features are subsequently aligned with visual attribute features using Contrastive Language-Image Pre-training (CLIP). Incorporating the visual attribute features as prior knowledge, enhances the navigation process. Experiments on AI2Thor with the ProcThor dataset demonstrate that the visual attribute features improve the agent's navigation performance and outperform the baseline methods commonly used in the VON and VLN task and methods with LLMs. The codes and demonstrations can be viewed at https://sites.google.com/view/demand-driven-navigation.
****

On the Convergence of No-Regret Learning Dynamics in Time-Varying Games

Ioannis Anagnostides, Ioannis Panageas, Gabriele Farina, Tuomas Sandholm

Most of the literature on learning in games has focused on the restrictive setting where the underlying repeated game does not change over time. Much less is known about the convergence of no-regret learning algorithms in dynamic multiagent settings. In this paper, we characterize the convergence of optimistic gradient descent (OGD) in time-varying games. Our framework yields sharp convergence bounds for the equilibrium gap of OGD in zero-sum games parameterized on natural variation measures of the sequence of games, subsuming known results for static games. Furthermore, we establish improved second-order variation bounds under strong convexity-concavity, as long as each game is repeated multiple times. Our results also apply to time-varying general-sum multi-player games via a bilinear formulation of correlated equilibria, which has novel implications for meta-learning and for obtaining refined variation-dependent regret bounds, addressing questions left open in prior papers. Finally, we leverage our framework to also provide new insights on dynamic regret guarantees in static games.
****

Getting ViT in Shape: Scaling Laws for Compute-Optimal Model Design

Ibrahim M. Alabdulmohsin, Xiaohua Zhai, Alexander Kolesnikov, Lucas Beyer

Scaling laws have been recently employed to derive compute-optimal model size (number of parameters) for a given compute duration. We advance and refine such methods to infer compute-optimal model shapes, such as width and depth, and successfully implement this in vision transformers. Our shape-optimized vision transformer, SoViT, achieves results competitive with models that exceed twice its size, despite being pre-trained with an equivalent amount of compute. For example, SoViT-400m/14 achieves 90.3% fine-tuning accuracy on ILSRCV2012, surpassing the much larger ViT-g/14 and approaching ViT-G/14 under identical settings, with also less than half the inference cost. We conduct a thorough evaluation across multiple tasks, such as image classification, captioning, VQA and zero-shot transfer, demonstrating the effectiveness of our model across a broad range of domains and identifying limitations. Overall, our findings challenge the prevailing approach of blindly scaling up vision models and pave a path for a more informed scaling.
****

Expressive Sign Equivariant Networks for Spectral Geometric Learning

Derek Lim, Joshua Robinson, Stefanie Jegelka, Haggai Maron

Recent work has shown the utility of developing machine learning models that respect the structure and symmetries of eigenvectors. These works promote sign invariance, since for any eigenvector v the negation -v is also an eigenvector. Howe

ver, we show that sign invariance is theoretically limited for tasks such as building orthogonally equivariant models and learning node positional encodings for link prediction in graphs. In this work, we demonstrate the benefits of sign equivariance for these tasks. To obtain these benefits, we develop novel sign equivariant neural network architectures. Our models are based on a new analytic characterization of sign equivariant polynomials and thus inherit provable expressiveness properties. Controlled synthetic experiments show that our networks can achieve the theoretically predicted benefits of sign equivariant models.
****

FLAIR : a Country-Scale Land Cover Semantic Segmentation Dataset From Multi-Source Optical Imagery

Anatol Garioud, Nicolas Gonthier, Loic Landrieu, Apolline De Wit, Marion Valette, Marc Poupée, Sebastien Giordano, boris Wattrelos

We introduce the French Land cover from Aerospace ImageRy (FLAIR), an extensive dataset from the French National Institute of Geographical and Forest Information (IGN) that provides a unique and rich resource for large-scale geospatial analysis. FLAIR contains high-resolution aerial imagery with a ground sample distance of 20 cm and over 20 billion individually labeled pixels for precise land-cover classification. The dataset also integrates temporal and spectral data from optical satellite time series. FLAIR thus combines data with varying spatial, spectral, and temporal resolutions across over 817 km² of acquisitions representing the full landscape diversity of France. This diversity makes FLAIR a valuable resource for the development and evaluation of novel methods for large-scale land-cover semantic segmentation and raises significant challenges in terms of computer vision, data fusion, and geospatial analysis. We also provide powerful uni- and multi-sensor baseline models that can be employed to assess algorithm's performance and for downstream applications.
****

Strategic Data Sharing between Competitors

Nikita Tsoy, Nikola Konstantinov

Collaborative learning techniques have significantly advanced in recent years, enabling private model training across multiple organizations. Despite this opportunity, firms face a dilemma when considering data sharing with competitors—while collaboration can improve a company’s machine learning model, it may also benefit competitors and hence reduce profits. In this work, we introduce a general framework for analyzing this data-sharing trade-off. The framework consists of three components, representing the firms’ production decisions, the effect of additional data on model quality, and the data-sharing negotiation process, respectively. We then study an instantiation of the framework, based on a conventional market model from economic theory, to identify key factors that affect collaboration incentives. Our findings indicate a profound impact of market conditions on the data-sharing incentives. In particular, we find that reduced competition, in terms of the similarities between the firms’ products, and harder learning tasks foster collaboration.
****

Optimal Time Complexities of Parallel Stochastic Optimization Methods Under a Fixed Computation Model

Alexander Tyurin, Peter Richtarik

Parallelization is a popular strategy for improving the performance of methods. Optimization methods are no exception: design of efficient parallel optimization methods and tight analysis of their theoretical properties are important research endeavors. While the minimax complexities are well known for sequential optimization methods, the theory of parallel optimization methods is less explored. In this paper, we propose a new protocol that generalizes the classical oracle framework approach. Using this protocol, we establish minimax complexities for parallel optimization methods that have access to an unbiased stochastic gradient oracle with bounded variance. We consider a fixed computation model characterized by each worker requiring a fixed but worker-dependent time to calculate stochastic gradient. We prove lower bounds and develop optimal algorithms that attain them. Our results have surprising consequences for the literature of asynchronous

optimization methods.
****
An ε-Best-Arm Identification Algorithm for Fixed-Confidence and Beyond
Marc Jourdan, Rémy Degenne, Emilie Kaufmann
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Debiased and Denoised Entity Recognition from Distant Supervision
Haobo Wang, Yiwen Dong, Ruixuan Xiao, Fei Huang, Gang Chen, Junbo Zhao
While distant supervision has been extensively explored and exploited in NLP tasks like named entity recognition, a major obstacle stems from the inevitable noisy distant labels tagged unsupervisedly. A few past works approach this problem by adopting a self-training framework with a sample-selection mechanism. In this work, we innovatively identify two types of biases that were omitted by prior work, and these biases lead to inferior performance of the distant-supervised NER setup. First, we characterize the noise concealed in the distant labels as highly structural rather than fully randomized. Second, the self-training framework would ubiquitously introduce an inherent bias that causes erroneous behavior in both sample selection and eventually prediction. To cope with these problems, we propose a novel self-training framework, dubbed DesERT. This framework augments the conventional NER predicative pathway to a dual form that effectively adapts the sample-selection process to conform to its innate distributional-bias structure. The other crucial component of DesERT composes a debiased module aiming to enhance the token representations, hence the quality of the pseudo-labels. Extensive experiments are conducted to validate the DesERT. The results show that our framework establishes a new state-of-art performance, it achieves a +2.22% average F1 score improvement on five standardized benchmarking datasets. Lastly, DesERT demonstrates its effectiveness under a new DSNER benchmark where additional distant supervision comes from the ChatGPT model.
****
Accelerated Training via Incrementally Growing Neural Networks using Variance Transfer and Learning Rate Adaptation
Xin Yuan, Pedro Savarese, Michael Maire
We develop an approach to efficiently grow neural networks, within which parameterization and optimization strategies are designed by considering their effects on the training dynamics. Unlike existing growing methods, which follow simple replication heuristics or utilize auxiliary gradient-based local optimization, we craft a parameterization scheme which dynamically stabilizes weight, activation, and gradient scaling as the architecture evolves, and maintains the inference functionality of the network. To address the optimization difficulty resulting from imbalanced training effort distributed to subnetworks fading in at different growth phases, we propose a learning rate adaption mechanism that rebalances the gradient contribution of these separate subcomponents. Experiments show that our method achieves comparable or better accuracy than training large fixed-size models, while saving a substantial portion of the original training computation budget. We demonstrate that these gains translate into real wall-clock training speedups.
****
Likelihood-Based Diffusion Language Models
Ishaan Gulrajani, Tatsunori B. Hashimoto
Despite a growing interest in diffusion-based language models, existing work has not shown that these models can attain nontrivial likelihoods on standard language modeling benchmarks. In this work, we take the first steps towards closing the likelihood gap between autoregressive and diffusion-based language models, with the goal of building and releasing a diffusion model which outperforms a small but widely-known autoregressive model. We pursue this goal through algorithmic improvements, scaling laws, and increased compute. On the algorithmic front, we

introduce several methodological improvements for the maximum-likelihood training of diffusion language models. We then study scaling laws for our diffusion models and find compute-optimal training regimes which differ substantially from autoregressive models. Using our methods and scaling analysis, we train and release Plaid 1B, a large diffusion language model which outperforms GPT-2 124M in likelihood on benchmark datasets and generates fluent samples in unconditional and zero-shot control settings.
****
Structural Pruning for Diffusion Models
Gongfan Fang, Xinyin Ma, Xinchao Wang
Generative modeling has recently undergone remarkable advancements, primarily propelled by the transformative implications of Diffusion Probabilistic Models (DPMs). The impressive capability of these models, however, often entails significant computational overhead during both training and inference. To tackle this challenge, we present Diff-Pruning, an efficient compression method tailored for learning lightweight diffusion models from pre-existing ones, without the need for extensive re-training. The essence of Diff-Pruning is encapsulated in a Taylor expansion over pruned timesteps, a process that disregards non-contributory diffusion steps and ensembles informative gradients to identify important weights. Our empirical assessment, undertaken across several datasets highlights two primary benefits of our proposed method: 1) Efficiency: it enables approximately a 50\% reduction in FLOPs at a mere 10% to 20% of the original training expenditure; 2) Consistency: the pruned diffusion models inherently preserve generative behavior congruent with their pre-trained models.
****
Fast and Regret Optimal Best Arm Identification: Fundamental Limits and Low-Complexity Algorithms
Qining Zhang, Lei Ying
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Simultaneous embedding of multiple attractor manifolds in a recurrent neural network using constrained gradient optimization
Haggai Agmon, Yoram Burak
The storage of continuous variables in working memory is hypothesized to be sustained in the brain by the dynamics of recurrent neural networks (RNNs) whose steady states form continuous manifolds. In some cases, it is thought that the synaptic connectivity supports multiple attractor manifolds, each mapped to a different context or task. For example, in hippocampal area CA3, positions in distinct environments are represented by distinct sets of population activity patterns, each forming a continuum. It has been argued that the embedding of multiple continuous attractors in a single RNN inevitably causes detrimental interference: quenched noise in the synaptic connectivity disrupts the continuity of each attractor, replacing it by a discrete set of steady states that can be conceptualized as lying on local minima of an abstract energy landscape. Consequently, population activity patterns exhibit systematic drifts towards one of these discrete minima, thereby degrading the stored memory over time. Here we show that it is possible to dramatically attenuate these detrimental interference effects by adjusting the synaptic weights. Synaptic weight adjustment are derived from a loss function that quantifies the roughness of the energy landscape along each of the embedded attractor manifolds. By minimizing this loss function, the stability of states can be dramatically improved, without compromising the capacity.
****
Enhancing Adversarial Contrastive Learning via Adversarial Invariant Regularization
Xilie Xu, Jingfeng ZHANG, Feng Liu, Masashi Sugiyama, Mohan S. Kankanhalli
Adversarial contrastive learning (ACL) is a technique that enhances standard contrastive learning (SCL) by incorporating adversarial data to learn a robust repr

esentation that can withstand adversarial attacks and common corruptions without
 requiring costly annotations. To improve transferability, the existing work int
roduced the standard invariant regularization (SIR) to impose style-independence
 property to SCL, which can exempt the impact of nuisance style factors in the s
tandard representation. However, it is unclear how the style-independence proper
ty benefits ACL-learned robust representations. In this paper, we leverage the t
echnique of causal reasoning to interpret the ACL and propose adversarial invari
ant regularization (AIR) to enforce independence from style factors. We regulate
 the ACL using both SIR and AIR to output the robust representation. Theoretical
ly, we show that AIR implicitly encourages the representational distance between
 different views of natural data and their adversarial variants to be independen
t of style factors. Empirically, our experimental results show that invariant re
gularization significantly improves the performance of state-of-the-art ACL meth
ods in terms of both standard generalization and robustness on downstream tasks.
 To the best of our knowledge, we are the first to apply causal reasoning to int
erpret ACL and develop AIR for enhancing ACL-learned robust representations. Our
 source code is at https://github.com/GodXuxilie/EnhancingACLvia_AIR.
****
Best Arm Identification with Fixed Budget: A Large Deviation Perspective
Po-An Wang, Ruo-Chun Tzeng, Alexandre Proutiere
We consider the problem of identifying the best arm in stochastic Multi-Armed Ba
ndits (MABs) using a fixed sampling budget. Characterizing the minimal instance-
specific error probability for this problem constitutes one of the important rem
aining open problems in MABs. When arms are selected using a static sampling str
ategy, the error probability decays exponentially with the number of samples at
a rate that can be explicitly derived via Large Deviation techniques. Analyzing
the performance of algorithms with adaptive sampling strategies is however much
more challenging. In this paper, we establish a connection between the Large Dev
iation Principle (LDP) satisfied by the empirical proportions of arm draws and t
hat satisfied by the empirical arm rewards. This connection holds for any adapti
ve algorithm, and is leveraged (i) to improve error probability upper bounds of
some existing algorithms, such as the celebrated SR (Successive Rejects) algorit
hm \cite{audibert2010best}, and (ii) to devise and analyze new algorithms. In pa
rticular, we present CR (Continuous Rejects), a truly adaptive algorithm that ca
n reject arms in {\it any} round based on the observed empirical gaps between th
e rewards of various arms. Applying our Large Deviation results, we prove that C
R enjoys better performance guarantees than existing algorithms, including SR. E
xtensive numerical experiments confirm this observation.
****
Full-Atom Protein Pocket Design via Iterative Refinement
ZAIXI ZHANG, Zepu Lu, Hao Zhongkai, Marinka Zitnik, Qi Liu
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Flow Matching for Scalable Simulation-Based Inference
Jonas Wildberger, Maximilian Dax, Simon Buchholz, Stephen Green, Jakob H Macke,
Bernhard Schölkopf
Neural posterior estimation methods based on discrete normalizing flows have bec
ome established tools for simulation-based inference (SBI), but scaling them to
high-dimensional problems can be challenging. Building on recent advances in gen
erative modeling, we here present flow matching posterior estimation (FMPE), a t
echnique for SBI using continuous normalizing flows. Like diffusion models, and
in contrast to discrete flows, flow matching allows for unconstrained architectu
res, providing enhanced flexibility for complex data modalities. Flow matching,
therefore, enables exact density evaluation, fast training, and seamless scalabi
lity to large architectures---making it ideal for SBI. We show that FMPE achieve
s competitive performance on an established SBI benchmark, and then demonstrate
its improved scalability on a challenging scientific problem: for gravitational-

wave inference, FMPE outperforms methods based on comparable discrete flows, red ucing training time by 30\% with substantially improved accuracy. Our work under scores the potential of FMPE to enhance performance in challenging inference sce narios, thereby paving the way for more advanced applications to scientific prob lems.
****

Learning DAGs from Data with Few Root Causes
Panagiotis Misiakos, Chris Wendler, Markus Püschel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Robust Learning for Smoothed Online Convex Optimization with Feedback Delay
Pengfei Li, Jianyi Yang, Adam Wierman, Shaolei Ren
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Scalarization for Multi-Task and Multi-Domain Learning at Scale
Amelie Royer, Tijmen Blankevoort, Babak Ehteshami Bejnordi
Training a single model on multiple input domains and/or output tasks allows for compressing information from multiple sources into a unified backbone hence imp roves model efficiency. It also enables potential positive knowledge transfer ac ross tasks/domains, leading to improved accuracy and data-efficient training. Ho wever, optimizing such networks is a challenge, in particular due to discrepanci es between the different tasks or domains: Despite several hypotheses and soluti ons proposed over the years, recent work has shown that uniform scalarization tr aining, i.e., simply minimizing the average of the task losses, yields on-par p erformance with more costly SotA optimization methods. This raises the issue of how well we understand the training dynamics of multi-task and multi-domain netw orks. In this work, we first devise a large-scale unified analysis of multi-doma in and multi-task learning to better understand the dynamics of scalarization ac ross varied task/domain combinations and model sizes. Following these insights, we then propose to leverage population-based training to efficiently search for the optimal scalarization weights when dealing with a large number of tasks or d omains.
****

Causal discovery from observational and interventional data across multiple envi ronments
Adam Li, Amin Jaber, Elias Bareinboim
A fundamental problem in many sciences is the learning of causal structure under lying a system, typically through observation and experimentation. Commonly, one even collects data across multiple domains, such as gene sequencing from differ ent labs, or neural recordings from different species. Although there exist meth ods for learning the equivalence class of causal diagrams from observational and experimental data, they are meant to operate in a single domain. In this paper, we develop a fundamental approach to structure learning in non-Markovian system s (i.e. when there exist latent confounders) leveraging observational and interv entional data collected from multiple domains. Specifically, we start by showing that learning from observational data in multiple domains is equivalent to lear ning from interventional data with unknown targets in a single domain. But there are also subtleties when considering observational and experimental data. Using causal invariances derived from do-calculus, we define a property called S-Mark ov that connects interventional distributions from multiple-domains to graphical criteria on a selection diagram. Leveraging the S-Markov property, we introduce a new constraint-based causal discovery algorithm, S-FCI, that can learn from o bservational and interventional data from different domains. We prove that the a lgorithm is sound and subsumes existing constraint-based causal discovery algori

thms.
****
Beyond Normal: On the Evaluation of Mutual Information Estimators
Pawe■ Czy■, Frederic Grabowski, Julia Vogt, Niko Beerenwinkel, Alexander Marx
Mutual information is a general statistical dependency measure which has found a
pplications in representation learning, causality, domain generalization and com
putational biology. However, mutual information estimators are typically evaluat
ed on simple families of probability distributions, namely multivariate normal d
istribution and selected distributions with one-dimensional random variables. In
 this paper, we show how to construct a diverse family of distributions with kno
wn ground-truth mutual information and propose a language-independent benchmarki
ng platform for mutual information estimators. We discuss the general applicabil
ity and limitations of classical and neural estimators in settings involving hig
h dimensions, sparse interactions, long-tailed distributions, and high mutual in
formation. Finally, we provide guidelines for practitioners on how to select app
ropriate estimator adapted to the difficulty of problem considered and issues on
e needs to consider when applying an estimator to a new data set.
****
Structured Semidefinite Programming for Recovering Structured Preconditioners
Arun Jambulapati, Jerry Li, Christopher Musco, Kirankumar Shiragur, Aaron Sidfor
d, Kevin Tian
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Certifiably Robust Graph Contrastive Learning
Minhua Lin, Teng Xiao, Enyan Dai, Xiang Zhang, Suhang Wang
Graph Contrastive Learning (GCL) has emerged as a popular unsupervised graph rep
resentation learning method. However, it has been shown that GCL is vulnerable t
o adversarial attacks on both the graph structure and node attributes. Although
empirical approaches have been proposed to enhance the robustness of GCL, the ce
rtifiable robustness of GCL is still remain unexplored. In this paper, we develo
p the first certifiably robust framework in GCL. Specifically, we first propose
a unified criteria to evaluate and certify the robustness of GCL. We then introd
uce a novel technique, RES (Randomized Edgedrop Smoothing), to ensure certifiabl
e robustness for any GCL model, and this certified robustness can be provably pr
eserved in downstream tasks. Furthermore, an effective training method is propos
ed for robust GCL. Extensive experiments on real-world datasets demonstrate the
effectiveness of our proposed method in providing effective certifiable robustne
ss and enhancing the robustness of any GCL model. The source code of RES is avai
lable at https://github.com/ventr1c/RES-GCL.
****
FACE: Evaluating Natural Language Generation with Fourier Analysis of Cross-Entr
opy
Zuhao Yang, Yingfang Yuan, Yang Xu, SHUO ZHAN, Huajun Bai, Kefan Chen
Measuring the distance between machine-produced and human language is a critical
 open problem. Inspired by empirical findings from psycholinguistics on the peri
odicity of entropy in language, we propose FACE, a set of metrics based on Fouri
er Analysis of the estimated Cross-Entropy of language, for measuring the simila
rity between model-generated and human-written languages. Based on an open-ended
 generation task and the experimental data from previous studies, we find that F
ACE can effectively identify the human-model gap, scales with model size, reflec
ts the outcomes of different sampling methods for decoding, correlates well with
 other evaluation metrics and with human judgment scores.
****
3D Copy-Paste: Physically Plausible Object Insertion for Monocular 3D Detection
Yunhao Ge, Hong-Xing Yu, Cheng Zhao, Yuliang Guo, Xinyu Huang, Liu Ren, Laurent
Itti, Jiajun Wu
A major challenge in monocular 3D object detection is the limited diversity and

quantity of objects in real datasets. While augmenting real scenes with virtual objects holds promise to improve both the diversity and quantity of the objects, it remains elusive due to the lack of an effective 3D object insertion method in complex real captured scenes. In this work, we study augmenting complex real indoor scenes with virtual objects for monocular 3D object detection. The main challenge is to automatically identify plausible physical properties for virtual assets (e.g., locations, appearances, sizes, etc.) in cluttered real scenes. To address this challenge, we propose a physically plausible indoor 3D object insertion approach to automatically copy virtual objects and paste them into real scenes. The resulting objects in scenes have 3D bounding boxes with plausible physical locations and appearances. In particular, our method first identifies physically feasible locations and poses for the inserted objects to prevent collisions with the existing room layout. Subsequently, it estimates spatially-varying illumination for the insertion location, enabling the immersive blending of the virtual objects into the original scene with plausible appearances and cast shadows. We show that our augmentation method significantly improves existing monocular 3D object models and achieves state-of-the-art performance. For the first time, we demonstrate that a physically plausible 3D object insertion, serving as a generative data augmentation technique, can lead to significant improvements for discriminative downstream tasks such as monocular 3D object detection. Project website: https://gyhandy.github.io/3D-Copy-Paste/.
****
Laughing Hyena Distillery: Extracting Compact Recurrences From Convolutions
Stefano Massaroli, Michael Poli, Dan Fu, Hermann Kumbong, Rom Parnichkun, David Romero, Aman Timalsina, Quinn McIntyre, Beidi Chen, Atri Rudra, Ce Zhang, Christopher Ré, Stefano Ermon, Yoshua Bengio
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Incomplete Multimodality-Diffused Emotion Recognition
Yuanzhi Wang, Yong Li, Zhen Cui
Human multimodal emotion recognition (MER) aims to perceive and understand human emotions via various heterogeneous modalities, such as language, vision, and acoustic. Compared with unimodality, the complementary information in the multimodalities facilitates robust emotion understanding. Nevertheless, in real-world scenarios, the missing modalities hinder multimodal understanding and result in degraded MER performance. In this paper, we propose an Incomplete Multimodality-Diffused emotion recognition (IMDer) method to mitigate the challenge of MER under incomplete multimodalities. To recover the missing modalities, IMDer exploits the score-based diffusion model that maps the input Gaussian noise into the desired distribution space of the missing modalities and recovers missing data abided by their original distributions. Specially, to reduce semantic ambiguity between the missing and the recovered modalities, the available modalities are embedded as the condition to guide and refine the diffusion-based recovering process. In contrast to previous work, the diffusion-based modality recovery mechanism in IMDer allows to simultaneously reach both distribution consistency and semantic disambiguation. Feature visualization of the recovered modalities illustrates the consistent modality-specific distribution and semantic alignment. Besides, quantitative experimental results verify that IMDer obtains state-of-the-art MER accuracy under various missing modality patterns.
****
Diffusion-Based Probabilistic Uncertainty Estimation for Active Domain Adaptation
Zhekai Du, Jingjing Li
Active Domain Adaptation (ADA) has emerged as an attractive technique for assisting domain adaptation by actively annotating a small subset of target samples. Most ADA methods focus on measuring the target representativeness beyond traditional active learning criteria to handle the domain shift problem, while leaving t

he uncertainty estimation to be performed by an uncalibrated deterministic model. In this work, we introduce a probabilistic framework that captures both data-level and prediction-level uncertainties beyond a point estimate. Specifically, we use variational inference to approximate the joint posterior distribution of latent representation and model prediction. The variational objective of labeled data can be formulated by a variational autoencoder and a latent diffusion classifier, and the objective of unlabeled data can be implemented in a knowledge distillation framework. We utilize adversarial learning to ensure an invariant latent space. The resulting diffusion classifier enables efficient sampling of all possible predictions for each individual to recover the predictive distribution. We then leverage a t-test-based criterion upon the sampling and select informative unlabeled target samples based on the p-value, which encodes both prediction variability and cross-category ambiguity. Experiments on both ADA and Source-Free ADA settings show that our method provides more calibrated predictions than previous ADA methods and achieves favorable performance on three domain adaptation datasets.
****

Augmented Memory Replay-based Continual Learning Approaches for Network Intrusion Detection

suresh kumar amalapuram, Sumohana Channappayya, Bheemarjuna Reddy Tamma

Intrusion detection is a form of anomalous activity detection in communication network traffic. Continual learning (CL) approaches to the intrusion detection task accumulate old knowledge while adapting to the latest threat knowledge. Previous works have shown the effectiveness of memory replay-based CL approaches for this task. In this work, we present two novel contributions to improve the performance of CL-based network intrusion detection in the context of class imbalance and scalability. First, we extend class balancing reservoir sampling (CBRS), a memory-based CL method, to address the problems of severe class imbalance for large datasets. Second, we propose a novel approach titled perturbation assistance for parameter approximation (PAPA) based on the Gaussian mixture model to reduce the number of \textit{virtual stochastic gradient descent (SGD) parameter} computations needed to discover maximally interfering samples for CL. We demonstrate that the proposed approaches perform remarkably better than the baselines on standard intrusion detection benchmarks created over shorter periods (KDDCUP'99, NSL-KDD, CICIDS-2017/2018, UNSW-NB15, and CTU-13) and a longer period with distribution shift (AnoShift). We also validated proposed approaches on standard continual learning benchmarks (SVHN, CIFAR-10/100, and CLEAR-10/100) and anomaly detection benchmarks (SMAP, SMD, and MSL). Further, the proposed PAPA approach significantly lowers the number of virtual SGD update operations, thus resulting in training time savings in the range of 12 to 40\% compared to the maximally interfered samples retrieval algorithm.
****

Selective Amnesia: A Continual Learning Approach to Forgetting in Deep Generative Models

Alvin Heng, Harold Soh

The recent proliferation of large-scale text-to-image models has led to growing concerns that such models may be misused to generate harmful, misleading, and inappropriate content. Motivated by this issue, we derive a technique inspired by continual learning to selectively forget concepts in pretrained deep generative models. Our method, dubbed Selective Amnesia, enables controllable forgetting where a user can specify how a concept should be forgotten. Selective Amnesia can be applied to conditional variational likelihood models, which encompass a variety of popular deep generative frameworks, including variational autoencoders and large-scale text-to-image diffusion models. Experiments across different models demonstrate that our approach induces forgetting on a variety of concepts, from entire classes in standard datasets to celebrity and nudity prompts in text-to-image models.
****

Structure from Duplicates: Neural Inverse Graphics from a Pile of Objects

Tianhang Cheng, Wei-Chiu Ma, Kaiyu Guan, Antonio Torralba, Shenlong Wang

Abstract Our world is full of identical objects (\emph{e.g.}, cans of coke, cars of same model). These duplicates, when seen together, provide additional and strong cues for us to effectively reason about 3D. Inspired by this observation, we introduce Structure from Duplicates (SfD), a novel inverse graphics framework that reconstructs geometry, material, and illumination from a single image containing multiple identical objects. SfD begins by identifying multiple instances of an object within an image, and then jointly estimates the 6DoF pose for all instances. An inverse graphics pipeline is subsequently employed to jointly reason about the shape, material of the object, and the environment light, while adhering to the shared geometry and material constraint across instances.Our primary contributions involve utilizing object duplicates as a robust prior for single-image inverse graphics and proposing an in-plane rotation-robust Structure from Motion (SfM) formulation for joint 6-DoF object pose estimation. By leveraging multi-view cues from a single image, SfD generates more realistic and detailed 3D reconstructions, significantly outperforming existing single image reconstruction models and multi-view reconstruction approaches with a similar or greater number of observations.

****

Weakly-Supervised Audio-Visual Segmentation

Shentong Mo, Bhiksha Raj

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

Adversarial Examples Are Not Real Features

Ang Li, Yifei Wang, Yiwen Guo, Yisen Wang

The existence of adversarial examples has been a mystery for years and attracted much interest. A well-known theory by \citet{ilyas2019adversarial} explains adversarial vulnerability from a data perspective by showing that one can extract non-robust features from adversarial examples and these features alone are useful for classification. However, the explanation remains quite counter-intuitive since non-robust features are mostly noise features to humans. In this paper, we re-examine the theory from a larger context by incorporating multiple learning paradigms. Notably, we find that contrary to their good usefulness under supervised learning, non-robust features attain poor usefulness when transferred to other self-supervised learning paradigms, such as contrastive learning, masked image modeling, and diffusion models. It reveals that non-robust features are not really as useful as robust or natural features that enjoy good transferability between these paradigms. Meanwhile, for robustness, we also show that naturally trained encoders from robust features are largely non-robust under AutoAttack. Our cross-paradigm examination suggests that the non-robust features are not really useful but more like paradigm-wise shortcuts, and robust features alone might be insufficient to attain reliable model robustness. Code is available at \url{https://github.com/PKU-ML/AdvNotRealFeatures}.

****

A Comprehensive Study on Text-attributed Graphs: Benchmarking and Rethinking

Hao Yan, Chaozhuo Li, Ruosong Long, Chao Yan, Jianan Zhao, Wenwen Zhuang, Jun Yin, Peiyan Zhang, Weihao Han, Hao Sun, Weiwei Deng, Qi Zhang, Lichao Sun, Xing Xie, Senzhang Wang

Text-attributed graphs (TAGs) are prevalent in various real-world scenarios, where each node is associated with a text description. The cornerstone of representation learning on TAGs lies in the seamless integration of textual semantics within individual nodes and the topological connections across nodes. Recent advancements in pre-trained language models (PLMs) and graph neural networks (GNNs) have facilitated effective learning on TAGs, garnering increased research interest. However, the absence of meaningful benchmark datasets and standardized evaluation procedures for TAGs has impeded progress in this field. In this paper, we propose CS-TAG, a comprehensive and diverse collection of challenging benchmark datasets for TAGs. The CS-TAG datasets are notably large in scale and encompass a

wide range of domains, spanning from citation networks to purchase graphs. In ad
dition to building the datasets, we conduct extensive benchmark experiments ove
r CS-TAG with various learning paradigms, including PLMs, GNNs, PLM-GNN co-train
ing methods, and the proposed novel topological pre-training of language models.
In a nutshell, we provide an overview of the CS-TAG datasets, standardized eval
uation procedures, and present baseline experiments. The entire CS-TAG project i
s publicly accessible at \url{https://github.com/sktsherlock/TAG-Benchmark}.
****
Online Ad Allocation with Predictions
Fabian Spaeh, Alina Ene
Display Ads and the generalized assignment problem are two well-studied online p
acking problems with important applications in ad allocation and other areas. In
both problems, ad impressions arrive online and have to be allocated immediatel
y to budget-constrained advertisers. Worst-case algorithms that achieve the idea
l competitive ratio are known for both problems, but might act overly conservati
ve given the predictable and usually tame nature of real-world input. Given this
discrepancy, we develop an algorithm for both problems that incorporate machine
-learned predictions and can thus improve the performance beyond the worst-case.
Our algorithm is based on the work of Feldman et al. (2009) and similar in natu
re to Mahdian et al. (2007) who were the first to develop a learning-augmented a
lgorithm for the related, but more structured Ad Words problem. We use a novel a
nalysis to show that our algorithm is able to capitalize on a good prediction, w
hile being robust against poor predictions. We experimentally evaluate our algor
ithm on synthetic and real-world data on a wide range of predictions. Our algori
thm is consistently outperforming the worst-case algorithm without predictions.
****
Transfer Learning with Affine Model Transformation
Shunya Minami, Kenji Fukumizu, Yoshihiro Hayashi, Ryo Yoshida
Supervised transfer learning has received considerable attention due to its pote
ntial to boost the predictive power of machine learning in scenarios where data
are scarce. Generally, a given set of source models and a dataset from a target
domain are used to adapt the pre-trained models to a target domain by statistica
lly learning domain shift and domain-specific factors. While such procedurally a
nd intuitively plausible methods have achieved great success in a wide range of
real-world applications, the lack of a theoretical basis hinders further methodo
logical development. This paper presents a general class of transfer learning re
gression called affine model transfer, following the principle of expected-squar
e loss minimization. It is shown that the affine model transfer broadly encompas
ses various existing methods, including the most common procedure based on neura
l feature extractors. Furthermore, the current paper clarifies theoretical prope
rties of the affine model transfer such as generalization error and excess risk.
Through several case studies, we demonstrate the practical benefits of modeling
and estimating inter-domain commonality and domain-specific factors separately
with the affine-type transfer models.
****
Towards Robust and Expressive Whole-body Human Pose and Shape Estimation
Hui En Pang, Zhongang Cai, Lei Yang, Qingyi Tao, Zhonghua Wu, Tianwei Zhang, Ziw
ei Liu
Whole-body pose and shape estimation aims to jointly predict different behaviors
(e.g., pose, hand gesture, facial expression) of the entire human body from a m
onocular image. Existing methods often exhibit suboptimal performance due to the
complexity of in-the-wild scenarios. We argue that the prediction accuracy of t
hese models is significantly affected by the quality of the bounding box, e.g.,
scale, alignment. The natural discrepancy between the ideal bounding box annotat
ions and model detection results is particularly detrimental to the performance
of whole-body pose and shape estimation.In this paper, we propose a novel framew
ork to enhance the robustness of whole-body pose and shape estimation. Our frame
work incorporates three new modules to address the above challenges from three p
erspectives: (1) a Localization Module enhances the model's awareness of the sub
ject's location and semantics within the image space; (2) a Contrastive Feature

Extraction Module encourages the model to be invariant to robust augmentations by incorporating a contrastive loss and positive samples; (3) a Pixel Alignment Module ensures the reprojected mesh from the predicted camera and body model parameters are more accurate and pixel-aligned. We perform comprehensive experiments to demonstrate the effectiveness of our proposed framework on body, hands, face and whole-body benchmarks.
****
Neural Lad: A Neural Latent Dynamics Framework for Times Series Modeling
ting li, Jianguo Li, Zhanxing Zhu
Neural ordinary differential equation (Neural ODE) is an elegant yet powerful framework to learn the temporal dynamics for time series modeling.However, we observe that existing Neural ODE forecasting models suffer from two disadvantages:i) controlling the latent states only through the linear transformation over the local change of the observed signals may be inadequate;ii) lacking the ability to capture the inherent periodical property in time series forecasting tasks;To overcome the two issues, we introduce a new neural ODE framework called \textbf{Neural Lad}, a \textbf{Neural} \textbf{La}tent \textbf{d}ynamics model in which the latent representations evolve with an ODE enhanced by the change of observed signal and seasonality-trend characterization. We incorporate the local change of input signal into the latent dynamics in an attention-based manner and design a residual architecture over basis expansion to depict the periodicity in the underlying dynamics. To accommodate the multivariate time series forecasting, we extend the Neural Lad through learning an adaptive relationship between multiple time series.   Experiments demonstrate that our model can achieve better or comparable performance against existing neural ODE families and transformer variants in various datasets. Remarkably, the empirical superiority of Neural Lad is consistent across short and long-horizon forecasting for both univariate, multivariate and even irregular sampled time series.
****
Weighted ROC Curve in Cost Space: Extending AUC to Cost-Sensitive Learning
HuiYang Shao, Qianqian Xu, Zhiyong Yang, Peisong Wen, Gao Peifeng, Qingming Huang
In this paper, we aim to tackle flexible cost requirements for long-tail datasets, where we need to construct a (a) cost-sensitive and (b) class-distribution robust learning framework. The misclassification cost and the area under the ROC curve (AUC) are popular metrics for (a) and (b), respectively. However, limited by their formulations, models trained with AUC cannot be applied to cost-sensitive decision problems, and models trained with fixed costs are sensitive to the class distribution shift. To address this issue, we present a new setting where costs are treated like a dataset to deal with arbitrarily unknown cost distributions. Moreover, we propose a novel weighted version of AUC where the cost distribution can be integrated into its calculation through decision thresholds. To formulate this setting, we propose a novel bilevel paradigm to bridge weighted AUC (WAUC) and cost. The inner-level problem approximates the optimal threshold from sampling costs, and the outer-level problem minimizes the WAUC loss over the optimal threshold distribution. To optimize this bilevel paradigm, we employ a stochastic optimization algorithm (SACCL) to optimize it. Finally, experiment results show that our algorithm performs better than existing cost-sensitive learning methods and two-stage AUC decisions approach.
****
Dynamic Non-monotone Submodular Maximization
Kiarash Banihashem, Leyla Biabani, Samira Goudarzi, MohammadTaghi Hajiaghayi, Peyman Jabbarzade, Morteza Monemizadeh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Semi-Implicit Denoising Diffusion Models (SIDDMs)
yanwu xu, Mingming Gong, Shaoan Xie, Wei Wei, Matthias Grundmann, Kayhan Batmang

helich, Tingbo Hou
Despite the proliferation of generative models, achieving fast sampling during inference without compromising sample diversity and quality remains challenging. Existing models such as Denoising Diffusion Probabilistic Models (DDPM) deliver high-quality, diverse samples but are slowed by an inherently high number of iterative steps. The Denoising Diffusion Generative Adversarial Networks (DDGAN) attempted to circumvent this limitation by integrating a GAN model for larger jumps in the diffusion process. However, DDGAN encountered scalability limitations when applied to large datasets. To address these limitations, we introduce a novel approach that tackles the problem by matching implicit and explicit factors. More specifically, our approach involves utilizing an implicit model to match the marginal distributions of noisy data and the explicit conditional distribution of the forward diffusion. This combination allows us to effectively match the joint denoising distributions. Unlike DDPM but similar to DDGAN, we do not enforce a parametric distribution for the reverse step, enabling us to take large steps during inference. Similar to the DDPM but unlike DDGAN, we take advantage of the exact form of the diffusion process. We demonstrate that our proposed method obtains comparable generative performance to diffusion-based models and vastly superior results to models with a small number of sampling steps.
****
Implicit Convolutional Kernels for Steerable CNNs
Maksim Zhdanov, Nico Hoffmann, Gabriele Cesa
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Block Low-Rank Preconditioner with Shared Basis for Stochastic Optimization
Jui-Nan Yen, Sai Surya Duvvuri, Inderjit Dhillon, Cho-Jui Hsieh
Adaptive methods with non-diagonal preconditioning have shown state-of-the-art results on various tasks. However, their computational complexity and memory requirement makes it challenging to scale these methods to modern neural network architectures. To address this challenge, some previous works have adopted block-diagonal preconditioners. However, the memory cost of storing the block-diagonal matrix remains substantial, leading to the use of smaller block sizes and ultimately resulting in suboptimal performance. To reduce the time and memory complexity without sacrificing performance, we propose approximating each diagonal block of the second moment matrix by low-rank matrices and enforcing the same basis for the blocks within each layer. We provide theoretical justification for such sharing and design an algorithm to efficiently maintain this shared-basis block low-rank approximation during training. Our results on a deep autoencoder and a transformer benchmark demonstrate that the proposed method outperforms first-order methods with slightly more time and memory usage, while also achieving competitive or superior performance compared to other second-order methods with less time and memory usage.
****
Learning in the Presence of Low-dimensional Structure: A Spiked Random Matrix Perspective
Jimmy Ba, Murat A. Erdogdu, Taiji Suzuki, Zhichao Wang, Denny Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Efficient Neural Music Generation
Max W. Y. Lam, Qiao Tian, Tang Li, Zongyu Yin, Siyuan Feng, Ming Tu, Yuliang Ji, Rui Xia, Mingbo Ma, Xuchen Song, Jitong Chen, Wang Yuping, Yuxuan Wang
Recent progress in music generation has been remarkably advanced by the state-of-the-art MusicLM, which comprises a hierarchy of three LMs, respectively, for semantic, coarse acoustic, and fine acoustic modelings. Yet, sampling with the Mus

icLM requires processing through these LMs one by one to obtain the fine-grained acoustic tokens, making it computationally expensive and prohibitive for a real-time generation. Efficient music generation with a quality on par with MusicLM remains a significant challenge.In this paper, we present MeLoDy (M for music; L for LM; D for diffusion), an LM-guided diffusion model that generates music audios of state-of-the-art quality meanwhile reducing 95.7\% to 99.6\% forward passes in MusicLM, respectively, for sampling 10s to 30s music. MeLoDy inherits the highest-level LM from MusicLM for semantic modeling, and applies a novel dual-path diffusion (DPD) model and an audio VAE-GAN to efficiently decode the conditioning semantic tokens into waveform. DPD is proposed to simultaneously model the coarse and fine acoustics by incorporating the semantic information into segments of latents effectively via cross-attention at each denoising step. Our experimental results suggest the superiority of MeLoDy, not only in its practical advantages on sampling speed and infinitely continuable generation, but also in its state-of-the-art musicality, audio quality, and text correlation.Our samples are available at https://Efficient-MeLoDy.github.io/.
****
Crystal Structure Prediction by Joint Equivariant Diffusion
Rui Jiao, Wenbing Huang, Peijia Lin, Jiaqi Han, Pin Chen, Yutong Lu, Yang Liu
Crystal Structure Prediction (CSP) is crucial in various scientific disciplines. While CSP can be addressed by employing currently-prevailing generative models (e.g. diffusion models), this task encounters unique challenges owing to the symmetric geometry of crystal structures---the invariance of translation, rotation, and periodicity. To incorporate the above symmetries, this paper proposes DiffCSP, a novel diffusion model to learn the structure distribution from stable crystals. To be specific, DiffCSP jointly generates the lattice and atom coordinates for each crystal by employing a periodic-E(3)-equivariant denoising model, to better model the crystal geometry. Notably, different from related equivariant generative approaches, DiffCSP leverages fractional coordinates other than Cartesian coordinates to represent crystals, remarkably promoting the diffusion and the generation process of atom positions. Extensive experiments verify that our DiffCSP remarkably outperforms existing CSP methods, with a much lower computation cost in contrast to DFT-based methods. Moreover, the superiority of DiffCSP is still observed when it is extended for ab initio crystal generation.
****
Understanding the detrimental class-level effects of data augmentation
Polina Kirichenko, Mark Ibrahim, Randall Balestriero, Diane Bouchacourt, Shanmukha Ramakrishna Vedantam, Hamed Firooz, Andrew G. Wilson
Data augmentation (DA) encodes invariance and provides implicit regularization critical to a model's performance in image classification tasks. However, while DA improves average accuracy, recent studies have shown that its impact can be highly class dependent: achieving optimal average accuracy comes at the cost of significantly hurting individual class accuracy by as much as 20% on ImageNet. There has been little progress in resolving class-level accuracy drops due to a limited understanding of these effects. In this work, we present a framework for understanding how DA interacts with class-level learning dynamics. Using higher-quality multi-label annotations on ImageNet, we systematically categorize the affected classes and find that the majority are inherently ambiguous, co-occur, or involve fine-grained distinctions, while DA controls the model's bias towards one of the closely related classes. While many of the previously reported performance drops are explained by multi-label annotations, we identify other sources of accuracy degradations by analyzing class confusions. We show that simple class-conditional augmentation strategies informed by our framework improve performance on the negatively affected classes.
****
Optimal Guarantees for Algorithmic Reproducibility and Gradient Complexity in Convex Optimization
Liang Zhang, Junchi YANG, Amin Karbasi, Niao He
Algorithmic reproducibility measures the deviation in outputs of machine learning algorithms upon minor changes in the training process. Previous work suggests

that first-order methods would need to trade-off convergence rate (gradient complexity) for better reproducibility. In this work, we challenge this perception and demonstrate that both optimal reproducibility and near-optimal convergence guarantees can be achieved for smooth convex minimization and smooth convex-concave minimax problems under various error-prone oracle settings. Particularly, given the inexact initialization oracle, our regularization-based algorithms achieve the best of both worlds -- optimal reproducibility and near-optimal gradient complexity -- for minimization and minimax optimization. With the inexact gradient oracle, the near-optimal guarantees also hold for minimax optimization. Additionally, with the stochastic gradient oracle, we show that stochastic gradient descent ascent is optimal in terms of both reproducibility and gradient complexity. We believe our results contribute to an enhanced understanding of the reproducibility-convergence trade-off in the context of convex optimization.
****

Diffusion-TTA: Test-time Adaptation of Discriminative Models via Generative Feedback

Mihir Prabhudesai, Tsung-Wei Ke, Alex Li, Deepak Pathak, Katerina Fragkiadaki

The advancements in generative modeling, particularly the advent of diffusion models, have sparked a fundamental question: how can these models be effectively used for discriminative tasks? In this work, we find that generative models can be great test-time adapters for discriminative models. Our method, Diffusion-TTA, adapts pre-trained discriminative models such as image classifiers, segmenters and depth predictors, to each unlabelled example in the test set using generative feedback from a diffusion model. We achieve this by modulating the conditioning of the diffusion model using the output of the discriminative model. We then maximize the image likelihood objective by backpropagating the gradients to discriminative model’s parameters. We show Diffusion-TTA significantly enhances the accuracy of various large-scale pre-trained discriminative models, such as, ImageNet classifiers, CLIP models, image pixel labellers and image depth predictors. Diffusion-TTA outperforms existing test-time adaptation methods, including TTT-MAE and TENT, and particularly shines in online adaptation setups, where the discriminative model is continually adapted to each example in the test set. We provide access to code, results, and visualizations on our website: diffusion-tta.github.io/
****

Fragment-based Pretraining and Finetuning on Molecular Graphs

Kha-Dinh Luong, Ambuj K Singh

Property prediction on molecular graphs is an important application of Graph Neural Networks (GNNs). Recently, unlabeled molecular data has become abundant, which facilitates the rapid development of self-supervised learning for GNNs in the chemical domain. In this work, we propose pretraining GNNs at the fragment level, a promising middle ground to overcome the limitations of node-level and graph-level pretraining. Borrowing techniques from recent work on principal subgraph mining, we obtain a compact vocabulary of prevalent fragments from a large pretraining dataset. From the extracted vocabulary, we introduce several fragment-based contrastive and predictive pretraining tasks. The contrastive learning task jointly pretrains two different GNNs: one on molecular graphs and the other on fragment graphs, which represents higher-order connectivity within molecules. By enforcing consistency between the fragment embedding and the aggregated embedding of the corresponding atoms from the molecular graphs, we ensure that the embeddings capture structural information at multiple resolutions. The structural information of fragment graphs is further exploited to extract auxiliary labels for graph-level predictive pretraining. We employ both the pretrained molecular-based and fragment-based GNNs for downstream prediction, thus utilizing the fragment information during finetuning. Our graph fragment-based pretraining (GraphFP) advances the performances on 5 out of 8 common molecular benchmarks and improves the performances on long-range biological benchmarks by at least 11.5%. Code is available at: https://github.com/lvkd84/GraphFP.
****

Characterizing Out-of-Distribution Error via Optimal Transport

Yuzhe Lu, Yilong Qin, Runtian Zhai, Andrew Shen, Ketong Chen, Zhenlin Wang, Soheil Kolouri, Simon Stepputtis, Joseph Campbell, Katia Sycara

Out-of-distribution (OOD) data poses serious challenges in deployed machine learning models,so methods of predicting a model's performance on OOD data without labels are important for machine learning safety.While a number of methods have been proposed by prior work, they often underestimate the actual error, sometimes by a large margin, which greatly impacts their applicability to real tasks. In this work, we identify pseudo-label shift, or the difference between the predicted and true OOD label distributions, as a key indicator of this underestimation. Based on this observation, we introduce a novel method for estimating model performance by leveraging optimal transport theory, Confidence Optimal Transport (COT), and show that it provably provides more robust error estimates in the presence of pseudo-label shift. Additionally, we introduce an empirically-motivated variant of COT, Confidence Optimal Transport with Thresholding (COTT), which applies thresholding to the individual transport costs and further improves the accuracy of COT's error estimates. We evaluate COT and COTT on a variety of standard benchmarks that induce various types of distribution shift -- synthetic, novel subpopulation, and natural -- and show that our approaches significantly outperform existing state-of-the-art methods with up to 3x lower prediction errors.

****

Unsupervised Video Domain Adaptation for Action Recognition: A Disentanglement Perspective

Pengfei Wei, Lingdong Kong, Xinghua Qu, Yi Ren, Zhiqiang Xu, Jing Jiang, Xiang Yin

Unsupervised video domain adaptation is a practical yet challenging task. In this work, for the first time, we tackle it from a disentanglement view. Our key idea is to handle the spatial and temporal domain divergence separately through disentanglement. Specifically, we consider the generation of cross-domain videos from two sets of latent factors, one encoding the static information and another encoding the dynamic information. A Transfer Sequential VAE (TranSVAE) framework is then developed to model such generation. To better serve for adaptation, we propose several objectives to constrain the latent factors. With these constraints, the spatial divergence can be readily removed by disentangling the static domain-specific information out, and the temporal divergence is further reduced from both frame- and video-levels through adversarial learning. Extensive experiments on the UCF-HMDB, Jester, and Epic-Kitchens datasets verify the effectiveness and superiority of TranSVAE compared with several state-of-the-art approaches.

****

Does Localization Inform Editing? Surprising Differences in Causality-Based Localization vs. Knowledge Editing in Language Models

Peter Hase, Mohit Bansal, Been Kim, Asma Ghandeharioun

Language models learn a great quantity of factual information during pretraining, and recent work localizes this information to specific model weights like mid-layer MLP weights. In this paper, we find that we can change how a fact is stored in a model by editing weights that are in a different location than where existing methods suggest that the fact is stored. This is surprising because we would expect that localizing facts to specific model parameters would tell us where to manipulate knowledge in models, and this assumption has motivated past work on model editing methods. Specifically, we show that localization conclusions from representation denoising (also known as Causal Tracing) do not provide any insight into which model MLP layer would be best to edit in order to override an existing stored fact with a new one. This finding raises questions about how past work relies on Causal Tracing to select which model layers to edit. Next, we consider several variants of the editing problem, including erasing and amplifying facts. For one of our editing problems, editing performance does relate to localization results from representation denoising, but we find that which layer we edit is a far better predictor of performance. Our results suggest, counterintuitively, that better mechanistic understanding of how pretrained language models work may not always translate to insights about how to best change their behavior.

****
The Geometry of Neural Nets' Parameter Spaces Under Reparametrization
Agustinus Kristiadi, Felix Dangel, Philipp Hennig
Model reparametrization, which follows the change-of-variable rule of calculus, is a popular way to improve the training of neural nets. But it can also be problematic since it can induce inconsistencies in, e.g., Hessian-based flatness measures, optimization trajectories, and modes of probability densities. This complicates downstream analyses: e.g. one cannot definitively relate flatness with generalization since arbitrary reparametrization changes their relationship. In this work, we study the invariance of neural nets under reparametrization from the perspective of Riemannian geometry. From this point of view, invariance is an inherent property of any neural net if one explicitly represents the metric and uses the correct associated transformation rules. This is important since although the metric is always present, it is often implicitly assumed as identity, and thus dropped from the notation, then lost under reparametrization. We discuss implications for measuring the flatness of minima, optimization, and for probability-density maximization. Finally, we explore some interesting directions where invariance is useful.
****
A Dataset of Relighted 3D Interacting Hands
Gyeongsik Moon, Shunsuke Saito, Weipeng Xu, Rohan Joshi, Julia Buffalini, Harley Bellan, Nicholas Rosen, Jesse Richardson, Mallorie Mize, Philippe De Bree, Tomas Simon, Bo Peng, Shubham Garg, Kevyn McPhail, Takaaki Shiratori
The two-hand interaction is one of the most challenging signals to analyze due to the self-similarity, complicated articulations, and occlusions of hands. Although several datasets have been proposed for the two-hand interaction analysis, all of them do not achieve 1) diverse and realistic image appearances and 2) diverse and large-scale groundtruth (GT) 3D poses at the same time. In this work, we propose Re:InterHand, a dataset of relighted 3D interacting hands that achieve the two goals. To this end, we employ a state-of-the-art hand relighting network with our accurately tracked two-hand 3D poses. We compare our Re:InterHand with existing 3D interacting hands datasets and show the benefit of it. Our Re:InterHand is available in https://mks0601.github.io/ReInterHand/
****
Multi-Objective Intrinsic Reward Learning for Conversational Recommender Systems
Zhendong Chu, Nan Wang, Hongning Wang
Conversational Recommender Systems (CRS) actively elicit user preferences to generate adaptive recommendations. Mainstream reinforcement learning-based CRS solutions heavily rely on handcrafted reward functions, which may not be aligned with user intent in CRS tasks. Therefore, the design of task-specific rewards is critical to facilitate CRS policy learning, which remains largely under-explored in the literature. In this work, we propose a novel approach to address this challenge by learning intrinsic rewards from interactions with users. Specifically, we formulate intrinsic reward learning as a multi-objective bi-level optimization problem. The inner level optimizes the CRS policy augmented by the learned intrinsic rewards, while the outer level drives the intrinsic rewards to optimize two CRS-specific objectives: maximizing the success rate and minimizing the number of turns to reach a successful recommendation}in conversations. To evaluate the effectiveness of our approach, we conduct extensive experiments on three public CRS benchmarks. The results show that our algorithm significantly improves CRS performance by exploiting informative learned intrinsic rewards.
****
Predict-then-Calibrate: A New Perspective of Robust Contextual LP
Chunlin Sun, Linyu Liu, Xiaocheng Li
Contextual optimization, also known as predict-then-optimize or prescriptive analytics, considers an optimization problem with the presence of covariates (context or side information). The goal is to learn a prediction model (from the training data) that predicts the objective function from the covariates, and then in the test phase, solve the optimization problem with the covariates but without the observation of the objective function. In this paper, we consider a risk-sens

itive version of the problem and propose a generic algorithm design paradigm called predict-then-calibrate. The idea is to first develop a prediction model without concern for the downstream risk profile or robustness guarantee, and then utilize calibration (or recalibration) methods to quantify the uncertainty of the prediction. While the existing methods suffer from either a restricted choice of the prediction model or strong assumptions on the underlying data, we show the disentangling of the prediction model and the calibration/uncertainty quantification has several advantages. First, it imposes no restriction on the prediction model and thus fully unleashes the potential of off-the-shelf machine learning methods. Second, the derivation of the risk and robustness guarantee can be made independent of the choice of the prediction model through a data-splitting idea. Third, our paradigm of predict-then-calibrate applies to both (risk-sensitive) robust and (risk-neutral) distributionally robust optimization (DRO) formulations. Theoretically, it gives new generalization bounds for the contextual LP problem and sheds light on the existing results of DRO for contextual LP. Numerical experiments further reinforce the advantage of the predict-then-calibrate paradigm in that an improvement on either the prediction model or the calibration model will lead to a better final performance.
****
INSPECT: A Multimodal Dataset for Patient Outcome Prediction of Pulmonary Embolisms
Shih-Cheng Huang, Zepeng Huo, Ethan Steinberg, Chia-Chun Chiang, Curtis Langlotz, Matthew Lungren, Serena Yeung, Nigam Shah, Jason Fries
Synthesizing information from various data sources plays a crucial role in the practice of modern medicine. Current applications of artificial intelligence in medicine often focus on single-modality data due to a lack of publicly available, multimodal medical datasets. To address this limitation, we introduce INSPECT, which contains de-identified longitudinal records from a large cohort of pulmonary embolism (PE) patients, along with ground truth labels for multiple outcomes. INSPECT contains data from 19,402 patients, including CT images, sections of radiology reports, and structured electronic health record (EHR) data (including demographics, diagnoses, procedures, and vitals). Using our provided dataset, we develop and release a benchmark for evaluating several baseline modeling approaches on a variety of important PE related tasks. We evaluate image-only, EHR-only, and fused models. Trained models and the de-identified dataset are made available for non-commercial use under a data use agreement. To the best our knowledge, INSPECT is the largest multimodal dataset for enabling reproducible research on strategies for integrating 3D medical imaging and EHR data.
****
What Makes Good Examples for Visual In-Context Learning?
Yuanhan Zhang, Kaiyang Zhou, Ziwei Liu
Large vision models with billions of parameters and trained on broad data have great potential in numerous downstream applications. However, these models are typically difficult to adapt due to their large parameter size and sometimes lack of accesss to their weights---entities able to develop large vision models often provide APIs only. In this paper, we study how to better utilize large vision models through the lens of in-context learning, a concept that has been well-known in natural language processing but has only been studied very recently in computer vision. In-context learning refers to the ability to perform inference on tasks never seen during training by simply conditioning on in-context examples (i.e., input-output pairs) without updating any internal model parameters. To demystify in-context learning in computer vision, we conduct an extensive research and identify a critical problem: downstream performance is highly sensitivie to the choice of visual in-context examples. To address this problem, we propose a prompt retrieval framework specifically for large vision models, allowing the selection of in-context examples to be fully automated. Concretely, we provide two implementations: (i) an unsupervised prompt retrieval method based on nearest example search using an off-the-shelf model, and (ii) a supervised prompt retrieval method, which trains a neural network to choose examples that directly maximize in-context learning performance. Both methods do not require access to the int

ernal weights of large vision models. Our results demonstrate that our methods can bring non-trivial improvements to visual in-context learning in comparison to the commonly-used random selection. Code and models will be released.
****

Parameterizing Context: Unleashing the Power of Parameter-Efficient Fine-Tuning and In-Context Tuning for Continual Table Semantic Parsing
Yongrui Chen, Shenyu Zhang, Guilin Qi, Xinnan Guo
Continual table semantic parsing aims to train a parser on a sequence of tasks, where each task requires the parser to translate natural language into SQL based on task-specific tables but only offers limited training examples. Conventional methods tend to suffer from overfitting with limited supervision, as well as catastrophic forgetting due to parameter updates.Despite recent advancements that partially alleviate these issues through semi-supervised data augmentation and retention of a few past examples, the performance is still limited by the volume of unsupervised data and stored examples.To overcome these challenges, this paper introduces a novel method integrating parameter-efficient fine-tuning (PEFT) and in-context tuning (ICT) for training a continual table semantic parser. Initially, we present a task-adaptive PEFT framework capable of fully circumventing catastrophic forgetting, which is achieved by freezing the pre-trained model backbone and fine-tuning small-scale prompts. Building on this, we propose a teacher-student framework-based solution. The teacher addresses the few-shot problem using ICT, which procures contextual information by demonstrating a few training examples. In turn, the student leverages the proposed PEFT framework to learn from the teacher's output distribution, and subsequently compresses and saves the contextual information to the prompts, eliminating the need to store any training examples.Experimental evaluations on two benchmarks affirm the superiority of our method over prevalent few-shot and continual learning baselines across various metrics.
****

Incentives in Federated Learning: Equilibria, Dynamics, and Mechanisms for Welfare Maximization
Aniket Murhekar, Zhuowen Yuan, Bhaskar Ray Chaudhury, Bo Li, Ruta Mehta
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

MKOR: Momentum-Enabled Kronecker-Factor-Based Optimizer Using Rank-1 Updates
Mohammad Mozaffari, Sikan Li, Zhao Zhang, Maryam Mehri Dehnavi
This work proposes a Momentum-Enabled Kronecker-Factor-Based Optimizer Using Rank-1 updates, called MKOR, that improves the training time and convergence properties of deep neural networks (DNNs). Second-order techniques, while enjoying higher convergence rates vs first-order counterparts, have cubic complexity with respect to either the model size and/or the training batch size. Hence they exhibit poor scalability and performance in transformer models, e.g. large language models (LLMs), because the batch sizes in these models scale by the attention mechanism sequence length, leading to large model size and batch sizes. MKOR's complexity is quadratic with respect to the model size, alleviating the computation bottlenecks in second-order methods. Because of their high computation complexity, state-of-the-art implementations of second-order methods can only afford to update the second order information infrequently, and thus do not fully exploit the promise of better convergence from these updates. By reducing the communication complexity of the second-order updates as well as achieving a linear communication complexity, MKOR increases the frequency of second order updates. We also propose a hybrid version of MKOR (called MKOR-H) that mid-training falls backs to a first order optimizer if the second order updates no longer accelerate convergence. Our experiments show that MKOR outperforms state -of-the-art first order methods, e.g. the LAMB optimizer, and best implementations of second-order methods, i.e. KAISA/KFAC, up to 2.57x and 1.85x respectively on BERT-Large-Uncased on 64 GPUs.

****
DFRD: Data-Free Robustness Distillation for Heterogeneous Federated Learning
kangyang Luo, Shuai Wang, Yexuan Fu, Xiang Li, Yunshi Lan, Ming Gao
Federated Learning (FL) is a privacy-constrained decentralized machine learning paradigm in which clients enable collaborative training without compromising private data. However, how to learn a robust global model in the data-heterogeneous and model-heterogeneous FL scenarios is challenging. To address it, we resort to data-free knowledge distillation to propose a new FL method (namely DFRD).DFRD equips a conditional generator on the server to approximate the training space of the local models uploaded by clients, and systematically investigates its training in terms of fidelity, transferability and diversity. To overcome the catastrophic forgetting of the global model caused by the distribution shifts of the generator across communication rounds, we maintain an exponential moving average copy of the generator on the server. Additionally, we propose dynamic weighting and label sampling to accurately extract knowledge from local models. Finally, our extensive experiments on various image classification tasks illustrate that DFRD achieves significant performance gains compared to SOTA baselines.
****
SG×P : A Sorghum Genotype × Phenotype Prediction Dataset and Benchmark
Zeyu Zhang, Robert Pless, Nadia Shakoor, Austin Carnahan, Abby Stylianou
Large scale field-phenotyping approaches have the potential to solve important questions about the relationship of plant genotype to plant phenotype. Computational approaches to measuring the phenotype (the observable plant features) are required to address the problem at a large scale, but machine learning approaches to extract phenotypes from sensor data have been hampered by limited access to (a) sufficiently large, organized multi-sensor datasets, (b) field trials that have a large scale and significant number of genotypes, (c) full genetic sequencing of those phenotypes, and (d) datasets sufficiently organized so that algorithm centered researchers can directly address the real biological problems. To address this, we present SGxP, a novel benchmark dataset from a large-scale field trial consisting of the complete genotype of over 300 sorghum varieties, and time sequences of imagery from several field plots growing each variety, taken with RGB and laser 3D scanner imaging. To lower the barrier to entry and facilitate further developments, we provide a set of well organized, multi-sensor imagery and corresponding genomic data. We implement baseline deep learning based phenotyping approaches to create baseline results for individual sensors and multi-sensor fusion for detecting genetic mutations with known impacts. We also provide and support an open-ended challenge by identifying thousands of genetic mutations whose phenotypic impacts are currently unknown. A web interface for machine learning researchers and practitioners to share approaches, visualizations and hypotheses supports engagement with plant biologists to further the understanding of the sorghum genotype x phenotype relationship. The full dataset, leaderboard (including baseline results) and discussion forums can be found at http://sorghumsnpbenchmark.com.
****
Rank-N-Contrast: Learning Continuous Representations for Regression
Kaiwen Zha, Peng Cao, Jeany Son, Yuzhe Yang, Dina Katabi
Deep regression models typically learn in an end-to-end fashion without explicitly emphasizing a regression-aware representation. Consequently, the learned representations exhibit fragmentation and fail to capture the continuous nature of sample orders, inducing suboptimal results across a wide range of regression tasks. To fill the gap, we propose Rank-N-Contrast (RNC), a framework that learns continuous representations for regression by contrasting samples against each other based on their rankings in the target space. We demonstrate, theoretically and empirically, that RNC guarantees the desired order of learned representations in accordance with the target orders, enjoying not only better performance but also significantly improved robustness, efficiency, and generalization. Extensive experiments using five real-world regression datasets that span computer vision, human-computer interaction, and healthcare verify that RNC achieves state-of-the-art performance, highlighting its intriguing properties including better data

efficiency, robustness to spurious targets and data corruptions, and generalization to distribution shifts.
****
OpenGSL: A Comprehensive Benchmark for Graph Structure Learning
Zhou Zhiyao, Sheng Zhou, Bochao Mao, Xuanyi Zhou, Jiawei Chen, Qiaoyu Tan, Daochen Zha, Yan Feng, Chun Chen, Can Wang
Graph Neural Networks (GNNs) have emerged as the de facto standard for representation learning on graphs, owing to their ability to effectively integrate graph topology and node attributes. However, the inherent suboptimal nature of node connections, resulting from the complex and contingent formation process of graphs, presents significant challenges in modeling them effectively. To tackle this issue, Graph Structure Learning (GSL), a family of data-centric learning approaches, has garnered substantial attention in recent years. The core concept behind GSL is to jointly optimize the graph structure and the corresponding GNN models. Despite the proposal of numerous GSL methods, the progress in this field remains unclear due to inconsistent experimental protocols, including variations in datasets, data processing techniques, and splitting strategies. In this paper, we introduce OpenGSL, the first comprehensive benchmark for GSL, aimed at addressing this gap. OpenGSL enables a fair comparison among state-of-the-art GSL methods by evaluating them across various popular datasets using uniform data processing and splitting strategies. Through extensive experiments, we observe that existing GSL methods do not consistently outperform vanilla GNN counterparts. We also find that there is no significant correlation between the homophily of the learned structure and task performance, challenging the common belief. Moreover, we observe that the learned graph structure demonstrates a strong generalization ability across different GNN models, despite the high computational and space consumption. We hope that our open-sourced library will facilitate rapid and equitable evaluation and inspire further innovative research in this field. The code of the benchmark can be found in https://github.com/OpenGSL/OpenGSL.
****
Global Optimality in Bivariate Gradient-based DAG Learning
Chang Deng, Kevin Bello, Pradeep Ravikumar, Bryon Aragam
Recently, a new class of non-convex optimization problems motivated by the statistical problem of learning an acyclic directed graphical model from data has attracted significant interest. While existing work uses standard first-order optimization schemes to solve this problem, proving the global optimality of such approaches has proven elusive. The difficulty lies in the fact that unlike other non-convex problems in the literature, this problem is not "benign", and possesses multiple spurious solutions that standard approaches can easily get trapped in. In this paper, we prove that a simple path-following optimization scheme globally converges to the global minimum of the population loss in the bivariate setting.
****
Perceptual adjustment queries and an inverted measurement paradigm for low-rank metric learning
Austin Xu, Andrew McRae, Jingyan Wang, Mark Davenport, Ashwin Pananjady
We introduce a new type of query mechanism for collecting human feedback, called the perceptual adjustment query (PAQ). Being both informative and cognitively lightweight, the PAQ adopts an inverted measurement scheme, and combines advantages from both cardinal and ordinal queries. We showcase the PAQ in the metric learning problem, where we collect PAQ measurements to learn an unknown Mahalanobis distance. This gives rise to a high-dimensional, low-rank matrix estimation problem to which standard matrix estimators cannot be applied. Consequently, we develop a two-stage estimator for metric learning from PAQs, and provide sample complexity guarantees for this estimator. We present numerical simulations demonstrating the performance of the estimator and its notable properties.
****
Joint processing of linguistic properties in brains and language models
SUBBAREDDY OOTA, Manish Gupta, Mariya Toneva
Language models have been shown to be very effective in predicting brain recordi

ngs of subjects experiencing complex language stimuli. For a deeper understandin g of this alignment, it is important to understand the correspondence between th e detailed processing of linguistic information by the human brain versus langua ge models. We investigate this correspondence via a direct approach, in which we eliminate information related to specific linguistic properties in the language model representations and observe how this intervention affects the alignment w ith fMRI brain recordings obtained while participants listened to a story. We in vestigate a range of linguistic properties (surface, syntactic, and semantic) an d find that the elimination of each one results in a significant decrease in bra in alignment. Specifically, we find that syntactic properties (i.e. Top Constitu ents and Tree Depth) have the largest effect on the trend of brain alignment acr oss model layers. These findings provide clear evidence for the role of specific linguistic information in the alignment between brain and language models, and open new avenues for mapping the joint information processing in both systems. W e make the code publicly available https://github.com/subbareddy248/lingprop-bra in-alignment.
****
S^3: Increasing GPU Utilization during Generative Inference for Higher Through put
Yunho Jin, Chun-Feng Wu, David Brooks, Gu-Yeon Wei
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Disentangling Cognitive Diagnosis with Limited Exercise Labels
Xiangzhi Chen, Le Wu, Fei Liu, Lei Chen, Kun Zhang, Richang Hong, Meng Wang
Cognitive diagnosis is an important task in intelligence education, which aims a t measuring students’ proficiency in specific knowledge concepts. Given a fully labeled exercise-concept matrix, most existing models focused on mining students ' response records for cognitive diagnosis. Despite their success, due to the hu ge cost of labeling exercises, a more practical scenario is that limited exercis es are labeled with concepts. Performing cognitive diagnosis with limited exerci se labels is under-explored and remains pretty much open. In this paper, we prop ose Disentanglement based Cognitive Diagnosis (DCD) to address the challenges of limited exercise labels. Specifically, we utilize students' response records to model student proficiency, exercise difficulty and exercise label distribution. Then, we introduce two novel modules - group-based disentanglement and limite d-labeled alignment modules - to disentangle the factors relevant to concepts an d align them with real limited labels. Particularly, we introduce the tree-like structure of concepts with negligible cost for group-based disentangling, as co ncepts of different levels exhibit different independence relationships.Extensiv e experiments on widely used benchmarks demonstrate the superiority of our propo sed model.
****
Energy-Based Sliced Wasserstein Distance
Khai Nguyen, Nhat Ho
The sliced Wasserstein (SW) distance has been widely recognized as a statistical ly effective and computationally efficient metric between two probability measur es. A key component of the SW distance is the slicing distribution. There are tw o existing approaches for choosing this distribution. The first approach is usin g a fixed prior distribution. The second approach is optimizing for the best dis tribution which belongs to a parametric family of distributions and can maximize the expected distance. However, both approaches have their limitations. A fixed prior distribution is non-informative in terms of highlighting projecting direc tions that can discriminate two general probability measures. Doing optimization for the best distribution is often expensive and unstable. Moreover, designing the parametric family of the candidate distribution could be easily misspecified . To address the issues, we propose to design the slicing distribution as an ene rgy-based distribution that is parameter-free and has the density proportional t

o an energy function of the projected one-dimensional Wasserstein distance. We then derive a novel sliced Wasserstein variant, energy-based sliced Waserstein (EBSW) distance, and investigate its topological, statistical, and computational properties via importance sampling, sampling importance resampling, and Markov Chain methods. Finally, we conduct experiments on point-cloud gradient flow, color transfer, and point-cloud reconstruction to show the favorable performance of the EBSW.
****
E2PNet: Event to Point Cloud Registration with Spatio-Temporal Representation Learning
Xiuhong Lin, Changjie Qiu, zhipeng cai, Siqi Shen, Yu Zang, Weiquan Liu, Xuesheng Bian, Matthias Müller, Cheng Wang
Event cameras have emerged as a promising vision sensor in recent years due to their unparalleled temporal resolution and dynamic range. While registration of 2D RGB images to 3D point clouds is a long-standing problem in computer vision, no prior work studies 2D-3D registration for event cameras. To this end, we propose E2PNet, the first learning-based method for event-to-point cloud registration.The core of E2PNet is a novel feature representation network called Event-Points-to-Tensor (EP2T), which encodes event data into a 2D grid-shaped feature tensor. This grid-shaped feature enables matured RGB-based frameworks to be easily used for event-to-point cloud registration, without changing hyper-parameters and the training procedure. EP2T treats the event input as spatio-temporal point clouds. Unlike standard 3D learning architectures that treat all dimensions of point clouds equally, the novel sampling and information aggregation modules in EP2T are designed to handle the inhomogeneity of the spatial and temporal dimensions. Experiments on the MVSEC and VECtor datasets demonstrate the superiority of E2PNet over hand-crafted and other learning-based methods. Compared to RGB-based registration, E2PNet is more robust to extreme illumination or fast motion due to the use of event data. Beyond 2D-3D registration, we also show the potential of EP2T for other vision tasks such as flow estimation, event-to-image reconstruction and object recognition. The source code can be found at: https://github.com/Xmu-qcj/E2PNet.
****
Pengi: An Audio Language Model for Audio Tasks
Soham Deshmukh, Benjamin Elizalde, Rita Singh, Huaming Wang
In the domain of audio processing, Transfer Learning has facilitated the rise of Self-Supervised Learning and Zero-Shot Learning techniques. These approaches have led to the development of versatile models capable of tackling a wide array of tasks, while delivering state-of-the-art performance. However, current models inherently lack the capacity to produce the requisite language for open-ended tasks, such as Audio Captioning or Audio Question Answering. We introduce Pengi, a novel Audio Language Model that leverages Transfer Learning by framing all audio tasks as text-generation tasks. It takes as input, an audio recording, and text, and generates free-form text as output. The input audio is represented as a sequence of continuous embeddings by an audio encoder. A text encoder does the same for the corresponding text input. Both sequences are combined as a prefix to prompt a pre-trained frozen language model. The unified architecture of Pengi enables open-ended tasks and close-ended tasks without any additional fine-tuning or task-specific extensions. When evaluated on 21 downstream tasks, our approach yields state-of-the-art performance in several of them. Our results show that connecting language models with audio models is a major step towards general-purpose audio understanding.
****
Unleashing the Power of Graph Data Augmentation on Covariate Distribution Shift
Yongduo Sui, Qitian Wu, Jiancan Wu, Qing Cui, Longfei Li, Jun Zhou, Xiang Wang, Xiangnan He
The issue of distribution shifts is emerging as a critical concern in graph representation learning. From the perspective of invariant learning and stable learning, a recently well-established paradigm for out-of-distribution generalization, stable features of the graph are assumed to causally determine labels, while e

nvironmental features tend to be unstable and can lead to the two primary types of distribution shifts. The correlation shift is often caused by the spurious correlation between environmental features and labels that differs between the training and test data; the covariate shift often stems from the presence of new environmental features in test data. However, most strategies, such as invariant learning or graph augmentation, typically struggle with limited training environments or perturbed stable features, thus exposing limitations in handling the problem of covariate shift. To address this challenge, we propose a simple-yet-effective data augmentation strategy, Adversarial Invariant Augmentation (AIA), to handle the covariate shift on graphs. Specifically, given the training data, AIA aims to extrapolate and generate new environments, while concurrently preserving the original stable features during the augmentation process. Such a design equips the graph classification model with an enhanced capability to identify stable features in new environments, thereby effectively tackling the covariate shift in data. Extensive experiments with in-depth empirical analysis demonstrate the superiority of our approach. The implementation codes are publicly available at https://github.com/yongduosui/AIA.
****
Adaptive recurrent vision performs zero-shot computation scaling to unseen difficulty levels
Vijay Veerabadran, Srinivas Ravishankar, Yuan Tang, Ritik Raina, Virginia de Sa
Humans solving algorithmic (or) reasoning problems typically exhibit solution times that grow as a function of problem difficulty. Adaptive recurrent neural networks have been shown to exhibit this property for various language-processing tasks. However, little work has been performed to assess whether such adaptive computation can also enable vision models to extrapolate solutions beyond their training distribution's difficulty level, with prior work focusing on very simple tasks. In this study, we investigate a critical functional role of such adaptive processing using recurrent neural networks: to dynamically scale computational resources conditional on input requirements that allow for zero-shot generalization to novel difficulty levels not seen during training using two challenging visual reasoning tasks: PathFinder and Mazes. We combine convolutional recurrent neural networks (ConvRNNs) with a learnable halting mechanism based on Graves (2016). We explore various implementations of such adaptive ConvRNNs (AdRNNs) ranging from tying weights across layers to more sophisticated biologically inspired recurrent networks that possess lateral connections and gating. We show that 1) AdRNNs learn to dynamically halt processing early (or late) to solve easier (or harder) problems, 2) these RNNs zero-shot generalize to more difficult problem settings not shown during training by dynamically increasing the number of recurrent iterations at test time. Our study provides modeling evidence supporting the hypothesis that recurrent processing enables the functional advantage of adaptively allocating compute resources conditional on input requirements and hence allowing generalization to harder difficulty levels of a visual reasoning problem without training.
****
NurViD: A Large Expert-Level Video Database for Nursing Procedure Activity Understanding
Ming Hu, Lin Wang, Siyuan Yan, Don Ma, Qingli Ren, Peng Xia, Wei Feng, Peibo Duan, Lie Ju, Zongyuan Ge
The application of deep learning to nursing procedure activity understanding has the potential to greatly enhance the quality and safety of nurse-patient interactions. By utilizing the technique, we can facilitate training and education, improve quality control, and enable operational compliance monitoring. However, the development of automatic recognition systems in this field is currently hindered by the scarcity of appropriately labeled datasets. The existing video datasets pose several limitations: 1) these datasets are small-scale in size to support comprehensive investigations of nursing activity; 2) they primarily focus on single procedures, lacking expert-level annotations for various nursing procedures and action steps; and 3) they lack temporally localized annotations, which prevents the effective localization of targeted actions within longer video sequence

s. To mitigate these limitations, we propose NurViD, a large video dataset with expert-level annotation for nursing procedure activity understanding. NurViD consists of over 1.5k videos totaling 144 hours, making it approximately four times longer than the existing largest nursing activity datasets. Notably, it encompasses 51 distinct nursing procedures and 177 action steps, providing a much more comprehensive coverage compared to existing datasets that primarily focus on limited procedures. To evaluate the efficacy of current deep learning methods on nursing activity understanding, we establish three benchmarks on NurViD: procedure recognition on untrimmed videos, procedure and action recognition on trimmed videos, and action detection. Our benchmark and code will be available at https://github.com/minghu0830/NurViD-benchmark.
****
The Pick-to-Learn Algorithm: Empowering Compression for Tight Generalization Bounds and Improved Post-training Performance
Dario Paccagnan, Marco Campi, Simone Garatti
Generalization bounds are valuable both for theory and applications. On the one hand, they shed light on the mechanisms that underpin the learning processes; on the other, they certify how well a learned model performs against unseen inputs. In this work we build upon a recent breakthrough in compression theory to develop a new framework yielding tight generalization bounds of wide practical applicability. The core idea is to embed any given learning algorithm into a suitably-constructed meta-algorithm (here called Pick-to-Learn, P2L) in order to instill desirable compression properties. When applied to the MNIST classification dataset and to a synthetic regression problem, P2L not only attains generalization bounds that compare favorably with the state of the art (test-set and PAC-Bayes bounds), but it also learns models with better post-training performance.
****
Orthogonal Non-negative Tensor Factorization based Multi-view Clustering
Jing Li, Quanxue Gao, QIANQIAN WANG, Ming Yang, Wei Xia
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
RVD: A Handheld Device-Based Fundus Video Dataset for Retinal Vessel Segmentation
MD WAHIDUZZAMAN KHAN, Hongwei Sheng, Hu Zhang, Heming Du, Sen Wang, Minas Coroneo, Farshid Hajati, Sahar Shariflou, Michael Kalloniatis, Jack Phu, Ashish Agar, Zi Huang, S.Mojtaba Golzan, Xin Yu
Retinal vessel segmentation is generally grounded in image-based datasets collected with bench-top devices. The static images naturally lose the dynamic characteristics of retina fluctuation, resulting in diminished dataset richness, and the usage of bench-top devices further restricts dataset scalability due to its limited accessibility. Considering these limitations, we introduce the first video-based retinal dataset by employing handheld devices for data acquisition. The dataset comprises 635 smartphone-based fundus videos collected from four different clinics, involving 415 patients from 50 to 75 years old. It delivers comprehensive and precise annotations of retinal structures in both spatial and temporal dimensions, aiming to advance the landscape of vasculature segmentation. Specifically, the dataset provides three levels of spatial annotations: binary vessel masks for overall retinal structure delineation, general vein-artery masks for distinguishing the vein and artery, and fine-grained vein-artery masks for further characterizing the granularities of each artery and vein. In addition, the dataset offers temporal annotations that capture the vessel pulsation characteristics, assisting in detecting ocular diseases that require fine-grained recognition of hemodynamic fluctuation. In application, our dataset exhibits a significant domain shift with respect to data captured by bench-top devices, thus posing great challenges to existing methods. Thanks to rich annotations and data scales, our dataset potentially paves the path for more advanced retinal analysis and accurate disease diagnosis. In the experiments, we provide evaluation metrics and be

nchmark results on our dataset, reflecting both the potential and challenges it
offers for vessel segmentation tasks. We hope this challenging dataset would sig
nificantly contribute to the development of eye disease diagnosis and early prev
ention.
****
LayoutGPT: Compositional Visual Planning and Generation with Large Language Mode
ls
Weixi Feng, Wanrong Zhu, Tsu-Jui Fu, Varun Jampani, Arjun Akula, Xuehai He, S Ba
su, Xin Eric Wang, William Yang Wang
Attaining a high degree of user controllability in visual generation often requi
res intricate, fine-grained inputs like layouts. However, such inputs impose a s
ubstantial burden on users when compared to simple text inputs. To address the i
ssue, we study how Large Language Models (LLMs) can serve as visual planners by
generating layouts from text conditions, and thus collaborate with visual genera
tive models. We propose LayoutGPT, a method to compose in-context visual demonst
rations in style sheet language to enhance visual planning skills of LLMs. We sh
ow that LayoutGPT can generate plausible layouts in multiple domains, ranging fr
om 2D images to 3D indoor scenes. LayoutGPT also shows superior performance in c
onverting challenging language concepts like numerical and spatial relations to
layout arrangements for faithful text-to-image generation. When combined with a
downstream image generation model, LayoutGPT outperforms text-to-image models/sy
stems by 20-40\% and achieves comparable performance as human users in designing
 visual layouts for numerical and spatial correctness. Lastly, LayoutGPT achieve
s comparable performance to supervised methods in 3D indoor scene synthesis, dem
onstrating its effectiveness and potential in multiple visual domains.
****
Data Pruning via Moving-one-Sample-out
Haoru Tan, Sitong Wu, Fei Du, Yukang Chen, Zhibin Wang, Fan Wang, Xiaojuan Qi
In this paper, we propose a novel data-pruning approach called moving-one-sample
-out (MoSo), which aims to identify and remove the least informative samples fro
m the training set. The core insight behind MoSo is to determine the importance
of each sample by assessing its impact on the optimal empirical risk. This is ac
hieved by measuring the extent to which the empirical risk changes when a partic
ular sample is excluded from the training set. Instead of using the computationa
lly expensive leaving-one-out-retraining procedure, we propose an efficient firs
t-order approximator that only requires gradient information from different trai
ning stages. The key idea behind our approximation is that samples with gradient
s that are consistently aligned with the average gradient of the training set ar
e more informative and should receive higher scores, which could be intuitively
understood as follows: if the gradient from a specific sample is consistent with
 the average gradient vector, it implies that optimizing the network using the s
ample will yield a similar effect on all remaining samples. Experimental result
s demonstrate that MoSo effectively mitigates severe performance degradation at
high pruning ratios and achieves satisfactory performance across various setting
s. Experimental results demonstrate that MoSo effectively mitigates severe perfo
rmance degradation at high pruning ratios and outperforms state-of-the-art metho
ds by a large margin across various settings.
****
Alternation makes the adversary weaker in two-player games
Volkan Cevher, Ashok Cutkosky, Ali Kavis, Georgios Piliouras, Stratis Skoulakis,
 Luca Viano
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Spuriosity Didn’t Kill the Classifier: Using Invariant Predictions to Harness Sp
urious Features
Cian Eastwood, Shashank Singh, Andrei L Nicolicioiu, Marin Vlastelica Pogan■i■,
Julius von Kügelgen, Bernhard Schölkopf

To avoid failures on out-of-distribution data, recent works have sought to extract features that have an invariant or stable relationship with the label across domains, discarding "spurious" or unstable features whose relationship with the label changes across domains. However, unstable features often carry complementary information that could boost performance if used correctly in the test domain. In this work, we show how this can be done without test-domain labels. In particular, we prove that pseudo-labels based on stable features provide sufficient guidance for doing so, provided that stable and unstable features are conditionally independent given the label. Based on this theoretical insight, we propose Stable Feature Boosting (SFB), an algorithm for: (i) learning a predictor that separates stable and conditionally-independent unstable features; and (ii) using the stable-feature predictions to adapt the unstable-feature predictions in the test domain. Theoretically, we prove that SFB can learn an asymptotically-optimal predictor without test-domain labels. Empirically, we demonstrate the effectiveness of SFB on real and synthetic data.
****
A Pseudo-Semantic Loss for Autoregressive Models with Logical Constraints
Kareem Ahmed, Kai-Wei Chang, Guy Van den Broeck
Neuro-symbolic AI bridges the gap between purely symbolic and neural approaches to learning. This often requires maximizing the likelihood of a symbolic constraint w.r.t the neural network's output distribution. Such output distributions are typically assumed to be fully-factorized. This limits the applicability of neuro-symbolic learning to the more expressive auto-regressive distributions, e.g., transformers. Under such distributions, computing the likelihood of even simple constraints is #P-hard. Instead of attempting to enforce the constraint on the entire likelihood distribution, we propose to do so on a random, local approximation thereof. More precisely, we approximate the likelihood of the constraint with the pseudolikelihood of the constraint centered around a model sample. Our approach is factorizable, allowing us to reuse solutions to sub-problems---a main tenet for the efficient computation of neuro-symbolic losses. It also provides a local, high fidelity approximation of the likelihood: it exhibits low entropy and KL-divergence around the model sample. We tested our approach on Sudoku and shortest-path prediction cast as auto-regressive generation, and observe that we greatly improve upon the base model's ability to predict logically-consistent outputs. We also tested our approach on the task of detoxifying large language models. We observe that using a simple constraint disallowing a list of toxic words, we are able to steer the model's outputs away from toxic generations, achieving SoTA compared to previous approaches.
****
Physics-Informed Bayesian Optimization of Variational Quantum Circuits
Kim Nicoli, Christopher J. Anders, Lena Funcke, Tobias Hartung, Karl Jansen, Stefan Kühn, Klaus-Robert Müller, Paolo Stornati, Pan Kessel, Shinichi Nakajima
In this paper, we propose a novel and powerful method to harness Bayesian optimization for variational quantum eigensolvers (VQEs) - a hybrid quantum-classical protocol used to approximate the ground state of a quantum Hamiltonian. Specifically, we derive a VQE-kernel which incorporates important prior information about quantum circuits: the kernel feature map of the VQE-kernel exactly matches the known functional form of the VQE's objective function and thereby significantly reduces the posterior uncertainty.Moreover, we propose a novel acquisition function for Bayesian optimization called \emph{Expected Maximum Improvement over Confident Regions} (EMICoRe) which can actively exploit the inductive bias of the VQE-kernel by treating regions with low predictive uncertainty as indirectly "observed". As a result, observations at as few as three points in the search domain are sufficient to determine the complete objective function along an entire one-dimensional subspace of the optimization landscape. Our numerical experiments demonstrate that our approach improves over state-of-the-art baselines.
****
Rubik's Cube: High-Order Channel Interactions with a Hierarchical Receptive Field
Naishan Zheng, man zhou, Chong Zhou, Chen Change Loy

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Closing the Computational-Statistical Gap in Best Arm Identification for Combinatorial Semi-bandits
Ruo-Chun Tzeng, Po-An Wang, Alexandre Proutiere, Chi-Jen Lu
We study the best arm identification problem in combinatorial semi-bandits in the fixed confidence setting. We present Perturbed Frank-Wolfe Sampling (P-FWS), an algorithm that (i) runs in polynomial time, (ii) achieves the instance-specific minimal sample complexity in the high confidence regime, and (iii) enjoys polynomial sample complexity guarantees in the moderate confidence regime. To our best knowledge, existing algorithms cannot achieve (ii) and (iii) simultaneously in vanilla bandits. With P-FWS, we close the computational-statistical gap in best arm identification in combinatorial semi-bandits. The design of P-FWS starts from the optimization problem that defines the information-theoretical and instance-specific sample complexity lower bound. P-FWS solves this problem in an online manner using, in each round, a single iteration of the Frank-Wolfe algorithm. Structural properties of the problem are leveraged to make the P-FWS successive updates computationally efficient. In turn, P-FWS only relies on a simple linear maximization oracle.
****
Imitation Learning from Imperfection: Theoretical Justifications and Algorithms
Ziniu Li, Tian Xu, Zeyu Qin, Yang Yu, Zhi-Quan Luo
Imitation learning (IL) algorithms excel in acquiring high-quality policies from expert data for sequential decision-making tasks. But, their effectiveness is hampered when faced with limited expert data. To tackle this challenge, a novel framework called (offline) IL with supplementary data has been proposed, which enhances learning by incorporating an additional yet imperfect dataset obtained inexpensively from sub-optimal policies. Nonetheless, learning becomes challenging due to the potential inclusion of out-of-expert-distribution samples. In this work, we propose a mathematical formalization of this framework, uncovering its limitations. Our theoretical analysis reveals that a naive approach—applying the behavioral cloning (BC) algorithm concept to the combined set of expert and supplementary data—may fall short of vanilla BC, which solely relies on expert data. This deficiency arises due to the distribution shift between the two data sources. To address this issue, we propose a new importance-sampling-based technique for selecting data within the expert distribution. We prove that the proposed method eliminates the gap of the naive approach, highlighting its efficacy when handling imperfect data. Empirical studies demonstrate that our method outperforms previous state-of-the-art methods in tasks including robotic locomotion control, Atari video games, and image classification. Overall, our work underscores the potential of improving IL by leveraging diverse data sources through effective data selection.
****
Detection Based Part-level Articulated Object Reconstruction from Single RGBD Image
Yuki Kawana, Tatsuya Harada
We propose an end-to-end trainable, cross-category method for reconstructing multiple man-made articulated objects from a single RGBD image, focusing on part-level shape reconstruction and pose and kinematics estimation. We depart from previous works that rely on learning instance-level latent space, focusing on man-made articulated objects with predefined part counts. Instead, we propose a novel alternative approach that employs part-level representation, representing instances as combinations of detected parts. While our detect-then-group approach effectively handles instances with diverse part structures and various part counts, it faces issues of false positives, varying part sizes and scales, and an increasing model size due to end-to-end training. To address these challenges, we propose 1) test-time kinematics-aware part fusion to improve detection performance w

hile suppressing false positives, 2) anisotropic scale normalization for part shape learning to accommodate various part sizes and scales, and 3) a balancing strategy for cross-refinement between feature space and output space to improve part detection while maintaining model size. Evaluation on both synthetic and real data demonstrates that our method successfully reconstructs variously structured multiple instances that previous works cannot handle, and outperforms prior works in shape reconstruction and kinematics estimation.
****

FlatMatch: Bridging Labeled Data and Unlabeled Data with Cross-Sharpness for Semi-Supervised Learning

Zhuo Huang, Li Shen, Jun Yu, Bo Han, Tongliang Liu

Semi-Supervised Learning (SSL) has been an effective way to leverage abundant unlabeled data with extremely scarce labeled data. However, most SSL methods are commonly based on instance-wise consistency between different data transformations. Therefore, the label guidance on labeled data is hard to be propagated to unlabeled data. Consequently, the learning process on labeled data is much faster than on unlabeled data which is likely to fall into a local minima that does not favor unlabeled data, leading to sub-optimal generalization performance. In this paper, we propose FlatMatch which minimizes a cross-sharpness measure to ensure consistent learning performance between the two datasets. Specifically, we increase the empirical risk on labeled data to obtain a worst-case model which is a failure case needing to be enhanced. Then, by leveraging the richness of unlabeled data, we penalize the prediction difference (i.e., cross-sharpness) between the worst-case model and the original model so that the learning direction is beneficial to generalization on unlabeled data. Therefore, we can calibrate the learning process without being limited to insufficient label information. As a result, the mismatched learning performance can be mitigated, further enabling the effective exploitation of unlabeled data and improving SSL performance. Through comprehensive validation, we show FlatMatch achieves state-of-the-art results in many SSL settings.
****

Neural Sculpting: Uncovering hierarchically modular task structure in neural networks through pruning and network analysis

Shreyas Malakarjun Patil, Loizos Michael, Constantine Dovrolis

Natural target functions and tasks typically exhibit hierarchical modularity -- they can be broken down into simpler sub-functions that are organized in a hierarchy. Such sub-functions have two important features: they have a distinct set of inputs (input-separability) and they are reused as inputs higher in the hierarchy (reusability). Previous studies have established that hierarchically modular neural networks, which are inherently sparse, offer benefits such as learning efficiency, generalization, multi-task learning, and transfer. However, identifying the underlying sub-functions and their hierarchical structure for a given task can be challenging. The high-level question in this work is: if we learn a task using a sufficiently deep neural network, how can we uncover the underlying hierarchy of sub-functions in that task? As a starting point, we examine the domain of Boolean functions, where it is easier to determine whether a task is hierarchically modular. We propose an approach based on iterative unit and edge pruning (during training), combined with network analysis for module detection and hierarchy inference. Finally, we demonstrate that this method can uncover the hierarchical modularity of a wide range of Boolean functions and two vision tasks based on the MNIST digits dataset.
****

Elastic Decision Transformer

Yueh-Hua Wu, Xiaolong Wang, Masashi Hamaya

This paper introduces Elastic Decision Transformer (EDT), a significant advancement over the existing Decision Transformer (DT) and its variants. Although DT purports to generate an optimal trajectory, empirical evidence suggests it struggles with trajectory stitching, a process involving the generation of an optimal or near-optimal trajectory from the best parts of a set of sub-optimal trajectories. The proposed EDT differentiates itself by facilitating trajectory stitching

during action inference at test time, achieved by adjusting the history length m
aintained in DT. Further, the EDT optimizes the trajectory by retaining a longer
 history when the previous trajectory is optimal and a shorter one when it is su
b-optimal, enabling it to "stitch" with a more optimal trajectory. Extensive exp
erimentation demonstrates EDT's ability to bridge the performance gap between DT
-based and Q Learning-based approaches. In particular, the EDT outperforms Q Lea
rning-based methods in a multi-task regime on the D4RL locomotion benchmark and
Atari games.
****
Asymptotically Optimal Quantile Pure Exploration for Infinite-Armed Bandits
Evelyn Xiao-Yue Gong, Mark Sellke
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Learning Probabilistic Symmetrization for Architecture Agnostic Equivariance
Jinwoo Kim, Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, Seunghoon Hong
We present a novel framework to overcome the limitations of equivariant architec
tures in learning functions with group symmetries. In contrary to equivariant ar
chitectures, we use an arbitrary base model such as an MLP or a transformer and
symmetrize it to be equivariant to the given group by employing a small equivari
ant network that parameterizes the probabilistic distribution underlying the sym
metrization. The distribution is end-to-end trained with the base model which ca
n maximize performance while reducing sample complexity of symmetrization. We sh
ow that this approach ensures not only equivariance to given group but also univ
ersal approximation capability in expectation. We implement our method on variou
s base models, including patch-based transformers that can be initialized from p
retrained vision transformers, and test them for a wide range of symmetry groups
 including permutation and Euclidean groups and their combinations. Empirical te
sts show competitive results against tailored equivariant architectures, suggest
ing the potential for learning equivariant functions for diverse groups using a
non-equivariant universal base architecture. We further show evidence of enhance
d learning in symmetric modalities, like graphs, when pretrained from non-symmet
ric modalities, like vision. Code is available at https://github.com/jw9730/lps.
****
Distributionally Robust Linear Quadratic Control
Bahar Taskesen, Dan Iancu, Ça■■l Koçyi■it, Daniel Kuhn
Linear-Quadratic-Gaussian (LQG) control is a fundamental control paradigm that i
s studied in various fields such as engineering, computer science, economics, an
d neuroscience. It involves controlling a system with linear dynamics and imperf
ect observations, subject to additive noise, with the goal of minimizing a quadr
atic cost function for the state and control variables. In this work, we conside
r a generalization of the discrete-time, finite-horizon LQG problem, where the n
oise distributions are unknown and belong to Wasserstein ambiguity sets centered
 at nominal (Gaussian) distributions. The objective is to minimize a worst-case
cost across all distributions in the ambiguity set, including non-Gaussian distr
ibutions. Despite the added complexity, we prove that a control policy that is l
inear in the observations is optimal for this problem, as in the classic LQG pro
blem. We propose a numerical solution method that efficiently characterizes this
 optimal control policy. Our method uses the Frank-Wolfe algorithm to identify t
he least-favorable distributions within the Wasserstein ambiguity sets and compu
tes the controller's optimal policy using Kalman filter estimation under these d
istributions.
****
Fully Dynamic k-Clustering in $\tilde O(k)$ Update Time
Sayan Bhattacharya, Martín Costa, Silvio Lattanzi, Nikos Parotsidis
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
FreeMask: Synthetic Images with Dense Annotations Make Stronger Segmentation Models
Lihe Yang, Xiaogang Xu, Bingyi Kang, Yinghuan Shi, Hengshuang Zhao
Semantic segmentation has witnessed tremendous progress due to the proposal of various advanced network architectures. However, they are extremely hungry for delicate annotations to train, and the acquisition is laborious and unaffordable. Therefore, we present FreeMask in this work, which resorts to synthetic images from generative models to ease the burden of both data collection and annotation procedures. Concretely, we first synthesize abundant training images conditioned on the semantic masks provided by realistic datasets. This yields extra well-aligned image-mask training pairs for semantic segmentation models. We surprisingly observe that, solely trained with synthetic images, we already achieve comparable performance with real ones (e.g., 48.3 vs. 48.5 mIoU on ADE20K, and 49.3 vs. 50.5 on COCO-Stuff). Then, we investigate the role of synthetic images by joint training with real images, or pre-training for real images. Meantime, we design a robust filtering principle to suppress incorrectly synthesized regions. In addition, we propose to inequally treat different semantic masks to prioritize those harder ones and sample more corresponding synthetic images for them. As a result, either jointly trained or pre-trained with our filtered and re-sampled synthesized images, segmentation models can be greatly enhanced, e.g., from 48.7 to 52.0 on ADE20K.
****
RS-Del: Edit Distance Robustness Certificates for Sequence Classifiers via Randomized Deletion
Zhuoqun Huang, Neil G Marchant, Keane Lucas, Lujo Bauer, Olga Ohrimenko, Benjamin Rubinstein
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Flow: Per-instance Personalized Federated Learning
Kunjal Panchal, Sunav Choudhary, Nisarg Parikh, Lijun Zhang, Hui Guan
Federated learning (FL) suffers from data heterogeneity, where the diverse data distributions across clients make it challenging to train a single global model effectively. Existing personalization approaches aim to address the data heterogeneity issue by creating a personalized model for each client from the global model that fits their local data distribution. However, these personalized models may achieve lower accuracy than the global model in some clients, resulting in limited performance improvement compared to that without personalization. To overcome this limitation, we propose a per-instance personalization FL algorithm Flow. Flow creates dynamic personalized models that are adaptive not only to each client’s data distributions but also to each client’s data instances. The personalized model allows each instance to dynamically determine whether it prefers the local parameters or its global counterpart to make correct predictions, thereby improving clients’accuracy. We provide theoretical analysis on the convergence of Flow and empirically demonstrate the superiority of Flow in improving clients’ accuracy compared to state-of-the-art personalization approaches on both vision and language-based tasks.
****
MM-Fi: Multi-Modal Non-Intrusive 4D Human Dataset for Versatile Wireless Sensing
Jianfei Yang, He Huang, Yunjiao Zhou, Xinyan Chen, Yuecong Xu, Shenghai Yuan, Han Zou, Chris Xiaoxuan Lu, Lihua Xie
4D human perception plays an essential role in a myriad of applications, such as home automation and metaverse avatar simulation. However, existing solutions which mainly rely on cameras and wearable devices are either privacy intrusive or inconvenient to use. To address these issues, wireless sensing has emerged as a promising alternative, leveraging LiDAR, mmWave radar, and WiFi signals for devi

ce-free human sensing. In this paper, we propose MM-Fi, the first multi-modal no
n-intrusive 4D human dataset with 27 daily or rehabilitation action categories,
to bridge the gap between wireless sensing and high-level human perception tasks
. MM-Fi consists of over 320k synchronized frames of five modalities from 40 hum
an subjects. Various annotations are provided to support potential sensing tasks
, e.g., human pose estimation and action recognition. Extensive experiments have
 been conducted to compare the sensing capacity of each or several modalities in
 terms of multiple tasks. We envision that MM-Fi can contribute to wireless sens
ing research with respect to action recognition, human pose estimation, multi-mo
dal learning, cross-modal supervision, and interdisciplinary healthcare research
.
****
Live Graph Lab: Towards Open, Dynamic and Real Transaction Graphs with NFT
Zhen Zhang, Bingqiao Luo, Shengliang Lu, Bingsheng He
Numerous studies have been conducted to investigate the properties of large-scal
e temporal graphs. Despite the ubiquity of these graphs in real-world scenarios,
 it's usually impractical for us to obtain the whole real-time graphs due to pri
vacy concerns and technical limitations. In this paper, we introduce the concept
 of {\it Live Graph Lab} for temporal graphs, which enables open, dynamic and re
al transaction graphs from blockchains. Among them, Non-fungible tokens (NFTs) h
ave become one of the most prominent parts of blockchain over the past several y
ears. With more than \$40 billion market capitalization, this decentralized ecos
ystem produces massive, anonymous and real transaction activities, which natural
ly forms a complicated transaction network. However, there is limited understand
ing about the characteristics of this emerging NFT ecosystem from a temporal gra
ph analysis perspective. To mitigate this gap, we instantiate a live graph with
NFT transaction network and investigate its dynamics to provide new observations
 and insights. Specifically, through downloading and parsing the NFT transaction
 activities, we obtain a temporal graph with more than 4.5 million nodes and 124
 million edges. Then, a series of measurements are presented to understand the p
roperties of the NFT ecosystem. Through comparisons with social, citation, and w
eb networks, our analyses give intriguing findings and point out potential direc
tions for future exploration. Finally, we also study machine learning models in
this live graph to enrich the current datasets and provide new opportunities for
 the graph community. The source codes and dataset are available at https://live
graphlab.github.io.
****
CMMA: Benchmarking Multi-Affection Detection in Chinese Multi-Modal Conversation
s
Yazhou Zhang, Yang Yu, Qing Guo, Benyou Wang, Dongming Zhao, Sagar Uprety, Dawei
 Song, Qiuchi Li, Jing Qin
Human communication has a multi-modal and multi-affection nature. The inter-rela
tedness of different emotions and sentiments poses a challenge to jointly detect
 multiple human affections with multi-modal clues. Recent advances in this field
 employed multi-task learning paradigms to render the inter-relatedness across t
asks, but the scarcity of publicly available resources sets a limit to the poten
tial of works. To fill this gap, we build the first Chinese Multi-modal Multi-Af
fection conversation (CMMA) dataset, which contains 3,000 multi-party conversati
ons and 21,795 multi-modal utterances collected from various styles of TV-series
. CMMA contains a wide variety of affection labels, including sentiment, emotion
, sarcasm and humor, as well as the novel inter-correlations values between cert
ain pairs of tasks. Moreover, it provides the topic and speaker information in c
onversations, which promotes better modeling of conversational context. On the d
ataset, we empirically analyze the influence of different data modalities and co
nversational contexts on different affection analysis tasks, and exhibit the pra
ctical benefit of inter-task correlations. The full dataset will be publicly ava
ilable for research\footnote{https://github.com/annoymity2022/Chinese-Dataset}
****
Inverse Preference Learning: Preference-based RL without a Reward Function
Joey Hejna, Dorsa Sadigh

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Matrix Compression via Randomized Low Rank and Low Precision Factorization
Rajarshi Saha, Varun Srivastava, Mert Pilanci
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

OpenLane-V2: A Topology Reasoning Benchmark for Unified 3D HD Mapping
Huijie Wang, Tianyu Li, Yang Li, Li Chen, Chonghao Sima, Zhenbo Liu, Bangjun Wang, Peijin Jia, Yuting Wang, Shengyin Jiang, Feng Wen, Hang Xu, Ping Luo, Junchi Yan, Wei Zhang, Hongyang Li
Accurately depicting the complex traffic scene is a vital component for autonomous vehicles to execute correct judgments. However, existing benchmarks tend to oversimplify the scene by solely focusing on lane perception tasks. Observing that human drivers rely on both lanes and traffic signals to operate their vehicles safely, we present OpenLane-V2, the first dataset on topology reasoning for traffic scene structure. The objective of the presented dataset is to advance research in understanding the structure of road scenes by examining the relationship between perceived entities, such as traffic elements and lanes. Leveraging existing datasets, OpenLane-V2 consists of 2,000 annotated road scenes that describe traffic elements and their correlation to the lanes. It comprises three primary sub-tasks, including the 3D lane detection inherited from OpenLane, accompanied by corresponding metrics to evaluate the model’s performance. We evaluate various state-of-the-art methods, and present their quantitative and qualitative results on OpenLane-V2 to indicate future avenues for investigating topology reasoning in traffic scenes.
****

Prompt-augmented Temporal Point Process for Streaming Event Sequence
Siqiao Xue, Yan Wang, Zhixuan Chu, Xiaoming Shi, Caigao JIANG, Hongyan Hao, Gangwei Jiang, Xiaoyun Feng, James Zhang, Jun Zhou
Neural Temporal Point Processes (TPPs) are the prevalent paradigm for modeling continuous-time event sequences, such as user activities on the web and financial transactions. In real world applications, the event data typically comes in a streaming manner, where the distribution of the patterns may shift over time. Under the privacy and memory constraints commonly seen in real scenarios, how to continuously monitor a TPP to learn the streaming event sequence is an important yet under-investigated problem. In this work, we approach this problem by adopting Continual Learning (CL), which aims to enable a model to continuously learn a sequence of tasks without catastrophic forgetting. While CL for event sequence is less well studied, we present a simple yet effective framework, PromptTPP, by integrating the base TPP with a continuous-time retrieval prompt pool. In our proposed framework, prompts are small learnable parameters, maintained in a memory space and jointly optimized with the base TPP so that the model is properly instructed to learn event streams arriving sequentially without buffering past examples or task-specific attributes. We formalize a novel and realistic experimental setup for modeling event streams, where PromptTPP consistently sets state-of-the-art performance across two real user behavior datasets.
****

Leveraging Locality and Robustness to Achieve Massively Scalable Gaussian Process Regression
Robert Allison, Anthony Stephenson, Samuel F, Edward O Pyzer-Knapp
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****
Building the Bridge of Schrödinger: A Continuous Entropic Optimal Transport Benc
hmark
Nikita Gushchin, Alexander Kolesov, Petr Mokrov, Polina Karpikova, Andrei Spirid
onov, Evgeny Burnaev, Alexander Korotin
Over the last several years, there has been significant progress in developing n
eural solvers for the Schrödinger Bridge (SB) problem and applying them to gener
ative modelling. This new research field is justifiably fruitful as it is interc
onnected with the practically well-performing diffusion models and theoretically
 grounded entropic optimal transport (EOT). Still, the area lacks non-trivial te
sts allowing a researcher to understand how well the methods solve SB or its equ
ivalent continuous EOT problem. We fill this gap and propose a novel way to crea
te pairs of probability distributions for which the ground truth OT solution is
known by the construction. Our methodology is generic and works for a wide range
 of OT formulations, in particular, it covers the EOT which is equivalent to SB
(the main interest of our study). This development allows us to create continuou
s benchmark distributions with the known EOT and SB solutions on high-dimensiona
l spaces such as spaces of images. As an illustration, we use these benchmark pa
irs to test how well existing neural EOT/SB solvers actually compute the EOT sol
ution. Our code for constructing benchmark pairs under different setups is avail
able at: https://github.com/ngushchin/EntropicOTBenchmark
****
Safety Gymnasium: A Unified Safe Reinforcement Learning Benchmark
Jiaming Ji, Borong Zhang, Jiayi Zhou, Xuehai Pan, Weidong Huang, Ruiyang Sun, Yi
ran Geng, Yifan Zhong, Josef Dai, Yaodong Yang
Artificial intelligence (AI) systems possess significant potential to drive soci
etal progress. However, their deployment often faces obstacles due to substantia
l safety concerns. Safe reinforcement learning (SafeRL) emerges as a solution to
 optimize policies while simultaneously adhering to multiple constraints, thereb
y addressing the challenge of integrating reinforcement learning in safety-criti
cal scenarios. In this paper, we present an environment suite called Safety-Gymn
asium, which encompasses safety-critical tasks in both single and multi-agent sc
enarios, accepting vector and vision-only input. Additionally, we offer a librar
y of algorithms named Safe Policy Optimization (SafePO), comprising 16 state-of-
the-art SafeRL algorithms. This comprehensive library can serve as a validation
tool for the research community. By introducing this benchmark, we aim to facili
tate the evaluation and comparison of safety performance, thus fostering the dev
elopment of reinforcement learning for safer, more reliable, and responsible rea
l-world applications. The website of this project can be accessed at https://sit
es.google.com/view/safety-gymnasium.
****
Direct Training of SNN using Local Zeroth Order Method
Bhaskar Mukhoty, Velibor Bojkovic, William de Vazelhes, Xiaohan Zhao, Giulia De
Masi, Huan Xiong, Bin Gu
Spiking neural networks are becoming increasingly popular for their low energy r
equirement in real-world tasks with accuracy comparable to traditional ANNs. SNN
 training algorithms face the loss of gradient information and non-differentiabi
lity due to the Heaviside function in minimizing the model loss over model param
eters. To circumvent this problem, the surrogate method employs a differentiable
 approximation of the Heaviside function in the backward pass, while the forward
 pass continues to use the Heaviside as the spiking function. We propose to use
the zeroth-order technique at the local or neuron level in training SNNs, motiva
ted by its regularizing and potential energy-efficient effects and establish a t
heoretical connection between it and the existing surrogate methods. We perform
experimental validation of the technique on standard static datasets (CIFAR-10,
CIFAR-100, ImageNet-100) and neuromorphic datasets (DVS-CIFAR-10, DVS-Gesture, N
-Caltech-101, NCARS) and obtain results that offer improvement over the state-of
-the-art results. The proposed method also lends itself to efficient implementat
ions of the back-propagation method, which could provide 3-4 times overall speed
up in training time. The code is available at \url{https://github.com/BhaskarMuk

hoty/LocalZO}.
****
Discover and Align Taxonomic Context Priors for Open-world Semi-Supervised Lear ning
Yu Wang, Zhun Zhong, Pengchong Qiao, Xuxin Cheng, Xiawu Zheng, Chang Liu, Nicu S ebe, Rongrong Ji, Jie Chen
Open-world Semi-Supervised Learning (OSSL) is a realistic and challenging task, aiming to classify unlabeled samples from both seen and novel classes using part ially labeled samples from the seen classes. Previous works typically explore th e relationship of samples as priors on the pre-defined single-granularity labels to help novel class recognition. In fact, classes follow a taxonomy and samples can be classified at multiple levels of granularity, which contains more underl ying relationships for supervision. We thus argue that learning with single-gran ularity labels results in sub-optimal representation learning and inaccurate pse udo labels, especially with unknown classes. In this paper, we take the initiati ve to explore and propose a uniformed framework, called Taxonomic context prIors Discovering and Aligning (TIDA), which exploits the relationship of samples und er various granularity. It allows us to discover multi-granularity semantic conc epts as taxonomic context priors (i.e., sub-class, target-class, and super-class), and then collaboratively leverage them to enhance representation learning and improve the quality of pseudo labels.Specifically, TIDA comprises two component s: i) A taxonomic context discovery module that constructs a set of hierarchical prototypes in the latent space to discover the underlying taxonomic context pri ors; ii) A taxonomic context-based prediction alignment module that enforces con sistency across hierarchical predictions to build the reliable relationship betw een classes among various granularity and provide additions supervision. We demo nstrate that these two components are mutually beneficial for an effective OSSL framework, which is theoretically explained from the perspective of the EM algor ithm. Extensive experiments on seven commonly used datasets show that TIDA can s ignificantly improve the performance and achieve a new state of the art. The sou rce codes are publicly available at https://github.com/rain305f/TIDA.
****
Curve Your Enthusiasm: Concurvity Regularization in Differentiable Generalized A dditive Models
Julien Siems, Konstantin Ditschuneit, Winfried Ripken, Alma Lindborg, Maximilian Schambach, Johannes Otterbach, Martin Genzel
Generalized Additive Models (GAMs) have recently experienced a resurgence in pop ularity due to their interpretability, which arises from expressing the target v alue as a sum of non-linear transformations of the features. Despite the current enthusiasm for GAMs, their susceptibility to concurvity — i.e., (possibly non-l inear) dependencies between the features — has hitherto been largely overlooked. Here, we demonstrate how concurvity can severly impair the interpretability of GAMs and propose a remedy: a conceptually simple, yet effective regularizer whic h penalizes pairwise correlations of the non-linearly transformed feature variab les. This procedure is applicable to any differentiable additive model, such as Neural Additive Models or NeuralProphet, and enhances interpretability by elimin ating ambiguities due to self-canceling feature contributions. We validate the e ffectiveness of our regularizer in experiments on synthetic as well as real-worl d datasets for time-series and tabular data. Our experiments show that concurvit y in GAMs can be reduced without significantly compromising prediction quality, improving interpretability and reducing variance in the feature importances.
****
Mutual Information Regularized Offline Reinforcement Learning
Xiao Ma, Bingyi Kang, Zhongwen Xu, Min Lin, Shuicheng Yan
The major challenge of offline RL is the distribution shift that appears when ou t-of-distribution actions are queried, which makes the policy improvement direct ion biased by extrapolation errors. Most existing methods address this problem b y penalizing the policy or value for deviating from the behavior policy during p olicy improvement or evaluation. In this work, we propose a novel MISA framework to approach offline RL from the perspective of Mutual Information between State

s and Actions in the dataset by directly constraining the policy improvement direction. MISA constructs lower bounds of mutual information parameterized by the policy and Q-values. We show that optimizing this lower bound is equivalent to maximizing the likelihood of a one-step improved policy on the offline dataset. Hence, we constrain the policy improvement direction to lie in the data manifold. The resulting algorithm simultaneously augments the policy evaluation and improvement by adding mutual information regularizations. MISA is a general framework that unifies conservative Q-learning (CQL) and behavior regularization methods (e.g., TD3+BC) as special cases. We introduce 3 different variants of MISA, and empirically demonstrate that tighter mutual information lower bound gives better offline RL performance. In addition, our extensive experiments show MISA significantly outperforms a wide range of baselines on various tasks of the D4RL benchmark, e.g., achieving 742.9 total points on gym-locomotion tasks. Our code is attached and will be released upon publication.
****
Have it your way: Individualized Privacy Assignment for DP-SGD
Franziska Boenisch, Christopher Mühl, Adam Dziedzic, Roy Rinberg, Nicolas Papernot
When training a machine learning model with differential privacy, one sets a privacy budget. This uniform budget represents an overall maximal privacy violation that any user is willing to face by contributing their data to the training set. We argue that this approach is limited because different users may have different privacy expectations. Thus, setting a uniform privacy budget across all points may be overly conservative for some users or, conversely, not sufficiently protective for others. In this paper, we capture these preferences through individualized privacy budgets. To demonstrate their practicality, we introduce a variant of Differentially Private Stochastic Gradient Descent (DP-SGD) which supports such individualized budgets. DP-SGD is the canonical approach to training models with differential privacy. We modify its data sampling and gradient noising mechanisms to arrive at our approach, which we call Individualized DP-SGD (IDP-SGD). Because IDP-SGD provides privacy guarantees tailored to the preferences of individual users and their data points, we empirically find it to improve privacy-utility trade-offs.
****
Penguin: Parallel-Packed Homomorphic Encryption for Fast Graph Convolutional Network Inference
Ran Ran, Nuo Xu, Tao Liu, Wei Wang, Gang Quan, Wujie Wen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Dynamic Attribute-factored World Models for Efficient Multi-object Reinforcement Learning
Fan Feng, Sara Magliacane
In many reinforcement learning tasks, the agent has to learn to interact with many objects of different types and generalize to unseen combinations and numbers of objects. Often a task is a composition of previously learned tasks (e.g. block stacking).These are examples of compositional generalization, in which we compose object-centric representations to solve complex tasks. Recent works have shown the benefits of object-factored representations and hierarchical abstractions for improving sample efficiency in these settings. On the other hand, these methods do not fully exploit the benefits of factorization in terms of object attributes. In this paper, we address this opportunity and introduce the Dynamic Attribute FacTored RL (DAFT-RL) framework. In DAFT-RL, we leverage object-centric representation learning to extract objects from visual inputs. We learn to classify them into classes and infer their latent parameters. For each class of object, we learn a class template graph that describes how the dynamics and reward of an object of this class factorize according to its attributes. We also learn an interaction pattern graph that describes how objects of different classes interac

t with each other at the attribute level. Through these graphs and a dynamic int
eraction graph that models the interactions between objects, we can learn a poli
cy that can then be directly applied in a new environment by estimating the inte
ractions and latent parameters.We evaluate DAFT-RL in three benchmark datasets a
nd show our framework outperforms the state-of-the-art in generalizing across un
seen objects with varying attributes and latent parameters, as well as in the co
mposition of previously learned tasks.
****
Statistical Insights into HSIC in High Dimensions
Tao Zhang, Yaowu Zhang, Tingyou Zhou
Measuring the nonlinear dependence between random vectors and testing for their
statistical independence is a fundamental problem in statistics. One of the most
 popular dependence measures is the Hilbert-Schmidt independence criterion (HSIC
), which has attracted increasing attention in recent years. However, most exist
ing works have focused on either fixed or very high-dimensional covariates. In t
his work, we bridge the gap between these two scenarios and provide statistical
insights into the performance of HSIC when the dimensions grow at different rate
s. We first show that, under the null hypothesis, the rescaled HSIC converges in
 distribution to a standard normal distribution. Then we provide a general condi
tion for the HSIC based tests to have nontrivial power in high dimensions. By de
composing this condition, we illustrate how the ability of HSIC to measure nonli
near dependence changes with increasing dimensions. Moreover, we demonstrate tha
t, depending on the sample size, the covariate dimensions and the dependence str
uctures within covariates, the HSIC can capture different types of associations
between random vectors. We also conduct extensive numerical studies to validate
our theoretical results.
****
Fair Adaptive Experiments
Waverly Wei, Xinwei Ma, Jingshen Wang
Randomized experiments have been the gold standard for assessing the effectivene
ss of a treatment, policy, or intervention, spanning various fields, including s
ocial sciences, biomedical studies, and e-commerce. The classical complete rando
mization approach assigns treatments based on a pre-specified probability and ma
y lead to inefficient use of data. Adaptive experiments improve upon complete ra
ndomization by sequentially learning and updating treatment assignment probabili
ties using accrued evidence during the experiment. Hence, they can help achieve
efficient data use and higher estimation efficiency. However, their application
can also raise fairness and equity concerns, as assignment probabilities may var
y drastically across groups of participants. Furthermore, when treatment is expe
cted to be extremely beneficial to certain groups of participants, it is more ap
propriate to expose many of these participants to favorable treatment. In respon
se to these challenges, we propose a fair adaptive experiment strategy that simu
ltaneously enhances data use efficiency, achieves an ``envy-free'' treatment ass
ignment guarantee, and improves the overall welfare of participants. An importan
t feature of our proposed strategy is that we do not impose parametric modeling
assumptions on the outcome variables, making it more versatile and applicable to
 a wider array of applications. Through our theoretical investigation, we charac
terize the convergence rate of the estimated treatment effects and the associate
d standard deviations at the group level and further prove that our adaptive tre
atment assignment algorithm, despite not having a closed-form expression, approa
ches the optimal allocation rule asymptotically. Our proof strategy takes into a
ccount the fact that the allocation decisions in our design depend on sequential
ly accumulated data, which poses a significant challenge in characterizing the p
roperties and conducting statistical inference of our method. We further provide
 simulation evidence and two synthetic data studies to showcase the performance
of our fair adaptive experiment strategy.
****
Alexa Arena: A User-Centric Interactive Platform for Embodied AI
Qiaozi Gao, Govind Thattai, Suhaila Shakiah, Xiaofeng Gao, Shreyas Pansare, Vasu
 Sharma, Gaurav Sukhatme, Hangjie Shi, Bofei Yang, Desheng Zhang, Lucy Hu, Karth

ika Arumugam, Shui Hu, Matthew Wen, Dinakar Guthy, Shunan Chung, Rohan Khanna, Osman Ipek, Leslie Ball, Kate Bland, Heather Rocker, Michael Johnston, Reza Ghanadan, Dilek Hakkani-Tur, Prem Natarajan
We introduce Alexa Arena, a user-centric simulation platform to facilitate research in building assistive conversational embodied agents. Alexa Arena features multi-room layouts and an abundance of interactable objects. With user-friendly graphics and control mechanisms, the platform supports the development of gamified robotic tasks readily accessible to general human users, allowing high-efficiency data collection and EAI system evaluation. Along with the platform, we introduce a dialog-enabled task completion benchmark with online human evaluations.
****
Synthetic Combinations: A Causal Inference Framework for Combinatorial Interventions
Abhineet Agarwal, Anish Agarwal, Suhas Vijaykumar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
PUCA: Patch-Unshuffle and Channel Attention for Enhanced Self-Supervised Image Denoising
Hyemi Jang, Junsung Park, Dahuin Jung, Jaihyun Lew, Ho Bae, Sungroh Yoon
Although supervised image denoising networks have shown remarkable performance on synthesized noisy images, they often fail in practice due to the difference between real and synthesized noise. Since clean-noisy image pairs from the real world are extremely costly to gather, self-supervised learning, which utilizes noisy input itself as a target, has been studied. To prevent a self-supervised denoising model from learning identical mapping, each output pixel should not be influenced by its corresponding input pixel; This requirement is known as J-invariance. Blind-spot networks (BSNs) have been a prevalent choice to ensure J-invariance in self-supervised image denoising. However, constructing variations of BSNs by injecting additional operations such as downsampling can expose blinded information, thereby violating J-invariance. Consequently, convolutions designed specifically for BSNs have been allowed only, limiting architectural flexibility. To overcome this limitation, we propose PUCA, a novel J-invariant U-Net architecture, for self-supervised denoising. PUCA leverages patch-unshuffle/shuffle to dramatically expand receptive fields while maintaining J-invariance and dilated attention blocks (DABs) for global context incorporation. Experimental results demonstrate that PUCA achieves state-of-the-art performance, outperforming existing methods in self-supervised image denoising.
****
Projection Regret: Reducing Background Bias for Novelty Detection via Diffusion Models
Sungik Choi, Hankook Lee, Honglak Lee, Moontae Lee
Novelty detection is a fundamental task of machine learning which aims to detect abnormal (i.e. out-of-distribution (OOD)) samples. Since diffusion models have recently emerged as the de facto standard generative framework with surprising generation results, novelty detection via diffusion models has also gained much attention. Recent methods have mainly utilized the reconstruction property of in-distribution samples. However, they often suffer from detecting OOD samples that share similar background information to the in-distribution data. Based on our observation that diffusion models can project any sample to an in-distribution sample with similar background information, we propose Projection Regret (PR), an efficient novelty detection method that mitigates the bias of non-semantic information. To be specific, PR computes the perceptual distance between the test image and its diffusion-based projection to detect abnormality. Since the perceptual distance often fails to capture semantic changes when the background information is dominant, we cancel out the background bias by comparing it against recursive projections. Extensive experiments demonstrate that PR outperforms the prior art of generative-model-based novelty detection methods by a significant margi

n.
****
Versatile Energy-Based Probabilistic Models for High Energy Physics
Taoli Cheng, Aaron C. Courville
As a classical generative modeling approach, energy-based models have the natural advantage of flexibility in the form of the energy function. Recently, energy-based models have achieved great success in modeling high-dimensional data in computer vision and natural language processing. In line with these advancements, we build a multi-purpose energy-based probabilistic model for High Energy Physics events at the Large Hadron Collider. This framework builds on a powerful generative model and describes higher-order inter-particle interactions. It suits different encoding architectures and builds on implicit generation. As for applicative aspects, it can serve as a powerful parameterized event generator for physics simulation, a generic anomalous signal detector free from spurious correlations, and an augmented event classifier for particle identification.
****
User-Level Differential Privacy With Few Examples Per User
Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Raghu Meka, Chiyuan Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Neural Lighting Simulation for Urban Scenes
Ava Pun, Gary Sun, Jingkang Wang, Yun Chen, Ze Yang, Sivabalan Manivasagam, Wei-Chiu Ma, Raquel Urtasun
Different outdoor illumination conditions drastically alter the appearance of urban scenes, and they can harm the performance of image-based robot perception systems if not seen during training. Camera simulation provides a cost-effective solution to create a large dataset of images captured under different lighting conditions. Towards this goal, we propose LightSim, a neural lighting camera simulation system that enables diverse, realistic, and controllable data generation. LightSim automatically builds lighting-aware digital twins at scale from collected raw sensor data and decomposes the scene into dynamic actors and static background with accurate geometry, appearance, and estimated scene lighting. These digital twins enable actor insertion, modification, removal, and rendering from a new viewpoint, all in a lighting-aware manner. LightSim then combines physically-based and learnable deferred rendering to perform realistic relighting of modified scenes, such as altering the sun location and modifying the shadows or changing the sun brightness, producing spatially- and temporally-consistent camera videos. Our experiments show that LightSim generates more realistic relighting results than prior work. Importantly, training perception models on data generated by LightSim can significantly improve their performance. Our project page is available at https://waabi.ai/lightsim/.
****
Learning to Compress Prompts with Gist Tokens
Jesse Mu, Xiang Li, Noah Goodman
Prompting is the primary way to utilize the multitask capabilities of language models (LMs), but prompts occupy valuable space in the input context window, and repeatedly encoding the same prompt is computationally inefficient. Finetuning and distillation methods allow for specialization of LMs without prompting, but require retraining the model for each task. To avoid this trade-off entirely, we present gisting, which trains an LM to compress prompts into smaller sets of "gist" tokens which can be cached and reused for compute efficiency. Gist models can be trained with no additional cost over standard instruction finetuning by simply modifying Transformer attention masks to encourage prompt compression. On decoder (LLaMA-7B) and encoder-decoder (FLAN-T5-XXL) LMs, gisting enables up to 26x compression of prompts, resulting in up to 40% FLOPs reductions, 4.2% wall time speedups, and storage savings, all with minimal loss in output quality.

****
A Heavy-Tailed Algebra for Probabilistic Programming
Feynman T. Liang, Liam Hodgkinson, Michael W. Mahoney
Despite the successes of probabilistic models based on passing noise through neural networks, recent work has identified that such methods often fail to capture tail behavior accurately---unless the tails of the base distribution are appropriately calibrated. To overcome this deficiency, we propose a systematic approach for analyzing the tails of random variables, and we illustrate how this approach can be used during the static analysis (before drawing samples) pass of a probabilistic programming language (PPL) compiler. To characterize how the tails change under various operations, we develop an algebra which acts on a three-parameter family of tail asymptotics and which is based on the generalized Gamma distribution. Our algebraic operations are closed under addition and multiplication; they are capable of distinguishing sub-Gaussians with differing scales; and they handle ratios sufficiently well to reproduce the tails of most important statistical distributions directly from their definitions. Our empirical results confirm that inference algorithms that leverage our heavy-tailed algebra attain superior performance across a number of density modeling and variational inference (VI) tasks.
****
AIMS: All-Inclusive Multi-Level Segmentation for Anything
Lu Qi, Jason Kuen, Weidong Guo, Jiuxiang Gu, Zhe Lin, Bo Du, Yu Xu, Ming-Hsuan Yang
Despite the progress of image segmentation for accurate visual entity segmentation, completing the diverse requirements of image editing applications for different-level region-of-interest selections remains unsolved. In this paper, we propose a new task, All-Inclusive Multi-Level Segmentation (AIMS), which segments visual regions into three levels: part, entity, and relation (two entities with some semantic relationships). We also build a unified AIMS model through multi-dataset multi-task training to address the two major challenges of annotation inconsistency and task correlation. Specifically, we propose task complementarity, association, and prompt mask encoder for three-level predictions. Extensive experiments demonstrate the effectiveness and generalization capacity of our method compared to other state-of-the-art methods on a single dataset or the concurrent work on segment anything. We will make our code and training model publicly available.
****
Performance Bounds for Policy-Based Average Reward Reinforcement Learning Algorithms
Yashaswini Murthy, Mehrdad Moharrami, R. Srikant
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Understanding Few-Shot Learning: Measuring Task Relatedness and Adaptation Difficulty via Attributes
Minyang Hu, Hong Chang, Zong Guo, Bingpeng MA, Shiguang Shan, Xilin Chen
Few-shot learning (FSL) aims to learn novel tasks with very few labeled samples by leveraging experience from \emph{related} training tasks. In this paper, we try to understand FSL by exploring two key questions: (1) How to quantify the relationship between \emph{ training} and \emph{novel} tasks? (2) How does the relationship affect the \emph{adaptation difficulty} on novel tasks for different models? To answer the first question, we propose Task Attribute Distance (TAD) as a metric to quantify the task relatedness via attributes. Unlike other metrics, TAD is independent of models, making it applicable to different FSL models. To address the second question, we utilize TAD metric to establish a theoretical connection between task relatedness and task adaptation difficulty. By deriving the generalization error bound on a novel task, we discover how TAD measures the adaptation difficulty on novel tasks for different models.

To validate our theoretical results, we conduct experiments on three benchmarks . Our experimental results confirm that TAD metric effectively quantifies the task relatedness and reflects the adaptation difficulty on novel tasks for various FSL methods, even if some of them do not learn attributes explicitly or human-annotated attributes are not provided. Our code is available at \href{https://github.com/hu-my/TaskAttributeDistance}{https://github.com/hu-my/TaskAttributeDistance}.
****
Locally Invariant Explanations: Towards Stable and Unidirectional Explanations through Local Invariant Learning
Amit Dhurandhar, Karthikeyan Natesan Ramamurthy, Kartik Ahuja, Vijay Arya
Locally interpretable model agnostic explanations (LIME) method is one of the most popular methods used to explain black-box models at a per example level. Although many variants have been proposed, few provide a simple way to produce high fidelity explanations that are also stable and intuitive. In this work, we provide a novel perspective by proposing a model agnostic local explanation method inspired by the invariant risk minimization (IRM) principle -- originally proposed for (global) out-of-distribution generalization -- to provide such high fidelity explanations that are also stable and unidirectional across nearby examples. Our method is based on a game theoretic formulation where we theoretically show that our approach has a strong tendency to eliminate features where the gradient of the black-box function abruptly changes sign in the locality of the example we want to explain, while in other cases it is more careful and will choose a more conservative (feature) attribution, a behavior which can be highly desirable for recourse. Empirically, we show on tabular, image and text data that the quality of our explanations with neighborhoods formed using random perturbations are much better than LIME and in some cases even comparable to other methods that use realistic neighbors sampled from the data manifold. This is desirable given that learning a manifold to either create realistic neighbors or to project explanations is typically expensive or may even be impossible. Moreover, our algorithm is simple and efficient to train, and can ascertain stable input features for local decisions of a black-box without access to side information such as a (partial) causal graph as has been seen in some recent works.
****
Quantification of Uncertainty with Adversarial Models
Kajetan Schweighofer, Lukas Aichberger, Mykyta Ielanskyi, Günter Klambauer, Sepp Hochreiter
Quantifying uncertainty is important for actionable predictions in real-world applications. A crucial part of predictive uncertainty quantification is the estimation of epistemic uncertainty, which is defined as an integral of the product between a divergence function and the posterior. Current methods such as Deep Ensembles or MC dropout underperform at estimating the epistemic uncertainty, since they primarily consider the posterior when sampling models. We suggest Quantification of Uncertainty with Adversarial Models (QUAM) to better estimate the epistemic uncertainty. QUAM identifies regions where the whole product under the integral is large, not just the posterior. Consequently, QUAM has lower approximation error of the epistemic uncertainty compared to previous methods. Models for which the product is large correspond to adversarial models (not adversarial examples!). Adversarial models have both a high posterior as well as a high divergence between their predictions and that of a reference model. Our experiments show that QUAM excels in capturing epistemic uncertainty for deep learning models and outperforms previous methods on challenging tasks in the vision domain.
****
NeuroGF: A Neural Representation for Fast Geodesic Distance and Path Queries
Qijian Zhang, Junhui Hou, Yohanes Adikusuma, Wenping Wang, Ying He
Geodesics play a critical role in many geometry processing applications. Traditional algorithms for computing geodesics on 3D mesh models are often inefficient and slow, which make them impractical for scenarios requiring extensive querying of arbitrary point-to-point geodesics. Recently, deep implicit functions have gained popularity for 3D geometry representation, yet there is still no research

on neural implicit representation of geodesics. To bridge this gap, we make the first attempt to represent geodesics using implicit learning frameworks. Specifically, we propose neural geodesic field (NeuroGF), which can be learned to encode all-pairs geodesics of a given 3D mesh model, enabling to efficiently and accurately answer queries of arbitrary point-to-point geodesic distances and paths. Evaluations on common 3D object models and real-captured scene-level meshes demonstrate our exceptional performances in terms of representation accuracy and querying efficiency. Besides, NeuroGF also provides a convenient way of jointly encoding both 3D geometry and geodesics in a unified representation. Moreover, the working mode of per-model overfitting is further extended to generalizable learning frameworks that can work on various input formats such as unstructured point clouds, which also show satisfactory performances for unseen shapes and categories. Our code and data are available at https://github.com/keeganhk/NeuroGF.
****

A Trichotomy for Transductive Online Learning
Steve Hanneke, Shay Moran, Jonathan Shafer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Evolutionary Neural Architecture Search for Transformer in Knowledge Tracing
Shangshang Yang, Xiaoshan Yu, Ye Tian, Xueming Yan, Haiping Ma, Xingyi Zhang
Knowledge tracing (KT) aims to trace students' knowledge states by predicting whether students answer correctly on exercises. Despite the excellent performance of existing Transformer-based KT approaches, they are criticized for the manually selected input features for fusion and the defect of single global context modelling to directly capture students' forgetting behavior in KT, when the related records are distant from the current record in terms of time. To address the issues, this paper first considers adding convolution operations to the Transformer to enhance its local context modelling ability used for students' forgetting behavior, then proposes an evolutionary neural architecture search approach to automate the input feature selection and automatically determine where to apply which operation for achieving the balancing of the local/global context modelling. In the search space, the original global path containing the attention module in Transformer is replaced with the sum of a global path and a local path that could contain different convolutions, and the selection of input features is also considered. To search the best architecture, we employ an effective evolutionary algorithm to explore the search space and also suggest a search space reduction strategy to accelerate the convergence of the algorithm. Experimental results on the two largest and most challenging education datasets demonstrate the effectiveness of the architecture found by the proposed approach.
****

Learning threshold neurons via edge of stability
Kwangjun Ahn, Sebastien Bubeck, Sinho Chewi, Yin Tat Lee, Felipe Suarez, Yi Zhang
Existing analyses of neural network training often operate under the unrealistic assumption of an extremely small learning rate. This lies in stark contrast to practical wisdom and empirical studies, such as the work of J. Cohen et al. (ICLR 2021), which exhibit startling new phenomena (the "edge of stability"' or "unstable convergence") and potential benefits for generalization in the large learning rate regime. Despite a flurry of recent works on this topic, however, the latter effect is still poorly understood. In this paper, we take a step towards understanding genuinely non-convex training dynamics with large learning rates by performing a detailed analysis of gradient descent for simplified models of two-layer neural networks. For these models, we provably establish the edge of stability phenomenon and discover a sharp phase transition for the step size below which the neural network fails to learn ``threshold-like'' neurons (i.e., neurons with a non-zero first-layer bias). This elucidates one possible mechanism by which the edge of stability can in fact lead to better generalization, as threshold

neurons are basic building blocks with useful inductive bias for many tasks.
****

k-Means Clustering with Distance-Based Privacy
Alessandro Epasto, Vahab Mirrokni, Shyam Narayanan, Peilin Zhong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

StyleTTS 2: Towards Human-Level Text-to-Speech through Style Diffusion and Adversarial Training with Large Speech Language Models
Yinghao Aaron Li, Cong Han, Vinay Raghavan, Gavin Mischler, Nima Mesgarani
In this paper, we present StyleTTS 2, a text-to-speech (TTS) model that leverages style diffusion and adversarial training with large speech language models (SLMs) to achieve human-level TTS synthesis. StyleTTS 2 differs from its predecessor by modeling styles as a latent random variable through diffusion models to generate the most suitable style for the text without requiring reference speech, achieving efficient latent diffusion while benefiting from the diverse speech synthesis offered by diffusion models. Furthermore, we employ large pre-trained SLMs, such as WavLM, as discriminators with our novel differentiable duration modeling for end-to-end training, resulting in improved speech naturalness. StyleTTS 2 surpasses human recordings on the single-speaker LJSpeech dataset and matches it on the multispeaker VCTK dataset as judged by native English speakers. Moreover, when trained on the LibriTTS dataset, our model outperforms previous publicly available models for zero-shot speaker adaptation. This work achieves the first human-level TTS on both single and multispeaker datasets, showcasing the potential of style diffusion and adversarial training with large SLMs. The audio demos and source code are available at https://styletts2.github.io/.
****

Large Language Models Are Zero-Shot Time Series Forecasters
Nate Gruver, Marc Finzi, Shikai Qiu, Andrew G. Wilson
By encoding time series as a string of numerical digits, we can frame time series forecasting as next-token prediction in text. Developing this approach, we find that large language models (LLMs) such as GPT-3 and LLaMA-2 can surprisingly zero-shot extrapolate time series at a level comparable to or exceeding the performance of purpose-built time series models trained on the downstream tasks. To facilitate this performance, we propose procedures for effectively tokenizing time series data and converting discrete distributions over tokens into highly flexible densities over continuous values. We argue the success of LLMs for time series stems from their ability to naturally represent multimodal distributions, in conjunction with biases for simplicity, and repetition, which align with the salient features in many time series, such as repeated seasonal trends. We also show how LLMs can naturally handle missing data without imputation through non-numerical text, accommodate textual side information, and answer questions to help explain predictions. While we find that increasing model size generally improves performance on time series, we show GPT-4 can perform worse than GPT-3 because of how it tokenizes numbers, and poor uncertainty calibration, which is likely the result of alignment interventions such as RLHF.
****

Learning Mixtures of Gaussians Using the DDPM Objective
Kulin Shah, Sitan Chen, Adam Klivans
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Graph Convolutional Kernel Machine versus Graph Convolutional Networks
Zhihao Wu, Zhao Zhang, Jicong Fan
Graph convolutional networks (GCN) with one or two hidden layers have been widely used in handling graph data that are prevalent in various disciplines. Many st

udies showed that the gain of making GCNs deeper is tiny or even negative. This implies that the complexity of graph data is often limited and shallow models are often sufficient to extract expressive features for various tasks such as node classification. Therefore, in this work, we present a framework called graph convolutional kernel machine (GCKM) for graph-based machine learning. GCKMs are built upon kernel functions integrated with graph convolution. An example is the graph convolutional kernel support vector machine (GCKSVM) for node classification, for which we analyze the generalization error bound and discuss the impact of the graph structure. Compared to GCNs, GCKMs require much less effort in architecture design, hyperparameter tuning, and optimization. More importantly, GCKMs are guaranteed to obtain globally optimal solutions and have strong generalization ability and high interpretability. GCKMs are composable, can be extended to large-scale data, and are applicable to various tasks (e.g., node or graph classification, clustering, feature extraction, dimensionality reduction). The numerical results on benchmark datasets show that, besides the aforementioned advantages, GCKMs have at least competitive accuracy compared to GCNs.
****
First Order Stochastic Optimization with Oblivious Noise
Ilias Diakonikolas, Sushrut Karmalkar, Jong Ho Park, Christos Tzamos
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CHAMMI: A benchmark for channel-adaptive models in microscopy imaging
Zitong Sam Chen, Chau Pham, Siqi Wang, Michael Doron, Nikita Moshkov, Bryan Plummer, Juan C. Caicedo
Most neural networks assume that input images have a fixed number of channels (three for RGB images). However, there are many settings where the number of channels may vary, such as microscopy images where the number of channels changes depending on instruments and experimental goals. Yet, there has not been a systemic attempt to create and evaluate neural networks that are invariant to the number and type of channels. As a result, trained models remain specific to individual studies and are hardly reusable for other microscopy settings. In this paper, we present a benchmark for investigating channel-adaptive models in microscopy imaging, which consists of 1) a dataset of varied-channel single-cell images, and 2) a biologically relevant evaluation framework. In addition, we adapted several existing techniques to create channel-adaptive models and compared their performance on this benchmark to fixed-channel, baseline models. We find that channel-adaptive models can generalize better to out-of-domain tasks and can be computationally efficient. We contribute a curated dataset and an evaluation API to facilitate objective comparisons in future research and applications.
****
A Theory of Link Prediction via Relational Weisfeiler-Leman on Knowledge Graphs
Xingyue Huang, Miguel Romero, Ismail Ceylan, Pablo Barceló
Graph neural networks are prominent models for representation learning over graph-structured data. While the capabilities and limitations of these models are well-understood for simple graphs, our understanding remains incomplete in the context of knowledge graphs. Our goal is to provide a systematic understanding of the landscape of graph neural networks for knowledge graphs pertaining to the prominent task of link prediction. Our analysis entails a unifying perspective on seemingly unrelated models and unlocks a series of other models. The expressive power of various models is characterized via a corresponding relational Weisfeiler-Leman algorithm. This analysis is extended to provide a precise logical characterization of the class of functions captured by a class of graph neural networks. The theoretical findings presented in this paper explain the benefits of some widely employed practical design choices, which are validated empirically.
****
Bayes beats Cross Validation: Efficient and Accurate Ridge Regression via Expectation Maximization

Shu Yu Tew, Mario Boley, Daniel Schmidt
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Segment Everything Everywhere All at Once
Xueyan Zou, Jianwei Yang, Hao Zhang, Feng Li, Linjie Li, Jianfeng Wang, Lijuan Wang, Jianfeng Gao, Yong Jae Lee
In this work, we present SEEM, a promotable and interactive model for segmenting everything everywhere all at once in an image. In SEEM, we propose a novel and versatile decoding mechanism that enables diverse prompting for all types of segmentation tasks, aiming at a universal interface that behaves like large language models (LLMs). More specifically, SEEM is designed with four desiderata:i) Versatility. We introduce a new visual prompt to unify different spatial queries including points, boxes, scribbles, and masks, which can further generalize to a different referring image; ii) Compositionality. We learn a joint visual-semantic space between text and visual prompts, which facilitates the dynamic composition of two prompt types required for various segmentation tasks, as shown in Fig. 1;iii) Interactivity. We further incorporate learnable memory prompts into the decoder to retain segmentation history through mask-guided cross-attention from the decoder to image features; iv) Semantic awareness. We use a text encoder to encode text queries and mask labels into the same semantic space for open-vocabulary segmentation. We conduct a comprehensive empirical study to validate the effectiveness of SEEM across diverse segmentation tasks. The results demonstrate that SEEM exhibits robust generalizing to unseen user intents as it learns to compose prompts of different types in a unified representation space. Our approach achieves competitive performance on interactive segmentation, generic segmentation, referring segmentation, and video object segmentation on 9 datasets with minimum 1/100 supervision in a single set of weights.
****
PUe: Biased Positive-Unlabeled Learning Enhancement by Causal Inference
Xutao Wang, Hanting Chen, Tianyu Guo, Yunhe Wang
Positive-Unlabeled (PU) learning aims to achieve high-accuracy binary classification with limited labeled positive examples and numerous unlabeled ones. Existing cost-sensitive-based methods often rely on strong assumptions that examples with an observed positive label were selected entirely at random. In fact, the uneven distribution of labels is prevalent in real-world PU problems, indicating that most actual positive and unlabeled data are subject to selection bias. In this paper, we propose a PU learning enhancement (PUe) algorithm based on causal inference theory, which employs normalized propensity scores and normalized inverse probability weighting (NIPW) techniques to reconstruct the loss function, thus obtaining a consistent, unbiased estimate of the classifier and enhancing the model's performance. Moreover, we investigate and propose a method for estimating propensity scores in deep learning using regularization techniques when the labeling mechanism is unknown. Our experiments on three benchmark datasets demonstrate the proposed PUe algorithm significantly improves the accuracy of classifiers on non-uniform label distribution datasets compared to advanced cost-sensitive PU methods. Codes are available at https://github.com/huawei-noah/Noah-research/tree/master/PUe and https://gitee.com/mindspore/models/tree/master/research/cv/PUe.
****
Sparse Modular Activation for Efficient Sequence Modeling
Liliang Ren, Yang Liu, Shuohang Wang, Yichong Xu, Chenguang Zhu, Cheng Xiang Zhai
Recent hybrid models combining Linear State Space Models (SSMs) with self-attention mechanisms have demonstrated impressive results across a range of sequence modeling tasks. However, current approaches apply attention modules statically and uniformly to all elements in the input sequences, leading to sub-optimal quality-efficiency trade-offs. To address this limitation, we introduce Sparse Modula

r Activation (SMA), a general mechanism enabling neural networks to sparsely and dynamically activate sub-modules for sequence elements in a differentiable manner. Through allowing each element to skip non-activated sub-modules, SMA reduces computation and memory consumption of neural networks at both training and inference stages. To validate the effectiveness of SMA on sequence modeling, we design a novel neural architecture, SeqBoat, which employs SMA to sparsely activate a Gated Attention Unit (GAU) based on the state representations learned from an SSM. By constraining the GAU to only conduct local attention on the activated inputs, SeqBoat can achieve linear inference complexity with theoretically infinite attention span, and provide substantially better quality-efficiency trade-off than the chunking-based models. With experiments on a wide range of tasks, including long sequence modeling, speech classification and language modeling, SeqBoat brings new state-of-the-art results among hybrid models with linear complexity, and reveals the amount of attention needed for each task through the learned sparse activation patterns. Our code is publicly available at https://github.com/renll/SeqBoat.
****
BuildingsBench: A Large-Scale Dataset of 900K Buildings and Benchmark for Short-Term Load Forecasting
Patrick Emami, Abhijeet Sahu, Peter Graf
Short-term forecasting of residential and commercial building energy consumption is widely used in power systems and continues to grow in importance. Data-driven short-term load forecasting (STLF), although promising, has suffered from a lack of open, large-scale datasets with high building diversity. This has hindered exploring the pretrain-then-fine-tune paradigm for STLF. To help address this, we present BuildingsBench, which consists of: 1) Buildings-900K, a large-scale dataset of 900K simulated buildings representing the U.S. building stock; and 2) an evaluation platform with over 1,900 real residential and commercial buildings from 7 open datasets. BuildingsBench benchmarks two under-explored tasks: zero-shot STLF, where a pretrained model is evaluated on unseen buildings without fine-tuning, and transfer learning, where a pretrained model is fine-tuned on a target building. The main finding of our benchmark analysis is that synthetically pretrained models generalize surprisingly well to real commercial buildings. An exploration of the effect of increasing dataset size and diversity on zero-shot commercial building performance reveals a power-law with diminishing returns. We also show that fine-tuning pretrained models on real commercial and residential buildings improves performance for a majority of target buildings. We hope that BuildingsBench encourages and facilitates future research on generalizable STLF. All datasets and code can be accessed from https://github.com/NREL/BuildingsBench.
****
Efficient Bayesian Learning Curve Extrapolation using Prior-Data Fitted Networks
Steven Adriaensen, Herilalaina Rakotoarison, Samuel Müller, Frank Hutter
Learning curve extrapolation aims to predict model performance in later epochs of training, based on the performance in earlier epochs.In this work, we argue that, while the inherent uncertainty in the extrapolation of learning curves warrants a Bayesian approach, existing methods are (i) overly restrictive, and/or (ii) computationally expensive. We describe the first application of prior-data fitted neural networks (PFNs) in this context. A PFN is a transformer, pre-trained on data generated from a prior, to perform approximate Bayesian inference in a single forward pass. We propose LC-PFN, a PFN trained to extrapolate 10 million artificial right-censored learning curves generated from a parametric prior proposed in prior art using MCMC. We demonstrate that LC-PFN can approximate the posterior predictive distribution more accurately than MCMC, while being over 10 000 times faster. We also show that the same LC-PFN achieves competitive performance extrapolating a total of 20 000 real learning curves from four learning curve benchmarks (LCBench, NAS-Bench-201, Taskset, and PD1) that stem from training a wide range of model architectures (MLPs, CNNs, RNNs, and Transformers) on 53 different datasets with varying input modalities (tabular, image, text, and protein data). Finally, we investigate its potential in the context of model selection

and find that a simple LC-PFN based predictive early stopping criterion obtains
2 - 6x speed-ups on 45 of these datasets, at virtually no overhead.
****
Unified Off-Policy Learning to Rank: a Reinforcement Learning Perspective
Zeyu Zhang, Yi Su, Hui Yuan, Yiran Wu, Rishab Balasubramanian, Qingyun Wu, Huazh
eng Wang, Mengdi Wang
Off-policy Learning to Rank (LTR) aims to optimize a ranker from data collected
by a deployed logging policy. However, existing off-policy learning to rank meth
ods often make strong assumptions about how users generate the click data, i.e.,
 the click model, and hence need to tailor their methods specifically under diff
erent click models. In this paper, we unified the ranking process under general
stochastic click models as a Markov Decision Process (MDP), and the optimal rank
ing could be learned with offline reinforcement learning (RL) directly. Building
 upon this, we leverage offline RL techniques for off-policy LTR and propose the
 Click Model-Agnostic Unified Off-policy Learning to Rank (CUOLR) method, which
could be easily applied to a wide range of click models. Through a dedicated for
mulation of the MDP, we show that offline RL algorithms can adapt to various cli
ck models without complex debiasing techniques and prior knowledge of the model.
 Results on various large-scale datasets demonstrate that CUOLR consistently out
performs the state-of-the-art off-policy learning to rank algorithms while maint
aining consistency and robustness under different click models.
****
Trust Region-Based Safe Distributional Reinforcement Learning for Multiple Const
raints
Dohyeong Kim, Kyungjae Lee, Songhwai Oh
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
The Contextual Lasso: Sparse Linear Models via Deep Neural Networks
Ryan Thompson, Amir Dezfouli, robert kohn
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
No Representation Rules Them All in Category Discovery
Sagar Vaze, Andrea Vedaldi, Andrew Zisserman
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
CS4ML: A general framework for active learning with arbitrary data based on Chri
stoffel functions
Juan M. Cardenas, Ben Adcock, Nick Dexter
We introduce a general framework for active learning in regression problems. Our
 framework extends the standard setup by allowing for general types of data, rat
her than merely pointwise samples of the target function. This generalization co
vers many cases of practical interest, such as data acquired in transform domain
s (e.g., Fourier data), vector-valued data (e.g., gradient-augmented data), data
 acquired along continuous curves, and, multimodal data (i.e., combinations of d
ifferent types of measurements). Our framework considers random sampling accord
ing to a finite number of sampling measures and arbitrary nonlinear approximatio
n spaces (model classes). We introduce the concept of \textit{generalized Christ
offel functions} and show how these can be used to optimize the sampling measure
s. We prove that this leads to near-optimal sample complexity in various importa
nt cases. This paper focuses on applications in scientific computing, where acti
ve learning is often desirable, since it is usually expensive to generate data.

We demonstrate the efficacy of our framework for gradient-augmented learning wit h polynomials, Magnetic Resonance Imaging (MRI) using generative models and adap tive sampling for solving PDEs using Physics-Informed Neural Networks (PINNs).
****
Two Heads are Better Than One: A Simple Exploration Framework for Efficient Mult i-Agent Reinforcement Learning
Jiahui Li, Kun Kuang, Baoxiang Wang, Xingchen Li, Fei Wu, Jun Xiao, Long Chen
Exploration strategy plays an important role in reinforcement learning, especial ly in sparse-reward tasks. In cooperative multi-agent reinforcement learning~(MA RL), designing a suitable exploration strategy is much more challenging due to t he large state space and the complex interaction among agents. Currently, mainst ream exploration methods in MARL either contribute to exploring the unfamiliar s tates which are large and sparse, or measuring the interaction among agents with high computational costs. We found an interesting phenomenon that different kin ds of exploration plays a different role in different MARL scenarios, and choosi ng a suitable one is often more effective than designing an exquisite algorithm. In this paper, we propose a exploration method that incorporate the \underline{ C}uri\underline{O}sity-based and \underline{IN}fluence-based exploration~(COIN) which is simple but effective in various situations. First, COIN measures the in fluence of each agent on the other agents based on mutual information theory and designs it as intrinsic rewards which are applied to each individual value func tion. Moreover, COIN computes the curiosity-based intrinsic rewards via predicti on errors which are added to the extrinsic reward. For integrating the two kinds of intrinsic rewards, COIN utilizes a novel framework in which they complement each other and lead to a sufficient and effective exploration on cooperative MAR L tasks. We perform extensive experiments on different challenging benchmarks, a nd results across different scenarios show the superiority of our method.
****
Cross-Scale MAE: A Tale of Multiscale Exploitation in Remote Sensing
Maofeng Tang, Andrei Cozma, Konstantinos Georgiou, Hairong Qi
Remote sensing images present unique challenges to image analysis due to the ext ensive geographic coverage, hardware limitations, and misaligned multi-scale ima ges. This paper revisits the classical multi-scale representation learning prob- lem but under the general framework of self-supervised learning for remote sens ing image understanding. We present Cross-Scale MAE, a self-supervised model bui lt upon the Masked Auto-Encoder (MAE). During pre-training, Cross-Scale MAE empl oys scale augmentation techniques and enforces cross-scale consistency constrain ts through both contrastive and generative losses to ensure consistent and meani ngful representations well-suited for a wide range of downstream tasks. Further, our implementation leverages the xFormers library to accelerate network pre-tra ining on a single GPU while maintaining the quality of learned represen- tations . Experimental evaluations demonstrate that Cross-Scale MAE exhibits superior pe rformance compared to standard MAE and other state-of-the-art remote sensing MAE methods.
****
MotionGPT: Human Motion as a Foreign Language
Biao Jiang, Xin Chen, Wen Liu, Jingyi Yu, Gang Yu, Tao Chen
Though the advancement of pre-trained large language models unfolds, the explora tion of building a unified model for language and other multimodal data, such as motion, remains challenging and untouched so far. Fortunately, human motion dis plays a semantic coupling akin to human language, often perceived as a form of b ody language. By fusing language data with large-scale motion models, motion-lan guage pre-training that can enhance the performance of motion-related tasks beco mes feasible. Driven by this insight, we propose MotionGPT, a unified, versatile , and user-friendly motion-language model to handle multiple motion-relevant tas ks. Specifically, we employ the discrete vector quantization for human motion an d transfer 3D motion into motion tokens, similar to the generation process of wo rd tokens. Building upon this "motion vocabulary", we perform language modeling on both motion and text in a unified manner, treating human motion as a specific language. Moreover, inspired by prompt learning, we pre-train MotionGPT with a

mixture of motion-language data and fine-tune it on prompt-based question-and-an swer tasks. Extensive experiments demonstrate that MotionGPT achieves state-of-t he-art performances on multiple motion tasks including text-driven motion genera tion, motion captioning, motion prediction, and motion in-between.
****

Model-Free Reinforcement Learning with the Decision-Estimation Coefficient
Dylan J Foster, Noah Golowich, Jian Qian, Alexander Rakhlin, Ayush Sekhari
We consider the problem of interactive decision making, encompassing structured bandits and reinforcementlearning with general function approximation. Recently, Foster et al. (2021) introduced theDecision-Estimation Coefficient, a measure o f statistical complexity that lower bounds the optimal regret for interactive de cisionmaking, as well as a meta-algorithm, Estimation-to-Decisions, which achiev es upperbounds in terms of the same quantity. Estimation-to-Decisions is a reduc tion, which liftsalgorithms for (supervised) online estimation into algorithms f ordecision making. In this paper, we show that by combining Estimation-to-Decisi ons witha specialized form of "optimistic" estimation introduced byZhang (2022), it is possible to obtain guaranteesthat improve upon those of Foster et al. (20 21) byaccommodating more lenient notions of estimation error. We use this approa ch to derive regret bounds formodel-free reinforcement learning with value funct ion approximation, and give structural results showing when it can and cannot he lp more generally.
****

FlowPG: Action-constrained Policy Gradient with Normalizing Flows
Janaka Brahmanage, Jiajing LING, Akshat Kumar
Action-constrained reinforcement learning (ACRL) is a popular approach for solvi ng safety-critical and resource-allocation related decision making problems. A m ajor challenge in ACRL is to ensure agent taking a valid action satisfying const raints in each RL step. Commonly used approach of using a projection layer on to p of the policy network requires solving an optimization program which can resul t in longer training time, slow convergence, and zero gradient problem. To addre ss this, first we use a normalizing flow model to learn an invertible, different iable mapping between the feasible action space and the support of a simple dist ribution on a latent variable, such as Gaussian. Second, learning the flow model requires sampling from the feasible action space, which is also challenging. We develop multiple methods, based on Hamiltonian Monte-Carlo and probabilistic se ntential decision diagrams for such action sampling for convex and non-convex co nstraints. Third, we integrate the learned normalizing flow with the DDPG algori thm. By design, a well-trained normalizing flow will transform policy output int o a valid action without requiring an optimization solver. Empirically, our appr oach results in significantly fewer constraint violations (upto an order-of-magn itude for several instances) and is multiple times faster on a variety of contin uous control tasks.
****

Distributionally Robust Bayesian Optimization with φ-divergences
Hisham Husain, Vu Nguyen, Anton van den Hengel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Connected Superlevel Set in (Deep) Reinforcement Learning and its Application to Minimax Theorems
Sihan Zeng, Thinh Doan, Justin Romberg
The aim of this paper is to improve the understanding of the optimization landsc ape for policy optimization problems in reinforcement learning. Specifically, we show that the superlevel set of the objective function with respect to the poli cy parameter is always a connected set both in the tabular setting and under pol icies represented by a class of neural networks. In addition, we show that the o ptimization objective as a function of the policy parameter and reward satisfies a stronger “equiconnectedness” property. To our best knowledge, these are novel

and previously unknown discoveries.We present an application of the connectedness of these superlevel sets to the derivation of minimax theorems for robust reinforcement learning. We show that any minimax optimization program which is convex on one side and is equiconnected on the other side observes the minimax equality (i.e. has a Nash equilibrium). We find that this exact structure is exhibited by an interesting class of robust reinforcement learning problems under an adversarial reward attack, and the validity of its minimax equality immediately follows. This is the first time such a result is established in the literature.
****
Towards Efficient and Accurate Winograd Convolution via Full Quantization
Tianqi Chen, Weixiang Xu, Weihan Chen, Peisong Wang, Jian Cheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Quantum Bayesian Optimization
Zhongxiang Dai, Gregory Kang Ruey Lau, Arun Verma, YAO SHU, Bryan Kian Hsiang Low, Patrick Jaillet
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Interpretable Reward Redistribution in Reinforcement Learning: A Causal Approach
Yudi Zhang, Yali Du, Biwei Huang, Ziyan Wang, Jun Wang, Meng Fang, Mykola Pechenizkiy
A major challenge in reinforcement learning is to determine which state-action pairs are responsible for future rewards that are delayed. Reward redistribution serves as a solution to re-assign credits for each time step from observed sequences. While the majority of current approaches construct the reward redistribution in an uninterpretable manner, we propose to explicitly model the contributions of state and action from a causal perspective, resulting in an interpretable reward redistribution and preserving policy invariance. In this paper, we start by studying the role of causal generative models in reward redistribution by characterizing the generation of Markovian rewards and trajectory-wise long-term return and further propose a framework, called Generative Return Decomposition (GRD), for policy optimization in delayed reward scenarios. Specifically, GRD first identifies the unobservable Markovian rewards and causal relations in the generative process. Then, GRD makes use of the identified causal generative model to form a compact representation to train policy over the most favorable subspace of the state space of the agent. Theoretically, we show that the unobservable Markovian reward function is identifiable, as well as the underlying causal structure and causal models. Experimental results show that our method outperforms state-of-the-art methods and the provided visualization further demonstrates the interpretability of our method.The project page is located at https://reedzyd.github.io/GenerativeReturnDecomposition/.
****
Guarantees for Self-Play in Multiplayer Games via Polymatrix Decomposability
Revan MacQueen, James Wright
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
VCC: Scaling Transformers to 128K Tokens or More by Prioritizing Important Tokens
Zhanpeng Zeng, Cole Hawkins, Mingyi Hong, Aston Zhang, Nikolaos Pappas, Vikas Singh, Shuai Zheng
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Greatness in Simplicity: Unified Self-Cycle Consistency for Parser-Free Virtual Try-On
Chenghu Du, junyin Wang, Shuqing Liu, Shengwu Xiong
Image-based virtual try-on tasks remain challenging, primarily due to inherent complexities associated with non-rigid garment deformation modeling and strong feature entanglement of clothing within human body. Recent groundbreaking formulations, such as in-painting, cycle consistency, and knowledge distillation, have facilitated self-supervised generation of try-on images. However, these paradigms necessitate the disentanglement of garment features within human body features through auxiliary tasks, such as leveraging 'teacher knowledge' and dual generators. The potential presence of irresponsible prior knowledge in the auxiliary task can serve as a significant bottleneck for the main generator (e.g., 'student model') in the downstream task. Moreover, existing garment deformation methods lack the ability to perceive the correlation between the garment and the human body in the real world, leading to unrealistic alignment effects. To tackle these limitations, we present a new parser-free virtual try-on network based on unified self-cycle consistency (USC-PFN), which enables robust translation between different garments using just a single generator, faithfully replicating non-rigid geometric deformation of garments in real-life scenarios. Specifically, we first propose a self-cycle consistency architecture with a circular mode. It utilizes real unpaired garment-person images exclusively as input for training, effectively eliminating the impact of irresponsible prior knowledge at the model input end. Additionally, we formulate a Markov Random Field to simulate a more natural and realistic garment deformation. Furthermore, USC-PFN can leverage a general generator for self-supervised cycle training. Experiments demonstrate that our method achieves state-of-the-art performance on a popular virtual try-on benchmark.
****
VPGTrans: Transfer Visual Prompt Generator across LLMs
Ao Zhang, Hao Fei, Yuan Yao, Wei Ji, Li Li, Zhiyuan Liu, Tat-Seng Chua
Since developing a new multimodal LLM (MLLM) by pre-training on tremendous image-text pairs from scratch can be exceedingly resource-consuming, connecting an existing LLM with a comparatively lightweight visual prompt generator (VPG) becomes a feasible paradigm. However, further tuning the VPG component of the MLLM still incurs significant computational costs, such as thousands of GPU hours and millions of training data points. An alternative solution is transferring an existing VPG from one MLLM to the target MLLM. In this work, we investigate VPG transferability across LLMs for the first time, aiming to reduce the cost of VPG training. Specifically, we explore VPG transfer across different LLM sizes (e.g., small-to-large) and types. We identify key factors to maximize transfer efficiency, based on which we develop a simple yet highly effective two-stage transfer framework, called VPGTrans. Notably, it enables VPG transfer from BLIP-2 OPT 2.7B to BLIP-2 OPT 6.7B with less than 10% of the GPU hours using only 10.7% of the training data compared to training a VPG for OPT 6.7B from scratch. Furthermore, we provide a series of intriguing findings and discuss potential explanations behind them. Finally, we showcase the practical value of our VPGTrans approach, by customizing two novel MLLMs, including VL-LLaMA and VL-Vicuna, with recently released LLaMA and Vicuna LLMs.
****
Nearest Neighbour with Bandit Feedback
Stephen Pasteris, Chris Hicks, Vasilios Mavroudis
In this paper we adapt the nearest neighbour rule to the contextual bandit problem. Our algorithm handles the fully adversarial setting in which no assumptions at all are made about the data-generation process. When combined with a sufficiently fast data-structure for (perhaps approximate) adaptive nearest neighbour search, such as a navigating net, our algorithm is extremely efficient - having a

per trial running time polylogarithmic in both the number of trials and actions, and taking only quasi-linear space. We give generic regret bounds for our algorithm and further analyse them when applied to the stochastic bandit problem in euclidean space. A side result of this paper is that, when applied to the online classification problem with stochastic labels, our algorithm can, under certain conditions, have sublinear regret whilst only finding a single nearest neighbour per trial - in stark contrast to the k-nearest neighbours algorithm.
****
Generative Neural Fields by Mixtures of Neural Implicit Functions
Tackgeun You, Mijeong Kim, Jungtaek Kim, Bohyung Han
We propose a novel approach to learning the generative neural fields represented by linear combinations of implicit basis networks. Our algorithm learns basis networks in the form of implicit neural representations and their coefficients in a latent space by either conducting meta-learning or adopting auto-decoding paradigms. The proposed method easily enlarges the capacity of generative neural fields by increasing the number of basis networks while maintaining the size of a network for inference to be small through their weighted model averaging. Consequently, sampling instances using the model is efficient in terms of latency and memory footprint. Moreover, we customize denoising diffusion probabilistic model for a target task to sample latent mixture coefficients, which allows our final model to generate unseen data effectively. Experiments show that our approach achieves competitive generation performance on diverse benchmarks for images, voxel data, and NeRF scenes without sophisticated designs for specific modalities and domains.
****
MAViL: Masked Audio-Video Learners
Po-Yao Huang, Vasu Sharma, Hu Xu, Chaitanya Ryali, haoqi fan, Yanghao Li, Shang-Wen Li, Gargi Ghosh, Jitendra Malik, Christoph Feichtenhofer
We present Masked Audio-Video Learners (MAViL) to learn audio-visual representations with three complementary forms of self-supervision: (1) reconstructing masked raw audio and video inputs, (2) intra-modal and inter-modal contrastive learning with masking, and (3) self-training to predict aligned and contextualized audio-video representations learned from the first two objectives. Empirically, MAViL achieves state-of-the-art audio-video classification performance on AudioSet (53.3 mAP) and VGGSound (67.1\% accuracy), surpassing recent self-supervised models and supervised models that utilize external labeled data. Notably, pre-training with MAViL not only enhances performance in multimodal classification and retrieval tasks, but it also improves the representations of each modality in isolation, without relying on information from the other modality during uni-modal fine-tuning or inference. The code and models are available at https://github.com/facebookresearch/MAViL.
****
Combating Representation Learning Disparity with Geometric Harmonization
Zhihan Zhou, Jiangchao Yao, Feng Hong, Ya Zhang, Bo Han, Yanfeng Wang
Self-supervised learning (SSL) as an effective paradigm of representation learning has achieved tremendous success on various curated datasets in diverse scenarios. Nevertheless, when facing the long-tailed distribution in real-world applications, it is still hard for existing methods to capture transferable and robust representation. The attribution is that the vanilla SSL methods that pursue the sample-level uniformity easily leads to representation learning disparity, where head classes with the huge sample number dominate the feature regime but tail classes with the small sample number passively collapse. To address this problem, we propose a novel Geometric Harmonization (GH) method to encourage the category-level uniformity in representation learning, which is more benign to the minority and almost does not hurt the majority under long-tailed distribution. Specially, GH measures the population statistics of the embedding space on top of self-supervised learning, and then infer an fine-grained instance-wise calibration to constrain the space expansion of head classes and avoid the passive collapse of tail classes. Our proposal does not alter the setting of SSL and can be easily integrated into existing methods in a low-cost manner. Extensive results on a

range of benchmark datasets show the effectiveness of \methodspace with high tolerance to the distribution skewness.
****
BioMassters: A Benchmark Dataset for Forest Biomass Estimation using Multi-modal Satellite Time-series
Andrea Nascetti, Ritu Yadav, Kirill Brodt, Qixun Qu, Hongwei Fan, Yuri Shendryk, Isha Shah, Christine Chung
Above Ground Biomass is an important variable as forests play a crucial role in mitigating climate change as they act as an efficient, natural and cost-effective carbon sink. Traditional field and airborne LiDAR measurements have been proven to provide reliable estimations of forest biomass. Nevertheless, the use of these techniques at a large scale can be challenging and expensive. Satellite data have been widely used as a valuable tool in estimating biomass on a global scale. However, the full potential of dense multi-modal satellite time series data, in combination with modern deep learning approaches, has yet to be fully explored. The aim of the "BioMassters" data challenge and benchmark dataset is to investigate the potential of multi-modal satellite data (Sentinel-1 SAR and Sentinel-2 MSI) to estimate forest biomass at a large scale using the Finnish Forest Centre's open forest and nature airborne LiDAR data as a reference. The performance of the top three baseline models shows the potential of deep learning to produce accurate and higher-resolution biomass maps. Our benchmark dataset is publically available at https://huggingface.co/datasets/nascetti-a/BioMassters (doi:10.57967/hf/1009) and the implementation of the top three winning models are available at https://github.com/drivendataorg/the-biomassters.
****
Online Inventory Problems: Beyond the i.i.d. Setting with Online Convex Optimization
Massil HIHAT, Stéphane Gaïffas, Guillaume Garrigos, Simon Bussy
We study multi-product inventory control problems where a manager makes sequential replenishment decisions based on partial historical information in order to minimize its cumulative losses. Our motivation is to consider general demands, losses and dynamics to go beyond standard models which usually rely on newsvendor-type losses, fixed dynamics, and unrealistic i.i.d. demand assumptions. We propose MaxCOSD, an online algorithm that has provable guarantees even for problems with non-i.i.d. demands and stateful dynamics, including for instance perishability. We consider what we call non-degeneracy assumptions on the demand process, and argue that they are necessary to allow learning.
****
On kernel-based statistical learning theory in the mean field limit
Christian Fiedler, Michael Herty, Sebastian Trimpe
In many applications of machine learning, a large number of variables are considered. Motivated by machine learning of interacting particle systems, we consider the situation when the number of input variables goes to infinity. First, we continue the recent investigation of the mean field limit of kernels and their reproducing kernel Hilbert spaces, completing the existing theory. Next, we provide results relevant for approximation with such kernels in the mean field limit, including a representer theorem. Finally, we use these kernels in the context of statistical learning in the mean field limit, focusing on Support Vector Machines. In particular, we show mean field convergence of empirical and infinite-sample solutions as well as the convergence of the corresponding risks. On the one hand, our results establish rigorous mean field limits in the context of kernel methods, providing new theoretical tools and insights for large-scale problems. On the other hand, our setting corresponds to a new form of limit of learning problems, which seems to have not been investigated yet in the statistical learning theory literature.
****
Benchmarking Encoder-Decoder Architectures for Biplanar X-ray to 3D Bone Shape Reconstruction
Mahesh Shakya, Bishesh Khanal
Various deep learning models have been proposed for 3D bone shape reconstruction

from two orthogonal (biplanar) X-ray images.However, it is unclear how these models compare against each other since they are evaluated on different anatomy, cohort and (often privately held) datasets.Moreover, the impact of the commonly optimized image-based segmentation metrics such as dice score on the estimation of clinical parameters relevant in 2D-3D bone shape reconstruction is not well known.To move closer toward clinical translation, we propose a benchmarking framework that evaluates tasks relevant to real-world clinical scenarios, including reconstruction of fractured bones, bones with implants, robustness to population shift, and error in estimating clinical parameters.Our open-source platform provides reference implementations of 8 models (many of whose implementations were not publicly available), APIs to easily collect and preprocess 6 public datasets, and the implementation of automatic clinical parameter and landmark extraction methods. We present an extensive evaluation of 8 2D-3D models on equal footing using 6 public datasets comprising images for four different anatomies.Our results show that attention-based methods that capture global spatial relationships tend to perform better across all anatomies and datasets; performance on clinically relevant subgroups may be overestimated without disaggregated reporting; ribs are substantially more difficult to reconstruct compared to femur, hip and spine; and the dice score improvement does not always bring corresponding improvement in the automatic estimation of clinically relevant parameters.
****
3D-LLM: Injecting the 3D World into Large Language Models
Yining Hong, Haoyu Zhen, Peihao Chen, Shuhong Zheng, Yilun Du, Zhenfang Chen, Chuang Gan
Large language models (LLMs) and Vision-Language Models (VLMs) have been proved to excel at multiple tasks, such as commonsense reasoning. Powerful as these models can be, they are not grounded in the 3D physical world, which involves richer concepts such as spatial relationships, affordances, physics, layout, and so on. In this work, we propose to inject the 3D world into large language models, and introduce a whole new family of 3D-LLMs. Specifically, 3D-LLMs can take 3D point clouds and their features as input and perform a diverse set of 3D-related tasks, including captioning, dense captioning, 3D question answering, task decomposition, 3Dgrounding, 3D-assisted dialog, navigation, and so on. Using three types of prompting mechanisms that we design, we are able to collect over 300k 3D-language data covering these tasks. To efficiently train 3D-LLMs, we first utilize a 3D feature extractor that obtains 3D features from rendered multi-view images. Then, we use 2D VLMs as our backbones to train our 3D-LLMs. By introducing a 3D localization mechanism, 3D-LLMs could better capture 3D spatial information. Experiments on ScanQA show that our model outperforms state-of-the-art baselines by a large margin (\textit{e.g.}, the BLEU-1 score surpasses state-of-the-art score by 9\%). Furthermore, experiments on our held-in datasets for 3D captioning, task composition, and 3D-assisted dialogue show that our model outperforms 2D VLMs. Qualitative examples also show that our model could perform more tasks beyond the scope of existing LLMs and VLMs. Our model and data will be publicly available.
****
An Optimal and Scalable Matrix Mechanism for Noisy Marginals under Convex Loss Functions
Yingtai Xiao, Guanlin He, Danfeng Zhang, Daniel Kifer
Noisy marginals are a common form of confidentiality-protecting data release and are useful for many downstream tasks such as contingency table analysis, construction of Bayesian networks, and even synthetic data generation. Privacy mechanisms that provide unbiased noisy answers to linear queries (such as marginals) are known as matrix mechanisms.We propose ResidualPlanner, a matrix mechanism for marginals with Gaussian noise that is both optimal and scalable. ResidualPlanner can optimize for many loss functions that can be written as a convex function of marginal variances (prior work was restricted to just one predefined objective function). ResidualPlanner can optimize the accuracy of marginals in large scale settings in seconds, even when the previous state of the art (HDMM) runs out of memory. It even runs on datasets with 100 attributes in a couple of minutes. F

urthermore ResidualPlanner can efficiently compute variance/covariance values for each marginal (prior methods quickly run out of memory, even for relatively small datasets).
****
Recovering Unbalanced Communities in the Stochastic Block Model with Application to Clustering with a Faulty Oracle
Chandra Sekhar Mukherjee, Pan Peng, Jiapeng Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Transition-constant Normalization for Image Enhancement
Jie Huang, man zhou, Jinghao Zhang, Gang Yang, Mingde Yao, Chongyi Li, Zhiwei Xiong, Feng Zhao
Normalization techniques that capture image style by statistical representation have become a popular component in deep neural networks.Although image enhancement can be considered as a form of style transformation, there has been little exploration of how normalization affect the enhancement performance. To fully leverage the potential of normalization, we present a novel Transition-Constant Normalization (TCN) for various image enhancement tasks.Specifically, it consists of two streams of normalization operations arranged under an invertible constraint, along with a feature sub-sampling operation that satisfies the normalization constraint.TCN enjoys several merits, including being parameter-free, plug-and-play, and incurring no additional computational costs.We provide various formats to utilize TCN for image enhancement, including seamless integration with enhancement networks, incorporation into encoder-decoder architectures for downsampling, and implementation of efficient architectures.Through extensive experiments on multiple image enhancement tasks, like low-light enhancement, exposure correction, SDR2HDR translation, and image dehazing, our TCN consistently demonstrates performance improvements.Besides, it showcases extensive ability in other tasks including pan-sharpening and medical segmentation.The code is available at \textit{\textcolor{blue}{https://github.com/huangkevinj/TCNorm}}.
****
Unexpected Improvements to Expected Improvement for Bayesian Optimization
Sebastian Ament, Samuel Daulton, David Eriksson, Maximilian Balandat, Eytan Bakshy
Expected Improvement (EI) is arguably the most popular acquisition function in Bayesian optimization and has found countless successful applications, but its performance is often exceeded by that of more recent methods. Notably, EI and its variants, including for the parallel and multi-objective settings, are challenging to optimize because their acquisition values vanish numerically in many regions. This difficulty generally increases as the number of observations, dimensionality of the search space, or the number of constraints grow, resulting in performance that is inconsistent across the literature and most often sub-optimal. Herein, we propose LogEI, a new family of acquisition functions whose members either have identical or approximately equal optima as their canonical counterparts, but are substantially easier to optimize numerically. We demonstrate that numerical pathologies manifest themselves in “classic” analytic EI, Expected Hypervolume Improvement (EHVI), as well as their constrained, noisy, and parallel variants, and propose corresponding reformulations that remedy these pathologies. Our empirical results show that members of the LogEI family of acquisition functions substantially improve on the optimization performance of their canonical counterparts and surprisingly, are on par with or exceed the performance of recent state-of-the-art acquisition functions, highlighting the understated role of numerical optimization in the literature.
****
Pseudo-Likelihood Inference
Theo Gruner, Boris Belousov, Fabio Muratore, Daniel Palenicek, Jan R. Peters
Simulation-Based Inference (SBI) is a common name for an emerging family of appr

oaches that infer the model parameters when the likelihood is intractable. Existing SBI methods either approximate the likelihood, such as Approximate Bayesian Computation (ABC) or directly model the posterior, such as Sequential Neural Posterior Estimation (SNPE). While ABC is efficient on low-dimensional problems, on higher-dimensional tasks, it is generally outperformed by SNPE, which leverages function approximation. In this paper, we propose Pseudo-Likelihood Inference (PLI), a new method that brings neural approximation into ABC, making it competitive on challenging Bayesian system identification tasks. By utilizing integral probability metrics, we introduce a smooth likelihood kernel with an adaptive bandwidth that is updated based on information-theoretic trust regions. Thanks to this formulation, our method (i) allows for optimizing neural posteriors via gradient descent, (ii) does not rely on summary statistics, and (iii) enables multiple observations as input. In comparison to SNPE, it leads to improved performance when more data is available. The effectiveness of PLI is evaluated on four classical SBI benchmark tasks and on a highly dynamic physical system, showing particular advantages on stochastic simulations and multi-modal posterior landscapes.

****

Calibrating “Cheap Signals” in Peer Review without a Prior

Yuxuan Lu, Yuqing Kong

Peer review lies at the core of the academic process, but even well-intentioned reviewers can still provide noisy ratings. While ranking papers by average ratings may reduce noise, varying noise levels and systematic biases stemming from ``cheap'' signals (e.g. author identity, proof length) can lead to unfairness. Detecting and correcting bias is challenging, as ratings are subjective and unverifiable. Unlike previous works relying on prior knowledge or historical data, we propose a one-shot noise calibration process without any prior information. We ask reviewers to predict others' scores and use these predictions for calibration. Assuming reviewers adjust their predictions according to the noise, we demonstrate that the calibrated score results in a more robust ranking compared to average ratings, even with varying noise levels and biases.In detail, we show that the error probability of the calibrated score approaches zero as the number of reviewers increases and is significantly lower compared to average ratings when the number of reviewers is small.

****

SnapFusion: Text-to-Image Diffusion Model on Mobile Devices within Two Seconds

Yanyu Li, Huan Wang, Qing Jin, Ju Hu, Pavlo Chemerys, Yun Fu, Yanzhi Wang, Sergey Tulyakov, Jian Ren

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

LinGCN: Structural Linearized Graph Convolutional Network for Homomorphically Encrypted Inference

Hongwu Peng, Ran Ran, Yukui Luo, Jiahui Zhao, Shaoyi Huang, Kiran Thorat, Tong Geng, Chenghong Wang, Xiaolin Xu, Wujie Wen, Caiwen Ding

The growth of Graph Convolution Network (GCN) model sizes has revolutionized numerous applications, surpassing human performance in areas such as personal healthcare and financial systems. The deployment of GCNs in the cloud raises privacy concerns due to potential adversarial attacks on client data. To address security concerns, Privacy-Preserving Machine Learning (PPML) using Homomorphic Encryption (HE) secures sensitive client data. However, it introduces substantial computational overhead in practical applications. To tackle those challenges, we present LinGCN, a framework designed to reduce multiplication depth and optimize the performance of HE based GCN inference. LinGCN is structured around three key elements: (1) A differentiable structural linearization algorithm, complemented by a parameterized discrete indicator function, co-trained with model weights to meet the optimization goal. This strategy promotes fine-grained node-level non-linear location selection, resulting in a model with minimized multiplication dep

th. (2) A compact node-wise polynomial replacement policy with a second-order trainable activation function, steered towards superior convergence by a two-level distillation approach from an all-ReLU based teacher model. (3) an enhanced HE solution that enables finer-grained operator fusion for node-wise activation functions, further reducing multiplication level consumption in HE-based inference. Our experiments on the NTU-XVIEW skeleton joint dataset reveal that LinGCN excels in latency, accuracy, and scalability for homomorphically encrypted inference, outperforming solutions such as CryptoGCN. Remarkably, LinGCN achieves a 14.2× latency speedup relative to CryptoGCN, while preserving an inference accuracy of ~75\% and notably reducing multiplication depth. Additionally, LinGCN proves scalable for larger models, delivering a substantial 85.78\% accuracy with 6371s latency, a 10.47\% accuracy improvement over CryptoGCN.
****
Spectral Evolution and Invariance in Linear-width Neural Networks
Zhichao Wang, Andrew Engel, Anand D Sarwate, Ioana Dumitriu, Tony Chiang
We investigate the spectral properties of linear-width feed-forward neural networks, where the sample size is asymptotically proportional to network width. Empirically, we show that the spectra of weight in this high dimensional regime are invariant when trained by gradient descent for small constant learning rates; we provide a theoretical justification for this observation and prove the invariance of the bulk spectra for both conjugate and neural tangent kernels. We demonstrate similar characteristics when training with stochastic gradient descent with small learning rates. When the learning rate is large, we exhibit the emergence of an outlier whose corresponding eigenvector is aligned with the training data structure. We also show that after adaptive gradient training, where a lower test error and feature learning emerge, both weight and kernel matrices exhibit heavy tail behavior. Simple examples are provided to explain when heavy tails can have better generalizations. We exhibit different spectral properties such as invariant bulk, spike, and heavy-tailed distribution from a two-layer neural network using different training strategies, and then correlate them to the feature learning. Analogous phenomena also appear when we train conventional neural networks with real-world data. We conclude that monitoring the evolution of the spectra during training is an essential step toward understanding the training dynamics and feature learning.
****
Paxion: Patching Action Knowledge in Video-Language Foundation Models
Zhenhailong Wang, Ansel Blume, Sha Li, Genglin Liu, Jaemin Cho, Zineng Tang, Mohit Bansal, Heng Ji
Action knowledge involves the understanding of textual, visual, and temporal aspects of actions. We introduce the Action Dynamics Benchmark (ActionBench) containing two carefully designed probing tasks: Action Antonym and Video Reversal, which targets multimodal alignment capabilities and temporal understanding skills of the model, respectively. Despite recent video-language models' (VidLM) impressive performance on various benchmark tasks, our diagnostic tasks reveal their surprising deficiency (near-random performance) in action knowledge, suggesting that current models rely on object recognition abilities as a shortcut for action understanding. To remedy this, we propose a novel framework, Paxion, along with a new Discriminative Video Dynamics Modeling (DVDM) objective. The Paxion framework utilizes a Knowledge Patcher network to encode new action knowledge and a Knowledge Fuser component to integrate the Patcher into frozen VidLMs without compromising their existing capabilities. Due to limitations of the widely-used Video-Text Contrastive (VTC) loss for learning action knowledge, we introduce the DVDM objective to train the Knowledge Patcher. DVDM forces the model to encode the correlation between the action text and the correct ordering of video frames. Our extensive analyses show that Paxion and DVDM together effectively fill the gap in action knowledge understanding (~50% $\rightarrow$ 80%), while maintaining or improving performance on a wide spectrum of both object- and action-centric downstream tasks.
****
ProPILE: Probing Privacy Leakage in Large Language Models

Siwon Kim, Sangdoo Yun, Hwaran Lee, Martin Gubri, Sungroh Yoon, Seong Joon Oh
The rapid advancement and widespread use of large language models (LLMs) have raised significant concerns regarding the potential leakage of personally identifiable information (PII). These models are often trained on vast quantities of web-collected data, which may inadvertently include sensitive personal data. This paper presents ProPILE, a novel probing tool designed to empower data subjects, or the owners of the PII, with awareness of potential PII leakage in LLM-based services. ProPILE lets data subjects formulate prompts based on their own PII to evaluate the level of privacy intrusion in LLMs. We demonstrate its application on the OPT-1.3B model trained on the publicly available Pile dataset. We show how hypothetical data subjects may assess the likelihood of their PII being included in the Pile dataset being revealed. ProPILE can also be leveraged by LLM service providers to effectively evaluate their own levels of PII leakage with more powerful prompts specifically tuned for their in-house models. This tool represents a pioneering step towards empowering the data subjects for their awareness and control over their own data on the web.
****

Mind the spikes: Benign overfitting of kernels and neural networks in fixed dimension
Moritz Haas, David Holzmüller, Ulrike Luxburg, Ingo Steinwart
The success of over-parameterized neural networks trained to near-zero training error has caused great interest in the phenomenon of benign overfitting, where estimators are statistically consistent even though they interpolate noisy training data. While benign overfitting in fixed dimension has been established for some learning methods, current literature suggests that for regression with typical kernel methods and wide neural networks, benign overfitting requires a high-dimensional setting, where the dimension grows with the sample size. In this paper, we show that the smoothness of the estimators, and not the dimension, is the key: benign overfitting is possible if and only if the estimator's derivatives are large enough. We generalize existing inconsistency results to non-interpolating models and more kernels to show that benign overfitting with moderate derivatives is impossible in fixed dimension. Conversely, we show that benign overfitting is possible for regression with a sequence of spiky-smooth kernels with large derivatives. Using neural tangent kernels, we translate our results to wide neural networks. We prove that while infinite-width networks do not overfit benignly with the ReLU activation, this can be fixed by adding small high-frequency fluctuations to the activation function. Our experiments verify that such neural networks, while overfitting, can indeed generalize well even on low-dimensional data sets.
****

The Goldilocks of Pragmatic Understanding: Fine-Tuning Strategy Matters for Implicature Resolution by LLMs
Laura Ruis, Akbir Khan, Stella Biderman, Sara Hooker, Tim Rocktäschel, Edward Grefenstette
Despite widespread use of LLMs as conversational agents, evaluations of performance fail to capture a crucial aspect of communication: interpreting language in context---incorporating its pragmatics. Humans interpret language using beliefs and prior knowledge about the world. For example, we intuitively understand the response "I wore gloves" to the question "Did you leave fingerprints?" as meaning "No". To investigate whether LLMs have the ability to make this type of inference, known as an implicature, we design a simple task and evaluate four categories of widely used state-of-the-art models. We find that, despite only evaluating on utterances that require a binary inference (yes or no), models in three of these categories perform close to random. However, LLMs instruction-tuned at the example-level perform significantly better. These results suggest that certain fine-tuning strategies are far better at inducing pragmatic understanding in models. We present our findings as the starting point for further research into evaluating how LLMs interpret language in context and to drive the development of more pragmatic and useful models of human discourse.
****

Slow and Weak Attractor Computation Embedded in Fast and Strong E-I Balanced Neural Dynamics
Xiaohan Lin, Liyuan Li, Boxin Shi, Tiejun Huang, Yuanyuan Mi, Si Wu
Attractor networks require neuronal connections to be highly structured in order to maintain attractor states that represent information, while excitation and inhibition balanced networks (E-INNs) require neuronal connections to be random and sparse to generate irregular neuronal firings. Despite being regarded as canonical models of neural circuits, both types of networks are usually studied in isolation, and it remains unclear how they coexist in the brain, given their very different structural demands. In this study, we investigate the compatibility of continuous attractor neural networks (CANNs) and E-INNs. In line with recent experimental data, we find that a neural circuit can exhibit both the traits of CANNs and E-INNs if the neuronal synapses consist of two sets: one set is strong and fast for irregular firing, and the other set is weak and slow for attractor dynamics. Our results from simulations and theoretical analysis reveal that the network also exhibits enhanced performance compared to the case of using only one set of synapses, with accelerated convergence of attractor states and retained E-I balanced condition for localized input. We also apply the network model to solve a real-world tracking problem and demonstrate that it can track fast-moving objects well. We hope that this study provides insight into how structured neural computations are realized by irregular firings of neurons.
****
Test-time Training for Matching-based Video Object Segmentation
Juliette Bertrand, Giorgos Kordopatis Zilos, Yannis Kalantidis, Giorgos Tolias
The video object segmentation (VOS) task involves the segmentation of an object over time based on a single initial mask. Current state-of-the-art approaches use a memory of previously processed frames and rely on matching to estimate segmentation masks of subsequent frames. Lacking any adaptation mechanism, such methods are prone to test-time distribution shifts. This work focuses on matching-based VOS under distribution shifts such as video corruptions, stylization, and sim-to-real transfer. We explore test-time training strategies that are agnostic to the specific task as well as strategies that are designed specifically for VOS. This includes a variant based on mask cycle consistency tailored to matching-based VOS methods. The experimental results on common benchmarks demonstrate that the proposed test-time training yields significant improvements in performance. In particular for the sim-to-real scenario and despite using only a single test video, our approach manages to recover a substantial portion of the performance gain achieved through training on real videos. Additionally, we introduce DAVIS-C, an augmented version of the popular DAVIS test set, featuring extreme distribution shifts like image-/video-level corruptions and stylizations. Our results illustrate that test-time training enhances performance even in these challenging cases.
****
Causal Effect Regularization: Automated Detection and Removal of Spurious Correlations
Abhinav Kumar, Amit Deshpande, Amit Sharma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Multi-resolution Spectral Coherence for Graph Generation with Score-based Diffusion
Hyuna Cho, Minjae Jeong, Sooyeon Jeon, Sungsoo Ahn, Won Hwa Kim
Successful graph generation depends on the accurate estimation of the joint distribution of graph components such as nodes and edges from training data. While recent deep neural networks have demonstrated sampling of realistic graphs together with diffusion models, however, they still suffer from oversmoothing problems which are inherited from conventional graph convolution and thus high-frequency characteristics of nodes and edges become intractable. To overcome such issues

and generate graphs with high fidelity, this paper introduces a novel approach that captures the dependency between nodes and edges at multiple resolutions in the spectral space. By modeling the joint distribution of node and edge signals in a shared graph wavelet space, together with a score-based diffusion model, we propose a Wavelet Graph Diffusion Model (Wave-GD) which lets us sample synthetic graphs with real-like frequency characteristics of nodes and edges. Experimental results on four representative benchmark datasets validate the superiority of the Wave-GD over existing approaches, highlighting its potential for a wide range of applications that involve graph data.
****
Real-World Image Super-Resolution as Multi-Task Learning
Wenlong Zhang, Xiaohui Li, Guangyuan SHI, Xiangyu Chen, Yu Qiao, Xiaoyun Zhang, Xiao-Ming Wu, Chao Dong
In this paper, we take a new look at real-world image super-resolution (real-SR) from a multi-task learning perspective. We demonstrate that the conventional formulation of real-SR can be viewed as solving multiple distinct degradation tasks using a single shared model. This poses a challenge known as task competition or task conflict in multi-task learning, where certain tasks dominate the learning process, resulting in poor performance on other tasks. This problem is exacerbated in the case of real-SR, due to the involvement of numerous degradation tasks. To address the issue of task competition in real-SR, we propose a task grouping approach. Our approach efficiently identifies the degradation tasks where a real-SR model falls short and groups these unsatisfactory tasks into multiple task groups. We then utilize the task groups to fine-tune the real-SR model in a simple way, which effectively mitigates task competition and facilitates knowledge transfer. Extensive experiments demonstrate our method achieves significantly enhanced performance across a wide range of degradation scenarios.
****
Exact Representation of Sparse Networks with Symmetric Nonnegative Embeddings
Sudhanshu Chanpuriya, Ryan Rossi, Anup B. Rao, Tung Mai, Nedim Lipka, Zhao Song, Cameron Musco
Graph models based on factorization of the adjacency matrix often fail to capture network structures related to links between dissimilar nodes (heterophily). We introduce a novel graph factorization model that leverages two nonnegative vectors per node to interpretably account for links between both similar and dissimilar nodes. We prove that our model can exactly represent any graph with low arboricity, a property that many real-world networks satisfy; our proof also applies to related models but has much greater scope than the closest prior bound, which is based on low max degree. Our factorization also has compelling properties besides expressiveness: due to its symmetric structure and nonnegativity, fitting the model inherently finds node communities, and the model's link predictions can be interpreted in terms of these communities. In experiments on real-world networks, we demonstrate our factorization's effectiveness on a variety of tasks, including community detection and link prediction.
****
Data-Centric Learning from Unlabeled Graphs with Diffusion Model
Gang Liu, Eric Inae, Tong Zhao, Jiaxin Xu, Tengfei Luo, Meng Jiang
Graph property prediction tasks are important and numerous. While each task offers a small size of labeled examples, unlabeled graphs have been collected from various sources and at a large scale. A conventional approach is training a model with the unlabeled graphs on self-supervised tasks and then fine-tuning the model on the prediction tasks. However, the self-supervised task knowledge could not be aligned or sometimes conflicted with what the predictions needed. In this paper, we propose to extract the knowledge underlying the large set of unlabeled graphs as a specific set of useful data points to augment each property prediction model. We use a diffusion model to fully utilize the unlabeled graphs and design two new objectives to guide the model's denoising process with each task's labeled data to generate task-specific graph examples and their labels. Experiments demonstrate that our data-centric approach performs significantly better than fifteen existing various methods on fifteen tasks. The performance improvement

brought by unlabeled data is visible as the generated labeled examples unlike the self-supervised learning.
****
Wasserstein Gradient Flows for Optimizing Gaussian Mixture Policies
Hanna Ziesche, Leonel Rozo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Change point detection and inference in multivariate non-parametric models under mixing conditions
Carlos Misael Madrid Padilla, Haotian Xu, Daren Wang, OSCAR HERNAN MADRID PADILLA, Yi Yu
This paper addresses the problem of localizing and inferring multiple change points, in non-parametric multivariate time series settings. Specifically, we consider a multivariate time series with potentially short-range dependence, whose underlying distributions have Hölder smooth densities and can change over time in a piecewise-constant manner. The change points, which correspond to the times when the distribution changes, are unknown. We present the limiting distributions of the change point estimators under the scenarios where the minimal jump size vanishes or remains constant. Such results have not been revealed in the literature in non-parametric change point settings. As byproducts, we develop a sharp estimator that can accurately localize the change points in multivariate non-parametric time series, and a consistent block-type long-run variance estimator. Numerical studies are provided to complement our theoretical findings.
****
Near-optimal learning with average Hölder smoothness
Guy Kornowski, Steve Hanneke, Aryeh Kontorovich
We generalize the notion of average Lipschitz smoothness proposed by Ashlagi et al. (COLT 2021) by extending it to Hölder smoothness. This measure of the "effective smoothness" of a function is sensitive to the underlying distribution and can be dramatically smaller than its classic "worst-case" Hölder constant.We consider both the realizable and the agnostic (noisy) regression settings, proving upper and lower risk bounds in terms of the average Hölder smoothness; these rates improve upon both previously known rates even in the special case of average Lipschitz smoothness.Moreover, our lower bound is tight in the realizable setting up to log factors, thus we establish the minimax rate.From an algorithmic perspective, since our notion of average smoothness is defined with respect to the unknown underlying distribution, the learner does not have an explicit representation of the function class, hence is unable to execute ERM. Nevertheless, we provide distinct learning algorithms that achieve both (nearly) optimal learning rates.Our results hold in any totally bounded metric space, and are stated in terms of its intrinsic geometry.Overall, our results show that the classic worst-case notion of Hölder smoothness can be essentially replaced by its average, yielding considerably sharper guarantees.
****
Neural-Logic Human-Object Interaction Detection
Liulei Li, Jianan Wei, Wenguan Wang, Yi Yang
The interaction decoder utilized in prevalent Transformer-based HOI detectors typically accepts pre-composed human-object pairs as inputs. Though achieving remarkable performance, such a paradigm lacks feasibility and cannot explore novel combinations over entities during decoding. We present LogicHOI, a new HOI detector that leverages neural-logic reasoning and Transformer to infer feasible interactions between. entities. Specifically, we modify. self-attention mechanism in the vanilla Transformer, enabling it to reason over the ■ human, action, object ■ triplet and constitute novel interactions. Meanwhile, such a reasoning process is guided by two crucial properties for understanding HOI: affordances (the potential actions an object can facilitate) and proxemics (the spatial relations between humans and objects). We formulate these two properties in first-order logi

c and ground them into continuous space to constrain the learning process of our approach, leading to improved performance and zero-shot generalization capabilities. We evaluate L OGIC HOI on V-COCO and HICO-DET under both normal and zero-shot setups, achieving significant improvements over existing methods.
****

Which Models have Perceptually-Aligned Gradients? An Explanation via Off-Manifold Robustness

Suraj Srinivas, Sebastian Bordt, Himabindu Lakkaraju

One of the remarkable properties of robust computer vision models is that their input-gradients are often aligned with human perception, referred to in the literature as perceptually-aligned gradients (PAGs). Despite only being trained for classification, PAGs cause robust models to have rudimentary generative capabilities, including image generation, denoising, and in-painting. However, the underlying mechanisms behind these phenomena remain unknown. In this work, we provide a first explanation of PAGs via \emph{off-manifold robustness}, which states that models must be more robust off- the data manifold than they are on-manifold. We first demonstrate theoretically that off-manifold robustness leads input gradients to lie approximately on the data manifold, explaining their perceptual alignment. We then show that Bayes optimal models satisfy off-manifold robustness, and confirm the same empirically for robust models trained via gradient norm regularization, randomized smoothing, and adversarial training with projected gradient descent. Quantifying the perceptual alignment of model gradients via their similarity with the gradients of generative models, we show that off-manifold robustness correlates well with perceptual alignment. Finally, based on the levels of on- and off-manifold robustness, we identify three different regimes of robustness that affect both perceptual alignment and model accuracy: weak robustness, bayes-aligned robustness, and excessive robustness. Code is available at https://github.com/tml-tuebingen/pags.
****

Inferring the Future by Imagining the Past

Kartik Chandra, Tony Chen, Tzu-Mao Li, Jonathan Ragan-Kelley, Josh Tenenbaum

A single panel of a comic book can say a lot: it can depict not only where the characters currently are, but also their motions, their motivations, their emotions, and what they might do next. More generally, humans routinely infer complex sequences of past and future events from a static snapshot of a dynamic scene, even in situations they have never seen before.In this paper, we model how humans make such rapid and flexible inferences. Building on a long line of work in cognitive science, we offer a Monte Carlo algorithm whose inferences correlate well with human intuitions in a wide variety of domains, while only using a small, cognitively-plausible number of samples. Our key technical insight is a surprising connection between our inference problem and Monte Carlo path tracing, which allows us to apply decades of ideas from the computer graphics community to this seemingly-unrelated theory of mind task.
****

The Grand Illusion: The Myth of Software Portability and Implications for ML Progress.

Fraser Mince, Dzung Dinh, Jonas Kgomo, Neil Thompson, Sara Hooker

Pushing the boundaries of machine learning often requires exploring different hardware and software combinations. However, this ability to experiment with different systems can be at odds with the drive for efficiency, which has produced increasingly specialized AI hardware and incentivized consolidation around a narrow set of ML frameworks. Exploratory research can be further restricted if software and hardware are co-evolving, making it even harder to stray away from a given tooling stack. While this friction increasingly impacts the rate of innovation in machine learning, to our knowledge the lack of portability in tooling has not been quantified. In this work we ask: How portable are popular ML software frameworks? We conduct a large scale study of the portability of mainstream ML frameworks across different hardware types. Our findings paint an uncomfortable picture -- frameworks can lose more than 40% of their key functions when ported to other hardware. Worse, even when functions are portable, the slowdown in their pe

rformance can be extreme. Collectively, our results reveal how costly straying from a narrow set of hardware-software combinations can be - and thus how specialization incurs an exploration cost that can impede innovation in machine learning research.
****

Computing Optimal Nash Equilibria in Multiplayer Games

Youzhi Zhang, Bo An, Venkatramanan Subrahmanian

Designing efficient algorithms to compute a Nash Equilibrium (NE) in multiplayer games is still an open challenge. In this paper, we focus on computing an NE that optimizes a given objective function. For example, when there is a team of players independently playing against an adversary in a game (e.g., several groups in a forest trying to interdict illegal loggers in green security games), these team members may need to find an NE minimizing the adversary’s utility. Finding an optimal NE in multiplayer games can be formulated as a mixed-integer bilinear program by introducing auxiliary variables to represent bilinear terms, leading to a huge number of bilinear terms, making it hard to solve. To overcome this challenge, we first propose a general framework for this formulation based on a set of correlation plans. We then develop a novel algorithm called CRM based on this framework, which uses correlation plans with their relations to strictly reduce the feasible solution space after the convex relaxation of bilinear terms while minimizing the number of correlation plans to significantly reduce the number of bilinear terms. We show that our techniques can significantly reduce the time complexity and CRM can be several orders of magnitude faster than the state-of-the-art baseline.
****

AND: Adversarial Neural Degradation for Learning Blind Image Super-Resolution

Fangzhou Luo, Xiaolin Wu, Yanhui Guo

Learnt deep neural networks for image super-resolution fail easily if the assumed degradation model in training mismatches that of the real degradation source at the inference stage. Instead of attempting to exhaust all degradation variants in simulation, which is unwieldy and impractical, we propose a novel adversarial neural degradation (AND) model that can, when trained in conjunction with a deep restoration neural network under a minmax criterion, generate a wide range of highly nonlinear complex degradation effects without any explicit supervision. The AND model has a unique advantage over the current state of the art in that it can generalize much better to unseen degradation variants and hence deliver significantly improved restoration performance on real-world images.
****

Into the LAION’s Den: Investigating Hate in Multimodal Datasets

Abeba Birhane, vinay prabhu, Sanghyun Han, Vishnu Boddeti, Sasha Luccioni

`Scale the model, scale the data, scale the compute' is the reigning sentiment in the world of generative AI today. While the impact of model scaling has been extensively studied, we are only beginning to scratch the surface of data scaling and its consequences. This is especially of critical importance in the context of vision-language datasets such as LAION. These datasets are continually growing in size and are built based on large-scale internet dumps such as the Common Crawl, which is known to have numerous drawbacks ranging from quality, legality, and content. The datasets then serve as the backbone for large generative models, contributing to the operationalization and perpetuation of harmful societal and historical biases and stereotypes. In this paper, we investigate the effect of scaling datasets on hateful content through a comparative audit of two datasets: LAION-400M and LAION-2B. Our results show that hate content increased by nearly 12% with dataset scale, measured both qualitatively and quantitatively using a metric that we term as Hate Content Rate (HCR). We also found that filtering dataset contents based on Not Safe For Work (NSFW) values calculated based on images alone does not exclude all the harmful content in alt-text. Instead, we found that trace amounts of hateful, targeted, and aggressive text remain even when carrying out conservative filtering. We end with a reflection and a discussion of the significance of our results for dataset curation and usage in the AI community.Code and the meta-data assets curated in this paper are publicly available a

t https://github.com/vinayprabhu/hate_scaling. Content warning: This paper contains examples of hateful text that might be disturbing, distressing, and/or offensive.
****
SE(3) Diffusion Model-based Point Cloud Registration for Robust 6D Object Pose Estimation
Haobo Jiang, Mathieu Salzmann, Zheng Dang, Jin Xie, Jian Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
RoboDepth: Robust Out-of-Distribution Depth Estimation under Corruptions
Lingdong Kong, Shaoyuan Xie, Hanjiang Hu, Lai Xing Ng, Benoit Cottereau, Wei Tsang Ooi
Depth estimation from monocular images is pivotal for real-world visual perception systems. While current learning-based depth estimation models train and test on meticulously curated data, they often overlook out-of-distribution (OoD) situations. Yet, in practical settings -- especially safety-critical ones like autonomous driving -- common corruptions can arise. Addressing this oversight, we introduce a comprehensive robustness test suite, RoboDepth, encompassing 18 corruptions spanning three categories: i) weather and lighting conditions; ii) sensor failures and movement; and iii) data processing anomalies. We subsequently benchmark 42 depth estimation models across indoor and outdoor scenes to assess their resilience to these corruptions. Our findings underscore that, in the absence of a dedicated robustness evaluation framework, many leading depth estimation models may be susceptible to typical corruptions. We delve into design considerations for crafting more robust depth estimation models, touching upon pre-training, augmentation, modality, model capacity, and learning paradigms. We anticipate our benchmark will establish a foundational platform for advancing robust OoD depth estimation.
****
Fed-CO$_{2}$: Cooperation of Online and Offline Models for Severe Data Heterogeneity in Federated Learning
Zhongyi Cai, Ye Shi, Wei Huang, Jingya Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Combating Bilateral Edge Noise for Robust Link Prediction
Zhanke Zhou, Jiangchao Yao, Jiaxu Liu, Xiawei Guo, Quanming Yao, LI He, Liang Wang, Bo Zheng, Bo Han
Although link prediction on graphs has achieved great success with the development of graph neural networks (GNNs), the potential robustness under the edge noise is still less investigated. To close this gap, we first conduct an empirical study to disclose that the edge noise bilaterally perturbs both input topology and target label, yielding severe performance degradation and representation collapse. To address this dilemma, we propose an information-theory-guided principle, Robust Graph Information Bottleneck (RGIB), to extract reliable supervision signals and avoid representation collapse. Different from the basic information bottleneck, RGIB further decouples and balances the mutual dependence among graph topology, target labels, and representation, building new learning objectives for robust representation against the bilateral noise. Two instantiations, RGIB-SSL and RGIB-REP, are explored to leverage the merits of different methodologies, i.e., self-supervised learning and data reparameterization, for implicit and explicit data denoising, respectively. Extensive experiments on six datasets and three GNNs with diverse noisy scenarios verify the effectiveness of our RGIB instantiations. The code is publicly available at: https://github.com/tmlr-group/RGIB.
****

SyncTREE: Fast Timing Analysis for Integrated Circuit Design through a Physics-informed Tree-based Graph Neural Network
Yuting Hu, Jiajie Li, Florian Klemme, Gi-Joon Nam, Tengfei Ma, Hussam Amrouch, Jinjun Xiong
Nowadays integrated circuits (ICs) are underpinning all major information technology innovations including the current trends of artificial intelligence (AI). Modern IC designs often involve analyses of complex phenomena (such as timing, noise, and power etc.) for tens of billions of electronic components, like resistance (R), capacitance (C), transistors and gates, interconnected in various complex structures. Those analyses often need to strike a balance between accuracy and speed as those analyses need to be carried out many times throughout the entire IC design cycles. With the advancement of AI, researchers also start to explore news ways in leveraging AI to improve those analyses. This paper focuses on one of the most important analyses, timing analysis for interconnects. Since IC interconnects can be represented as an RC-tree, a specialized graph as tree, we design a novel tree-based graph neural network, SyncTREE, to speed up the timing analysis by incorporating both the structural and physical properties of electronic circuits. Our major innovations include (1) a two-pass message-passing (bottom-up and top-down) for graph embedding, (2) a tree contrastive loss to guide learning, and (3) a closed formular-based approach to conduct fast timing. Our experiments show that, compared to conventional GNN models, SyncTREE achieves the best timing prediction in terms of both delays and slews, all in reference to the industry golden numerical analyses results on real IC design data.
****
Hierarchical Open-vocabulary Universal Image Segmentation
Xudong Wang, Shufan Li, Konstantinos Kallidromitis, Yusuke Kato, Kazuki Kozuka, Trevor Darrell
Open-vocabulary image segmentation aims to partition an image into semantic regions according to arbitrary text descriptions. However, complex visual scenes can be naturally decomposed into simpler parts and abstracted at multiple lev4 els of granularity, introducing inherent segmentation ambiguity. Unlike existing methods that typically sidestep this ambiguity and treat it as an external factor, our approach actively incorporates a hierarchical representation encompassing different semantic-levels into the learning process. We propose a decoupled text-image fusion mechanism and representation learning modules for both “things” and “stuff”. Additionally, we systematically examine the differences that exist in the textual and visual features between these types of categories. Our resulting model, named HIPIE, tackles HIerarchical, oPen-vocabulary, and unIvErsal segmentation tasks within a unified framework. Benchmarked on diverse datasets, e.g., ADE20K,COCO, Pascal-VOC Part, and RefCOCO/RefCOCOg, HIPIE achieves the state-of14 the-art results at various levels of image comprehension, including semantic-level (e.g., semantic segmentation), instance-level (e.g., panoptic/referring segmentationand object detection), as well as part-level (e.g., part/subpart segmentation) tasks.
****
Fairly Recommending with Social Attributes: A Flexible and Controllable Optimization Approach
Jinqiu Jin, Haoxuan Li, Fuli Feng, Sihao Ding, Peng Wu, Xiangnan He
Item-side group fairness (IGF) requires a recommendation model to treat different item groups similarly, and has a crucial impact on information diffusion, consumption activity, and market equilibrium. Previous IGF notions only focus on the direct utility of the item exposures, i.e., the exposure numbers across different item groups. Nevertheless, the item exposures also facilitate utility gained from the neighboring users via social influence, called social utility, such as information sharing on the social media. To fill this gap, this paper introduces two social attribute-aware IGF metrics, which require similar user social attributes on the exposed items across the different item groups. In light of the trade-off between the direct utility and social utility, we formulate a new multi-objective optimization problem for training recommender models with flexible trade-off while ensuring controllable accuracy. To solve this problem, we develop a

gradient-based optimization algorithm and theoretically show that the proposed a lgorithm can find Pareto optimal solutions with varying trade-off and guaranteed accuracy. Extensive experiments on two real-world datasets validate the effecti veness of our approach.
****
Look Ma, No Hands! Agent-Environment Factorization of Egocentric Videos
Matthew Chang, Aditya Prakash, Saurabh Gupta
The analysis and use of egocentric videos for robotics tasks is made challenging by occlusion and the visual mismatch between the human hand and a robot end-eff ector. Past work views the human hand as a nuisance and removes it from the scen e. However, the hand also provides a valuable signal for learning. In this work, we propose to extract a factored representation of the scene that separates the agent (human hand) and the environment. This alleviates both occlusion and mism atch while preserving the signal, thereby easing the design of models for downst ream robotics tasks. At the heart of this factorization is our proposed Video In painting via Diffusion Model (VIDM) that leverages both a prior on real-world im ages (through a large-scale pre-trained diffusion model) and the appearance of t he object in earlier frames of the video (through attention). Our experiments de monstrate the effectiveness of VIDM at improving the in-painting quality in egoc entric videos and the power of our factored representation for numerous tasks: o bject detection, 3D reconstruction of manipulated objects, and learning of rewar d functions, policies, and affordances from videos.
****
Generating Images with Multimodal Language Models
Jing Yu Koh, Daniel Fried, Russ R. Salakhutdinov
We propose a method to fuse frozen text-only large language models (LLMs) with p re-trained image encoder and decoder models, by mapping between their embedding spaces. Our model demonstrates a wide suite of multimodal capabilities: image re trieval, novel image generation, and multimodal dialogue. Ours is the first appr oach capable of conditioning on arbitrarily interleaved image and text inputs to generate coherent image (and text) outputs. To achieve strong performance on im age generation, we propose an efficient mapping network to ground the LLM to an off-the-shelf text-to-image generation model. This mapping network translates hi dden representations of text into the embedding space of the visual models, enab ling us to leverage the strong text representations of the LLM for visual output s. Our approach outperforms baseline generation models on tasks with longer and more complex language. In addition to novel image generation, our model is also capable of image retrieval from a prespecified dataset, and decides whether to r etrieve or generate at inference time. This is done with a learnt decision modul e which conditions on the hidden representations of the LLM. Our model exhibits a wider range of capabilities compared to prior multimodal language models. It c an process image-and-text inputs, and produce retrieved images, generated images , and generated text — outperforming non-LLM based generation models across seve ral text-to-image tasks that measure context dependence.
****
MoVie: Visual Model-Based Policy Adaptation for View Generalization
Sizhe Yang, Yanjie Ze, Huazhe Xu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Does Visual Pretraining Help End-to-End Reasoning?
Chen Sun, Calvin Luo, Xingyi Zhou, Anurag Arnab, Cordelia Schmid
We aim to investigate whether end-to-end learning of visual reasoning can be ach ieved with general-purpose neural networks, with the help of visual pretraining. A positive result would refute the common belief that explicit visual abstracti on (e.g. object detection) is essential for compositional generalization on visu al reasoning, and confirm the feasibility of a neural network ''generalist'' to solve visual recognition and reasoning tasks. We propose a simple and general se

lf-supervised framework which ''compresses'' each video frame into a small set of tokens with a transformer network, and reconstructs the remaining frames based on the compressed temporal context. To minimize the reconstruction loss, the network must learn a compact representation for each image, as well as capture temporal dynamics and object permanence from temporal context. We perform evaluation on two visual reasoning benchmarks, CATER and ACRE. We observe that pretraining is essential to achieve compositional generalization for end-to-end visual reasoning. Our proposed framework outperforms traditional supervised pretraining, including image classification and explicit object detection, by large margins.
****
Newton–Cotes Graph Neural Networks: On the Time Evolution of Dynamic Systems
Lingbing Guo, Weiqing Wang, Zhuo Chen, Ningyu Zhang, Zequn Sun, Yixuan Lai, Qiang Zhang, Huajun Chen
Reasoning system dynamics is one of the most important analytical approaches for many scientific studies. With the initial state of a system as input, the recent graph neural networks (GNNs)-based methods are capable of predicting the future state distant in time with high accuracy. Although these methods have diverse designs in modeling the coordinates and interacting forces of the system, we show that they actually share a common paradigm that learns the integration of the velocity over the interval between the initial and terminal coordinates. However, their integrand is constant w.r.t. time. Inspired by this observation, we propose a new approach to predict the integration based on several velocity estimations with Newton–Cotes formulas and prove its effectiveness theoretically. Extensive experiments on several benchmarks empirically demonstrate consistent and significant improvement compared with the state-of-the-art methods.
****
Is Your Code Generated by ChatGPT Really Correct? Rigorous Evaluation of Large Language Models for Code Generation
Jiawei Liu, Chunqiu Steven Xia, Yuyao Wang, LINGMING ZHANG
Program synthesis has been long studied with recent approaches focused on directly using the power of Large Language Models (LLMs) to generate code. Programming benchmarks, with curated synthesis problems and test-cases, are used to measure the performance of various LLMs on code synthesis. However, these test-cases can be limited in both quantity and quality for fully assessing the functional correctness of the generated code. Such limitation in the existing benchmarks begs the following question: In the era of LLMs, is the code generated really correct? To answer this, we propose EvalPlus – a code synthesis evaluation framework to rigorously benchmark the functional correctness of LLM-synthesized code. EvalPlus augments a given evaluation dataset with large amounts of test-cases newly produced by an automatic test input generator, powered by both LLM- and mutation-based strategies. While EvalPlus is general, we extend the test-cases of the popular HumanEval benchmark by 80x to build HumanEval+. Our extensive evaluation across 26 popular LLMs (e.g., GPT-4 and ChatGPT) demonstrates that HumanEval+ is able to catch significant amounts of previously undetected wrong code synthesized by LLMs, reducing the pass@k by up-to 19.3-28.9%. We also surprisingly found that test insufficiency can lead to mis-ranking. For example, both WizardCoder-CodeLlama and Phind-CodeLlama now outperform ChatGPT on HumanEval+, while none of them could on HumanEval. Our work not only indicates that prior popular code synthesis evaluation results do not accurately reflect the true performance of LLMs for code synthesis, but also opens up a new direction to improve such programming benchmarks through automated testing. We have open-sourced our tools, enhanced datasets as well as all LLM-generated code at https://github.com/evalplus/evalplus to facilitate and accelerate future LLM-for-code research.
****
LeanDojo: Theorem Proving with Retrieval-Augmented Language Models
Kaiyu Yang, Aidan Swope, Alex Gu, Rahul Chalamala, Peiyang Song, Shixing Yu, Saad Godil, Ryan J Prenger, Animashree Anandkumar
Large language models (LLMs) have shown promise in proving formal theorems using proof assistants such as Lean. However, existing methods are difficult to reproduce or build on, due to private code, data, and large compute requirements. Thi

s has created substantial barriers to research on machine learning methods for theorem proving. This paper removes these barriers by introducing LeanDojo: an open-source Lean playground consisting of toolkits, data, models, and benchmarks. LeanDojo extracts data from Lean and enables interaction with the proof environment programmatically. It contains fine-grained annotations of premises in proofs, providing valuable data for premise selection—a key bottleneck in theorem proving. Using this data, we develop ReProver (Retrieval-Augmented Prover): an LLM-based prover augmented with retrieval for selecting premises from a vast math library. It is inexpensive and needs only one GPU week of training. Our retriever leverages LeanDojo's program analysis capability to identify accessible premises and hard negative examples, which makes retrieval much more effective. Furthermore, we construct a new benchmark consisting of 98,734 theorems and proofs extracted from Lean's math library. It features challenging data split requiring the prover to generalize to theorems relying on novel premises that are never used in training. We use this benchmark for training and evaluation, and experimental results demonstrate the effectiveness of ReProver over non-retrieval baselines and GPT-4. We thus provide the first set of open-source LLM-based theorem provers without any proprietary datasets and release it under a permissive MIT license to facilitate further research.
****
Cognitive Steering in Deep Neural Networks via Long-Range Modulatory Feedback Connections
Talia Konkle, George Alvarez
Given the rich visual information available in each glance, humans can internally direct their visual attention to enhance goal-relevant information---a capacity often absent in standard vision models. Here we introduce cognitively and biologically-inspired long-range modulatory pathways to enable `cognitive steering’ in vision models. First, we show that models equipped with these feedback pathways naturally show improved image recognition, adversarial robustness, and increased brain alignment, relative to baseline models. Further, these feedback projections from the final layer of the vision backbone provide a meaningful steering interface, where goals can be specified as vectors in the output space. We show that there are effective ways to steer the model that dramatically improve recognition of categories in composite images of multiple categories, succeeding where baseline feed-forward models without flexible steering fail. And, our multiplicative modulatory motif prevents rampant hallucination of the top-down goal category, dissociating what the model is looking for, from what it is looking at. Thus, these long-range modulatory pathways enable new behavioral capacities for goal-directed visual encoding, offering a flexible communication interface between cognitive and visual systems.
****
Neuro-symbolic Learning Yielding Logical Constraints
Zenan Li, Yunpeng Huang, Zhaoyu Li, Yuan Yao, Jingwei Xu, Taolue Chen, Xiaoxing Ma, Jian Lu
Neuro-symbolic systems combine the abilities of neural perception and logical reasoning. However, end-to-end learning of neuro-symbolic systems is still an unsolved challenge. This paper proposes a natural framework that fuses neural network training, symbol grounding, and logical constraint synthesis into a coherent and efficient end-to-end learning process. The capability of this framework comes from the improved interactions between the neural and the symbolic parts of the system in both the training and inference stages. Technically, to bridge the gap between the continuous neural network and the discrete logical constraint, we introduce a difference-of-convex programming technique to relax the logical constraints while maintaining their precision. We also employ cardinality constraints as the language for logical constraint learning and incorporate a trust region method to avoid the degeneracy of logical constraint in learning. Both theoretical analyses and empirical evaluations substantiate the effectiveness of the proposed framework.
****
Exploiting Connections between Lipschitz Structures for Certifiably Robust Deep

Equilibrium Models
Aaron Havens, Alexandre Araujo, Siddharth Garg, Farshad Khorrami, Bin Hu
Recently, deep equilibrium models (DEQs) have drawn increasing attention from th
e machine learning community. However, DEQs are much less understood in terms of
 certified robustness than their explicit network counterparts. In this paper, w
e advance the understanding of certified robustness of DEQs via exploiting the c
onnections between various Lipschitz network parameterizations for both explicit
 and implicit models. Importantly, we show that various popular Lipschitz networ
k structures, including convex potential layers (CPL), SDP-based Lipschitz layer
s (SLL), almost orthogonal layers (AOL), Sandwich layers, and monotone DEQs (Mon
DEQ) can all be reparameterized as special cases of the Lipschitz-bounded equili
brium networks (LBEN) without changing the prescribed Lipschitz constant in the
original network parameterization. A key feature of our reparameterization techn
ique is that it preserves the Lipschitz prescription used in different structure
s. This opens the possibility of achieving improved certified robustness of DEQs
 via a combination of network reparameterization, structure-preserving regulariz
ation, and LBEN-based fine-tuning. We also support our theoretical understanding
 with new empirical results, which show that our proposed method improves the ce
rtified robust accuracy of DEQs on classification tasks. All codes and experimen
ts are made available at \url{https://github.com/AaronHavens/ExploitingLipschitz
DEQ}.
****
A Combinatorial Algorithm for Approximating the Optimal Transport in the Paralle
l and MPC Settings
Nathaniel Lahn, Sharath Raghvendra, Kaiyi Zhang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
RegBN: Batch Normalization of Multimodal Data with Regularization
Morteza Ghahremani Boozandani, Christian Wachinger
Recent years have witnessed a surge of interest in integrating high-dimensional
data captured by multisource sensors, driven by the impressive success of neural
 networks in integrating multimodal data. However, the integration of heterogene
ous multimodal data poses a significant challenge, as confounding effects and de
pendencies among such heterogeneous data sources introduce unwanted variability
and bias, leading to suboptimal performance of multimodal models. Therefore, it
becomes crucial to normalize the low- or high-level features extracted from data
 modalities before their fusion takes place. This paper introduces RegBN, a nove
l approach for multimodal Batch Normalization with REGularization. RegBN uses th
e Frobenius norm as a regularizer term to address the side effects of confounder
s and underlying dependencies among different data sources. The proposed method
generalizes well across multiple modalities and eliminates the need for learnabl
e parameters, simplifying training and inference. We validate the effectiveness
of RegBN on eight databases from five research areas, encompassing diverse modal
ities such as language, audio, image, video, depth, tabular, and 3D MRI. The pro
posed method demonstrates broad applicability across different architectures suc
h as multilayer perceptrons, convolutional neural networks, and vision transform
ers, enabling effective normalization of both low- and high-level features in mu
ltimodal neural networks. RegBN is available at https://mogvision.github.io/RegB
N.
****
LLM-Pruner: On the Structural Pruning of Large Language Models
Xinyin Ma, Gongfan Fang, Xinchao Wang
Large language models (LLMs) have shown remarkable capabilities in language unde
rstanding and generation. However, such impressive capability typically comes wi
th a substantial model size, which presents significant challenges in both the d
eployment, inference, and training stages. With LLM being a general-purpose task
 solver, we explore its compression in a task-agnostic manner, which aims to pre

serve the multi-task solving and language generation ability of the original LLM. One challenge to achieving this is the enormous size of the training corpus of LLM, which makes both data transfer and model post-training over-burdensome. Thus, we tackle the compression of LLMs within the bound of two constraints: being task-agnostic and minimizing the reliance on the original training dataset. Our method, named LLM-pruner, adopts structural pruning that selectively removes non-critical coupled structures based on gradient information, maximally preserving the majority of the LLM's functionality. To this end, the performance of pruned models can be efficiently recovered through tuning techniques, LoRA, in merely 3 hours, requiring only 50K data. We validate the LLM-Pruner on three LLMs, including LLaMA, Vicuna, and ChatGLM, and demonstrate that the compressed models still exhibit satisfactory capabilities in zero-shot classification and generation. The code will be made public.
****
Nearly Optimal VC-Dimension and Pseudo-Dimension Bounds for Deep Neural Network Derivatives
Yahong Yang, Haizhao Yang, Yang Xiang
This paper addresses the problem of nearly optimal Vapnik--Chervonenkis dimension (VC-dimension) and pseudo-dimension estimations of the derivative functions of deep neural networks (DNNs). Two important applications of these estimations include: 1) Establishing a nearly tight approximation result of DNNs in the Sobolev space; 2) Characterizing the generalization error of machine learning methods with loss functions involving function derivatives. This theoretical investigation fills the gap of learning error estimations for a wide range of physics-informed machine learning models and applications including generative models, solving partial differential equations, operator learning, network compression, distillation, regularization, etc.
****
ClimateSet: A Large-Scale Climate Model Dataset for Machine Learning
Julia Kaltenborn, Charlotte Lange, Venkatesh Ramesh, Philippe Brouillard, Yaniv Gurwicz, Chandni Nagda, Jakob Runge, Peer Nowack, David Rolnick
Climate models have been key for assessing the impact of climate change and simulating future climate scenarios. The machine learning (ML) community has taken an increased interest in supporting climate scientists’ efforts on various tasks such as climate model emulation, downscaling, and prediction tasks. Many of those tasks have been addressed on datasets created with single climate models. However, both the climate science and ML communities have suggested that to address those tasks at scale, we need large, consistent, and ML-ready climate model datasets. Here, we introduce ClimateSet, a dataset containing the inputs and outputs of 36 climate models from the Input4MIPs and CMIP6 archives. In addition, we provide a modular dataset pipeline for retrieving and preprocessing additional climate models and scenarios. We showcase the potential of our dataset by using it as a benchmark for ML-based climate model emulation. We gain new insights about the performance and generalization capabilities of the different ML models by analyzing their performance across different climate models. Furthermore, the dataset can be used to train an ML emulator on several climate models instead of just one. Such a “super emulator” can quickly project new climate change scenarios, complementing existing scenarios already provided to policymakers. We believe ClimateSet will create the basis needed for the ML community to tackle climate-related tasks at scale.
****
Near-Optimal Bounds for Learning Gaussian Halfspaces with Random Classification Noise
Ilias Diakonikolas, Jelena Diakonikolas, Daniel Kane, Puqian Wang, Nikos Zarifis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Explain Any Concept: Segment Anything Meets Concept-Based Explanation

Ao Sun, Pingchuan Ma, Yuanyuan Yuan, Shuai Wang
EXplainable AI (XAI) is an essential topic to improve human understanding of deep neural networks (DNNs) given their black-box internals. For computer vision tasks, mainstream pixel-based XAI methods explain DNN decisions by identifying important pixels, and emerging concept-based XAI explore forming explanations with concepts (e.g., a head in an image). However, pixels are generally hard to interpret and sensitive to the imprecision of XAI methods, whereas “concepts” in prior works require human annotation or are limited to pre-defined concept sets. On the other hand, driven by large-scale pre-training, Segment Anything Model (SAM) has been demonstrated as a powerful and promotable framework for performing precise and comprehensive instance segmentation, enabling automatic preparation of concept sets from a given image. This paper for the first time explores using SAM to augment concept-based XAI. We offer an effective and flexible concept-based explanation method, namely Explain Any Concept (EAC), which explains DNN decisions with any concept. While SAM is highly effective and offers an “out-of-the-box” instance segmentation, it is costly when being integrated into defacto XAI pipelines. We thus propose a lightweight per-input equivalent (PIE) scheme, enabling efficient explanation with a surrogate model. Our evaluation over two popular datasets (ImageNet and COCO) illustrate the highly encouraging performance of EAC over commonly-used XAI methods.
****
Data-Driven Network Neuroscience: On Data Collection and Benchmark
Jiaxing Xu, Yunhan Yang, David Huang, Sophi Shilpa Gururajapathy, Yiping Ke, Miao Qiao, Alan Wang, Haribalan Kumar, Josh McGeown, Eryn Kwon
This paper presents a comprehensive and quality collection of functional human brain network data for potential research in the intersection of neuroscience, machine learning, and graph analytics. Anatomical and functional MRI images have been used to understand the functional connectivity of the human brain and are particularly important in identifying underlying neurodegenerative conditions such as Alzheimer's, Parkinson's, and Autism. Recently, the study of the brain in the form of brain networks using machine learning and graph analytics has become increasingly popular, especially to predict the early onset of these conditions. A brain network, represented as a graph, retains rich structural and positional information that traditional examination methods are unable to capture. However, the lack of publicly accessible brain network data prevents researchers from data-driven explorations. One of the main difficulties lies in the complicated domain-specific preprocessing steps and the exhaustive computation required to convert the data from MRI images into brain networks. We bridge this gap by collecting a large amount of MRI images from public databases and a private source, working with domain experts to make sensible design choices, and preprocessing the MRI images to produce a collection of brain network datasets. The datasets originate from 6 different sources, cover 4 brain conditions, and consist of a total of 2,702 subjects. We test our graph datasets on 12 machine learning models to provide baselines and validate the data quality on a recent graph analysis model. To lower the barrier to entry and promote the research in this interdisciplinary field, we release our brain network data and complete preprocessing details including codes at https://doi.org/10.17608/k6.auckland.21397377 and https://github.com/brainnetuoa/datadrivennetwork_neuroscience.
****
No-Regret Learning with Unbounded Losses: The Case of Logarithmic Pooling
Eric Neyman, Tim Roughgarden
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
PanoGen: Text-Conditioned Panoramic Environment Generation for Vision-and-Language Navigation
Jialu Li, Mohit Bansal
Vision-and-Language Navigation requires the agent to follow language instruction

s to navigate through 3D environments. One main challenge in Vision-and-Language Navigation is the limited availability of photorealistic training environments, which makes it hard to generalize to new and unseen environments. To address this problem, we propose PanoGen, a generation method that can potentially create an infinite number of diverse panoramic environments conditioned on text. Specifically, we collect room descriptions by captioning the room images in existing Matterport3D environments, and leverage a state-of-the-art text-to-image diffusion model to generate the new panoramic environments. We use recursive outpainting over the generated images to create consistent 360-degree panorama views. Our new panoramic environments share similar semantic information with the original environments by conditioning on text descriptions, which ensures the co-occurrence of objects in the panorama follows human intuition, and creates enough diversity in room appearance and layout with image outpainting. Lastly, we explore two ways of utilizing PanoGen in VLN pre-training and fine-tuning. We generate instructions for paths in our PanoGen environments with a speaker built on a pre-trained vision-and-language model for VLN pre-training, and augment the visual observation with our panoramic environments during agents' fine-tuning to avoid overfitting to seen environments. Empirically, learning with our PanoGen environments achieves the new state-of-the-art on the Room-to-Room, Room-for-Room, and CVDN datasets. Besides, we find that pre-training with our PanoGen speaker data is especially effective for CVDN, which has under-specified instructions and needs commonsense knowledge to reach the target. Lastly, we show that the agent can benefit from training with more generated panoramic environments, suggesting promising results for scaling up the PanoGen environments to enhance agents' generalization to unseen environments.
****

Scaling laws for language encoding models in fMRI
Richard Antonello, Aditya Vaidya, Alexander Huth
Representations from transformer-based unidirectional language models are known to be effective at predicting brain responses to natural language. However, most studies comparing language models to brains have used GPT-2 or similarly sized language models. Here we tested whether larger open-source models such as those from the OPT and LLaMA families are better at predicting brain responses recorded using fMRI. Mirroring scaling results from other contexts, we found that brain prediction performance scales logarithmically with model size from 125M to 30B parameter models, with ~15% increased encoding performance as measured by correlation with a held-out test set across 3 subjects. Similar log-linear behavior was observed when scaling the size of the fMRI training set. We also characterized scaling for acoustic encoding models that use HuBERT, WavLM, and Whisper, and we found comparable improvements with model size. A noise ceiling analysis of these large, high-performance encoding models showed that performance is nearing the theoretical maximum for brain areas such as the precuneus and higher auditory cortex. These results suggest that increasing scale in both models and data will yield incredibly effective models of language processing in the brain, enabling better scientific understanding as well as applications such as decoding.
****

Optimal Rates for Bandit Nonstochastic Control
Y. Jennifer Sun, Stephen Newman, Elad Hazan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Flow-Attention-based Spatio-Temporal Aggregation Network for 3D Mask Detection
Yuxin Cao, Yian Li, Yumeng Zhu, Derui Wang, Minhui Xue
Anti-spoofing detection has become a necessity for face recognition systems due to the security threat posed by spoofing attacks. Despite great success in traditional attacks, most deep-learning-based methods perform poorly in 3D masks, which can highly simulate real faces in appearance and structure, suffering generalizability insufficiency while focusing only on the spatial domain with single fr

ame input. This has been mitigated by the recent introduction of a biomedical te
chnology called rPPG (remote photoplethysmography). However, rPPG-based methods
are sensitive to noisy interference and require at least one second (> 25 frames
) of observation time, which induces high computational overhead. To address the
se challenges, we propose a novel 3D mask detection framework, called FASTEN (Fl
ow-Attention-based Spatio-Temporal aggrEgation Network). We tailor the network f
or focusing more on fine-grained details in large movements, which can eliminate
 redundant spatio-temporal feature interference and quickly capture splicing tra
ces of 3D masks in fewer frames. Our proposed network contains three key modules
: 1) a facial optical flow network to obtain non-RGB inter-frame flow informatio
n; 2) flow attention to assign different significance to each frame; 3) spatio-t
emporal aggregation to aggregate high-level spatial features and temporal transi
tion features. Through extensive experiments, FASTEN only requires five frames o
f input and outperforms eight competitors for both intra-dataset and cross-datas
et evaluations in terms of multiple detection metrics. Moreover, FASTEN has been
 deployed in real-world mobile devices for practical 3D mask detection.
****
On the Last-iterate Convergence in Time-varying Zero-sum Games: Extra Gradient S
ucceeds where Optimism Fails
Yi Feng, Hu Fu, Qun Hu, Ping Li, Ioannis Panageas, bo peng, Xiao Wang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Taking the neural sampling code very seriously: A data-driven approach for evalu
ating generative models of the visual system
Suhas Shrinivasan, Konstantin-Klemens Lurz, Kelli Restivo, George Denfield, Andr
eas Tolias, Edgar Walker, Fabian Sinz
Prevailing theories of perception hypothesize that the brain implements percepti
on via Bayesian inference in a generative model of the world.One prominent theor
y, the Neural Sampling Code (NSC), posits that neuronal responses to a stimulus
represent samples from the posterior distribution over latent world state variab
les that cause the stimulus.Although theoretically elegant, NSC does not specify
 the exact form of the generative model or prescribe how to link the theory to r
ecorded neuronal activity.Previous works assume simple generative models and tes
t their qualitative agreement with neurophysiological data.Currently, there is n
o precise alignment of the normative theory with neuronal recordings, especially
 in response to natural stimuli, and a quantitative, experimental evaluation of
models under NSC has been lacking.Here, we propose a novel formalization of NSC,
 that (a) allows us to directly fit NSC generative models to recorded neuronal a
ctivity in response to natural images, (b) formulate richer and more flexible ge
nerative models, and (c) employ standard metrics to quantitatively evaluate diff
erent generative models under NSC.Furthermore, we derive a stimulus-conditioned
predictive model of neuronal responses from the trained generative model using o
ur formalization that we compare to neural system identification models.We demon
strate our approach by fitting and comparing classical- and flexible deep learni
ng-based generative models on population recordings from the macaque primary vis
ual cortex (V1) to natural images, and show that the flexible models outperform
classical models in both their generative- and predictive-model performance.Over
all, our work is an important step towards a quantitative evaluation of NSC. It
provides a framework that lets us \textit{learn} the generative model directly f
rom neuronal population recordings, paving the way for an experimentally-informe
d understanding of probabilistic computational principles underlying perception
and behavior.
****
Can semi-supervised learning use all the data effectively? A lower bound perspec
tive
Alexandru Tifrea, Gizem Yüce, Amartya Sanyal, Fanny Yang
Prior theoretical and empirical works have established that semi-supervised lear

ning algorithms can leverage the unlabeled data to improve over the labeled sample complexity of supervised learning (SL) algorithms. However, existing theoretical work focuses on regimes where the unlabeled data is sufficient to learn a good decision boundary using unsupervised learning (UL) alone. This begs the question: Can SSL algorithms simultaneously improve upon both UL and SL? To this end, we derive a tight lower bound for 2-Gaussian mixture models that explicitly depends on the labeled and the unlabeled dataset size as well as the signal-to-noise ratio of the mixture distribution. Surprisingly, our result implies that no SSL algorithm improves upon the minimax-optimal statistical error rates of SL or UL algorithms for these distributions. Nevertheless, in our real-world experiments, SSL algorithms can often outperform UL and SL algorithms. In summary, our work suggests that while it is possible to prove the performance gains of SSL algorithms, this would require careful tracking of constants in the theoretical analysis.
****
Evolving Standardization for Continual Domain Generalization over Temporal Drift
Mixue Xie, Shuang Li, Longhui Yuan, Chi Liu, Zehui Dai
The capability of generalizing to out-of-distribution data is crucial for the deployment of machine learning models in the real world. Existing domain generalization (DG) mainly embarks on offline and discrete scenarios, where multiple source domains are simultaneously accessible and the distribution shift among domains is abrupt and violent. Nevertheless, such setting may not be universally applicable to all real-world applications, as there are cases where the data distribution gradually changes over time due to various factors, e.g., the process of aging. Additionally, as the domain constantly evolves, new domains will continually emerge. Re-training and updating models with both new and previous domains using existing DG methods can be resource-intensive and inefficient. Therefore, in this paper, we present a problem formulation for Continual Domain Generalization over Temporal Drift (CDGTD). CDGTD addresses the challenge of gradually shifting data distributions over time, where domains arrive sequentially and models can only access the data of the current domain. The goal is to generalize to unseen domains that are not too far into the future. To this end, we propose an Evolving Standardization (EvoS) method, which characterizes the evolving pattern of feature distribution and mitigates the distribution shift by standardizing features with generated statistics of corresponding domain. Specifically, inspired by the powerful ability of transformers to model sequence relations, we design a multi-scale attention module (MSAM) to learn the evolving pattern under sliding time windows of different lengths. MSAM can generate statistics of current domain based on the statistics of previous domains and the learned evolving pattern. Experiments on multiple real-world datasets including images and texts validate the efficacy of our EvoS.
****
Learning the Efficient Frontier
Philippe Chatigny, Ivan Sergienko, Ryan Ferguson, Jordan Weir, Maxime Bergeron
The efficient frontier (EF) is a fundamental resource allocation problem where one has to find an optimal portfolio maximizing a reward at a given level of risk. This optimal solution is traditionally found by solving a convex optimization problem. In this paper, we introduce NeuralEF: a fast neural approximation framework that robustly forecasts the result of the EF convex optimizations problems with respect to heterogeneous linear constraints and variable number of optimization inputs. By reformulating an optimization problem as a sequence to sequence problem, we show that NeuralEF is a viable solution to accelerate large-scale simulation while handling discontinuous behavior.
****
Dissecting Chain-of-Thought: Compositionality through In-Context Filtering and Learning
Yingcong Li, Kartik Sreenivasan, Angeliki Giannou, Dimitris Papailiopoulos, Samet Oymak
Chain-of-thought (CoT) is a method that enables language models to handle complex reasoning tasks by decomposing them into simpler steps. Despite its success, t

he underlying mechanics of CoT are not yet fully understood. In an attempt to sh
ed light on this, our study investigates the impact of CoT on the ability of tra
nsformers to in-context learn a simple to study, yet general family of compositi
onal functions: multi-layer perceptrons (MLPs). In this setting, we find that th
e success of CoT can be attributed to breaking down in-context learning of a com
positional function into two distinct phases: focusing on and filtering data rel
ated to each step of the composition and in-context learning the single-step com
position function. Through both experimental and theoretical evidence, we demons
trate how CoT significantly reduces the sample complexity of in-context learning
 (ICL) and facilitates the learning of complex functions that non-CoT methods st
ruggle with. Furthermore, we illustrate how transformers can transition from van
illa in-context learning to mastering a compositional function with CoT by simpl
y incorporating additional layers that perform the necessary data-filtering for
CoT via the attention mechanism. In addition to these test-time benefits, we sho
w CoT helps accelerate pretraining by learning shortcuts to represent complex fu
nctions and filtering plays an important role in this process. These findings co
llectively provide insights into the mechanics of CoT, inviting further investig
ation of its role in complex reasoning tasks.
****
Improving multimodal datasets with image captioning
Thao Nguyen, Samir Yitzhak Gadre, Gabriel Ilharco, Sewoong Oh, Ludwig Schmidt
Massive web datasets play a key role in the success of large vision-language mod
els like CLIP and Flamingo. However, the raw web data is noisy, and existing fil
tering methods to reduce noise often come at the expense of data diversity. Our
work focuses on caption quality as one major source of noise, and studies how ge
nerated captions can increase the utility of web-scraped datapoints with nondesc
ript text. Through exploring different mixing strategies for raw and generated c
aptions, we outperform the best filtering method proposed by the DataComp benchm
ark by 2% on ImageNet and 4% on average across 38 tasks, given a candidate pool
of 128M image-text pairs. Our best approach is also 2x better at Flickr and MS-C
OCO retrieval. We then analyze what makes synthetic captions an effective source
 of text supervision. In experimenting with different image captioning models, w
e also demonstrate that the performance of a model on standard image captioning
benchmarks (e.g., NoCaps CIDEr) is not a reliable indicator of the utility of th
e captions it generates for multimodal training. Finally, our experiments with u
sing generated captions at DataComp's large scale (1.28B image-text pairs) offer
 insights into the limitations of synthetic text, as well as the importance of i
mage curation with increasing training data quantity. The synthetic captions use
d in our experiments are now available on HuggingFace.
****
ClimSim: A large multi-scale dataset for hybrid physics-ML climate emulation
Sungduk Yu, Walter Hannah, Liran Peng, Jerry Lin, Mohamed Aziz Bhouri, Ritwik Gu
pta, Björn Lütjens, Justus C. Will, Gunnar Behrens, Julius Busecke, Nora Loose,
Charles Stern, Tom Beucler, Bryce Harrop, Benjamin Hillman, Andrea Jenney, Savan
nah L. Ferretti, Nana Liu, Animashree Anandkumar, Noah Brenowitz, Veronika Eyrin
g, Nicholas Geneva, Pierre Gentine, Stephan Mandt, Jaideep Pathak, Akshay Subram
aniam, Carl Vondrick, Rose Yu, Laure Zanna, Tian Zheng, Ryan Abernathey, Fiaz Ah
med, David Bader, Pierre Baldi, Elizabeth Barnes, Christopher Bretherton, Peter
Caldwell, Wayne Chuang, Yilun Han, YU HUANG, Fernando Iglesias-Suarez, Sanket Ja
ntre, Karthik Kashinath, Marat Khairoutdinov, Thorsten Kurth, Nicholas Lutsko, P
o-Lun Ma, Griffin Mooers, J. David Neelin, David Randall, Sara Shamekh, Mark Tay
lor, Nathan Urban, Janni Yuval, Guang Zhang, Mike Pritchard
Modern climate projections lack adequate spatial and temporal resolution due to
computational constraints. A consequence is inaccurate and imprecise predictions
 of critical processes such as storms. Hybrid methods that combine physics with
machine learning (ML) have introduced a new generation of higher fidelity climat
e simulators that can sidestep Moore's Law by outsourcing compute-hungry, short,
 high-resolution simulations to ML emulators. However, this hybrid ML-physics si
mulation approach requires domain-specific treatment and has been inaccessible t
o ML experts because of lack of training data and relevant, easy-to-use workflow

s. We present ClimSim, the largest-ever dataset designed for hybrid ML-physics research. It comprises multi-scale climate simulations, developed by a consortium of climate scientists and ML researchers. It consists of 5.7 billion pairs of multivariate input and output vectors that isolate the influence of locally-nested, high-resolution, high-fidelity physics on a host climate simulator's macro-scale physical state.The dataset is global in coverage, spans multiple years at high sampling frequency, and is designed such that resulting emulators are compatible with downstream coupling into operational climate simulators. We implement a range of deterministic and stochastic regression baselines to highlight the ML challenges and their scoring. The data (https://huggingface.co/datasets/LEAP/ClimSim_high-res) and code (https://leap-stc.github.io/ClimSim) are released openly to support the development of hybrid ML-physics and high-fidelity climate simulations for the benefit of science and society.
****
Relative Entropic Optimal Transport: a (Prior-aware) Matching Perspective to (Unbalanced) Classification
Liangliang Shi, Haoyu Zhen, Gu Zhang, Junchi Yan
Classification is a fundamental problem in machine learning, and considerable efforts have been recently devoted to the demanding long-tailed setting due to its prevalence in nature. Departure from the Bayesian framework, this paper rethinks classification from a matching perspective by studying the matching probability between samples and labels with optimal transport (OT) formulation. Specifically, we first propose a new variant of optimal transport, called Relative Entropic Optimal Transport (RE-OT), which guides the coupling solution to a known prior information matrix. We gives some theoretical results and their proof for RE-OT and surprisingly find RE-OT can help to deblur for barycenter images. Then we adopt inverse RE-OT for training long-tailed data and find that the loss derived from RE-OT has a similar form to Softmax-based cross-entropy loss, indicating a close connection between optimal transport and classification and the potential for transferring concepts between these two academic fields, such as barycentric projection in OT, which can map the labels back to the feature space. We further derive an epoch-varying RE-OT loss, and do the experiments on unbalanced image classification, molecule classification, instance segmentation and representation learning. Experimental results show its effectiveness.
****
Connecting Multi-modal Contrastive Representations
Zehan Wang, Yang Zhao, Xize ■, Haifeng Huang, Jiageng Liu, Aoxiong Yin, Li Tang, Linjun Li, Yongqi Wang, Ziang Zhang, Zhou Zhao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Boosting Learning for LDPC Codes to Improve the Error-Floor Performance
Hee-Youl Kwak, Dae-Young Yun, Yongjune Kim, Sang-Hyo Kim, Jong-Seon No
Low-density parity-check (LDPC) codes have been successfully commercialized in communication systems due to their strong error correction capabilities and simple decoding process. However, the error-floor phenomenon of LDPC codes, in which the error rate stops decreasing rapidly at a certain level, presents challenges for achieving extremely low error rates and deploying LDPC codes in scenarios demanding ultra-high reliability. In this work, we propose training methods for neural min-sum (NMS) decoders to eliminate the error-floor effect. First, by leveraging the boosting learning technique of ensemble networks, we divide the decoding network into two neural decoders and train the post decoder to be specialized for uncorrected words that the first decoder fails to correct. Secondly, to address the vanishing gradient issue in training, we introduce a block-wise training schedule that locally trains a block of weights while retraining the preceding block. Lastly, we show that assigning different weights to unsatisfied check nodes effectively lowers the error-floor with a minimal number of weights. By applying these training methods to standard LDPC codes, we achieve the best error-fl

oor performance compared to other decoding methods. The proposed NMS decoder, optimized solely through novel training methods without additional modules, can be integrated into existing LDPC decoders without incurring extra hardware costs. The source code is available at https://github.com/ghy1228/LDPCErrorFloor.
****
Learning Score-based Grasping Primitive for Human-assisting Dexterous Grasping
Tianhao Wu, Mingdong Wu, Jiyao Zhang, Yunchong Gan, Hao Dong
The use of anthropomorphic robotic hands for assisting individuals in situations where human hands may be unavailable or unsuitable has gained significant importance. In this paper, we propose a novel task called human-assisting dexterous grasping that aims to train a policy for controlling a robotic hand's fingers to assist users in grasping objects. Unlike conventional dexterous grasping, this task presents a more complex challenge as the policy needs to adapt to diverse user intentions, in addition to the object's geometry. We address this challenge by proposing an approach consisting of two sub-modules: a hand-object-conditional grasping primitive called Grasping Gradient Field (GraspGF), and a history-conditional residual policy. GraspGF learns 'how' to grasp by estimating the gradient of a synthesised success grasping example set, while the residual policy determines 'when' and at what speed the grasping action should be executed based on the trajectory history. Experimental results demonstrate the superiority of our proposed method compared to baselines, highlighting the user-awareness and practicality in real-world applications. The codes and demonstrations can be viewed at https://sites.google.com/view/graspgf.
****
Maximize to Explore: One Objective Function Fusing Estimation, Planning, and Exploration
Zhihan Liu, Miao Lu, WEI XIONG, Han Zhong, Hao Hu, Shenao Zhang, Sirui Zheng, Zhuoran Yang, Zhaoran Wang
In reinforcement learning (RL), balancing exploration and exploitation is crucial for achieving an optimal policy in a sample-efficient way. To this end, existing sample- efficient algorithms typically consist of three components: estimation, planning, and exploration. However, to cope with general function approximators, most of them involve impractical algorithmic components to incentivize exploration, such as data-dependent level-set constraints or complicated sampling procedures. To address this challenge, we propose an easy-to-implement RL framework called Maximize to Explore (MEX), which only needs to optimize unconstrainedly a single objective that integrates the estimation and planning components while balancing exploration and exploitation automatically. Theoretically, we prove that the MEX achieves a sublinear regret with general function approximators and is extendable to the zero-sum Markov game setting. Meanwhile, we adapt deep RL baselines to design practical versions of MEX in both the model-based and model-free settings, which outperform baselines in various MuJoCo environments with sparse reward by a stable margin. Compared with existing sample-efficient algorithms with general function approximators, MEX achieves similar sample efficiency while also enjoying a lower computational cost and is more compatible with modern deep RL methods.
****
Hokoff: Real Game Dataset from Honor of Kings and its Offline Reinforcement Learning Benchmarks
Yun Qu, Boyuan Wang, Jianzhun Shao, Yuhang Jiang, Chen Chen, Zhenbin Ye, Liu Linc, Yang Feng, Lin Lai, Hongyang Qin, Minwen Deng, Juchao Zhuo, Deheng Ye, Qiang Fu, YANG GUANG, Wei Yang, Lanxiao Huang, Xiangyang Ji
The advancement of Offline Reinforcement Learning (RL) and Offline Multi-Agent Reinforcement Learning (MARL) critically depends on the availability of high-quality, pre-collected offline datasets that represent real-world complexities and practical applications. However, existing datasets often fall short in their simplicity and lack of realism. To address this gap, we propose Hokoff, a comprehensive set of pre-collected datasets that covers both offline RL and offline MARL, accompanied by a robust framework, to facilitate further research. This data is derived from Honor of Kings, a recognized Multiplayer Online Battle Arena (MOBA)

game known for its intricate nature, closely resembling real-life situations. Utilizing this framework, we benchmark a variety of offline RL and offline MARL algorithms. We also introduce a novel baseline algorithm tailored for the inherent hierarchical action space of the game. We reveal the incompetency of current offline RL approaches in handling task complexity, generalization and multi-task learning.
****
Learning and Collusion in Multi-unit Auctions
Simina Branzei, Mahsa Derakhshan, Negin Golrezaei, Yanjun Han
In a carbon auction, licenses for CO2 emissions are allocated among multiple interested players. Inspired by this setting, we consider repeated multi-unit auctions with uniform pricing, which are widely used in practice. Our contribution is to analyze these auctions in both the offline and online settings, by designing efficient bidding algorithms with low regret and giving regret lower bounds. We also analyze the quality of the equilibria in two main variants of the auction, finding that one variant is susceptible to collusion among the bidders while the other is not.
****
One-2-3-45: Any Single Image to 3D Mesh in 45 Seconds without Per-Shape Optimization
Minghua Liu, Chao Xu, Haian Jin, Linghao Chen, Mukund Varma T, Zexiang Xu, Hao Su
Single image 3D reconstruction is an important but challenging task that requires extensive knowledge of our natural world. Many existing methods solve this problem by optimizing a neural radiance field under the guidance of 2D diffusion models but suffer from lengthy optimization time, 3D inconsistency results, and poor geometry. In this work, we propose a novel method that takes a single image of any object as input and generates a full 360-degree 3D textured mesh in a single feed-forward pass. Given a single image, we first use a view-conditioned 2D diffusion model, Zero123, to generate multi-view images for the input view, and then aim to lift them up to 3D space. Since traditional reconstruction methods struggle with inconsistent multi-view predictions, we build our 3D reconstruction module upon an SDF-based generalizable neural surface reconstruction method and propose several critical training strategies to enable the reconstruction of 360-degree meshes. Without costly optimizations, our method reconstructs 3D shapes in significantly less time than existing methods. Moreover, our method favors better geometry, generates more 3D consistent results, and adheres more closely to the input image. We evaluate our approach on both synthetic data and in-the-wild images and demonstrate its superiority in terms of both mesh quality and runtime. In addition, our approach can seamlessly support the text-to-3D task by integrating with off-the-shelf text-to-image diffusion models.
****
VeriX: Towards Verified Explainability of Deep Neural Networks
Min Wu, Haoze Wu, Clark Barrett
We present VeriX (Verified eXplainability), a system for producing optimal robust explanations and generating counterfactuals along decision boundaries of machine learning models. We build such explanations and counterfactuals iteratively using constraint solving techniques and a heuristic based on feature-level sensitivity ranking. We evaluate our method on image recognition benchmarks and a real-world scenario of autonomous aircraft taxiing.
****
Generalized test utilities for long-tail performance in extreme multi-label classification
Erik Schultheis, Marek Wydmuch, Wojciech Kotlowski, Rohit Babbar, Krzysztof Dembczynski
Extreme multi-label classification (XMLC) is the task of selecting a small subset of relevant labels from a very large set of possible labels. As such, it is characterized by long-tail labels, i.e., most labels have very few positive instances. With standard performance measures such as precision@k, a classifier can ignore tail labels and still report good performance. However, it is often argued

that correct predictions in the tail are more "interesting" or "rewarding," but the community has not yet settled on a metric capturing this intuitive concept. The existing propensity-scored metrics fall short on this goal by confounding the problems of long-tail and missing labels. In this paper, we analyze generalized metrics budgeted "at k" as an alternative solution. To tackle the challenging problem of optimizing these metrics, we formulate it in the expected test utility (ETU) framework, which aims to optimize the expected performance on a given test set. We derive optimal prediction rules and construct their computationally efficient approximations with provable regret guarantees and being robust against model misspecification. Our algorithm, based on block coordinate descent, scales effortlessly to XMLC problems and obtains promising results in terms of long-tail performance.
****

Compositional Foundation Models for Hierarchical Planning

Anurag Ajay, Seungwook Han, Yilun Du, Shuang Li, Abhi Gupta, Tommi Jaakkola, Josh Tenenbaum, Leslie Kaelbling, Akash Srivastava, Pulkit Agrawal

To make effective decisions in novel environments with long-horizon goals, it is crucial to engage in hierarchical reasoning across spatial and temporal scales. This entails planning abstract subgoal sequences, visually reasoning about the underlying plans, and executing actions in accordance with the devised plan through visual-motor control. We propose Compositional Foundation Models for Hierarchical Planning (HiP), a foundation model which leverages multiple expert foundation model trained on language, vision and action data individually jointly together to solve long-horizon tasks. We use a large language model to construct symbolic plans that are grounded in the environment through a large video diffusion model. Generated video plans are then grounded to visual-motor control, through an inverse dynamics model that infers actions from generated videos. To enable effective reasoning within this hierarchy, we enforce consistency between the models via iterative refinement. We illustrate the efficacy and adaptability of our approach in three different long-horizon table-top manipulation tasks.
****

Diffusion Model for Graph Inverse Problems: Towards Effective Source Localization on Complex Networks

Xin Yan, Hui Fang, Qiang He

Information diffusion problems, such as the spread of epidemics or rumors, are widespread in society. The inverse problems of graph diffusion, which involve locating the sources and identifying the paths of diffusion based on currently observed diffusion graphs, are crucial to controlling the spread of information. The problem of localizing the source of diffusion is highly ill-posed, presenting a major obstacle in accurately assessing the uncertainty involved. Besides, while comprehending how information diffuses through a graph is crucial, there is a scarcity of research on reconstructing the paths of information propagation. To tackle these challenges, we propose a probabilistic model called DDMSL (Discrete Diffusion Model for Source Localization). Our approach is based on the natural diffusion process of information propagation over complex networks, which can be formulated using a message-passing function. First, we model the forward diffusion of information using Markov chains. Then, we design a reversible residual network to construct a denoising-diffusion model in discrete space for both source localization and reconstruction of information diffusion paths. We provide rigorous theoretical guarantees for DDMSL and demonstrate its effectiveness through extensive experiments on five real-world datasets.
****

UniT: A Unified Look at Certified Robust Training against Text Adversarial Perturbation

Muchao Ye, Ziyi Yin, Tianrong Zhang, Tianyu Du, Jinghui Chen, Ting Wang, Fenglong Ma

Recent years have witnessed a surge of certified robust training pipelines against text adversarial perturbation constructed by synonym substitutions. Given a base model, existing pipelines provide prediction certificates either in the discrete word space or the continuous latent space. However, they are isolated from

each other with a structural gap. We observe that existing training frameworks need unification to provide stronger certified robustness. Additionally, they mainly focus on building the certification process but neglect to improve the robustness of the base model. To mitigate the aforementioned limitations, we propose a unified framework named UniT that enables us to train flexibly in either fashion by working in the word embedding space. It can provide a stronger robustness guarantee obtained directly from the word embedding space without extra modules. In addition, we introduce the decoupled regularization (DR) loss to improve the robustness of the base model, which includes two separate robustness regularization terms for the feature extraction and classifier modules. Experimental results on widely used text classification datasets further demonstrate the effectiveness of the designed unified framework and the proposed DR loss for improving the certified robust accuracy.
****
Convergence of Alternating Gradient Descent for Matrix Factorization
Rachel Ward, Tamara Kolda
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SPRING: Studying Papers and Reasoning to play Games
Yue Wu, So Yeon Min, Shrimai Prabhumoye, Yonatan Bisk, Russ R. Salakhutdinov, Amos Azaria, Tom M. Mitchell, Yuanzhi Li
Open-world survival games pose significant challenges for AI algorithms due to their multi-tasking, deep exploration, and goal prioritization requirements. Despite reinforcement learning (RL) being popular for solving games, its high sample complexity limits its effectiveness in complex open-world games like Crafter or Minecraft. We propose a novel approach, SPRING, to read Crafter's original academic paper and use the knowledge learned to reason and play the game through a large language model (LLM).Prompted with the LaTeX source as game context and a description of the agent's current observation, our SPRING framework employs a directed acyclic graph (DAG) with game-related questions as nodes and dependencies as edges. We identify the optimal action to take in the environment by traversing the DAG and calculating LLM responses for each node in topological order, with the LLM's answer to final node directly translating to environment actions.In our experiments, we study the quality of in-context "reasoning" induced by different forms of prompts under the setting of the Crafter environment. Our experiments suggest that LLMs, when prompted with consistent chain-of-thought, have great potential in completing sophisticated high-level trajectories. Quantitatively, SPRING with GPT-4 outperforms all state-of-the-art RL baselines, trained for 1M steps, without any training. Finally, we show the potential of Crafter as a test bed for LLMs. Code at github.com/holmeswww/SPRING
****
Hybrid Search for Efficient Planning with Completeness Guarantees
Kalle Kujanpää, Joni Pajarinen, Alexander Ilin
Solving complex planning problems has been a long-standing challenge in computer science. Learning-based subgoal search methods have shown promise in tackling these problems, but they often suffer from a lack of completeness guarantees, meaning that they may fail to find a solution even if one exists. In this paper, we propose an efficient approach to augment a subgoal search method to achieve completeness in discrete action spaces. Specifically, we augment the high-level search with low-level actions to execute a multi-level (hybrid) search, which we call complete subgoal search. This solution achieves the best of both worlds: the practical efficiency of high-level search and the completeness of low-level search. We apply the proposed search method to a recently proposed subgoal search algorithm and evaluate the algorithm trained on offline data on complex planning problems. We demonstrate that our complete subgoal search not only guarantees completeness but can even improve performance in terms of search expansions for instances that the high-level could solve without low-level augmentations. Our appr

oach makes it possible to apply subgoal-level planning for systems where complet
eness is a critical requirement.
****
Diversified Outlier Exposure for Out-of-Distribution Detection via Informative E
xtrapolation
Jianing Zhu, Yu Geng, Jiangchao Yao, Tongliang Liu, Gang Niu, Masashi Sugiyama,
Bo Han
Out-of-distribution (OOD) detection is important for deploying reliable machine learning models on real-world applications. Recent advances in outlier exposure have shown promising results on OOD detection via fine-tuning model with informatively sampled auxiliary outliers. However, previous methods assume that the collected outliers can be sufficiently large and representative to cover the boundary between ID and OOD data, which might be impractical and challenging. In this work, we propose a novel framework, namely, Diversified Outlier Exposure (DivOE), for effective OOD detection via informative extrapolation based on the given auxiliary outliers. Specifically, DivOE introduces a new learning objective, which diversifies the auxiliary distribution by explicitly synthesizing more informative outliers for extrapolation during training. It leverages a multi-step optimization method to generate novel outliers beyond the original ones, which is compatible with many variants of outlier exposure. Extensive experiments and analyses have been conducted to characterize and demonstrate the effectiveness of the proposed DivOE. The code is publicly available at: https://github.com/tmlr-group/DivOE.
****
Attacks on Online Learners: a Teacher-Student Analysis
Riccardo Giuseppe Margiotta, Sebastian Goldt, Guido Sanguinetti
Machine learning models are famously vulnerable to adversarial attacks: small ad-hoc perturbations of the data that can catastrophically alter the model predictions. While a large literature has studied the case of test-time attacks on pre-trained models, the important case of attacks in an online learning setting has received little attention so far. In this work, we use a control-theoretical perspective to study the scenario where an attacker may perturb data labels to manipulate the learning dynamics of an online learner. We perform a theoretical analysis of the problem in a teacher-student setup, considering different attack strategies, and obtaining analytical results for the steady state of simple linear learners. These results enable us to prove that a discontinuous transition in the learner's accuracy occurs when the attack strength exceeds a critical threshold. We then study empirically attacks on learners with complex architectures using real data, confirming the insights of our theoretical analysis. Our findings show that greedy attacks can be extremely efficient, especially when data stream in small batches.
****
Delayed Algorithms for Distributed Stochastic Weakly Convex Optimization
Wenzhi Gao, Qi Deng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Grounding Neural Inference with Satisfiability Modulo Theories
Zifan Wang, Saranya Vijayakumar, Kaiji Lu, Vijay Ganesh, Somesh Jha, Matt Fredrikson
Recent techniques that integrate solver layers into Deep Neural Networks (DNNs) have shown promise in bridging a long-standing gap between inductive learning and symbolic reasoning techniques. In this paper we present a set of techniques for integrating Satisfiability Modulo Theories (SMT) solvers into the forward and backward passes of a deep network layer, called SMTLayer.Using this approach, one can encode rich domain knowledge into the network in the form of mathematical formulas.In the forward pass, the solver uses symbols produced by prior layers, along with these formulas, to construct inferences; in the backward pass, the so

lver informs updates to the network, driving it towards representations that are compatible with the solver's theory.Notably, the solver need not be differentiable. We implement SMTLayer as a Pytorch module, and our empirical results show that it leads to models that 1) require fewer training samples than conventional models, 2) that are robust to certain types of covariate shift, and 3) that ultimately learn representations that are consistent with symbolic knowledge, and thus naturally interpretable.
****
D^{2}CSG: Unsupervised Learning of Compact CSG Trees with Dual Complements and Dropouts
Fenggen Yu, Qimin Chen, Maham Tanveer, Ali Mahdavi Amiri, Hao Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Fine-grained Late-interaction Multi-modal Retrieval for Retrieval Augmented Visual Question Answering
Weizhe Lin, Jinghong Chen, Jingbiao Mei, Alexandru Coca, Bill Byrne
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Iteratively Learn Diverse Strategies with State Distance Information
Wei Fu, Weihua Du, Jingwei Li, Sunli Chen, Jingzhao Zhang, YI WU
In complex reinforcement learning (RL) problems, policies with similar rewards may have substantially different behaviors. It remains a fundamental challenge to optimize rewards while also discovering as many diverse strategies as possible, which can be crucial in many practical applications. Our study examines two design choices for tackling this challenge, i.e., diversity measure and computation framework. First, we find that with existing diversity measures, visually indistinguishable policies can still yield high diversity scores. To accurately capture the behavioral difference, we propose to incorporate the state-space distance information into the diversity measure. In addition, we examine two common computation frameworks for this problem, i.e., population-based training (PBT) and iterative learning (ITR). We show that although PBT is the precise problem formulation, ITR can achieve comparable diversity scores with higher computation efficiency, leading to improved solution quality in practice. Based on our analysis, we further combine ITR with two tractable realizations of the state-distance-based diversity measures and develop a novel diversity-driven RL algorithm, State-based Intrinsic-reward Policy Optimization (SIPO), with provable convergence properties. We empirically examine SIPO across three domains from robot locomotion to multi-agent games. In all of our testing environments, SIPO consistently produces strategically diverse and human-interpretable policies that cannot be discovered by existing baselines.
****
Neural Fields with Hard Constraints of Arbitrary Differential Order
Fangcheng Zhong, Kyle Fogarty, Param Hanji, Tianhao Wu, Alejandro Sztrajman, Andrew Spielberg, Andrea Tagliasacchi, Petra Bosilj, Cengiz Oztireli
While deep learning techniques have become extremely popular for solving a broad range of optimization problems, methods to enforce hard constraints during optimization, particularly on deep neural networks, remain underdeveloped. Inspired by the rich literature on meshless interpolation and its extension to spectral collocation methods in scientific computing, we develop a series of approaches for enforcing hard constraints on neural fields, which we refer to as Constrained Neural Fields (CNF). The constraints can be specified as a linear operator applied to the neural field and its derivatives. We also design specific model representations and training strategies for problems where standard models may encounter difficulties, such as conditioning of the system, memory consumption, and cap

acity of the network when being constrained. Our approaches are demonstrated in a wide range of real-world applications. Additionally, we develop a framework that enables highly efficient model and constraint specification, which can be readily applied to any downstream task where hard constraints need to be explicitly satisfied during optimization.
****
Thinker: Learning to Plan and Act
Stephen Chung, Ivan Anokhin, David Krueger
We propose the Thinker algorithm, a novel approach that enables reinforcement learning agents to autonomously interact with and utilize a learned world model. The Thinker algorithm wraps the environment with a world model and introduces new actions designed for interacting with the world model. These model-interaction actions enable agents to perform planning by proposing alternative plans to the world model before selecting a final action to execute in the environment. This approach eliminates the need for handcrafted planning algorithms by enabling the agent to learn how to plan autonomously and allows for easy interpretation of the agent's plan with visualization. We demonstrate the algorithm's effectiveness through experimental results in the game of Sokoban and the Atari 2600 benchmark, where the Thinker algorithm achieves state-of-the-art performance and competitive results, respectively. Visualizations of agents trained with the Thinker algorithm demonstrate that they have learned to plan effectively with the world model to select better actions. Thinker is the first work showing that an RL agent can learn to plan with a learned world model in complex environments.
****
Near-Optimal k-Clustering in the Sliding Window Model
David Woodruff, Peilin Zhong, Samson Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SynMob: Creating High-Fidelity Synthetic GPS Trajectory Dataset for Urban Mobility Analysis
Yuanshao Zhu, Yongchao Ye, Ying Wu, Xiangyu Zhao, James Yu
Urban mobility analysis has been extensively studied in the past decade using a vast amount of GPS trajectory data, which reveals hidden patterns in movement and human activity within urban landscapes. Despite its significant value, the availability of such datasets often faces limitations due to privacy concerns, proprietary barriers, and quality inconsistencies. To address these challenges, this paper presents a synthetic trajectory dataset with high fidelity, offering a general solution to these data accessibility issues. Specifically, the proposed dataset adopts a diffusion model as its synthesizer, with the primary aim of accurately emulating the spatial-temporal behavior of the original trajectory data. These synthesized data can retain the geo-distribution and statistical properties characteristic of real-world datasets. Through rigorous analysis and case studies, we validate the high similarity and utility between the proposed synthetic trajectory dataset and real-world counterparts. Such validation underscores the practicality of synthetic datasets for urban mobility analysis and advocates for its wider acceptance within the research community. Finally, we publicly release the trajectory synthesizer and datasets, aiming to enhance the quality and availability of synthetic trajectory datasets and encourage continued contributions to this rapidly evolving field. The dataset is released for public online availability https://github.com/Applied-Machine-Learning-Lab/SynMob.
****
Window-Based Distribution Shift Detection for Deep Neural Networks
Guy Bar-Shalom, Yonatan Geifman, Ran El-Yaniv
To deploy and operate deep neural models in production, the quality of their predictions, which might be contaminated benignly or manipulated maliciously by input distributional deviations, must be monitored and assessed. Specifically, we study the case of monitoring the healthy operation of a deep neural network (DNN)

receiving a stream of data, with the aim of detecting input distributional deviations over which the quality of the network's predictions is potentially damaged. Using selective prediction principles, we propose a distribution deviation detection method for DNNs. The proposed method is derived from a tight coverage generalization bound computed over a sample of instances drawn from the true underlying distribution. Based on this bound, our detector continuously monitors the operation of the network over a test window and fires off an alarm whenever a deviation is detected. Our novel detection method performs on-par or better than the state-of-the-art, while consuming substantially lower computation time (five orders of magnitude reduction) and space complexity. Unlike previous methods, which require at least linear dependence on the size of the source distribution for each detection, rendering them inapplicable to ``Google-Scale'' datasets, our approach eliminates this dependence, making it suitable for real-world applications. Code is available at https://github.com/BarSGuy/Window-Based-Distribution-Shift-Detection.
****
Towards Label Position Bias in Graph Neural Networks
Haoyu Han, Xiaorui Liu, Feng Shi, MohamadAli Torkamani, Charu Aggarwal, Jiliang Tang
Graph Neural Networks (GNNs) have emerged as a powerful tool for semi-supervised node classification tasks. However, recent studies have revealed various biases in GNNs stemming from both node features and graph topology. In this work, we uncover a new bias - label position bias, which indicates that the node closer to the labeled nodes tends to perform better. We introduce a new metric, the Label Proximity Score, to quantify this bias, and find that it is closely related to performance disparities. To address the label position bias, we propose a novel optimization framework for learning a label position unbiased graph structure, which can be applied to existing GNNs. Extensive experiments demonstrate that our proposed method not only outperforms backbone methods but also significantly mitigates the issue of label position bias in GNNs.
****
Label Robust and Differentially Private Linear Regression: Computational and Statistical Efficiency
Xiyang Liu, Prateek Jain, Weihao Kong, Sewoong Oh, Arun Suggala
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Explainable and Efficient Randomized Voting Rules
Soroush Ebadian, Aris Filos-Ratsikas, Mohamad Latifian, Nisarg Shah
With a rapid growth in the deployment of AI tools for making critical decisions (or aiding humans in doing so), there is a growing demand to be able to explain to the stakeholders how these tools arrive at a decision. Consequently, voting is frequently used to make such decisions due to its inherent explainability. Recent work suggests that using randomized (as opposed to deterministic) voting rules can lead to significant efficiency gains measured via the distortion framework. However, rules that use intricate randomization can often become too complex to explain to the stakeholders; losing explainability can eliminate the key advantage of voting over black-box AI tools, which may outweigh the efficiency gains.We study the efficiency gains which can be unlocked by using voting rules that add a simple randomization step to a deterministic rule, thereby retaining explainability. We focus on two such families of rules, randomized positional scoring rules and random committee member rules, and show, theoretically and empirically, that they indeed achieve explainability and efficiency simultaneously to some extent.
****
Conformal PID Control for Time Series Prediction
Anastasios Angelopoulos, Emmanuel Candes, Ryan J. Tibshirani
We study the problem of uncertainty quantification for time series prediction, w

ith the goal of providing easy-to-use algorithms with formal guarantees. The al
gorithms we present build upon ideas from conformal prediction and control theor
y, are able to prospectively model conformal scores in an online setting, and ad
apt to the presence of systematic errors due to seasonality, trends, and general
 distribution shifts. Our theory both simplifies and strengthens existing analys
es in online conformal prediction. Experiments on 4-week-ahead forecasting of st
atewide COVID-19 death counts in the U.S. show an improvement in coverage over t
he ensemble forecaster used inofficial CDC communications. We also run experimen
ts on predicting electricity demand, market returns, and temperature using autor
egressive, Theta, Prophet, and Transformer models. We provide an extendable code
base for testing our methods and for the integration of new algorithms, data set
s, and forecasting rules at this link.
****
LLMScore: Unveiling the Power of Large Language Models in Text-to-Image Synthesi
s Evaluation
Yujie Lu, Xianjun Yang, Xiujun Li, Xin Eric Wang, William Yang Wang
Existing automatic evaluation on text-to-image synthesis can only provide an ima
ge-text matching score, without considering the object-level compositionality, w
hich results in poor correlation with human judgments. In this work, we propose
LLMScore, a new framework that offers evaluation scores with multi-granularity c
ompositionality. LLMScore leverages the large language models (LLMs) to evaluate
 text-to-image models. Initially, it transforms the image into image-level and o
bject-level visual descriptions. Then an evaluation instruction is fed into the
LLMs to measure the alignment between the synthesized image and the text, ultima
tely generating a score accompanied by a rationale. Our substantial analysis rev
eals the highest correlation of LLMScore with human judgments on a wide range of
 datasets (Attribute Binding Contrast, Concept Conjunction, MSCOCO, DrawBench, P
aintSkills). Notably, our LLMScore achieves Kendall's tau correlation with human
 evaluations that is 58.8% and 31.2% higher than the commonly-used text-image ma
tching metrics CLIP and BLIP, respectively.
****
Dynamically Masked Discriminator for GANs
Wentian Zhang, Haozhe Liu, Bing Li, Jinheng Xie, Yawen Huang, Yuexiang Li, Yefen
g Zheng, Bernard Ghanem
Training Generative Adversarial Networks (GANs) remains a challenging problem. T
he discriminator trains the generator by learning the distribution of real/gener
ated data. However, the distribution of generated data changes throughout the tr
aining process, which is difficult for the discriminator to learn. In this paper
, we propose a novel method for GANs from the viewpoint of online continual lear
ning. We observe that the discriminator model, trained on historically generated
 data, often slows down its adaptation to the changes in the new arrival generat
ed data, which accordingly decreases the quality of generated results. By treati
ng the generated data in training as a stream, we propose to detect whether the
discriminator slows down the learning of new knowledge in generated data. Theref
ore, we can explicitly enforce the discriminator to learn new knowledge fast. Pa
rticularly, we propose a new discriminator, which automatically detects its reta
rdation and then dynamically masks its features, such that the discriminator can
 adaptively learn the temporally-vary distribution of generated data. Experiment
al results show our method outperforms the state-of-the-art approaches.
****
Diverse Conventions for Human-AI Collaboration
Bidipta Sarkar, Andy Shih, Dorsa Sadigh
Conventions are crucial for strong performance in cooperative multi-agent games,
 because they allow players to coordinate on a shared strategy without explicit
communication. Unfortunately, standard multi-agent reinforcement learning techni
ques, such as self-play, converge to conventions that are arbitrary and non-dive
rse, leading to poor generalization when interacting with new partners. In this
work, we present a technique for generating diverse conventions by (1) maximizin
g their rewards during self-play, while (2) minimizing their rewards when playin
g with previously discovered conventions (cross-play), stimulating conventions t

o be semantically different. To ensure that learned policies act in good faith d
espite the adversarial optimization of cross-play, we introduce mixed-play, wher
e an initial state is randomly generated by sampling self-play and cross-play tr
ansitions and the player learns to maximize the self-play reward from this initi
al state. We analyze the benefits of our technique on various multi-agent collab
orative games, including Overcooked, and find that our technique can adapt to th
e conventions of humans, surpassing human-level performance when paired with rea
l users.
****
Self-Supervised Learning of Representations for Space Generates Multi-Modular Gr
id Cells
Rylan Schaeffer, Mikail Khona, Tzuhsuan Ma, Cristobal Eyzaguirre, Sanmi Koyejo,
Ila Fiete
To solve the spatial problems of mapping, localization and navigation, the mamma
lian lineage has developed striking spatial representations. One important spati
al representation is the Nobel-prize winning grid cells: neurons that represent
self-location, a local and aperiodic quantity, with seemingly bizarre non-local
and spatially periodic activity patterns of a few discrete periods. Why has the
mammalian lineage learnt this peculiar grid representation? Mathematical analysi
s suggests that this multi-periodic representation has excellent properties as a
n algebraic code with high capacity and intrinsic error-correction, but to date,
 synthesis of multi-modular grid cells in deep recurrent neural networks remains
 absent. In this work, we begin by identifying key insights from four families o
f approaches to answering the grid cell question: dynamical systems, coding theo
ry, function optimization and supervised deep learning. We then leverage our ins
ights to propose a new approach that elegantly combines the strengths of all fou
r approaches. Our approach is a self-supervised learning (SSL) framework - inclu
ding data, data augmentations, loss functions and a network architecture - motiv
ated from a normative perspective, with no access to supervised position informa
tion. Without making assumptions about internal or readout representations, we s
how that multiple grid cell modules can emerge in networks trained on our SSL fr
amework and that the networks generalize significantly beyond their training dis
tribution. This work contains insights for neuroscientists interested in the ori
gins of grid cells as well as machine learning researchers interested in novel S
SL frameworks.
****
A Guide Through the Zoo of Biased SGD
Yury Demidovich, Grigory Malinovsky, Igor Sokolov, Peter Richtarik
Stochastic Gradient Descent (SGD) is arguably the most important single algorith
m in modern machine learning. Although SGD with unbiased gradient estimators has
 been studied extensively over at least half a century, SGD variants relying on
biased estimators are rare. Nevertheless, there has been an increased interest i
n this topic in recent years. However, existing literature on SGD with biased es
timators lacks coherence since each new paper relies on a different set of assum
ptions, without any clear understanding of how they are connected, which may lea
d to confusion. We address this gap by establishing connections among the existi
ng assumptions, and presenting a comprehensive map of the underlying relationshi
ps. Additionally, we introduce a new set of assumptions that is provably weaker
than all previous assumptions, and use it to present a thorough analysis of Bias
edSGD in both convex and non-convex settings, offering advantages over previous
results. We also provide examples where biased estimators outperform their unbia
sed counterparts or where unbiased versions are simply not available. Finally, w
e demonstrate the effectiveness of our framework through experimental results th
at validate our theoretical findings.
****
Construction of Hierarchical Neural Architecture Search Spaces based on Context-
free Grammars
Simon Schrodi, Danny Stoll, Binxin Ru, Rhea Sukthanker, Thomas Brox, Frank Hutte
r
The discovery of neural architectures from simple building blocks is a long-stan

ding goal of Neural Architecture Search (NAS). Hierarchical search spaces are a promising step towards this goal but lack a unifying search space design framework and typically only search over some limited aspect of architectures. In this work, we introduce a unifying search space design framework based on context-free grammars that can naturally and compactly generate expressive hierarchical search spaces that are 100s of orders of magnitude larger than common spaces from the literature. By enhancing and using their properties, we effectively enable search over the complete architecture and can foster regularity. Further, we propose an efficient hierarchical kernel design for a Bayesian Optimization search strategy to efficiently search over such huge spaces. We demonstrate the versatility of our search space design framework and show that our search strategy can be superior to existing NAS approaches. Code is available at https://github.com/automl/hierarchicalnasconstruction.
****

Data-Informed Geometric Space Selection
Shuai Zhang, Wenqi Jiang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Prioritizing Samples in Reinforcement Learning with Reducible Loss
Shivakanth Sujit, Somjit Nath, Pedro Braga, Samira Ebrahimi Kahou
Most reinforcement learning algorithms take advantage of an experience replay buffer to repeatedly train on samples the agent has observed in the past. Not all samples carry the same amount of significance and simply assigning equal importance to each of the samples is a naïve strategy. In this paper, we propose a method to prioritize samples based on how much we can learn from a sample. We define the learn-ability of a sample as the steady decrease of the training loss associated with this sample over time. We develop an algorithm to prioritize samples with high learn-ability, while assigning lower priority to those that are hard-to-learn, typically caused by noise or stochasticity. We empirically show that across multiple domains our method is more robust than random sampling and also better than just prioritizing with respect to the training loss, i.e. the temporal difference loss, which is used in prioritized experience replay.
****

Intensity Profile Projection: A Framework for Continuous-Time Representation Learning for Dynamic Networks
Alexander Modell, Ian Gallagher, Emma Ceccherini, Nick Whiteley, Patrick Rubin-Delanchy
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Understanding Contrastive Learning via Distributionally Robust Optimization
Junkang Wu, Jiawei Chen, Jiancan Wu, Wentao Shi, Xiang Wang, Xiangnan He
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

K-Nearest-Neighbor Local Sampling Based Conditional Independence Testing
Shuai Li, Yingjie Zhang, Hongtu Zhu, Christina Wang, Hai Shu, Ziqi Chen, Zhuoran Sun, Yanfeng Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Learning Large Graph Property Prediction via Graph Segment Training
Kaidi Cao, Mangpo Phothilimthana, Sami Abu-El-Haija, Dustin Zelle, Yanqi Zhou, Charith Mendis, Jure Leskovec, Bryan Perozzi
Learning to predict properties of large graphs is challenging because each prediction requires the knowledge of an entire graph, while the amount of memory available during training is bounded. Here we propose Graph Segment Training (GST), a general framework that utilizes a divide-and-conquer approach to allow learning large graph property prediction with a constant memory footprint. GST first divides a large graph into segments and then backpropagates through only a few segments sampled per training iteration. We refine the GST paradigm by introducing a historical embedding table to efficiently obtain embeddings for segments not sampled for backpropagation. To mitigate the staleness of historical embeddings, we design two novel techniques. First, we finetune the prediction head to fix the input distribution shift. Second, we introduce Stale Embedding Dropout to drop some stale embeddings during training to reduce bias. We evaluate our complete method GST-EFD (with all the techniques together) on two large graph property prediction benchmarks: MalNet and TpuGraphs. Our experiments show that GST-EFD is both memory-efficient and fast, while offering a slight boost on test accuracy over a typical full graph training regime.
****
Online Nonstochastic Model-Free Reinforcement Learning
Udaya Ghai, Arushi Gupta, Wenhan Xia, Karan Singh, Elad Hazan
We investigate robust model-free reinforcement learning algorithms designed for environments that may be dynamic or even adversarial. Traditional state-based policies often struggle to accommodate the challenges imposed by the presence of unmodeled disturbances in such settings. Moreover, optimizing linear state-based policies pose an obstacle for efficient optimization, leading to nonconvex objectives, even in benign environments like linear dynamical systems.Drawing inspiration from recent advancements in model-based control, we intro- duce a novel class of policies centered on disturbance signals. We define several categories of these signals, which we term pseudo-disturbances, and develop corresponding policy classes based on them. We provide efficient and practical algorithms for optimizing these policies.Next, we examine the task of online adaptation of reinforcement learning agents in the face of adversarial disturbances. Our methods seamlessly integrate with any black-box model-free approach, yielding provable regret guarantees when dealing with linear dynamics. These regret guarantees unconditionally improve the best-known results for bandit linear control in having no dependence on the state-space dimension. We evaluate our method over various standard RL benchmarks and demonstrate improved robustness.
****
Time-Reversed Dissipation Induces Duality Between Minimizing Gradient Norm and Function Value
Jaeyeon Kim, Asuman Ozdaglar, Chanwoo Park, Ernest Ryu
In convex optimization, first-order optimization methods efficiently minimizing function values have been a central subject study since Nesterov's seminal work of 1983. Recently, however, Kim and Fessler's OGM-G and Lee et al.'s FISTA-G have been presented as alternatives that efficiently minimize the gradient magnitude instead. In this paper, we present H-duality, which represents a surprising one-to-one correspondence between methods efficiently minimizing function values and methods efficiently minimizing gradient magnitude. In continuous-time formulations, H-duality corresponds to reversing the time dependence of the dissipation/friction term. To the best of our knowledge, H-duality is different from Lagrange/Fenchel duality and is distinct from any previously known duality or symmetry relations. Using H-duality, we obtain a clearer understanding of the symmetry between Nesterov's method and OGM-G, derive a new class of methods efficiently reducing gradient magnitudes of smooth convex functions, and find a new composite minimization method that is simpler and faster than FISTA-G.
****
Cascading Contextual Assortment Bandits
Hyun-jun Choi, Rajan Udwani, Min-hwan Oh

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Dynamic Tensor Decomposition via Neural Diffusion-Reaction Processes
Zheng Wang, Shikai Fang, Shibo Li, Shandian Zhe
Tensor decomposition is an important tool for multiway data analysis. In practice, the data is often sparse yet associated with rich temporal information. Existing methods, however, often under-use the time information and ignore the structural knowledge within the sparsely observed tensor entries. To overcome these limitations and to better capture the underlying temporal structure, we propose Dynamic EMbedIngs fOr dynamic Tensor dEcomposition (DEMOTE). We develop a neural diffusion-reaction process to estimate dynamic embeddings for the entities in each tensor mode. Specifically, based on the observed tensor entries, we build a multi-partite graph to encode the correlation between the entities. We construct a graph diffusion process to co-evolve the embedding trajectories of the correlated entities and use a neural network to construct a reaction process for each individual entity. In this way, our model can capture both the commonalities and personalities during the evolution of the embeddings for different entities. We then use a neural network to model the entry value as a nonlinear function of the embedding trajectories. For model estimation, we combine ODE solvers to develop a stochastic mini-batch learning algorithm. We propose a stratified sampling method to balance the cost of processing each mini-batch so as to improve the overall efficiency. We show the advantage of our approach in both simulation studies and real-world applications. The code is available at https://github.com/wzhut/Dynamic-Tensor-Decomposition-via-Neural-Diffusion-Reaction-Processes.
****

CSMeD: Bridging the Dataset Gap in Automated Citation Screening for Systematic Literature Reviews
Wojciech Kusa, Oscar E. Mendoza, Matthias Samwald, Petr Knoth, Allan Hanbury
Systematic literature reviews (SLRs) play an essential role in summarising, synthesising and validating scientific evidence. In recent years, there has been a growing interest in using machine learning techniques to automate the identification of relevant studies for SLRs. However, the lack of standardised evaluation datasets makes comparing the performance of such automated literature screening systems difficult. In this paper, we analyse the citation screening evaluation datasets, revealing that many of the available datasets are either too small, suffer from data leakage or have limited applicability to systems treating automated literature screening as a classification task, as opposed to, for example, a retrieval or question-answering task. To address these challenges, we introduce CSMED, a meta-dataset consolidating nine publicly released collections, providing unified access to 325 SLRs from the fields of medicine and computer science. CSMED serves as a comprehensive resource for training and evaluating the performance of automated citation screening models. Additionally, we introduce CSMED-FT, a new dataset designed explicitly for evaluating the full text publication screening task. To demonstrate the utility of CSMED, we conduct experiments and establish baselines on new datasets.
****

Sample based Explanations via Generalized Representers
Che-Ping Tsai, Chih-Kuan Yeh, Pradeep Ravikumar
We propose a general class of sample based explanations of machine learning models, which we term generalized representers. To measure the effect of a training sample on a model's test prediction, generalized representers use two components: a global sample importance that quantifies the importance of the training point to the model and is invariant to test samples, and a local sample importance that measures similarity between the training sample and the test point with a kernel. A key contribution of the paper is to show that generalized representers are the only class of sample based explanations satisfying a natural set of axiomatic properties. We discuss approaches to extract global importances given a ker

nel, and also natural choices of kernels given modern non-linear models. As we s
how, many popular existing sample based explanations could be cast as generalize
d representers with particular choices of kernels and approaches to extract glob
al importances. Additionally, we conduct empirical comparisons of different gene
ralized representers on two image classification datasets.
****
Open Visual Knowledge Extraction via Relation-Oriented Multimodality Model Promp
ting
Hejie Cui, Xinyu Fang, Zihan Zhang, Ran Xu, Xuan Kan, Xin Liu, Yue Yu, Manling L
i, Yangqiu Song, Carl Yang
Images contain rich relational knowledge that can help machines understand the w
orld. Existing methods on visual knowledge extraction often rely on the pre-defi
ned format (e.g., sub-verb-obj tuples) or vocabulary (e.g., relation types), res
tricting the expressiveness of the extracted knowledge. In this work, we take a
first exploration to a new paradigm of open visual knowledge extraction. To achi
eve this, we present OpenVik which consists of an open relational region detecto
r to detect regions potentially containing relational knowledge and a visual kno
wledge generator that generates format-free knowledge by prompting the large mul
timodality model with the detected region of interest. We also explore two data
enhancement techniques for diversifying the generated format-free visual knowled
ge. Extensive knowledge quality evaluations highlight the correctness and unique
ness of the extracted open visual knowledge by OpenVik. Moreover, integrating ou
r extracted knowledge across various visual reasoning applications shows consist
ent improvements, indicating the real-world applicability of OpenVik.
****
Continuous Parametric Optical Flow
Jianqin Luo, Zhexiong Wan, yuxin mao, Bo Li, Yuchao Dai
In this paper, we present continuous parametric optical flow, a parametric repre
sentation of dense and continuous motion over arbitrary time interval. In contra
st to existing discrete-time representations (i.e., flow in between consecutive
frames), this new representation transforms the frame-to-frame pixel corresponde
nces to dense continuous flow. In particular, we present a temporal-parametric m
odel that employs B-splines to fit point trajectories using a limited number of
frames. To further improve the stability and robustness of the trajectories, we
also add an encoder with a neural ordinary differential equation (NODE) to repre
sent features associated with specific times. We also contribute a synthetic dat
aset and introduce two evaluation perspectives to measure the accuracy and robus
tness of continuous flow estimation. Benefiting from the combination of explicit
 parametric modeling and implicit feature optimization, our model focuses on mot
ion continuity and outperforms the flow-based and point-tracking approaches for
fitting long-term and variable sequences.
****
Reusable Slotwise Mechanisms
Trang Nguyen, Amin Mansouri, Kanika Madan, Khuong Duy Nguyen, Kartik Ahuja, Dian
bo Liu, Yoshua Bengio
Agents with the ability to comprehend and reason about the dynamics of objects w
ould be expected to exhibit improved robustness and generalization in novel scen
arios. However, achieving this capability necessitates not only an effective sce
ne representation but also an understanding of the mechanisms governing interact
ions among object subsets. Recent studies have made significant progress in repr
esenting scenes using object slots. In this work, we introduce Reusable Slotwise
 Mechanisms, or RSM, a framework that models object dynamics by leveraging commu
nication among slots along with a modular architecture capable of dynamically se
lecting reusable mechanisms for predicting the future states of each object slot
. Crucially, RSM leverages the Central Contextual Information (CCI), enabling se
lected mechanisms to access the remaining slots through a bottleneck, effectivel
y allowing for modeling of higher order and complex interactions that might requ
ire a sparse subset of objects. Experimental results demonstrate the superior pe
rformance of RSM compared to state-of-the-art methods across various future pred
iction and related downstream tasks, including Visual Question Answering and act

ion planning. Furthermore, we showcase RSM’s Out-of-Distribution generalization ability to handle scenes in intricate scenarios.
****
Improved Bayesian Regret Bounds for Thompson Sampling in Reinforcement Learning
Ahmadreza Moradipari, Mohammad Pedramfar, Modjtaba Shokrian Zini, Vaneet Aggarwal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Evaluating and Improving Tool-Augmented Computation-Intensive Math Reasoning
Beichen Zhang, Kun Zhou, Xilin Wei, Xin Zhao, Jing Sha, Shijin Wang, Ji-Rong Wen
Chain-of-thought prompting (CoT) and tool augmentation have been validated in recent work as effective practices for improving large language models (LLMs) to perform step-by-step reasoning on complex math-related tasks.However, most existing math reasoning datasets may not be able to fully evaluate and analyze the ability of LLMs in manipulating tools and performing reasoning, as they often only require very few invocations of tools or miss annotations for evaluating intermediate reasoning steps, thus supporting only outcome evaluation.To address the issue, we construct CARP, a new Chinese dataset consisting of 4,886 computation-intensive algebra problems with formulated annotations on intermediate steps, facilitating the evaluation of the intermediate reasoning process.In CARP, we test four LLMs with CoT prompting, and find that they are all prone to make mistakes at the early steps of the solution, leading to incorrect answers.Based on this finding, we propose a new approach that can facilitate the deliberation on reasoning steps with tool interfaces, namely DELI.In DELI, we first initialize a step-by-step solution based on retrieved exemplars, then iterate two deliberation procedures that check and refine the intermediate steps of the generated solution, from both tool manipulation and natural language reasoning perspectives, until solutions converge or the maximum iteration is achieved.Experimental results on CARP and six other datasets show that the proposed DELI mostly outperforms competitive baselines, and can further boost the performance of existing CoT methods.Our data and code are available at https://github.com/RUCAIBox/CARP.
****
Moment Matching Denoising Gibbs Sampling
Mingtian Zhang, Alex Hawkins-Hooker, Brooks Paige, David Barber
Energy-Based Models (EBMs) offer a versatile framework for modelling complex data distributions. However, training and sampling from EBMs continue to pose significant challenges. The widely-used Denoising Score Matching (DSM) method for scalable EBM training suffers from inconsistency issues, causing the energy model to learn a noisy data distribution. In this work, we propose an efficient sampling framework: (pseudo)-Gibbs sampling with moment matching, which enables effective sampling from the underlying clean model when given a noisy model that has been well-trained via DSM. We explore the benefits of our approach compared to related methods and demonstrate how to scale the method to high-dimensional datasets.
****
Bottleneck Structure in Learned Features: Low-Dimension vs Regularity Tradeoff
Arthur Jacot
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Noise-Adaptive Thompson Sampling for Linear Contextual Bandits
Ruitu Xu, Yifei Min, Tianhao Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Regularization properties of adversarially-trained linear regression
Antonio Ribeiro, Dave Zachariah, Francis Bach, Thomas Schön
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Toolkit for Reliable Benchmarking and Research in Multi-Objective Reinforcement Learning
Florian Felten, Lucas N. Alegre, Ann Nowe, Ana Bazzan, El Ghazali Talbi, Grégoire Danoy, Bruno C. da Silva
Multi-objective reinforcement learning algorithms (MORL) extend standard reinforcement learning (RL) to scenarios where agents must optimize multiple---potentially conflicting---objectives, each represented by a distinct reward function. To facilitate and accelerate research and benchmarking in multi-objective RL problems, we introduce a comprehensive collection of software libraries that includes: (i) MO-Gymnasium, an easy-to-use and flexible API enabling the rapid construction of novel MORL environments. It also includes more than 20 environments under this API. This allows researchers to effortlessly evaluate any algorithms on any existing domains; (ii) MORL-Baselines, a collection of reliable and efficient implementations of state-of-the-art MORL algorithms, designed to provide a solid foundation for advancing research. Notably, all algorithms are inherently compatible with MO-Gymnasium; and(iii) a thorough and robust set of benchmark results and comparisons of MORL-Baselines algorithms, tested across various challenging MO-Gymnasium environments. These benchmarks were constructed to serve as guidelines for the research community, underscoring the properties, advantages, and limitations of each particular state-of-the-art method.
****
$\mathbf{\mathbb{E}^{FWI}}$: Multiparameter Benchmark Datasets for Elastic Full Waveform Inversion of Geophysical Properties
Shihang Feng, Hanchen Wang, Chengyuan Deng, Yinan Feng, Yanhua Liu, Min Zhu, Peng Jin, Yinpeng Chen, Youzuo Lin
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Complex-valued Neurons Can Learn More but Slower than Real-valued Neurons via Gradient Descent
Jin-Hui Wu, Shao-Qun Zhang, Yuan Jiang, Zhi-Hua Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning a Neuron by a Shallow ReLU Network: Dynamics and Implicit Bias for Correlated Inputs
Dmitry Chistikov, Matthias Englert, Ranko Lazic
We prove that, for the fundamental regression task of learning a single neuron, training a one-hidden layer ReLU network of any width by gradient flow from a small initialisation converges to zero loss and is implicitly biased to minimise the rank of network parameters. By assuming that the training points are correlated with the teacher neuron, we complement previous work that considered orthogonal datasets. Our results are based on a detailed non-asymptotic analysis of the dynamics of each hidden neuron throughout the training. We also show and characterise a surprising distinction in this setting between interpolator networks of minimal rank and those of minimal Euclidean norm. Finally we perform a range of numerical experiments, which corroborate our theoretical findings.

****
Separable Physics-Informed Neural Networks
Junwoo Cho, Seungtae Nam, Hyunmo Yang, Seok-Bae Yun, Youngjoon Hong, Eunbyung Park
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Beyond Invariance: Test-Time Label-Shift Adaptation for Addressing "Spurious" Correlations
Qingyao Sun, Kevin P. Murphy, Sayna Ebrahimi, Alexander D'Amour
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SwiftSage: A Generative Agent with Fast and Slow Thinking for Complex Interactive Tasks
Bill Yuchen Lin, Yicheng Fu, Karina Yang, Faeze Brahman, Shiyu Huang, Chandra Bhagavatula, Prithviraj Ammanabrolu, Yejin Choi, Xiang Ren
We introduce SwiftSage, a novel agent framework inspired by the dual-process theory of human cognition, designed to excel in action planning for complex interactive reasoning tasks. SwiftSage integrates the strengths of behavior cloning and prompting large language models (LLMs) to enhance task completion performance. The framework comprises two primary modules: the Swift module, representing fast and intuitive thinking, and the Sage module, emulating deliberate thought processes. The Swift module is a small encoder-decoder LM fine-tuned on the oracle agent's action trajectories, while the Sage module employs LLMs such as GPT-4 for subgoal planning and grounding. We develop a heuristic method to harmoniously integrate the two modules, resulting in a more efficient and robust problem-solving process. In 30 tasks from the ScienceWorld benchmark, SwiftSage significantly outperforms other methods such as SayCan, ReAct, and Reflexion, demonstrating its effectiveness in solving complex interactive tasks.
****
InterCode: Standardizing and Benchmarking Interactive Coding with Execution Feedback
John Yang, Akshara Prabhakar, Karthik Narasimhan, Shunyu Yao
Humans write code in a fundamentally interactive manner and rely on constant execution feedback to correct errors, resolve ambiguities, and decompose tasks. While LLMs have recently exhibited promising coding capabilities, current coding benchmarks mostly consider a static instruction-to-code sequence transduction process, which has the potential for error propagation and a disconnect between the generated code and its final execution environment. To address this gap, we introduce InterCode, a lightweight, flexible, and easy-to-use framework of interactive coding as a standard reinforcement learning (RL) environment, with code as actions and execution feedback as observations. Our framework is language and platform agnostic, uses self-contained Docker environments to provide safe and reproducible execution, and is compatible out-of-the-box with traditional seq2seq coding methods, while enabling the development of new methods for interactive code generation. We use InterCode to create three interactive code environments with Bash, SQL, and Python as action spaces, leveraging data from the static NL2Bash, Spider, and MBPP datasets. We demonstrate InterCode’s viability as a testbed by evaluating multiple state-of-the-art LLMs configured with different prompting strategies such as ReAct and Plan & Solve. Our results showcase the benefits of interactive code generation and demonstrate that InterCode can serve as a challenging benchmark for advancing code understanding and generation capabilities. InterCode is designed to be easily extensible and can even be used to create new tasks such as Capture the Flag, a popular coding puzzle that is inherently multi-step and involves multiple programming languages.

****
Gradient-Free Kernel Stein Discrepancy
Matthew Fisher, Chris J. Oates
Stein discrepancies have emerged as a powerful statistical tool, being applied to fundamental statistical problems including parameter inference, goodness-of-fit testing, and sampling. The canonical Stein discrepancies require the derivatives of a statistical model to be computed, and in return provide theoretical guarantees of convergence detection and control. However, for complex statistical models, the stable numerical computation of derivatives can require bespoke algorithmic development and render Stein discrepancies impractical. This paper focuses on posterior approximation using Stein discrepancies, and introduces a collection of non-canonical Stein discrepancies that are gradient-free, meaning that derivatives of the statistical model are not required. Sufficient conditions for convergence detection and control are established, and applications to sampling and variational inference are presented.
****
ConDaFormer: Disassembled Transformer with Local Structure Enhancement for 3D Point Cloud Understanding
Lunhao Duan, Shanshan Zhao, Nan Xue, Mingming Gong, Gui-Song Xia, Dacheng Tao
Transformers have been recently explored for 3D point cloud understanding with impressive progress achieved. A large number of points, over 0.1 million, make the global self-attention infeasible for point cloud data. Thus, most methods propose to apply the transformer in a local region, e.g., spherical or cubic window. However, it still contains a large number of Query-Key pairs, which requires high computational costs. In addition, previous methods usually learn the query, key, and value using a linear projection without modeling the local 3D geometric structure. In this paper, we attempt to reduce the costs and model the local geometry prior by developing a new transformer block, named ConDaFormer. Technically, ConDaFormer disassembles the cubic window into three orthogonal 2D planes, leading to fewer points when modeling the attention in a similar range. The disassembling operation is beneficial to enlarging the range of attention without increasing the computational complexity, but ignores some contexts. To provide a remedy, we develop a local structure enhancement strategy that introduces a depth-wise convolution before and after the attention. This scheme can also capture the local geometric information. Taking advantage of these designs, ConDaFormer captures both long-range contextual information and local priors. The effectiveness is demonstrated by experimental results on several 3D point cloud understanding benchmarks. Our code will be available.
****
Variational Monte Carlo on a Budget — Fine-tuning pre-trained Neural Wavefunctions
Michael Scherbela, Leon Gerard, Philipp Grohs
Obtaining accurate solutions to the Schrödinger equation is the key challenge in computational quantum chemistry. Deep-learning-based Variational Monte Carlo (DL-VMC) has recently outperformed conventional approaches in terms of accuracy, but only at large computational cost.Whereas in many domains models are trained once and subsequently applied for inference, accurate DL-VMC so far requires a full optimization for every new problem instance, consuming thousands of GPUhs even for small molecules.We instead propose a DL-VMC model which has been pre-trained using self-supervised wavefunction optimization on a large and chemically diverse set of molecules. Applying this model to new molecules without any optimization, yields wavefunctions and absolute energies that outperform established methods such as CCSD(T)-2Z.To obtain accurate relative energies, only few fine-tuning steps of this base model are required.We accomplish this with a fully end-to-end machine-learned model, consisting of an improved geometry embedding architecture and an existing SE(3)-equivariant model to represent molecular orbitals. Combining this architecture with continuous sampling of geometries, we improve zero-shot accuracy by two orders of magnitude compared to the state of the art.We extensively evaluate the accuracy, scalability and limitations of our base model on a wide variety of test systems.

****
ReDS: Offline RL With Heteroskedastic Datasets via Support Constraints
Anikait Singh, Aviral Kumar, Quan Vuong, Yevgen Chebotar, Sergey Levine
Offline reinforcement learning (RL) learns policies entirely from static datasets. Practical applications of offline RL will inevitably require learning from datasets where the variability of demonstrated behaviors changes non-uniformly across the state space. For example, at a red light, nearly all human drivers behave similarly by stopping, but when merging onto a highway, some drivers merge quickly, efficiently, and safely, while many hesitate or merge dangerously. Both theoretically and empirically, we show that typical offline RL methods, which are based on distribution constraints fail to learn from data with such non-uniform variability, due to the requirement to stay close to the behavior policy to the same extent across the state space. Ideally, the learned policy should be free to choose per state how closely to follow the behavior policy to maximize long-term return, as long as the learned policy stays within the support of the behavior policy. To instantiate this principle, we reweight the data distribution in conservative Q-learning (CQL) to obtain an approximate support constraint formulation. The reweighted distribution is a mixture of the current policy and an additional policy trained to mine poor actions that are likely under the behavior policy. Our method, CQL (ReDS), is theoretically motivated, and improves performance across a wide range of offline RL problems in games, navigation, and pixel-based manipulation.
****
A Graph-Theoretic Framework for Understanding Open-World Semi-Supervised Learning
Yiyou Sun, Zhenmei Shi, Yixuan Li
Open-world semi-supervised learning aims at inferring both known and novel classes in unlabeled data, by harnessing prior knowledge from a labeled set with known classes. Despite its importance, there is a lack of theoretical foundations for this problem. This paper bridges the gap by formalizing a graph-theoretic framework tailored for the open-world setting, where the clustering can be theoretically characterized by graph factorization. Our graph-theoretic framework illuminates practical algorithms and provides guarantees. In particular, based on our graph formulation, we apply the algorithm called Spectral Open-world Representation Learning (SORL), and show that minimizing our loss is equivalent to performing spectral decomposition on the graph. Such equivalence allows us to derive a provable error bound on the clustering performance for both known and novel classes, and analyze rigorously when labeled data helps. Empirically, SORL can match or outperform several strong baselines on common benchmark datasets, which is appealing for practical usage while enjoying theoretical guarantees.
****
Near Optimal Reconstruction of Spherical Harmonic Expansions
Amir Zandieh, Insu Han, Haim Avron
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Lexinvariant Language Models
Qian Huang, Eric Zelikman, Sarah Chen, Yuhuai Wu, Gregory Valiant, Percy S. Liang
Token embeddings, a mapping from discrete lexical symbols to continuous vectors, are at the heart of any language model (LM). However, lexical symbol meanings can also be determined and even redefined by their structural role in a long context. In this paper, we ask: is it possible for a language model to be performant without \emph{any} fixed token embeddings? Such a language model would have to rely entirely on the co-occurence and repetition of tokens in the context rather than the \textit{a priori} identity of any token. To answer this, we study \textit{lexinvariant}language models that are invariant to lexical symbols and therefore do not need fixed token embeddings in practice. First, we prove that we can

construct a lexinvariant LM to converge to the true language model at a uniform rate that is polynomial in terms of the context length, with a constant factor that is sublinear in the vocabulary size. Second, to build a lexinvariant LM, we simply encode tokens using random Gaussian vectors, such that each token maps to the same representation within each sequence but different representations across sequences. Empirically, we demonstrate that it can indeed attain perplexity comparable to that of a standard language model, given a sufficiently long context. We further explore two properties of the lexinvariant language models: First, given text generated from a substitution cipher of English, it implicitly implements Bayesian in-context deciphering and infers the mapping to the underlying real tokens with high accuracy. Second, it has on average 4X better accuracy over synthetic in-context reasoning tasks. Finally, we discuss regularizing standard language models towards lexinvariance and potential practical applications.
****

REFINE: A Fine-Grained Medication Recommendation System Using Deep Learning and Personalized Drug Interaction Modeling

Suman Bhoi, Mong Li Lee, Wynne Hsu, Ngiap Chuan Tan

Patients with co-morbidities often require multiple medications to manage their conditions. However, existing medication recommendation systems only offer class-level medications and regard all interactions among drugs to have the same level of severity. This limits their ability to provide personalized and safe recommendations tailored to individual needs. In this work, we introduce a deep learning-based fine-grained medication recommendation system called REFINE, which is designed to improve treatment outcomes and minimize adverse drug interactions. In order to better characterize patients' health conditions, we model the trend in medication dosage titrations and lab test responses, and adapt the vision transformer to obtain effective patient representations. We also model drug interaction severity levels as weighted graphs to learn safe drug combinations and design a balanced loss function to avoid overly conservative recommendations and miss medications that might be needed for certain conditions. Extensive experiments on two real-world datasets show that REFINE outperforms state-of-the-art techniques.
****

Bayesian Extensive-Rank Matrix Factorization with Rotational Invariant Priors

Farzad Pourkamali, Nicolas Macris

We consider a statistical model for matrix factorization in a regime where the rank of the two hidden matrix factors grows linearly with their dimension and their product is corrupted by additive noise. Despite various approaches, statistical and algorithmic limits of such problems have remained elusive. We study a Bayesian setting with the assumptions that (a) one of the matrix factors is symmetric, (b) both factors as well as the additive noise have rotational invariant priors, (c) the priors are known to the statistician. We derive analytical formulas for Rotation Invariant Estimators to reconstruct the two matrix factors, and conjecture that these are optimal in the large-dimension limit, in the sense that they minimize the average mean-square-error. We provide numerical checks which confirm the optimality conjecture when confronted to Oracle Estimators which are optimal by definition, but involve the ground-truth. Our derivation relies on a combination of tools, namely random matrix theory transforms, spherical integral formulas, and the replica method from statistical mechanics.
****

Optimal Transport Model Distributional Robustness

Van-Anh Nguyen, Trung Le, Anh Bui, Thanh-Toan Do, Dinh Phung

Distributional robustness is a promising framework for training deep learning models that are less vulnerable to adversarial examples and data distribution shifts. Previous works have mainly focused on exploiting distributional robustness in the data space. In this work, we explore an optimal transport-based distributional robustness framework in model spaces. Specifically, we examine a model distribution within a Wasserstein ball centered on a given model distribution that maximizes the loss. We have developed theories that enable us to learn the optimal robust center model distribution. Interestingly, our developed theories allow

us to flexibly incorporate the concept of sharpness awareness into training, whether it's a single model, ensemble models, or Bayesian Neural Networks, by considering specific forms of the center model distribution. These forms include a Dirac delta distribution over a single model, a uniform distribution over several models, and a general Bayesian Neural Network. Furthermore, we demonstrate that Sharpness-Aware Minimization (SAM) is a specific case of our framework when using a Dirac delta distribution over a single model, while our framework can be seen as a probabilistic extension of SAM. To validate the effectiveness of our framework in the aforementioned settings, we conducted extensive experiments, and the results reveal remarkable improvements compared to the baselines.
****
Language Semantic Graph Guided Data-Efficient Learning
Wenxuan Ma, Shuang Li, lincan Cai, Jingxuan Kang
Developing generalizable models that can effectively learn from limited data and with minimal reliance on human supervision is a significant objective within the machine learning community, particularly in the era of deep neural networks. Therefore, to achieve data-efficient learning, researchers typically explore approaches that can leverage more related or unlabeled data without necessitating additional manual labeling efforts, such as Semi-Supervised Learning (SSL), Transfer Learning (TL), and Data Augmentation (DA).SSL leverages unlabeled data in the training process, while TL enables the transfer of expertise from related data distributions. DA broadens the dataset by synthesizing new data from existing examples. However, the significance of additional knowledge contained within labels has been largely overlooked in research. In this paper, we propose a novel perspective on data efficiency that involves exploiting the semantic information contained in the labels of the available data. Specifically, we introduce a Language Semantic Graph (LSG) which is constructed from labels manifest as natural language descriptions. Upon this graph, an auxiliary graph neural network is trained to extract high-level semantic relations and then used to guide the training of the primary model, enabling more adequate utilization of label knowledge. Across image, video, and audio modalities, we utilize the LSG method in both TL and SSL scenarios and illustrate its versatility in significantly enhancing performance compared to other data-efficient learning approaches. Additionally, our in-depth analysis shows that the LSG method also expedites the training process.
****
Learning Efficient Coding of Natural Images with Maximum Manifold Capacity Representations
Thomas Yerxa, Yilun Kuang, Eero Simoncelli, SueYeon Chung
The efficient coding hypothesis proposes that the response properties of sensory systems are adapted to the statistics of their inputs such that they capture maximal information about the environment, subject to biological constraints. While elegant, information theoretic properties are notoriously difficult to measure in practical settings or to employ as objective functions in optimization. This difficulty has necessitated that computational models designed to test the hypothesis employ several different information metrics ranging from approximations and lower bounds to proxy measures like reconstruction error. Recent theoretical advances have characterized a novel and ecologically relevant efficiency metric, the ``manifold capacity,” which is the number of object categories that may be represented in a linearly separable fashion. However, calculating manifold capacity is a computationally intensive iterative procedure that until now has precluded its use as an objective. Here we outline the simplifying assumptions that allow manifold capacity to be optimized directly, yielding Maximum Manifold Capacity Representations (MMCR). The resulting method is closely related to and inspired by advances in the field of self supervised learning (SSL), and we demonstrate that MMCRs are competitive with state of the art results on standard SSL benchmarks. Empirical analyses reveal differences between MMCRs and representations learned by other SSL frameworks, and suggest a mechanism by which manifold compression gives rise to class separability. Finally we evaluate a set of SSL methods on a suite of neural predicitivity benchmarks, and find MMCRs are higly competitive as models of the ventral stream.

****
Understanding the Latent Space of Diffusion Models through the Lens of Riemannia n Geometry
Yong-Hyun Park, Mingi Kwon, Jaewoong Choi, Junghyo Jo, Youngjung Uh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Single-Stage Visual Query Localization in Egocentric Videos
Hanwen Jiang, Santhosh Kumar Ramakrishnan, Kristen Grauman
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Hyper-Skin: A Hyperspectral Dataset for Reconstructing Facial Skin-Spectra from RGB Images
Pai Chet Ng, Zhixiang Chi, Yannick Verdie, Juwei Lu, Konstantinos N Plataniotis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Generalizing Importance Weighting to A Universal Solver for Distribution Shift P roblems
Tongtong Fang, Nan Lu, Gang Niu, Masashi Sugiyama
Distribution shift (DS) may have two levels: the distribution itself changes, an d the support (i.e., the set where the probability density is non-zero) also cha nges. When considering the support change between the training and test distribu tions, there can be four cases: (i) they exactly match; (ii) the training suppor t is wider (and thus covers the test support); (iii) the test support is wider; (iv) they partially overlap. Existing methods are good at cases (i) and (ii), wh ile cases (iii) and (iv) are more common nowadays but still under-explored. In t his paper, we generalize importance weighting (IW), a golden solver for cases (i) and (ii), to a universal solver for all cases. Specifically, we first investig ate why IW might fail in cases (iii) and (iv); based on the findings, we propose generalized IW (GIW) that could handle cases (iii) and (iv) and would reduce to IW in cases (i) and (ii). In GIW, the test support is split into an in-training (IT) part and an out-of-training (OOT) part, and the expected risk is decompose d into a weighted classification term over the IT part and a standard classifica tion term over the OOT part, which guarantees the risk consistency of GIW. Then, the implementation of GIW consists of three components: (a) the split of valida tion data is carried out by the one-class support vector machine, (b) the first term of the empirical risk can be handled by any IW algorithm given training dat a and IT validation data, and (c) the second term just involves OOT validation d ata. Experiments demonstrate that GIW is a universal solver for DS problems, out performing IW methods in cases (iii) and (iv).
****
Improved Convergence in High Probability of Clipped Gradient Methods with Heavy Tailed Noise
Ta Duy Nguyen, Thien H Nguyen, Alina Ene, Huy Nguyen

****
Refining Diffusion Planner for Reliable Behavior Synthesis by Automatic Detectio n of Infeasible Plans
Kyowoon Lee, Seongun Kim, Jaesik Choi
Diffusion-based planning has shown promising results in long-horizon, sparse-rew ard tasks by training trajectory diffusion models and conditioning the sampled t

rajectories using auxiliary guidance functions. However, due to their nature as generative models, diffusion models are not guaranteed to generate feasible plan s, resulting in failed execution and precluding planners from being useful in sa fety-critical applications. In this work, we propose a novel approach to refine unreliable plans generated by diffusion models by providing refining guidance to error-prone plans. To this end, we suggest a new metric named restoration gap f or evaluating the quality of individual plans generated by the diffusion model. A restoration gap is estimated by a gap predictor which produces restoration gap guidance to refine a diffusion planner. We additionally present an attribution map regularizer to prevent adversarial refining guidance that could be generated from the sub-optimal gap predictor, which enables further refinement of infeasi ble plans. We demonstrate the effectiveness of our approach on three different b enchmarks in offline control settings that require long-horizon planning. We als o illustrate that our approach presents explainability by presenting the attribu tion maps of the gap predictor and highlighting error-prone transitions, allowin g for a deeper understanding of the generated plans.
****
Generate What You Prefer: Reshaping Sequential Recommendation via Guided Diffusi on
Zhengyi Yang, Jiancan Wu, Zhicai Wang, Xiang Wang, Yancheng Yuan, Xiangnan He
Sequential recommendation aims to recommend the next item that matches a user'si nterest, based on the sequence of items he/she interacted with before. Scrutiniz ingprevious studies, we can summarize a common learning-to-classify paradigm—giv en a positive item, a recommender model performs negative sampling to addnegativ e items and learns to classify whether the user prefers them or not, based onhis /her historical interaction sequence. Although effective, we reveal two inherent limitations: (1) it may differ from human behavior in that a user could imaginea n oracle item in mind and select potential items matching the oracle; and (2)the classification is limited in the candidate pool with noisy or easy supervisionf rom negative samples, which dilutes the preference signals towards the oracleite m. Yet, generating the oracle item from the historical interaction sequence ismo stly unexplored. To bridge the gap, we reshape sequential recommendationas a lea rning-to-generate paradigm, which is achieved via a guided diffusionmodel, terme d DreamRec. Specifically, for a sequence of historical items, itapplies a Transf ormer encoder to create guidance representations. Noising targetitems explores t he underlying distribution of item space; then, with the guidance ofhistorical i nteractions, the denoising process generates an oracle item to recoverthe positi ve item, so as to cast off negative sampling and depict the true preferenceof th e user directly. We evaluate the effectiveness of DreamRec through extensiveexpe riments and comparisons with existing methods. Codes and data are open-sourcedat https://github.com/YangZhengyi98/DreamRec.
****
Conditional score-based diffusion models for Bayesian inference in infinite dime nsions
Lorenzo Baldassari, Ali Siahkoohi, Josselin Garnier, Knut Solna, Maarten V. de H oop
Since their initial introduction, score-based diffusion models (SDMs) have been successfully applied to solve a variety of linear inverse problems in finite-dim ensional vector spaces due to their ability to efficiently approximate the poste rior distribution. However, using SDMs for inverse problems in infinite-dimensio nal function spaces has only been addressed recently, primarily through methods that learn the unconditional score. While this approach is advantageous for some inverse problems, it is mostly heuristic and involves numerous computationally costly forward operator evaluations during posterior sampling. To address these limitations, we propose a theoretically grounded method for sampling from the po sterior of infinite-dimensional Bayesian linear inverse problems based on amorti zed conditional SDMs. In particular, we prove that one of the most successful ap proaches for estimating the conditional score in finite dimensions—the condition al denoising estimator—can also be applied in infinite dimensions. A significant part of our analysis is dedicated to demonstrating that extending infinite-dime

nsional SDMs to the conditional setting requires careful consideration, as the conditional score typically blows up for small times, contrarily to the unconditional score. We conclude by presenting stylized and large-scale numerical examples that validate our approach, offer additional insights, and demonstrate that our method enables large-scale, discretization-invariant Bayesian inference.
****
Provable Advantage of Curriculum Learning on Parity Targets with Mixed Inputs
Emmanuel Abbe, Elisabetta Cornacchia, Aryo Lotfi
Experimental results have shown that curriculum learning, i.e., presenting simpler examples before more complex ones, can improve the efficiency of learning. Some recent theoretical results also showed that changing the sampling distribution can help neural networks learn parities, with formal results only for large learning rates and one-step arguments. Here we show a separation result in the number of training steps with standard (bounded) learning rates on a common sample distribution: if the data distribution is a mixture of sparse and dense inputs, there exists a regime in which a 2-layer ReLU neural network trained by a curriculum noisy-GD (or SGD) algorithm that uses sparse examples first, can learn parities of sufficiently large degree, while any fully connected neural network of possibly larger width or depth trained by noisy-GD on the unordered samples cannot learn without additional steps. We also provide experimental results supporting the qualitative separation beyond the specific regime of the theoretical results.
****
Event Stream GPT: A Data Pre-processing and Modeling Library for Generative, Pre-trained Transformers over Continuous-time Sequences of Complex Events
Matthew McDermott, Bret Nestor, Peniel Argaw, Isaac S Kohane
Generative, pre-trained transformers (GPTs, a type of "Foundation Models") have reshaped natural language processing (NLP) through their versatility in diverse downstream tasks. However, their potential extends far beyond NLP. This paper provides a software utility to help realize this potential, extending the applicability of GPTs to continuous-time sequences of complex events with internal dependencies, such as medical record datasets. Despite their potential, the adoption of foundation models in these domains has been hampered by the lack of suitable tools for model construction and evaluation. To bridge this gap, we introduce Event Stream GPT (ESGPT), an open-source library designed to streamline the end-to-end process for building GPTs for continuous-time event sequences. ESGPT allows users to (1) build flexible, foundation-model scale input datasets by specifying only a minimal configuration file, (2) leverage a Hugging Face compatible modeling API for GPTs over this modality that incorporates intra-event causal dependency structures and autoregressive generation capabilities, and (3) evaluate models via standardized processes that can assess few and even zero-shot performance of pre-trained models on user-specified fine-tuning tasks.
****
Modeling Human Visual Motion Processing with Trainable Motion Energy Sensing and a Self-attention Network
Zitang Sun, Yen-Ju Chen, Yung-Hao Yang, Shin'ya Nishida
Visual motion processing is essential for humans to perceive and interact with dynamic environments. Despite extensive research in cognitive neuroscience, image-computable models that can extract informative motion flow from natural scenes in a manner consistent with human visual processing have yet to be established. Meanwhile, recent advancements in computer vision (CV), propelled by deep learning, have led to significant progress in optical flow estimation, a task closely related to motion perception. Here we propose an image-computable model of human motion perception by bridging the gap between biological and CV models. Specifically, we introduce a novel two-stages approach that combines trainable motion energy sensing with a recurrent self-attention network for adaptive motion integration and segregation. This model architecture aims to capture the computations in V1-MT, the core structure for motion perception in the biological visual system, while providing the ability to derive informative motion flow for a wide range of stimuli, including complex natural scenes. In silico neurophysiology revea

ls that our model's unit responses are similar to mammalian neural recordings re garding motion pooling and speed tuning. The proposed model can also replicate h uman responses to a range of stimuli examined in past psychophysical studies. Th e experimental results on the Sintel benchmark demonstrate that our model predic ts human responses better than the ground truth, whereas the state-of-the-art CV models show the opposite. Our study provides a computational architecture consi stent with human visual motion processing, although the physiological correspond ence may not be exact.
****
Self-Supervised Visual Acoustic Matching
Arjun Somayazulu, Changan Chen, Kristen Grauman
Acoustic matching aims to re-synthesize an audio clip to sound as if it were rec orded in a target acoustic environment. Existing methods assume access to paired training data, where the audio is observed in both source and target environmen ts, but this limits the diversity of training data or requires the use of simula ted data or heuristics to create paired samples. We propose a self-supervised ap proach to visual acoustic matching where training samples include only the targe t scene image and audio---without acoustically mismatched source audio for refer ence. Our approach jointly learns to disentangle room acoustics and re-synthesiz e audio into the target environment, via a conditional GAN framework and a novel metric that quantifies the level of residual acoustic information in the de-bia sed audio. Training with either in-the-wild web data or simulated data, we demon strate it outperforms the state-of-the-art on multiple challenging datasets and a wide variety of real-world audio and environments.
****
Optimal Excess Risk Bounds for Empirical Risk Minimization on p-Norm Linear Re gression
Ayoub El Hanchi, Murat A. Erdogdu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Failure-Aware Gaussian Process Optimization with Regret Bounds
Shogo Iwazaki, Shion Takeno, Tomohiko Tanabe, Mitsuru Irie
Real-world optimization problems often require black-box optimization with obser vation failure, where we can obtain the objective function value if we succeed, otherwise, we can only obtain a fact of failure. Moreover, this failure region c an be complex by several latent constraints, whose number is also unknown. For t his problem, we propose a failure-aware Gaussian process upper confidence bound (F-GP-UCB), which only requires a mild assumption for the observation failure th at an optimal solution lies on an interior of a feasible region. Furthermore, we show that the number of successful observations grows linearly, by which we pro vide the first regret upper bounds and the convergence of F-GP-UCB. We demonstra te the effectiveness of F-GP-UCB in several benchmark functions, including the s imulation function motivated by material synthesis experiments.
****
Efficient Adversarial Attacks on Online Multi-agent Reinforcement Learning
Guanlin Liu, Lifeng LAI
Due to the broad range of applications of multi-agent reinforcement learning (MA RL), understanding the effects of adversarial attacks against MARL model is esse ntial for the safe applications of this model. Motivated by this, we investigate the impact of adversarial attacks on MARL. In the considered setup, there is an exogenous attacker who is able to modify the rewards before the agents receive them or manipulate the actions before the environment receives them. The attacke r aims to guide each agent into a target policy or maximize the cumulative rewar ds under some specific reward function chosen by the attacker, while minimizing the amount of the manipulation on feedback and action. We first show the limitat ions of the action poisoning only attacks and the reward poisoning only attacks. We then introduce a mixed attack strategy with both the action poisoning and re

ward poisoning. We show that the mixed attack strategy can efficiently attack MARL agents even if the attacker has no prior information about the underlying environment and the agents' algorithms.
****

Similarity-based cooperative equilibrium
Caspar Oesterheld, Johannes Treutlein, Roger B. Grosse, Vincent Conitzer, Jakob Foerster
As machine learning agents act more autonomously in the world, they will increasingly interact with each other. Unfortunately, in many social dilemmas like the one-shot Prisoner's Dilemma, standard game theory predicts that ML agents will fail to cooperate with each other. Prior work has shown that one way to enable cooperative outcomes in the one-shot Prisoner's Dilemma is to make the agents mutually transparent to each other, i.e., to allow them to access one another's source code (Rubinstein, 1998; Tennenholtz, 2004) – or weights in the case of ML agents. However, full transparency is often unrealistic, whereas partial transparency is commonplace. Moreover, it is challenging for agents to learn their way to cooperation in the full transparency setting. In this paper, we introduce a more realistic setting in which agents only observe a single number indicating how similar they are to each other. We prove that this allows for the same set of cooperative outcomes as the full transparency setting. We also demonstrate experimentally that cooperation can be learned using simple ML methods.
****

Preference-grounded Token-level Guidance for Language Model Fine-tuning
Shentao Yang, Shujian Zhang, Congying Xia, Yihao Feng, Caiming Xiong, Mingyuan Zhou
Aligning language models (LMs) with preferences is an important problem in natural language generation. A key challenge is that preferences are typically provided at the sequence level while LM training and generation both occur at the token level. There is, therefore, a granularity mismatch between the preference and the LM training losses, which may complicate the learning problem. In this paper, we address this issue by developing an alternate training process, where we iterate between grounding the sequence-level preference into token-level training guidance, and improving the LM with the learned guidance. For guidance learning, we design a framework that extends the pairwise-preference learning in imitation learning to both variable-length LM generation and the utilization of the preference among multiple generations. For LM training, based on the amount of supervised data, we present two minimalist learning objectives that utilize the learned guidance. In experiments, our method performs competitively on two distinct representative LM tasks --- discrete-prompt generation and text summarization.
****

Joint Feature and Differentiable $ k $-NN Graph Learning using Dirichlet Energy
Lei Xu, Lei Chen, Rong Wang, Feiping Nie, Xuelong Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Transformers learn through gradual rank increase
Enric Boix-Adsera, Etai Littwin, Emmanuel Abbe, Samy Bengio, Joshua Susskind
We identify incremental learning dynamics in transformers, where the difference between trained and initial weights progressively increases in rank. We rigorously prove this occurs under the simplifying assumptions of diagonal weight matrices and small initialization. Our experiments support the theory and also show that phenomenon can occur in practice without the simplifying assumptions.
****

SiT Dataset: Socially Interactive Pedestrian Trajectory Dataset for Social Navigation Robots
Jong Wook Bae, Jungho Kim, Junyong Yun, Changwon Kang, Jeongseon Choi, Chanhyeok Kim, Junho Lee, Jungwook Choi, Jun Won Choi
To ensure secure and dependable mobility in environments shared by humans and ro

bots, social navigation robots should possess the capability to accurately perceive and predict the trajectories of nearby pedestrians. In this paper, we present a novel dataset of pedestrian trajectories, referred to as Social Interactive Trajectory (SiT) dataset, which can be used to train pedestrian detection, tracking, and trajectory prediction models needed to design social navigation robots. Our dataset includes sequential raw data captured by two 3D LiDARs and five cameras covering a 360-degree view, two inertial measurement unit (IMU) sensors, and real-time kinematic positioning (RTK), as well as annotations including 2D & 3D boxes, object classes, and object IDs. Thus far, various human trajectory datasets have been introduced to support the development of pedestrian motion forecasting models. Our SiT dataset differs from these datasets in the following two respects. First, whereas the pedestrian trajectory data in other datasets was obtained from static scenes, our data was collected while the robot navigates in a crowded environment, capturing human-robot interactive scenarios in motion. Second, our dataset has been carefully organized to facilitate training and evaluation of end-to-end prediction models encompassing 3D detection, 3D multi-object tracking, and trajectory prediction. This design allows for an end-to-end unified modular approach across different tasks. We have introduced a comprehensive benchmark for assessing models across all aforementioned tasks, and have showcased the performance of multiple baseline models as part of our evaluation. Our dataset provides a strong foundation for future research in pedestrian trajectory prediction, which could expedite the development of safe and agile social navigation robots. The SiT dataset, devkit, and pre-trained models are publicly released at: https://spalaboratory.github.io/SiT
****

Prototype-based Aleatoric Uncertainty Quantification for Cross-modal Retrieval
Hao Li, Jingkuan Song, Lianli Gao, Xiaosu Zhu, Hengtao Shen

Cross-modal Retrieval methods build similarity relations between vision and language modalities by jointly learning a common representation space. However, the predictions are often unreliable due to the Aleatoric uncertainty, which is induced by low-quality data, e.g., corrupt images, fast-paced videos, and non-detailed texts. In this paper, we propose a novel Prototype-based Aleatoric Uncertainty Quantification (PAU) framework to provide trustworthy predictions by quantifying the uncertainty arisen from the inherent data ambiguity. Concretely, we first construct a set of various learnable prototypes for each modality to represent the entire semantics subspace. Then Dempster-Shafer Theory and Subjective Logic Theory are utilized to build an evidential theoretical framework by associating evidence with Dirichlet Distribution parameters. The PAU model induces accurate uncertainty and reliable predictions for cross-modal retrieval. Extensive experiments are performed on four major benchmark datasets of MSR-VTT, MSVD, DiDeMo, and MS-COCO, demonstrating the effectiveness of our method. The code is accessible at https://github.com/leolee99/PAU.
****

A-NeSI: A Scalable Approximate Method for Probabilistic Neurosymbolic Inference
Emile van Krieken, Thiviyan Thanapalasingam, Jakub Tomczak, Frank van Harmelen, Annette Ten Teije

We study the problem of combining neural networks with symbolic reasoning. Recently introduced frameworks for Probabilistic Neurosymbolic Learning (PNL), such as DeepProbLog, perform exponential-time exact inference, limiting the scalability of PNL solutions. We introduce Approximate Neurosymbolic Inference (A-NeSI): a new framework for PNL that uses neural networks for scalable approximate inference. A-NeSI 1) performs approximate inference in polynomial time without changing the semantics of probabilistic logics; 2) is trained using data generated by the background knowledge; 3) can generate symbolic explanations of predictions; and 4) can guarantee the satisfaction of logical constraints at test time, which is vital in safety-critical applications. Our experiments show that A-NeSI is the first end-to-end method to solve three neurosymbolic tasks with exponential combinatorial scaling. Finally, our experiments show that A-NeSI achieves explainability and safety without a penalty in performance.
****

Global Convergence Analysis of Local SGD for Two-layer Neural Network without Overparameterization
Yajie Bao, Amarda Shehu, Mingrui Liu
Local SGD, a cornerstone algorithm in federated learning, is widely used in training deep neural networks and shown to have strong empirical performance. A theoretical understanding of such performance on nonconvex loss landscapes is currently lacking. Analysis of the global convergence of SGD is challenging, as the noise depends on the model parameters. Indeed, many works narrow their focus to GD and rely on injecting noise to enable convergence to the local or global optimum. When expanding the focus to local SGD, existing analyses in the nonconvex case can only guarantee finding stationary points or assume the neural network is overparameterized so as to guarantee convergence to the global minimum through neural tangent kernel analysis. In this work, we provide the first global convergence analysis of the vanilla local SGD for two-layer neural networks \emph{without overparameterization} and \textit{without injecting noise}, when the input data is Gaussian. The main technical ingredients of our proof are \textit{a self-correction mechanism} and \textit{a new exact recursive characterization of the direction of global model parameters}. The self-correction mechanism guarantees the algorithm reaches a good region even if the initialization is in a bad region. A good (bad) region means updating the model by gradient descent will move closer to (away from) the optimal solution. The main difficulty in establishing a self-correction mechanism is to cope with the gradient dependency between two layers. To address this challenge, we divide the landscape of the objective into several regions to carefully control the interference of two layers during the correction process. As a result, we show that local SGD can correct the two layers and enter the good region in polynomial time. After that, we establish a new exact recursive characterization of the direction of global parameters, which is the key to showing convergence to the global minimum with linear speedup in the number of machines and reduced communication rounds. Experiments on synthetic data confirm theoretical results.
****
MuSe-GNN: Learning Unified Gene Representation From Multimodal Biological Graph Data
Tianyu Liu, Yuge Wang, Rex Ying, Hongyu Zhao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
BeaverTails: Towards Improved Safety Alignment of LLM via a Human-Preference Dataset
Jiaming Ji, Mickel Liu, Josef Dai, Xuehai Pan, Chi Zhang, Ce Bian, Boyuan Chen, Ruiyang Sun, Yizhou Wang, Yaodong Yang
In this paper, we introduce the BeaverTails dataset, aimed at fostering research on safety alignment in large language models (LLMs). This dataset uniquely separates annotations of helpfulness and harmlessness for question-answering pairs, thus offering distinct perspectives on these crucial attributes. In total, we have gathered safety meta-labels for 333,963 question-answer (QA) pairs and 361,903 pairs of expert comparison data for both the helpfulness and harmlessness metrics. We further showcase applications of BeaverTails in content moderation and reinforcement learning with human feedback (RLHF), emphasizing its potential for practical safety measures in LLMs. We believe this dataset provides vital resources for the community, contributing towards the safe development and deployment of LLMs. Our project page is available at the following URL: https://sites.google.com/view/pku-beavertails.
****
Reconstructing the Mind's Eye: fMRI-to-Image with Contrastive Learning and Diffusion Priors
Paul Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, ethan cohen, Aidan Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kennet

h Norman, Tanishq Abraham
We present MindEye, a novel fMRI-to-image approach to retrieve and reconstruct viewed images from brain activity. Our model comprises two parallel submodules that are specialized for retrieval (using contrastive learning) and reconstruction (using a diffusion prior). MindEye can map fMRI brain activity to any high dimensional multimodal latent space, like CLIP image space, enabling image reconstruction using generative models that accept embeddings from this latent space. We comprehensively compare our approach with other existing methods, using both qualitative side-by-side comparisons and quantitative evaluations, and show that MindEye achieves state-of-the-art performance in both reconstruction and retrieval tasks. In particular, MindEye can retrieve the exact original image even among highly similar candidates indicating that its brain embeddings retain fine-grained image-specific information. This allows us to accurately retrieve images even from large-scale databases like LAION-5B. We demonstrate through ablations that MindEye's performance improvements over previous methods result from specialized submodules for retrieval and reconstruction, improved training techniques, and training models with orders of magnitude more parameters. Furthermore, we show that MindEye can better preserve low-level image features in the reconstructions by using img2img, with outputs from a separate autoencoder. All code is available on GitHub.
****
Exploring Why Object Recognition Performance Degrades Across Income Levels and Geographies with Factor Annotations
Laura Gustafson, Megan Richards, Melissa Hall, Caner Hazirbas, Diane Bouchacourt, Mark Ibrahim
Despite impressive advances in object-recognition, deep learning systems’ performance degrades significantly across geographies and lower income levels---raising pressing concerns of inequity. Addressing such performance gaps remains a challenge, as little is understood about why performance degrades across incomes or geographies.We take a step in this direction by annotating images from Dollar Street, a popular benchmark of geographically and economically diverse images, labeling each image with factors such as color, shape, and background. These annotations unlock a new granular view into how objects differ across incomes/regions. We then use these object differences to pinpoint model vulnerabilities across incomes and regions.We study a range of modern vision models, finding that performance disparities are most associated with differences in texture, occlusion, and images with darker lighting.We illustrate how insights from our factor labels can surface mitigations to improve models' performance disparities.As an example, we show that mitigating a model's vulnerability to texture can improve performance on the lower income level.We release all the factor annotations along with an interactive dashboardto facilitate research into more equitable vision systems.
****
Improving Few-Shot Generalization by Exploring and Exploiting Auxiliary Data
Alon Albalak, Colin A. Raffel, William Yang Wang
Few-shot learning is valuable in many real-world applications, but learning a generalizable model without overfitting to the few labeled datapoints is challenging.In this work, we focus on Few-shot Learning with Auxiliary Data (FLAD), a training paradigm that assumes access to auxiliary data during few-shot learning in hopes of improving generalization.Previous works have proposed automated methods for mixing auxiliary and target data, but these methods typically scale linearly (or worse) with the number of auxiliary datasets, limiting their practicality.In this work we relate FLAD to the explore-exploit dilemma that is central to the multi-armed bandit setting and derive algorithms whose computational complexity is independent of the number of auxiliary datasets, allowing us to scale to 100x more auxiliary datasets than prior methods.We propose two algorithms -- EXP3-FLAD and UCB1-FLAD -- and compare them with prior FLAD methods that either explore or exploit, finding that the combination of exploration and exploitation is crucial.Through extensive experimentation we find that our methods outperform all pre-existing FLAD methods by 4% and lead to the first 3 billion parameter lang

uage models that outperform the 175 billion parameter GPT-3.Overall, our work suggests that the discovery of better, more efficient mixing strategies for FLAD may provide a viable path towards substantially improving generalization in few-shot learning.
****
Outlier-Robust Gromov-Wasserstein for Graph Data
Lemin Kong, Jiajin Li, Jianheng Tang, Anthony Man-Cho So
Gromov-Wasserstein (GW) distance is a powerful tool for comparing and aligning probability distributions supported on different metric spaces. Recently, GW has become the main modeling technique for aligning heterogeneous data for a wide range of graph learning tasks. However, the GW distance is known to be highly sensitive to outliers, which can result in large inaccuracies if the outliers are given the same weight as other samples in the objective function. To mitigate this issue, we introduce a new and robust version of the GW distance called RGW. RGW features optimistically perturbed marginal constraints within a Kullback-Leibler divergence-based ambiguity set. To make the benefits of RGW more accessible in practice, we develop a computationally efficient and theoretically provable procedure using Bregman proximal alternating linearized minimization algorithm. Through extensive experimentation, we validate our theoretical results and demonstrate the effectiveness of RGW on real-world graph learning tasks, such as subgraph matching and partial shape correspondence.
****
Labeling Neural Representations with Inverse Recognition
Kirill Bykov, Laura Kopf, Shinichi Nakajima, Marius Kloft, Marina Höhne
Deep Neural Networks (DNNs) demonstrate remarkable capabilities in learning complex hierarchical data representations, but the nature of these representations remains largely unknown. Existing global explainability methods, such as Network Dissection, face limitations such as reliance on segmentation masks, lack of statistical significance testing, and high computational demands. We propose Inverse Recognition (INVERT), a scalable approach for connecting learned representations with human-understandable concepts by leveraging their capacity to discriminate between these concepts. In contrast to prior work, INVERT is capable of handling diverse types of neurons, exhibits less computational complexity, and does not rely on the availability of segmentation masks. Moreover, INVERT provides an interpretable metric assessing the alignment between the representation and its corresponding explanation and delivering a measure of statistical significance. We demonstrate the applicability of INVERT in various scenarios, including the identification of representations affected by spurious correlations, and the interpretation of the hierarchical structure of decision-making within the models.
****
Cross-modal Active Complementary Learning with Self-refining Correspondence
Yang Qin, Yuan Sun, Dezhong Peng, Joey Tianyi Zhou, Xi Peng, Peng Hu
Recently, image-text matching has attracted more and more attention from academia and industry, which is fundamental to understanding the latent correspondence across visual and textual modalities. However, most existing methods implicitly assume the training pairs are well-aligned while ignoring the ubiquitous annotation noise, a.k.a noisy correspondence (NC), thereby inevitably leading to a performance drop. Although some methods attempt to address such noise, they still face two challenging problems: excessive memorizing/overfitting and unreliable correction for NC, especially under high noise. To address the two problems, we propose a generalized Cross-modal Robust Complementary Learning framework (CRCL), which benefits from a novel Active Complementary Loss (ACL) and an efficient Self-refining Correspondence Correction (SCC) to improve the robustness of existing methods. Specifically, ACL exploits active and complementary learning losses to reduce the risk of providing erroneous supervision, leading to theoretically and experimentally demonstrated robustness against NC. SCC utilizes multiple self-refining processes with momentum correction to enlarge the receptive field for correcting correspondences, thereby alleviating error accumulation and achieving accurate and stable corrections. We carry out extensive experiments on three image-text benchmarks, i.e., Flickr30K, MS-COCO, and CC152K, to verify the superio

r robustness of our CRCL against synthetic and real-world noisy correspondences.
****

Cinematic Mindscapes: High-quality Video Reconstruction from Brain Activity
Zijiao Chen, Jiaxin Qing, Juan Helen Zhou
Reconstructing human vision from brain activities has been an appealing task that helps to understand our cognitive process. Even though recent research has seen great success in reconstructing static images from non-invasive brain recordings, work on recovering continuous visual experiences in the form of videos is limited. In this work, we propose Mind-Video that learns spatiotemporal information from continuous fMRI data of the cerebral cortex progressively through masked brain modeling, multimodal contrastive learning with spatiotemporal attention, and co-training with an augmented Stable Diffusion model that incorporates network temporal inflation. We show that high-quality videos of arbitrary frame rates can be reconstructed with Mind-Video using adversarial guidance. The recovered videos were evaluated with various semantic and pixel-level metrics. We achieved an average accuracy of 85% in semantic classification tasks and 0.19 in structural similarity index (SSIM), outperforming the previous state-of-the-art by 45%. We also show that our model is biologically plausible and interpretable, reflecting established physiological processes.
****

Retrieval-Augmented Multiple Instance Learning
Yufei CUI, Ziquan Liu, Yixin Chen, Yuchen Lu, Xinyue Yu, Xue (Steve) Liu, Tei-Wei Kuo, Miguel Rodrigues, Chun Jason Xue, Antoni Chan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Multi-task Graph Neural Architecture Search with Task-aware Collaboration and Curriculum
Yijian Qin, Xin Wang, Ziwei Zhang, Hong Chen, Wenwu Zhu
Graph neural architecture search (GraphNAS) has shown great potential for automatically designing graph neural architectures for graph related tasks. However, multi-task GraphNAS capable of handling multiple tasks simultaneously has been largely unexplored in literature, posing great challenges to capture the complex relations and influences among different tasks. To tackle this problem, we propose a novel multi-task graph neural architecture search with task-aware collaboration and curriculum (MTGC3), which is able to simultaneously discover optimal architectures for different tasks and learn the collaborative relationships among different tasks in a joint manner. Specifically, we design the layer-wise disentangled supernet capable of managing multiple architectures in a unified framework, which combines with our proposed soft task-collaborative module to learn the transferability relationships between tasks. We further develop the task-wise curriculum training strategy to improve the architecture search procedure via reweighing the influence of different tasks based on task difficulties. Extensive experiments show that our proposed MTGC3 model achieves state-of-the-art performance against several baselines in multi-task scenarios, demonstrating its ability to discover effective architectures and capture the collaborative relationships for multiple tasks.
****

The Impact of Positional Encoding on Length Generalization in Transformers
Amirhossein Kazemnejad, Inkit Padhi, Karthikeyan Natesan Ramamurthy, Payel Das, Siva Reddy
Length generalization, the ability to generalize from small training context sizes to larger ones, is a critical challenge in the development of Transformer-based language models. Positional encoding (PE) has been identified as a major factor influencing length generalization, but the exact impact of different PE schemes on extrapolation in downstream tasks remains unclear. In this paper, we conduct a systematic empirical study comparing the length generalization performance of decoder-only Transformers with five different position encoding approaches in

cluding Absolute Position Embedding (APE), T5's Relative PE, ALiBi, and Rotary, in addition to Transformers without positional encoding (NoPE). Our evaluation encompasses a battery of reasoning and mathematical tasks. Our findings reveal that the most commonly used positional encoding methods, such as ALiBi, Rotary, and APE, are not well suited for length generalization in downstream tasks. More importantly, NoPE outperforms other explicit positional encoding methods while requiring no additional computation. We theoretically demonstrate that NoPE can represent both absolute and relative PEs, but when trained with SGD, it mostly resembles T5's relative PE attention patterns. Finally, we find that scratchpad is not always helpful to solve length generalization and its format highly impacts the model's performance. Overall, our work suggests that explicit position embeddings are not essential for decoder-only Transformers to generalize well to longer sequences.
****
Attention as Implicit Structural Inference
Ryan Singh, Christopher L Buckley
Attention mechanisms play a crucial role in cognitive systems by allowing them to flexibly allocate cognitive resources. Transformers, in particular, have become a dominant architecture in machine learning, with attention as their central innovation. However, the underlying intuition and formalism of attention in Transformers is based on ideas of keys and queries in database management systems. In this work, we pursue a structural inference perspective, building upon, and bringing together, previous theoretical descriptions of attention such as; Gaussian Mixture Models, alignment mechanisms and Hopfield Networks. Specifically, we demonstrate that attention can be viewed as inference over an implicitly defined set of possible adjacency structures in a graphical model, revealing the generality of such a mechanism. This perspective unifies different attentional architectures in machine learning and suggests potential modifications and generalizations of attention. Here we investigate two and demonstrate their behaviour on explanatory toy problems: (a) extending the value function to incorporate more nodes of a graphical model yielding a mechanism with a bias toward attending multiple tokens; (b) introducing a geometric prior (with conjugate hyper-prior) over the adjacency structures producing a mechanism which dynamically scales the context window depending on input. Moreover, by describing a link between structural inference and precision-regulation in Predictive Coding Networks, we discuss how this framework can bridge the gap between attentional mechanisms in machine learning and Bayesian conceptions of attention in Neuroscience. We hope by providing a new lens on attention architectures our work can guide the development of new and improved attentional mechanisms.
****
Nearly Tight Bounds For Differentially Private Multiway Cut
Mina Dalirrooyfard, Slobodan Mitrovic, Yuriy Nevmyvaka
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Permutation Equivariant Neural Functionals
Allan Zhou, Kaien Yang, Kaylee Burns, Adriano Cardace, Yiding Jiang, Samuel Sokota, J. Zico Kolter, Chelsea Finn
This work studies the design of neural networks that can process the weights or gradients of other neural networks, which we refer to as neural functional networks (NFNs). Despite a wide range of potential applications, including learned optimization, processing implicit neural representations, network editing, and policy evaluation, there are few unifying principles for designing effective architectures that process the weights of other networks. We approach the design of neural functionals through the lens of symmetry, in particular by focusing on the permutation symmetries that arise in the weights of deep feedforward networks because hidden layer neurons have no inherent order. We introduce a framework for building permutation equivariant neural functionals, whose architectures encode

these symmetries as an inductive bias. The key building blocks of this framework are NF-Layers (neural functional layers) that we constrain to be permutation equivariant through an appropriate parameter sharing scheme. In our experiments, we find that permutation equivariant neural functionals are effective on a diverse set of tasks that require processing the weights of MLPs and CNNs, such as predicting classifier generalization, producing "winning ticket" sparsity masks for initializations, and classifying or editing implicit neural representations (INRs). In addition, we provide code for our models and experiments at https://github.com/AllanYangZhou/nfn.

****

Fine-Grained Visual Prompting

Lingfeng Yang, Yueze Wang, Xiang Li, Xinlong Wang, Jian Yang

Vision-Language Models (VLMs), such as CLIP, have demonstrated impressive zero-shot transfer capabilities in image-level visual perception. However, these models have shown limited performance in instance-level tasks that demand precise localization and recognition. Previous works have suggested that incorporating visual prompts, such as colorful boxes or circles, can improve the ability of models to recognize objects of interest. Nonetheless, compared to language prompting, visual prompting designs are rarely explored. Existing approaches, which employ coarse visual cues such as colorful boxes or circles, often result in sub-optimal performance due to the inclusion of irrelevant and noisy pixels. In this paper, we carefully study the visual prompting designs by exploring more fine-grained markings, such as segmentation masks and their variations. In addition, we introduce a new zero-shot framework that leverages pixel-level annotations acquired from a generalist segmentation model for fine-grained visual prompting. Consequently, our investigation reveals that a straightforward application of blur outside the target mask, referred to as the Blur Reverse Mask, exhibits exceptional effectiveness. This proposed prompting strategy leverages the precise mask annotations to reduce focus on weakly related regions while retaining spatial coherence between the target and the surrounding background. Our Fine-Grained Visual Prompting (FGVP) demonstrates superior performance in zero-shot comprehension of referring expressions on the RefCOCO, RefCOCO+, and RefCOCOg benchmarks. It outperforms prior methods by an average margin of 3.0\% to 4.6\%, with a maximum improvement of 12.5\% on the RefCOCO+ testA subset. The part detection experiments conducted on the PACO dataset further validate the preponderance of FGVP over existing visual prompting techniques. Code is available at https://github.com/ylingfeng/FGVP.

****

A Multi-modal Global Instance Tracking Benchmark (MGIT): Better Locating Target in Complex Spatio-temporal and Causal Relationship

Shiyu Hu, Dailing Zhang, wu meiqi, Xiaokun Feng, Xuchen Li, Xin Zhao, Kaiqi Huang

Tracking an arbitrary moving target in a video sequence is the foundation for high-level tasks like video understanding. Although existing visual-based trackers have demonstrated good tracking capabilities in short video sequences, they always perform poorly in complex environments, as represented by the recently proposed global instance tracking task, which consists of longer videos with more complicated narrative content. Recently, several works have introduced natural language into object tracking, desiring to address the limitations of relying only on a single visual modality. However, these selected videos are still short sequences with uncomplicated spatio-temporal and causal relationships, and the provided semantic descriptions are too simple to characterize video content.To address these issues, we (1) first propose a new multi-modal global instance tracking benchmark named MGIT. It consists of 150 long video sequences with a total of 2.03 million frames, aiming to fully represent the complex spatio-temporal and causal relationships coupled in longer narrative content. (2) Each video sequence is annotated with three semantic grains (i.e., action, activity, and story) to model the progressive process of human cognition. We expect this multi-granular annotation strategy can provide a favorable environment for multi-modal object tracking research and long video understanding. (3) Besides, we execute comparative

experiments on existing multi-modal object tracking benchmarks, which not only e xplore the impact of different annotation methods, but also validate that our an notation method is a feasible solution for coupling human understanding into sem antic labels. (4) Additionally, we conduct detailed experimental analyses on MGI T, and hope the explored performance bottlenecks of existing algorithms can supp ort further research in multi-modal object tracking. The proposed benchmark, exp erimental results, and toolkit will be released gradually on http://videocube.a itestunion.com/.
****
Integration-free Training for Spatio-temporal Multimodal Covariate Deep Kernel P oint Processes
YIXUAN ZHANG, Quyu Kong, Feng Zhou
In this study, we propose a novel deep spatio-temporal point process model, Deep Kernel Mixture Point Processes (DKMPP), that incorporates multimodal covariate information. DKMPP is an enhanced version of Deep Mixture Point Processes (DMPP) , which uses a more flexible deep kernel to model complex relationships between events and covariate data, improving the model's expressiveness. To address the intractable training procedure of DKMPP due to the non-integrable deep kernel, w e utilize an integration-free method based on score matching, and further improv e efficiency by adopting a scalable denoising score matching method. Our experim ents demonstrate that DKMPP and its corresponding score-based estimators outperf orm baseline models, showcasing the advantages of incorporating covariate inform ation, utilizing a deep kernel, and employing score-based estimators.
****
Does progress on ImageNet transfer to real-world datasets?
Alex Fang, Simon Kornblith, Ludwig Schmidt
Does progress on ImageNet transfer to real-world datasets? We investigate this q uestion by evaluating ImageNet pre-trained models with varying accuracy (57% - 8 3%) on six practical image classification datasets. In particular, we study data sets collected with the goal of solving real-world tasks (e.g., classifying imag es from camera traps or satellites), as opposed to web-scraped benchmarks collec ted for comparing models. On multiple datasets, models with higher ImageNet accu racy do not consistently yield performance improvements. For certain tasks, inte rventions such as data augmentation improve performance even when architectures do not. We hope that future benchmarks will include more diverse datasets to enc ourage a more comprehensive approach to improving learning algorithms.
****
EmbodiedGPT: Vision-Language Pre-Training via Embodied Chain of Thought
Yao Mu, Qinglong Zhang, Mengkang Hu, Wenhai Wang, Mingyu Ding, Jun Jin, Bin Wang , Jifeng Dai, Yu Qiao, Ping Luo
Embodied AI is a crucial frontier in robotics, capable of planning and executing action sequences for robots to accomplish long-horizon tasks in physical enviro nments.In this work, we introduce EmbodiedGPT, an end-to-end multi-modal foundat ion model for embodied AI, empowering embodied agents with multi-modal understan ding and execution capabilities. To achieve this, we have made the following eff orts: (i) We craft a large-scale embodied planning dataset, termed EgoCOT. The d ataset consists of carefully selected videos from the Ego4D dataset, along with corresponding high-quality language instructions. Specifically, we generate a se quence of sub-goals with the "Chain of Thoughts" mode for effective embodied pla nning.(ii) We introduce an efficient training approach to EmbodiedGPT for high-q uality plan generation, by adapting a 7B large language model (LLM) to the EgoCO T dataset via prefix tuning. (iii) We introduce a paradigm for extracting task-r elated features from LLM-generated planning queries to form a closed loop betwee n high-level planning and low-level control.Extensive experiments show the effec tiveness of EmbodiedGPT on embodied tasks, including embodied planning, embodied control, visual captioning, and visual question answering.Notably, EmbodiedGPT significantly enhances the success rate of the embodied control task by extracti ng more effective features. It has achieved a remarkable 1.6 times increase in s uccess rate on the Franka Kitchen benchmark and a 1.3 times increase on the Meta -World benchmark, compared to the BLIP-2 baseline fine-tuned with the Ego4D data

set.
****
Conditional Matrix Flows for Gaussian Graphical Models
Marcello Massimo Negri, Fabricio Arend Torres, Volker Roth
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Additive Decoders for Latent Variables Identification and Cartesian-Product Extrapolation
Sébastien Lachapelle, Divyat Mahajan, Ioannis Mitliagkas, Simon Lacoste-Julien
We tackle the problems of latent variables identification and "out-of-support'' image generation in representation learning. We show that both are possible for a class of decoders that we call additive, which are reminiscent of decoders used for object-centric representation learning (OCRL) and well suited for images that can be decomposed as a sum of object-specific images. We provide conditions under which exactly solving the reconstruction problem using an additive decoder is guaranteed to identify the blocks of latent variables up to permutation and block-wise invertible transformations. This guarantee relies only on very weak assumptions about the distribution of the latent factors, which might present statistical dependencies and have an almost arbitrarily shaped support. Our result provides a new setting where nonlinear independent component analysis (ICA) is possible and adds to our theoretical understanding of OCRL methods. We also show theoretically that additive decoders can generate novel images by recombining observed factors of variations in novel ways, an ability we refer to as Cartesian-product extrapolation. We show empirically that additivity is crucial for both identifiability and extrapolation on simulated data.
****
How2comm: Communication-Efficient and Collaboration-Pragmatic Multi-Agent Perception
Dingkang Yang, Kun Yang, Yuzheng Wang, Jing Liu, Zhi Xu, Rongbin Yin, Peng Zhai, Lihua Zhang
Multi-agent collaborative perception has recently received widespread attention as an emerging application in driving scenarios. Despite the advancements in previous efforts, challenges remain due to various noises in the perception procedure, including communication redundancy, transmission delay, and collaboration heterogeneity. To tackle these issues, we propose \textit{How2comm}, a collaborative perception framework that seeks a trade-off between perception performance and communication bandwidth. Our novelties lie in three aspects. First, we devise a mutual information-aware communication mechanism to maximally sustain the informative features shared by collaborators. The spatial-channel filtering is adopted to perform effective feature sparsification for efficient communication. Second, we present a flow-guided delay compensation strategy to predict future characteristics from collaborators and eliminate feature misalignment due to temporal asynchrony. Ultimately, a pragmatic collaboration transformer is introduced to integrate holistic spatial semantics and temporal context clues among agents. Our framework is thoroughly evaluated on several LiDAR-based collaborative detection datasets in real-world and simulated scenarios. Comprehensive experiments demonstrate the superiority of How2comm and the effectiveness of all its vital components. The code will be released at https://github.com/ydk122024/How2comm.
****
LANCE: Stress-testing Visual Models by Generating Language-guided Counterfactual Images
Viraj Prabhu, Sriram Yenamandra, Prithvijit Chattopadhyay, Judy Hoffman
We propose an automated algorithm to stress-test a trained visual model by generating language-guided counterfactual test images (LANCE). Our method leverages recent progress in large language modeling and text-based image editing to augment an IID test set with a suite of diverse, realistic, and challenging test images without altering model weights. We benchmark the performance of a diverse set

of pre-trained models on our generated data and observe significant and consistent performance drops. We further analyze model sensitivity across different types of edits, and demonstrate its applicability at surfacing previously unknown class-level model biases in ImageNet. Code is available at https://github.com/virajprabhu/lance.
****
Refined Mechanism Design for Approximately Structured Priors via Active Regression
Christos Boutsikas, Petros Drineas, Marios Mertzanidis, Alexandros Psomas, Paritosh Verma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Most Neural Networks Are Almost Learnable
Amit Daniely, Nati Srebro, Gal Vardi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Bounded rationality in structured density estimation
Tianyuan Teng, Kevin Li, Hang Zhang
Learning to accurately represent environmental uncertainty is crucial for adaptive and optimal behaviors in various cognitive tasks. However, it remains unclear how the human brain, constrained by finite cognitive resources, constructs an internal model from an infinite space of probability distributions. In this study, we explore how these learned distributions deviate from the ground truth, resulting in observable inconsistency in a novel structured density estimation task. During each trial, human participants were asked to form and report the latent probability distribution functions underlying sequentially presented independent observations. As the number of observations increased, the reported predictive density became closer to the ground truth. Nevertheless, we observed an intriguing inconsistency in human structure estimation, specifically a large error in the number of reported clusters. Such inconsistency is invariant to the scale of the distribution and persists across stimulus modalities. We modeled uncertainty learning as approximate Bayesian inference in a nonparametric mixture prior of distributions. Human reports were best explained under resource rationality embodied in a decaying tendency towards model expansion. Our study offers insights into human cognitive processes under uncertainty and lays the groundwork for further exploration of resource-rational representations in the brain under more complex tasks.
****
RayDF: Neural Ray-surface Distance Fields with Multi-view Consistency
Zhuoman Liu, Bo Yang, Yan Luximon, Ajay Kumar, Jinxi Li
In this paper, we study the problem of continuous 3D shape representations. The majority of existing successful methods are coordinate-based implicit neural representations. However, they are inefficient to render novel views or recover explicit surface points. A few works start to formulate 3D shapes as ray-based neural functions, but the learned structures are inferior due to the lack of multi-view geometry consistency. To tackle these challenges, we propose a new framework called RayDF. It consists of three major components: 1) the simple ray-surface distance field, 2) the novel dual-ray visibility classifier, and 3) a multi-view consistency optimization module to drive the learned ray-surface distances to be multi-view geometry consistent. We extensively evaluate our method on three public datasets, demonstrating remarkable performance in 3D surface point reconstruction on both synthetic and challenging real-world 3D scenes, clearly surpassing existing coordinate-based and ray-based baselines. Most notably, our method achieves a 1000x faster speed than coordinate-based methods to render an 800x800 d

epth image, showing the superiority of our method for 3D shape representation. Our code and data are available at https://github.com/vLAR-group/RayDF
****
Motion-X: A Large-scale 3D Expressive Whole-body Human Motion Dataset
Jing Lin, Ailing Zeng, Shunlin Lu, Yuanhao Cai, Ruimao Zhang, Haoqian Wang, Lei Zhang
In this paper, we present Motion-X, a large-scale 3D expressive whole-body motion dataset. Existing motion datasets predominantly contain body-only poses, lacking facial expressions, hand gestures, and fine-grained pose descriptions. Moreover, they are primarily collected from limited laboratory scenes with textual descriptions manually labeled, which greatly limits their scalability. To overcome these limitations, we develop a whole-body motion and text annotation pipeline, which can automatically annotate motion from either single- or multi-view videos and provide comprehensive semantic labels for each video and fine-grained whole-body pose descriptions for each frame. This pipeline is of high precision, cost-effective, and scalable for further research. Based on it, we construct Motion-X, which comprises 15.6M precise 3D whole-body pose annotations (i.e., SMPL-X) covering 81.1K motion sequences from massive scenes. Besides, Motion-X provides 15.6M frame-level whole-body pose descriptions and 81.1K sequence-level semantic labels. Comprehensive experiments demonstrate the accuracy of the annotation pipeline and the significant benefit of Motion-X in enhancing expressive, diverse, and natural motion generation, as well as 3D whole-body human mesh recovery.
****
Blocked Collaborative Bandits: Online Collaborative Filtering with Per-Item Budget Constraints
Soumyabrata Pal, Arun Suggala, Karthikeyan Shanmugam, Prateek Jain
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Efficient Online Clustering with Moving Costs
Dimitrios Christou, Stratis Skoulakis, Volkan Cevher
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SEGA: Instructing Text-to-Image Models using Semantic Guidance
Manuel Brack, Felix Friedrich, Dominik Hintersdorf, Lukas Struppek, Patrick Schramowski, Kristian Kersting
Text-to-image diffusion models have recently received a lot of interest for their astonishing ability to produce high-fidelity images from text only. However, achieving one-shot generation that aligns with the user’s intent is nearly impossible, yet small changes to the input prompt often result in very different images. This leaves the user with little semantic control. To put the user in control, we show how to interact with the diffusion process to flexibly steer it along semantic directions. This semantic guidance (SEGA) generalizes to any generative architecture using classifier-free guidance. More importantly, it allows for subtle and extensive edits, composition and style changes, and optimizing the overall artistic conception. We demonstrate SEGA’s effectiveness on both latent and pixel-based diffusion models such as Stable Diffusion, Paella, and DeepFloyd-IF using a variety of tasks, thus providing strong evidence for its versatility and flexibility.
****
Learning Sample Difficulty from Pre-trained Models for Reliable Prediction
Peng Cui, Dan Zhang, Zhijie Deng, Yinpeng Dong, Jun Zhu
Large-scale pre-trained models have achieved remarkable success in many applications, but how to leverage them to improve the prediction reliability of downstream models is undesirably under-explored. Moreover, modern neural networks have b

een found to be poorly calibrated and make overconfident predictions regardless of inherent sample difficulty and data uncertainty. To address this issue, we propose to utilize large-scale pre-trained models to guide downstream model training with sample difficulty-aware entropy regularization. Pre-trained models that have been exposed to large-scale datasets and do not overfit the downstream training classes enable us to measure each training sample's difficulty via feature-space Gaussian modeling and relative Mahalanobis distance computation. Importantly, by adaptively penalizing overconfident prediction based on the sample difficulty, we simultaneously improve accuracy and uncertainty calibration across challenging benchmarks (e.g., +0.55% ACC and −3.7% ECE on ImageNet1k using ResNet34), consistently surpassing competitive baselines for reliable prediction. The improved uncertainty estimate further improves selective classification (abstaining from erroneous predictions) and out-of-distribution detection.
****
Asynchronous Proportional Response Dynamics: Convergence in Markets with Adversarial Scheduling
Yoav Kolumbus, Menahem Levy, Noam Nisan
We study Proportional Response Dynamics (PRD) in linear Fisher markets, where participants act asynchronously. We model this scenario as a sequential process in which at each step, an adversary selects a subset of the players to update their bids, subject to liveness constraints. We show that if every bidder individually applies the PRD update rule whenever they are included in the group of bidders selected by the adversary, then, in the generic case, the entire dynamic converges to a competitive equilibrium of the market. Our proof technique reveals additional properties of linear Fisher markets, such as the uniqueness of the market equilibrium for generic parameters and the convergence of associated no swap regret dynamics and best response dynamics under certain conditions.
****
Seeing is not always believing: Benchmarking Human and Model Perception of AI-Generated Images
Zeyu Lu, Di Huang, LEI BAI, Jingjing Qu, Chengyue Wu, Xihui Liu, Wanli Ouyang
Photos serve as a way for humans to record what they experience in their daily lives, and they are often regarded as trustworthy sources of information. However, there is a growing concern that the advancement of artificial intelligence (AI) technology may produce fake photos, which can create confusion and diminish trust in photographs. This study aims to comprehensively evaluate agents for distinguishing state-of-the-art AI-generated visual content. Our study benchmarks both human capability and cutting-edge fake image detection AI algorithms, using a newly collected large-scale fake image dataset Fake2M. In our human perception evaluation, titled HPBench, we discovered that humans struggle significantly to distinguish real photos from AI-generated ones, with a misclassification rate of 38.7\%. Along with this, we conduct the model capability of AI-Generated images detection evaluation MPBench and the top-performing model from MPBench achieves a 13\% failure rate under the same setting used in the human evaluation.We hope that our study can raise awareness of the potential risks of AI-generated images and facilitate further research to prevent the spread of false information. More information can refer to https://github.com/Inf-imagine/Sentry.
****
FORB: A Flat Object Retrieval Benchmark for Universal Image Embedding
Pengxiang Wu, Siman Wang, Kevin Dela Rosa, Derek Hu
Image retrieval is a fundamental task in computer vision. Despite recent advances in this field, many techniques have been evaluated on a limited number of domains, with a small number of instance categories. Notably, most existing works only consider domains like 3D landmarks, making it difficult to generalize the conclusions made by these works to other domains, e.g., logo and other 2D flat objects. To bridge this gap, we introduce a new dataset for benchmarking visual search methods on flat images with diverse patterns. Our flat object retrieval benchmark (FORB) supplements the commonly adopted 3D object domain, and more importantly, it serves as a testbed for assessing the image embedding quality on out-of-distribution domains. In this benchmark we investigate the retrieval accuracy of

representative methods in terms of candidate ranks, as well as matching score margin, a viewpoint which is largely ignored by many works. Our experiments not only highlight the challenges and rich heterogeneity of FORB, but also reveal the hidden properties of different retrieval strategies. The proposed benchmark is a growing project and we expect to expand in both quantity and variety of objects. The dataset and supporting codes are available at https://github.com/pxiangwu/FORB/.
****

Intra-Modal Proxy Learning for Zero-Shot Visual Categorization with CLIP

Qi Qian, Yuanhong Xu, Juhua Hu

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Overcoming Recency Bias of Normalization Statistics in Continual Learning: Balance and Adaptation

Yilin Lyu, Liyuan Wang, Xingxing Zhang, Zicheng Sun, Hang Su, Jun Zhu, Liping Jing

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

The Simplicity Bias in Multi-Task RNNs: Shared Attractors, Reuse of Dynamics, and Geometric Representation

Elia Turner, Omri Barak

How does a single interconnected neural population perform multiple tasks, each with its own dynamical requirements? The relation between task requirements and neural dynamics in Recurrent Neural Networks (RNNs) has been investigated for single tasks. The forces shaping joint dynamics of multiple tasks, however, are largely unexplored. In this work, we first construct a systematic framework to study multiple tasks in RNNs, minimizing interference from input and output correlations with the hidden representation. This allows us to reveal how RNNs tend to share attractors and reuse dynamics, a tendency we define as the "simplicity bias".We find that RNNs develop attractors sequentially during training, preferentially reusing existing dynamics and opting for simple solutions when possible. This sequenced emergence and preferential reuse encapsulate the simplicity bias. Through concrete examples, we demonstrate that new attractors primarily emerge due to task demands or architectural constraints, illustrating a balance between simplicity bias and external factors.We examine the geometry of joint representations within a single attractor, by constructing a family of tasks from a set of functions. We show that the steepness of the associated functions controls their alignment within the attractor. This arrangement again highlights the simplicity bias, as points with similar input spacings undergo comparable transformations to reach the shared attractor.Our findings propose compelling applications. The geometry of shared attractors might allow us to infer the nature of unknown tasks. Furthermore, the simplicity bias implies that without specific incentives, modularity in RNNs may not spontaneously emerge, providing insights into the conditions required for network specialization.
****

Optimize Planning Heuristics to Rank, not to Estimate Cost-to-Goal

Leah Chrestien, Stefan Edelkamp, Antonin Komenda, Tomas Pevny

In imitation learning for planning, parameters of heuristic functions are optimized against a set of solved problem instances. This work revisits the necessary and sufficient conditions of strictly optimally efficient heuristics for forward search algorithms, mainly A* and greedy best-first search, which expand only states on the returned optimal path. It then proposes a family of loss functions based on ranking tailored for a given variant of the forward search algorithm. Furthermore, from a learning theory point of view, it discusses why optimizing cos

t-to-goal h* is unnecessarily difficult. The experimental comparison on a diverse set of problems unequivocally supports the derived theory.
****
Goal-Conditioned Predictive Coding for Offline Reinforcement Learning
Zilai Zeng, Ce Zhang, Shijie Wang, Chen Sun
Recent work has demonstrated the effectiveness of formulating decision making as supervised learning on offline-collected trajectories. Powerful sequence models, such as GPT or BERT, are often employed to encode the trajectories. However, the benefits of performing sequence modeling on trajectory data remain unclear. In this work, we investigate whether sequence modeling has the ability to condense trajectories into useful representations that enhance policy learning. We adopt a two-stage framework that first leverages sequence models to encode trajectory-level representations, and then learns a goal-conditioned policy employing the encoded representations as its input. This formulation allows us to consider many existing supervised offline RL methods as specific instances of our framework. Within this framework, we introduce Goal-Conditioned Predictive Coding (GCPC), a sequence modeling objective that yields powerful trajectory representations and leads to performant policies. Through extensive empirical evaluations on AntMaze, FrankaKitchen and Locomotion environments, we observe that sequence modeling can have a significant impact on challenging decision making tasks. Furthermore, we demonstrate that GCPC learns a goal-conditioned latent representation encoding the future trajectory, which enables competitive performance on all three benchmarks.
****
Exposing Attention Glitches with Flip-Flop Language Modeling
Bingbin Liu, Jordan Ash, Surbhi Goel, Akshay Krishnamurthy, Cyril Zhang
Why do large language models sometimes output factual inaccuracies and exhibit erroneous reasoning? The brittleness of these models, particularly when executing long chains of reasoning, currently seems to be an inevitable price to pay for their advanced capabilities of coherently synthesizing knowledge, pragmatics, and abstract thought. Towards making sense of this fundamentally unsolved problem, this work identifies and analyzes the phenomenon of attention glitches, in which the Transformer architecture's inductive biases intermittently fail to capture robust reasoning. To isolate the issue, we introduce flip-flop language modeling (FFLM), a parametric family of synthetic benchmarks designed to probe the extrapolative behavior of neural language models. This simple generative task requires a model to copy binary symbols over long-range dependencies, ignoring the tokens in between. We find that Transformer FFLMs suffer from a long tail of sporadic reasoning errors, some of which we can eliminate using various regularization techniques. Our preliminary mechanistic analyses show why the remaining errors may be very difficult to diagnose and resolve. We hypothesize that attention glitches account for (some of) the closed-domain hallucinations in natural LLMs.
****
Information Design in Multi-Agent Reinforcement Learning
Yue Lin, Wenhao Li, Hongyuan Zha, Baoxiang Wang
Reinforcement learning (RL) is inspired by the way human infants and animals learn from the environment. The setting is somewhat idealized because, in actual tasks, other agents in the environment have their own goals and behave adaptively to the ego agent. To thrive in those environments, the agent needs to influence other agents so their actions become more helpful and less harmful. Research in computational economics distills two ways to influence others directly: by providing tangible goods (mechanism design) and by providing information (information design). This work investigates information design problems for a group of RL agents. The main challenges are two-fold. One is the information provided will immediately affect the transition of the agent trajectories, which introduces additional non-stationarity. The other is the information can be ignored, so the sender must provide information that the receiver is willing to respect. We formulate the Markov signaling game, and develop the notions of signaling gradient and the extended obedience constraints that address these challenges. Our algorithm is efficient on various mixed-motive tasks and provides further insights into co

mputational economics. Our code is publicly available at https://github.com/YueLin301/InformationDesignMARL.
****
Gaussian Mixture Solvers for Diffusion Models
Hanzhong Guo, Cheng Lu, Fan Bao, Tianyu Pang, Shuicheng Yan, Chao Du, Chongxuan LI
Recently, diffusion models have achieved great success in generative tasks. Sampling from diffusion models is equivalent to solving the reverse diffusion stochastic differential equations (SDEs) or the corresponding probability flow ordinary differential equations (ODEs). In comparison, SDE-based solvers can generate samples of higher quality and are suited for image translation tasks like stroke-based synthesis. During inference, however, existing SDE-based solvers are severely constrained by the efficiency-effectiveness dilemma. Our investigation suggests that this is because the Gaussian assumption in the reverse transition kernel is frequently violated (even in the case of simple mixture data) given a limited number of discretization steps. To overcome this limitation, we introduce a novel class of SDE-based solvers called \emph{Gaussian Mixture Solvers (GMS)} for diffusion models. Our solver estimates the first three-order moments and optimizes the parameters of a Gaussian mixture transition kernel using generalized methods of moments in each step during sampling. Empirically, our solver outperforms numerous SDE-based solvers in terms of sample quality in image generation and stroke-based synthesis in various diffusion models, which validates the motivation and effectiveness of GMS. Our code is available at https://github.com/Guohanzhong/GMS.
****
Trade-off Between Efficiency and Consistency for Removal-based Explanations
Yifan Zhang, Haowei He, Zhiquan Tan, Yang Yuan
In the current landscape of explanation methodologies, most predominant approaches, such as SHAP and LIME, employ removal-based techniques to evaluate the impact of individual features by simulating various scenarios with specific features omitted. Nonetheless, these methods primarily emphasize efficiency in the original context, often resulting in general inconsistencies. In this paper, we demonstrate that such inconsistency is an inherent aspect of these approaches by establishing the Impossible Trinity Theorem, which posits that interpretability, efficiency, and consistency cannot hold simultaneously. Recognizing that the attainment of an ideal explanation remains elusive, we propose the utilization of interpretation error as a metric to gauge inefficiencies and inconsistencies. To this end, we present two novel algorithms founded on the standard polynomial basis, aimed at minimizing interpretation error. Our empirical findings indicate that the proposed methods achieve a substantial reduction in interpretation error, up to 31.8 times lower when compared to alternative techniques.
****
QuACK: Accelerating Gradient-Based Quantum Optimization with Koopman Operator Learning
Di Luo, Jiayu Shen, Rumen Dangovski, Marin Soljacic
Quantum optimization, a key application of quantum computing, has traditionally been stymied by the linearly increasing complexity of gradient calculations with an increasing number of parameters. This work bridges the gap between Koopman operator theory, which has found utility in applications because it allows for a linear representation of nonlinear dynamical systems, and natural gradient methods in quantum optimization, leading to a significant acceleration of gradient-based quantum optimization. We present Quantum-circuit Alternating Controlled Koopman learning (QuACK), a novel framework that leverages an alternating algorithm for efficient prediction of gradient dynamics on quantum computers. We demonstrate QuACK's remarkable ability to accelerate gradient-based optimization across a range of applications in quantum optimization and machine learning. In fact, our empirical studies, spanning quantum chemistry, quantum condensed matter, quantum machine learning, and noisy environments, have shown accelerations of more than 200x speedup in the overparameterized regime, 10x speedup in the smooth regime, and 3x speedup in the non-smooth regime. With QuACK, we offer a robust advanc

ement that harnesses the advantage of gradient-based quantum optimization for practical benefits.
****
Provably Robust Temporal Difference Learning for Heavy-Tailed Rewards
Semih Cayci, Atilla Eryilmaz
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Train Faster, Perform Better: Modular Adaptive Training in Over-Parameterized Models
Yubin Shi, Yixuan Chen, Mingzhi Dong, Xiaochen Yang, Dongsheng Li, Yujiang Wang, Robert Dick, Qin Lv, Yingying Zhao, Fan Yang, Tun Lu, Ning Gu, Li Shang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Class-Distribution-Aware Pseudo-Labeling for Semi-Supervised Multi-Label Learning
Ming-Kun Xie, Jiahao Xiao, Hao-Zhe Liu, Gang Niu, Masashi Sugiyama, Sheng-Jun Huang
Pseudo-labeling has emerged as a popular and effective approach for utilizing unlabeled data. However, in the context of semi-supervised multi-label learning (SSMLL), conventional pseudo-labeling methods encounter difficulties when dealing with instances associated with multiple labels and an unknown label count. These limitations often result in the introduction of false positive labels or the neglect of true positive ones. To overcome these challenges, this paper proposes a novel solution called Class-Aware Pseudo-Labeling (CAP) that performs pseudo-labeling in a class-aware manner. The proposed approach introduces a regularized learning framework incorporating class-aware thresholds, which effectively control the assignment of positive and negative pseudo-labels for each class. Notably, even with a small proportion of labeled examples, our observations demonstrate that the estimated class distribution serves as a reliable approximation. Motivated by this finding, we develop a class-distribution-aware thresholding strategy to ensure the alignment of pseudo-label distribution with the true distribution. The correctness of the estimated class distribution is theoretically verified, and a generalization error bound is provided for our proposed method. Extensive experiments on multiple benchmark datasets confirm the efficacy of CAP in addressing the challenges of SSMLL problems.
****
Adaptive Data Analysis in a Balanced Adversarial Model
Kobbi Nissim, Uri Stemmer, Eliad Tsfadia
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Accelerating Molecular Graph Neural Networks via Knowledge Distillation
Filip Ekström Kelvinius, Dimitar Georgiev, Artur Toshev, Johannes Gasteiger
Recent advances in graph neural networks (GNNs) have enabled more comprehensive modeling of molecules and molecular systems, thereby enhancing the precision of molecular property prediction and molecular simulations. Nonetheless, as the field has been progressing to bigger and more complex architectures, state-of-the-art GNNs have become largely prohibitive for many large-scale applications. In this paper, we explore the utility of knowledge distillation (KD) for accelerating molecular GNNs. To this end, we devise KD strategies that facilitate the distillation of hidden representations in directional and equivariant GNNs, and evaluate their performance on the regression task of energy and force prediction. We v

alidate our protocols across different teacher-student configurations and datasets, and demonstrate that they can consistently boost the predictive accuracy of student models without any modifications to their architecture. Moreover, we conduct comprehensive optimization of various components of our framework, and investigate the potential of data augmentation to further enhance performance. All in all, we manage to close the gap in predictive accuracy between teacher and student models by as much as 96.7\% and 62.5\% for energy and force prediction respectively, while fully preserving the inference throughput of the more lightweight models.
****
No Train No Gain: Revisiting Efficient Training Algorithms For Transformer-based Language Models
Jean Kaddour, Oscar Key, Piotr Nawrot, Pasquale Minervini, Matt J. Kusner
The computation necessary for training Transformer-based language models has skyrocketed in recent years.This trend has motivated research on efficient training algorithms designed to improve training, validation, and downstream performance faster than standard training. In this work, we revisit three categories of such algorithms: dynamic architectures (layer stacking, layer dropping), batch selection (selective backprop., RHO-loss), and efficient optimizers (Lion, Sophia). When pre-training BERT and T5 with a fixed computation budget using such methods, we find that their training, validation, and downstream gains vanish compared to a baseline with a fully-decayed learning rate. We define an evaluation protocol that enables computation to be done on arbitrary machines by mapping all computation time to a reference machine which we call reference system time. We discuss the limitations of our proposed protocol and release our code to encourage rigorous research in efficient training procedures: https://github.com/JeanKaddour/NoTrainNoGain.
****
Layer-Neighbor Sampling --- Defusing Neighborhood Explosion in GNNs
Muhammed Fatih Balin, Ümit Çatalyürek
Graph Neural Networks (GNNs) have received significant attention recently, but training them at a large scale remains a challenge.Mini-batch training coupled with sampling is used to alleviate this challenge.However, existing approaches either suffer from the neighborhood explosion phenomenon or have suboptimal performance. To address these issues, we propose a new sampling algorithm called LAyer-neighBOR sampling (LABOR). It is designed to be a direct replacement for Neighbor Sampling (NS) with the same fanout hyperparameter while sampling up to 7 times fewer vertices, without sacrificing quality.By design, the variance of the estimator of each vertex matches NS from the point of view of a single vertex.Moreover, under the same vertex sampling budget constraints, LABOR converges faster than existing layer sampling approaches and can use up to 112 times larger batch sizes compared to NS.
****
Undirected Probabilistic Model for Tensor Decomposition
Zerui Tao, Toshihisa Tanaka, Qibin Zhao
Tensor decompositions (TDs) serve as a powerful tool for analyzing multiway data. Traditional TDs incorporate prior knowledge about the data into the model, such as a directed generative process from latent factors to observations. In practice, selecting proper structural or distributional assumptions beforehand is crucial for obtaining a promising TD representation. However, since such prior knowledge is typically unavailable in real-world applications, choosing an appropriate TD model can be challenging. This paper aims to address this issue by introducing a flexible TD framework that discards the structural and distributional assumptions, in order to learn as much information from the data. Specifically, we construct a TD model that captures the joint probability of the data and latent tensor factors through a deep energy-based model (EBM). Neural networks are then employed to parameterize the joint energy function of tensor factors and tensor entries. The flexibility of EBM and neural networks enables the learning of underlying structures and distributions. In addition, by designing the energy function, our model unifies the learning process of different types of tensors, such

as static tensors and dynamic tensors with time stamps. The resulting model presents a doubly intractable nature due to the presence of latent tensor factors and the unnormalized probability function. To efficiently train the model, we derive a variational upper bound of the conditional noise-contrastive estimation objective that learns the unnormalized joint probability by distinguishing data from conditional noises. We show advantages of our model on both synthetic and several real-world datasets.
****
Rethinking Tokenizer and Decoder in Masked Graph Modeling for Molecules
ZHIYUAN LIU, Yaorui Shi, An Zhang, Enzhi Zhang, Kenji Kawaguchi, Xiang Wang, Tat-Seng Chua
Masked graph modeling excels in the self-supervised representation learning of molecular graphs. Scrutinizing previous studies, we can reveal a common scheme consisting of three key components: (1) graph tokenizer, which breaks a molecular graph into smaller fragments (\ie subgraphs) and converts them into tokens; (2) graph masking, which corrupts the graph with masks; (3) graph autoencoder, which first applies an encoder on the masked graph to generate the representations, and then employs a decoder on the representations to recover the tokens of the original graph. However, the previous MGM studies focus extensively on graph masking and encoder, while there is limited understanding of tokenizer and decoder. To bridge the gap, we first summarize popular molecule tokenizers at the granularity of node, edge, motif, and Graph Neural Networks (GNNs), and then examine their roles as the MGM's reconstruction targets. Further, we explore the potential of adopting an expressive decoder in MGM. Our results show that a subgraph-level tokenizer and a sufficiently expressive decoder with remask decoding have a \yuan{large impact on the encoder's representation learning}. Finally, we propose a novel MGM method SimSGT, featuring a Simple GNN-based Tokenizer (SGT) and an effective decoding strategy. We empirically validate that our method outperforms the existing molecule self-supervised learning methods. Our codes and checkpoints are available at https://github.com/syr-cn/SimSGT.
****
Shared Adversarial Unlearning: Backdoor Mitigation by Unlearning Shared Adversarial Examples
Shaokui Wei, Mingda Zhang, Hongyuan Zha, Baoyuan Wu
Backdoor attacks are serious security threats to machine learning models where an adversary can inject poisoned samples into the training set, causing a backdoored model which predicts poisoned samples with particular triggers to particular target classes, while behaving normally on benign samples. In this paper, we explore the task of purifying a backdoored model using a small clean dataset. By establishing the connection between backdoor risk and adversarial risk, we derive a novel upper bound for backdoor risk, which mainly captures the risk on the shared adversarial examples (SAEs) between the backdoored model and the purified model. This upper bound further suggests a novel bi-level optimization problem for mitigating backdoor using adversarial training techniques. To solve it, we propose Shared Adversarial Unlearning (SAU). Specifically, SAU first generates SAEs, and then, unlearns the generated SAEs such that they are either correctly classified by the purified model and/or differently classified by the two models, such that the backdoor effect in the backdoored model will be mitigated in the purified model. Experiments on various benchmark datasets and network architectures show that our proposed method achieves state-of-the-art performance for backdoor defense. The code is available at https://github.com/SCLBD/BackdoorBench (PyTorch) and https://github.com/shawkui/MindTrojan (MindSpore).
****
Calibration by Distribution Matching: Trainable Kernel Calibration Metrics
Charlie Marx, Sofian Zalouk, Stefano Ermon
Calibration ensures that probabilistic forecasts meaningfully capture uncertainty by requiring that predicted probabilities align with empirical frequencies. However, many existing calibration methods are specialized for post-hoc recalibration, which can worsen the sharpness of forecasts. Drawing on the insight that calibration can be viewed as a distribution matching task, we introduce kernel-bas

ed calibration metrics that unify and generalize popular forms of calibration for both classification and regression. These metrics admit differentiable sample estimates, making it easy to incorporate a calibration objective into empirical risk minimization. Furthermore, we provide intuitive mechanisms to tailor calibration metrics to a decision task, and enforce accurate loss estimation and no regret decisions. Our empirical evaluation demonstrates that employing these metrics as regularizers enhances calibration, sharpness, and decision-making across a range of regression and classification tasks, outperforming methods relying solely on post-hoc recalibration.
****
Polynomially Over-Parameterized Convolutional Neural Networks Contain Structured Strong Winning Lottery Tickets
Arthur da Cunha, Francesco D'Amore, Natale
The Strong Lottery Ticket Hypothesis (SLTH) states that randomly-initialised neural networks likely contain subnetworks that perform well without any training. Although unstructured pruning has been extensively studied in this context, its structured counterpart, which can deliver significant computational and memory efficiency gains, has been largely unexplored. One of the main reasons for this gap is the limitations of the underlying mathematical tools used in formal analyses of the SLTH.In this paper, we overcome these limitations: we leverage recent advances in the multidimensional generalisation of the Random Subset-Sum Problem and obtain a variant that admits the stochastic dependencies that arise when addressing structured pruning in the SLTH. We apply this result to prove, for a wide class of random Convolutional Neural Networks, the existence of structured subnetworks that can approximate any sufficiently smaller network.This result provides the first sub-exponential bound around the SLTH for structured pruning, opening up new avenues for further research on the hypothesis and contributing to the understanding of the role of over-parameterization in deep learning.
****
Robustifying Generalizable Implicit Shape Networks with a Tunable Non-Parametric Model
Amine Ouasfi, Adnane Boukhayma
Feedforward generalizable models for implicit shape reconstruction from unoriented point cloud present multiple advantages, including high performance and inference speed. However, they still suffer from generalization issues, ranging from underfitting the input point cloud, to misrepresenting samples outside of the training data distribution, or with toplogies unseen at training. We propose here an efficient mechanism to remedy some of these limitations at test time. We combine the inter-shape data prior of the network with an intra-shape regularization prior of a Nyström Kernel Ridge Regression, that we further adapt by fitting its hyperprameters to the current shape. The resulting shape function defined in a shape specific Reproducing Kernel Hilbert Space benefits from desirable stability and efficiency properties and grants a shape adaptive expressiveness-robustness trade-off. We demonstrate the improvement obtained through our method with respect to baselines and the state-of-the-art using synthetic and real data.
****
Segment Anything in 3D with NeRFs
Jiazhong Cen, Zanwei Zhou, Jiemin Fang, chen yang, Wei Shen, Lingxi Xie, Dongsheng Jiang, XIAOPENG ZHANG, Qi Tian
Recently, the Segment Anything Model (SAM) emerged as a powerful vision foundation model which is capable to segment anything in 2D images. This paper aims to generalize SAM to segment 3D objects. Rather than replicating the data acquisition and annotation procedure which is costly in 3D, we design an efficient solution, leveraging the Neural Radiance Field (NeRF) as a cheap and off-the-shelf prior that connects multi-view 2D images to the 3D space. We refer to the proposed solution as SA3D, for Segment Anything in 3D. It is only required to provide a manual segmentation prompt (e.g., rough points) for the target object in a single view, which is used to generate its 2D mask in this view with SAM. Next, SA3D alternately performs mask inverse rendering and cross-view self-prompting across various views to iteratively complete the 3D mask of the target object constructe

d with voxel grids. The former projects the 2D mask obtained by SAM in the current view onto 3D mask with guidance of the density distribution learned by the NeRF; The latter extracts reliable prompts automatically as the input to SAM from the NeRF-rendered 2D mask in another view. We show in experiments that SA3D adapts to various scenes and achieves 3D segmentation within minutes. Our research offers a generic and efficient methodology to lift a 2D vision foundation model to 3D, as long as the 2D model can steadily address promptable segmentation across multiple views.

****

Every Parameter Matters: Ensuring the Convergence of Federated Learning with Dynamic Heterogeneous Models Reduction

Hanhan Zhou, Tian Lan, Guru Prasadh Venkataramani, Wenbo Ding

Cross-device Federated Learning (FL) faces significant challenges where low-end clients that could potentially make unique contributions are excluded from training large models due to their resource bottlenecks. Recent research efforts have focused on model-heterogeneous FL, by extracting reduced-size models from the global model and applying them to local clients accordingly. Despite the empirical success, general theoretical guarantees of convergence on this method remain an open question. This paper presents a unifying framework for heterogeneous FL algorithms with online model extraction and provides a general convergence analysis for the first time. In particular, we prove that under certain sufficient conditions and for both IID and non-IID data, these algorithms converge to a stationary point of standard FL for general smooth cost functions. Moreover, we introduce the concept of minimum coverage index, together with model reduction noise, which will determine the convergence of heterogeneous federated learning, and therefore we advocate for a holistic approach that considers both factors to enhance the efficiency of heterogeneous federated learning.

****

Small batch deep reinforcement learning

Johan Obando Ceron, Marc Bellemare, Pablo Samuel Castro

In value-based deep reinforcement learning with replay memories, the batch size parameter specifies how many transitions to sample for each gradient update. Although critical to the learning process, this value is typically not adjusted when proposing new algorithms. In this work we present a broad empirical study that suggests reducing the batch size can result in a number of significant performance gains; this is surprising, as the general tendency when training neural networks is towards larger batch sizes for improved performance. We complement our experimental findings with a set of empirical analyses towards better understanding this phenomenon.

****

A Deep Instance Generative Framework for MILP Solvers Under Limited Data Availability

Zijie Geng, Xijun Li, Jie Wang, Xiao Li, Yongdong Zhang, Feng Wu

In the past few years, there has been an explosive surge in the use of machine learning (ML) techniques to address combinatorial optimization (CO) problems, especially mixed-integer linear programs (MILPs). Despite the achievements, the limited availability of real-world instances often leads to sub-optimal decisions and biased solver assessments, which motivates a suite of synthetic MILP instance generation techniques. However, existing methods either rely heavily on expert-designed formulations or struggle to capture the rich features of real-world instances. To tackle this problem, we propose G2MILP, the first deep generative framework for MILP instances. Specifically, G2MILP represents MILP instances as bipartite graphs, and applies a masked variational autoencoder to iteratively corrupt and replace parts of the original graphs to generate new ones. The appealing feature of G2MILP is that it can learn to generate novel and realistic MILP instances without prior expert-designed formulations, while preserving the structures and computational hardness of real-world datasets, simultaneously. Thus the generated instances can facilitate downstream tasks for enhancing MILP solvers under limited data availability. We design a suite of benchmarks to evaluate the quality of the generated MILP instances. Experiments demonstrate that our method c

an produce instances that closely resemble real-world datasets in terms of both structures and computational hardness. The deliverables are released at https://miralab-ustc.github.io/L2O-G2MILP.
****

WordScape: a Pipeline to extract multilingual, visually rich Documents with Layout Annotations from Web Crawl Data
Maurice Weber, Carlo Siebenschuh, Rory Butler, Anton Alexandrov, Valdemar Thanner, Georgios Tsolakis, Haris Jabbar, Ian Foster, Bo Li, Rick Stevens, Ce Zhang
We introduce WordScape, a novel pipeline for the creation of cross-disciplinary, multilingual corpora comprising millions of pages with annotations for document layout detection. Relating visual and textual items on document pages has gained further significance with the advent of multimodal models. Various approaches proved effective for visual question answering or layout segmentation. However, the interplay of text, tables, and visuals remains challenging for a variety of document understanding tasks. In particular, many models fail to generalize well to diverse domains and new languages due to insufficient availability of training data. WordScape addresses these limitations. Our automatic annotation pipeline parses the Open XML structure of Word documents obtained from the web, jointly providing layout-annotated document images and their textual representations. In turn, WordScape offers unique properties as it (1) leverages the ubiquity of the Word file format on the internet, (2) is readily accessible through the Common Crawl web corpus, (3) is adaptive to domain-specific documents, and (4) offers culturally and linguistically diverse document pages with natural semantic structure and high-quality text. Together with the pipeline, we will additionally release 9.5M urls to word documents which can be processed using WordScape to create a dataset of over 40M pages. Finally, we investigate the quality of text and layout annotations extracted by WordScape, assess the impact on document understanding benchmarks, and demonstrate that manual labeling costs can be substantially reduced.
****

Multi-Swap k-Means++
Lorenzo Beretta, Vincent Cohen-Addad, Silvio Lattanzi, Nikos Parotsidis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

A Unified Discretization Framework for Differential Equation Approach with Lyapunov Arguments for Convex Optimization
Kansei Ushiyama, Shun Sato, Takayasu Matsuo
The differential equation (DE) approach for convex optimization, which relates optimization methods to specific continuous DEs with rate-revealing Lyapunov functionals, has gained increasing interest since the seminal paper by Su--Boyd--Candès (2014).However, the approach still lacks a crucial component to make it truly useful: there is no general, consistent way to transition back to discrete optimization methods. Consequently, even if we derive insights from continuous DEs, we still need to perform individualized and tedious calculations for the analysis of each method.This paper aims to bridge this gap by introducing a new concept called ``weak discrete gradient'' (wDG), which consolidates the conditions required for discrete versions of gradients in the DE approach arguments.We then define abstract optimization methods using wDG and provide abstract convergence theories that parallel those in continuous DEs.We demonstrate that many typical optimization methods and their convergence rates can be derived as special cases of this abstract theory.The proposed unified discretization framework for the differential equation approach to convex optimization provides an easy environment for developing new optimization methods and achieving competitive convergence rates with state-of-the-art methods, such as Nesterov's accelerated gradient.
****

SARAMIS: Simulation Assets for Robotic Assisted and Minimally Invasive Surgery
Nina Montana-Brown, Shaheer U. Saeed, Ahmed Abdulaal, Thomas Dowrick, Yakup Kili

c, Sophie Wilkinson, Jack Gao, Meghavi Mashar, Chloe He, Alkisti Stavropoulou, Emma Thomson, Zachary MC Baum, Simone Foti, Brian Davidson, Yipeng Hu, Matthew Clarkson
Minimally-invasive surgery (MIS) and robot-assisted minimally invasive (RAMIS) surgery offer well-documented benefits to patients such as reduced post-operative pain and shorter hospital stays.However, the automation of MIS and RAMIS through the use of AI has been slow due to difficulties in data acquisition and curation, partially caused by the ethical considerations of training, testing and deploying AI models in medical environments.We introduce \texttt{SARAMIS}, the first large-scale dataset of anatomically derived 3D rendering assets of the human abdominal anatomy.Using previously existing, open-source CT datasets of the human anatomy, we derive novel 3D meshes, tetrahedral volumes, textures and diffuse maps for over 104 different anatomical targets in the human body, representing the largest, open-source dataset of 3D rendering assets for synthetic simulation of vision tasks in MIS+RAMIS, increasing the availability of openly available 3D meshes in the literature by three orders of magnitude.We supplement our dataset with a series of GPU-enabled rendering environments, which can be used to generate datasets for realistic MIS/RAMIS tasks.Finally, we present an example of the use of \texttt{SARAMIS} assets for an autonomous navigation task in colonoscopy from CT abdomen-pelvis scans for the first time in the literature.\texttt{SARAMIS} is publically made available at https://github.com/NMontanaBrown/saramis/, with assets released under a CC-BY-NC-SA license.
****
Free-Bloom: Zero-Shot Text-to-Video Generator with LLM Director and LDM Animator
Hanzhuo Huang, Yufan Feng, Cheng Shi, Lan Xu, Jingyi Yu, Sibei Yang
Text-to-video is a rapidly growing research area that aims to generate a semantic, identical, and temporal coherence sequence of frames that accurately align with the input text prompt. This study focuses on zero-shot text-to-video generation considering the data- and cost-efficient. To generate a semantic-coherent video, exhibiting a rich portrayal of temporal semantics such as the whole process of flower blooming rather than a set of ``moving images'', we propose a novel Free-Bloom pipeline that harnesses large language models (LLMs) as the director to generate a semantic-coherence prompt sequence, while pre-trained latent diffusion models (LDMs) as the animator to generate the high fidelity frames. Furthermore, to ensure temporal and identical coherence while maintaining semantic coherence, we propose a series of annotative modifications to adapting LDMs in the reverse process, including joint noise sampling, step-aware attention shift, and dual-path interpolation. Without any video data and training requirements, Free-Bloom generates vivid and high-quality videos, awe-inspiring in generating complex scenes with semantic meaningful frame sequences. In addition, Free-Bloom is naturally compatible with LDMs-based extensions.
****
NeRF Revisited: Fixing Quadrature Instability in Volume Rendering
Mikaela Angelina Uy, Kiyohiro Nakayama, Guandao Yang, Rahul Thomas, Leonidas J. Guibas, Ke Li
Neural radiance fields (NeRF) rely on volume rendering to synthesize novel views. Volume rendering requires evaluating an integral along each ray, which is numerically approximated with a finite sum that corresponds to the exact integral along the ray under piecewise constant volume density. As a consequence, the rendered result is unstable w.r.t. the choice of samples along the ray, a phenomenon that we dub quadrature instability. We propose a mathematically principled solution by reformulating the sample-based rendering equation so that it corresponds to the exact integral under piecewise linear volume density. This simultaneously resolves multiple issues: conflicts between samples along different rays, imprecise hierarchical sampling, and non-differentiability of quantiles of ray termination distances w.r.t. model parameters. We demonstrate several benefits over the classical sample-based rendering equation, such as sharper textures, better geometric reconstruction, and stronger depth supervision. Our proposed formulation can be also be used as a drop-in replacement to the volume rendering equation of existing NeRF-based methods. Our project page can be found at pl-nerf.github.i

o.
****
Practical Sharpness-Aware Minimization Cannot Converge All the Way to Optima
Dongkuk Si, Chulhee Yun
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Online Convex Optimization with Unbounded Memory
Raunak Kumar, Sarah Dean, Robert Kleinberg
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Making Scalable Meta Learning Practical
Sang Choe, Sanket Vaibhav Mehta, Hwijeen Ahn, Willie Neiswanger, Pengtao Xie, Emma Strubell, Eric Xing
Despite its flexibility to learn diverse inductive biases in machine learning programs, meta learning (i.e.,\ learning to learn) has long been recognized to suffer from poor scalability due to its tremendous compute/memory costs, training instability, and a lack of efficient distributed training support. In this work, we focus on making scalable meta learning practical by introducing SAMA, which combines advances in both implicit differentiation algorithms and systems. Specifically, SAMA is designed to flexibly support a broad range of adaptive optimizers in the base level of meta learning programs, while reducing computational burden by avoiding explicit computation of second-order gradient information, and exploiting efficient distributed training techniques implemented for first-order gradients. Evaluated on multiple large-scale meta learning benchmarks, SAMA showcases up to 1.7/4.8x increase in throughput and 2.0/3.8x decrease in memory consumption respectively on single-/multi-GPU setups compared to other baseline meta learning algorithms. Furthermore, we show that SAMA-based data optimization leads to consistent improvements in text classification accuracy with BERT and RoBERTa large language models, and achieves state-of-the-art results in both small- and large-scale data pruning on image classification tasks, demonstrating the practical applicability of scalable meta learning across language and vision domains.
****
Dynamic Prompt Learning: Addressing Cross-Attention Leakage for Text-Based Image Editing
kai wang, Fei Yang, Shiqi Yang, Muhammad Atif Butt, Joost van de Weijer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Brant: Foundation Model for Intracranial Neural Signal
Daoze Zhang, Zhizhang Yuan, YANG YANG, Junru Chen, Jingjing Wang, Yafeng Li
We propose a foundation model named Brant for modeling intracranial recordings, which learns powerful representations of intracranial neural signals by pre-training, providing a large-scale, off-the-shelf model for medicine. Brant is the largest model in the field of brain signals and is pre-trained on a large corpus of intracranial data collected by us. The design of Brant is to capture long-term temporal dependency and spatial correlation from neural signals, combining the information in both time and frequency domains. As a foundation model, Brant achieves SOTA performance on various downstream tasks (i.e. neural signal forecasting, frequency-phase forecasting, imputation and seizure detection), showing the generalization ability to a broad range of tasks. The low-resource label analysis and representation visualization further illustrate the effectiveness of our p

re-training strategy. In addition, we explore the effect of model size to show that a larger model with a higher capacity can lead to performance improvements on our dataset. The source code and pre-trained weights are available at: https://zju-brainnet.github.io/Brant.github.io/.
****
Wasserstein distributional robustness of neural networks
Xingjian Bai, Guangyi He, Yifan Jiang, Jan Obloj
Deep neural networks are known to be vulnerable to adversarial attacks (AA). For an image recognition task, this means that a small perturbation of the original can result in the image being misclassified. Design of such attacks as well as methods of adversarial training against them are subject of intense research. We re-cast the problem using techniques of Wasserstein distributionally robust optimization (DRO) and obtain novel contributions leveraging recent insights from DRO sensitivity analysis. We consider a set of distributional threat models. Unlike the traditional pointwise attacks, which assume a uniform bound on perturbation of each input data point, distributional threat models allow attackers to perturb inputs in a non-uniform way. We link these more general attacks with questions of out-of-sample performance and Knightian uncertainty. To evaluate the distributional robustness of neural networks, we propose a first-order AA algorithm and its multistep version. Our attack algorithms include Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) as special cases. Furthermore, we provide a new asymptotic estimate of the adversarial accuracy against distributional threat models. The bound is fast to compute and first-order accurate, offering new insights even for the pointwise AA. It also naturally yields out-of-sample performance guarantees. We conduct numerical experiments on CIFAR-10, CIFAR-100, ImageNet datasets using DNNs on RobustBench to illustrate our theoretical results. Our code is available at https://github.com/JanObloj/W-DRO-Adversarial-Methods.
****
High dimensional, tabular deep learning with an auxiliary knowledge graph
Camilo Ruiz, Hongyu Ren, Kexin Huang, Jure Leskovec
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Space-Time Continuous Latent Neural PDEs from Partially Observed States
Valerii Iakovlev, Markus Heinonen, Harri Lähdesmäki
We introduce a novel grid-independent model for learning partial differential equations (PDEs) from noisy and partial observations on irregular spatiotemporal grids. We propose a space-time continuous latent neural PDE model with an efficient probabilistic framework and a novel encoder design for improved data efficiency and grid independence. The latent state dynamics are governed by a PDE model that combines the collocation method and the method of lines. We employ amortized variational inference for approximate posterior estimation and utilize a multiple shooting technique for enhanced training speed and stability. Our model demonstrates state-of-the-art performance on complex synthetic and real-world datasets, overcoming limitations of previous approaches and effectively handling partially-observed data. The proposed model outperforms recent methods, showing its potential to advance data-driven PDE modeling and enabling robust, grid-independent modeling of complex partially-observed dynamic processes across various domains.
****
Adaptive SGD with Polyak stepsize and Line-search: Robust Convergence and Variance Reduction
Xiaowen Jiang, Sebastian U. Stich
The recently proposed stochastic Polyak stepsize (SPS) and stochastic line-search (SLS) for SGD have shown remarkable effectiveness when training over-parameterized models. However, two issues remain unsolved in this line of work. First, in non-interpolation settings, both algorithms only guarantee convergence to a nei

ghborhood of a solution which may result in a worse output than the initial guess. While artificially decreasing the adaptive stepsize has been proposed to address this issue (Orvieto et al.), this approach results in slower convergence rates under interpolation. Second, intuitive line-search methods equipped with variance-reduction (VR) fail to converge (Dubois-Taine et al.). So far, no VR methods successfully accelerate these two stepsizes with a convergence guarantee.In this work, we make two contributions:Firstly, we propose two new robust variants of SPS and SLS, called AdaSPS and AdaSLS, which achieve optimal asymptotic rates in both strongly-convex or convex and interpolation or non-interpolation settings, except for the case when we have both strong convexity and non-interpolation. AdaSLS requires no knowledge of problem-dependent parameters, and AdaSPS requires only a lower bound of the optimal function value as input. Secondly, we propose a novel VR method that can use Polyak stepsizes or line-search to achieve acceleration. When it is equipped with AdaSPS or AdaSLS, the resulting algorithms obtain the optimal ratefor optimizing convex smooth functions. Finally, numerical experiments on synthetic and real datasets validate our theory and demonstrate the effectiveness and robustness of our algorithms.
****
SOC: Semantic-Assisted Object Cluster for Referring Video Object Segmentation
Zhuoyan Luo, Yicheng Xiao, Yong Liu, Shuyan Li, Yitong Wang, Yansong Tang, Xiu Li, Yujiu Yang
This paper studies referring video object segmentation (RVOS) by boosting video-level visual-linguistic alignment. Recent approaches model the RVOS task as a sequence prediction problem and perform multi-modal interaction as well as segmentation for each frame separately. However, the lack of a global view of video content leads to difficulties in effectively utilizing inter-frame relationships and understanding textual descriptions of object temporal variations. To address this issue, we propose Semantic-assisted Object Cluster (SOC), which aggregates video content and textual guidance for unified temporal modeling and cross-modal alignment. By associating a group of frame-level object embeddings with language tokens, SOC facilitates joint space learning across modalities and time steps. Moreover, we present multi-modal contrastive supervision to help construct well-aligned joint space at the video level. We conduct extensive experiments on popular RVOS benchmarks, and our method outperforms state-of-the-art competitors on all benchmarks by a remarkable margin. Besides, the emphasis on temporal coherence enhances the segmentation stability and adaptability of our method in processing text expressions with temporal variations. Code is available at https://github.com/RobertLuo1/NeurIPS2023_SOC.
****
Posthoc privacy guarantees for collaborative inference with modified Propose-Test-Release
Abhishek Singh, Praneeth Vepakomma, Vivek Sharma, Ramesh Raskar
Cloud-based machine learning inference is an emerging paradigm where users query by sending their data through a service provider who runs an ML model on that data and returns back the answer. Due to increased concerns over data privacy, recent works have proposed Collaborative Inference (CI) to learn a privacy-preserving encoding of sensitive user data before it is shared with an untrusted service provider. Existing works so far evaluate the privacy of these encodings through empirical reconstruction attacks. In this work, we develop a new framework that provides formal privacy guarantees for an arbitrarily trained neural network by linking its local Lipschitz constant with its local sensitivity. To guarantee privacy using local sensitivity, we extend the Propose-Test-Release (PTR) framework to make it tractable for neural network queries. We verify the efficacy of our framework experimentally on real-world datasets and elucidate the role of Adversarial Representation Learning (ARL) in improving the privacy-utility trade-off.
****
Strategic Classification under Unknown Personalized Manipulation
Han Shao, Avrim Blum, Omar Montasser
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
EPIC Fields: Marrying 3D Geometry and Video Understanding
Vadim Tschernezki, Ahmad Darkhalil, Zhifan Zhu, David Fouhey, Iro Laina, Diane Larlus, Dima Damen, Andrea Vedaldi
Neural rendering is fuelling a unification of learning, 3D geometry and video understanding that has been waiting for more than two decades. Progress, however, is still hampered by a lack of suitable datasets and benchmarks. To address this gap, we introduce EPIC Fields, an augmentation of EPIC-KITCHENS with 3D camera information. Like other datasets for neural rendering, EPIC Fields removes the complex and expensive step of reconstructing cameras using photogrammetry, and allows researchers to focus on modelling problems. We illustrate the challenge of photogrammetry in egocentric videos of dynamic actions and propose innovations to address them. Compared to other neural rendering datasets, EPIC Fields is better tailored to video understanding because it is paired with labelled action segments and the recent VISOR segment annotations. To further motivate the community, we also evaluate two benchmark tasks in neural rendering and segmenting dynamic objects, with strong baselines that showcase what is not possible today. We also highlight the advantage of geometry in semi-supervised video object segmentations on the VISOR annotations. EPIC Fields reconstructs 96\% of videos in EPIC-KITCHENS, registering 19M frames in 99 hours recorded in 45 kitchens, and is available from: http://epic-kitchens.github.io/epic-fields
****
On the Ability of Graph Neural Networks to Model Interactions Between Vertices
Noam Razin, Tom Verbin, Nadav Cohen
Graph neural networks (GNNs) are widely used for modeling complex interactions between entities represented as vertices of a graph. Despite recent efforts to theoretically analyze the expressive power of GNNs, a formal characterization of their ability to model interactions is lacking. The current paper aims to address this gap. Formalizing strength of interactions through an established measure known as separation rank, we quantify the ability of certain GNNs to model interaction between a given subset of vertices and its complement, i.e. between the sides of a given partition of input vertices. Our results reveal that the ability to model interaction is primarily determined by the partition's walk index --- a graph-theoretical characteristic defined by the number of walks originating from the boundary of the partition. Experiments with common GNN architectures corroborate this finding. As a practical application of our theory, we design an edge sparsification algorithm named Walk Index Sparsification (WIS), which preserves the ability of a GNN to model interactions when input edges are removed. WIS is simple, computationally efficient, and in our experiments has markedly outperformed alternative methods in terms of induced prediction accuracy. More broadly, it showcases the potential of improving GNNs by theoretically analyzing the interactions they can model.
****
Learning from Both Structural and Textual Knowledge for Inductive Knowledge Graph Completion
Kunxun Qi, Jianfeng Du, Hai Wan
Learning rule-based systems plays a pivotal role in knowledge graph completion (KGC). Existing rule-based systems restrict the input of the system to structural knowledge only, which may omit some useful knowledge for reasoning, e.g., textual knowledge. In this paper, we propose a two-stage framework that imposes both structural and textual knowledge to learn rule-based systems. In the first stage, we compute a set of triples with confidence scores (called \emph{soft triples}) from a text corpus by distant supervision, where a textual entailment model with multi-instance learning is exploited to estimate whether a given triple is entailed by a set of sentences. In the second stage, these soft triples are used to learn a rule-based model for KGC. To mitigate the negative impact of noise from soft triples, we propose a new formalism for rules to be learnt, named \emph{t

ext enhanced rules} or \emph{TE-rules} for short. To effectively learn TE-rules, we propose a neural model that simulates the inference of TE-rules. We theoretically show that any set of TE-rules can always be interpreted by a certain parameter assignment of the neural model. We introduce three new datasets to evaluate the effectiveness of our method. Experimental results demonstrate that the introduction of soft triples and TE-rules results in significant performance improvements in inductive link prediction.
****
Sorting with Predictions
Xingjian Bai, Christian Coester
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Posterior Sampling for Competitive RL: Function Approximation and Partial Observation
Shuang Qiu, Ziyu Dai, Han Zhong, Zhaoran Wang, Zhuoran Yang, Tong Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Towards Test-Time Refusals via Concept Negation
Peiran Dong, Song Guo, Junxiao Wang, Bingjie WANG, Jiewei Zhang, Ziming Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
LAMM: Language-Assisted Multi-Modal Instruction-Tuning Dataset, Framework, and Benchmark
Zhenfei Yin, Jiong Wang, Jianjian Cao, Zhelun Shi, Dingning Liu, Mukai Li, Xiaoshui Huang, Zhiyong Wang, Lu Sheng, LEI BAI, Jing Shao, Wanli Ouyang
Large language models have emerged as a promising approach towards achieving general-purpose AI agents. The thriving open-source LLM community has greatly accelerated the development of agents that support human-machine dialogue interaction through natural language processing. However, human interaction with the world extends beyond only text as a modality, and other modalities such as vision are also crucial. Recent works on multi-modal large language models, such as GPT-4V and Bard, have demonstrated their effectiveness in handling visual modalities. However, the transparency of these works is limited and insufficient to support academic research. To the best of our knowledge, we present one of the very first open-source endeavors in the field, LAMM, encompassing a Language-Assisted Multi-Modal instruction tuning dataset, framework, and benchmark. Our aim is to establish LAMM as a growing ecosystem for training and evaluating MLLMs, with a specific focus on facilitating AI agents capable of bridging the gap between ideas and execution, thereby enabling seamless human-AI interaction. Our main contribution is three-fold: 1) We present a comprehensive dataset and benchmark, which cover a wide range of vision tasks for 2D and 3D vision. Extensive experiments validate the effectiveness of our dataset and benchmark. 2) We outline the detailed methodology of constructing multi-modal instruction tuning datasets and benchmarks for MLLMs, enabling rapid scaling and extension of MLLM research to diverse domains, tasks, and modalities. 3) We provide a primary but potential MLLM training framework optimized for modality extension. We also provide baseline models, comprehensive experimental observations, and analysis to accelerate future research. Our baseline model is trained within 24 A100 GPU hours, framework supports training with V100 and RTX3090 is available thanks to the open-source society. Codes and data are now available at https://openlamm.github.io.
****

Likelihood Ratio Confidence Sets for Sequential Decision Making
Nicolas Emmenegger, Mojmir Mutny, Andreas Krause
Certifiable, adaptive uncertainty estimates for unknown quantities are an essent ial ingredient of sequential decision-making algorithms. Standard approaches rel y on problem-dependent concentration results and are limited to a specific combi nation of parameterization, noise family, and estimator. In this paper, we revis it the likelihood-based inference principle and propose to use \emph{likelihood ratios} to construct \emph{any-time valid} confidence sequences without requirin g specialized treatment in each application scenario. Our method is especially s uitable for problems with well-specified likelihoods, and the resulting sets alw ays maintain the prescribed coverage in a model-agnostic manner. The size of the sets depends on a choice of estimator sequence in the likelihood ratio. We disc uss how to provably choose the best sequence of estimators and shed light on con nections to online convex optimization with algorithms such as Follow-the-Regula rized-Leader. To counteract the initially large bias of the estimators, we propo se a reweighting scheme that also opens up deployment in non-parametric settings such as RKHS function classes. We provide a \emph{non-asymptotic} analysis of t he likelihood ratio confidence sets size for generalized linear models, using in sights from convex duality and online learning. We showcase the practical streng th of our method on generalized linear bandit problems, survival analysis, and b andits with various additive noise distributions.
****
Uncertainty Quantification over Graph with Conformalized Graph Neural Networks
Kexin Huang, Ying Jin, Emmanuel Candes, Jure Leskovec
Graph Neural Networks (GNNs) are powerful machine learning prediction models on graph-structured data. However, GNNs lack rigorous uncertainty estimates, limiti ng their reliable deployment in settings where the cost of errors is significant . We propose conformalized GNN (CF-GNN), extending conformal prediction (CP) to graph-based models for guaranteed uncertainty estimates. Given an entity in the graph, CF-GNN produces a prediction set/interval that provably contains the true label with pre-defined coverage probability (e.g. 90%). We establish a permutat ion invariance condition that enables the validity of CP on graph data and provi de an exact characterization of the test-time coverage. Moreover, besides valid coverage, it is crucial to reduce the prediction set size/interval length for pr actical use. We observe a key connection between non-conformity scores and netwo rk structures, which motivates us to develop a topology-aware output correction model that learns to update the prediction and produces more efficient predictio n sets/intervals. Extensive experiments show that CF-GNN achieves any pre-define d target marginal coverage while significantly reducing the prediction set/inter val size by up to 74% over the baselines. It also empirically achieves satisfact ory conditional coverage over various raw and network features.
****
EMBERSim: A Large-Scale Databank for Boosting Similarity Search in Malware Analy sis
Dragos Georgian Corlatescu, Alexandru Dinu, Mihaela Petruta Gaman, Paul Sumedrea
In recent years there has been a shift from heuristics based malware detection t owards machine learning, which proves to be more robust in the current heavily a dversarial threat landscape. While we acknowledge machine learning to be better equipped to mine for patterns in the increasingly high amounts of similar-lookin g files, we also note a remarkable scarcity of the data available for similarity targeted research. Moreover, we observe that the focus in the few related works falls on quantifying similarity in malware, often overlooking the clean data. T his one-sided quantification is especially dangerous in the context of detection bypass. We propose to address the deficiencies in the space of similarity resea rch on binary files, starting from EMBER — one of the largest malware classifica tion datasets. We enhance EMBER with similarity information as well as malware c lass tags, to enable further research in the similarity space. Our contribution is threefold: (1) we publish EMBERSim, an augmented version of EMBER, that inclu des similarity informed tags; (2) we enrich EMBERSim with automatically determin ed malware class tags using the open-source tool AVClass on VirusTotal data and

(3) we describe and share the implementation for our class scoring technique and leaf similarity method.
****
VPP: Efficient Conditional 3D Generation via Voxel-Point Progressive Representation
Zekun Qi, Muzhou Yu, Runpei Dong, Kaisheng Ma
Conditional 3D generation is undergoing a significant advancement, enabling the free creation of 3D content from inputs such as text or 2D images. However, previous approaches have suffered from low inference efficiency, limited generation categories, and restricted downstream applications. In this work, we revisit the impact of different 3D representations on generation quality and efficiency. We propose a progressive generation method through Voxel-Point Progressive Representation (VPP). VPP leverages structured voxel representation in the proposed Voxel Semantic Generator and the sparsity of unstructured point representation in the Point Upsampler, enabling efficient generation of multi-category objects. VPP can generate high-quality 8K point clouds within 0.2 seconds. Additionally, the masked generation Transformer allows for various 3D downstream tasks, such as generation, editing, completion, and pre-training. Extensive experiments demonstrate that VPP efficiently generates high-fidelity and diverse 3D shapes across different categories, while also exhibiting excellent representation transfer performance. Codes will be released at https://github.com/qizekun/VPP.
****
Multi Time Scale World Models
Vaisakh Shaj Kumar, SALEH GHOLAM ZADEH, Ozan Demir, Luiz Douat, Gerhard Neumann
Intelligent agents use internal world models to reason and make predictions about different courses of their actions at many scales. Devising learning paradigms and architectures that allow machines to learn world models that operate at multiple levels of temporal abstractions while dealing with complex uncertainty predictions is a major technical hurdle. In this work, we propose a probabilistic formalism to learn multi-time scale world models which we call the Multi Time Scale State Space (MTS3) model. Our model uses a computationally efficient inference scheme on multiple time scales for highly accurate long-horizon predictions and uncertainty estimates over several seconds into the future. Our experiments, which focus on action conditional long horizon future predictions, show that MTS3 outperforms recent methods on several system identification benchmarks including complex simulated and real-world dynamical systems. Code is available at this repository:https://github.com/ALRhub/MTS3.
****
How many samples are needed to leverage smoothness?
Vivien Cabannes, Stefano Vigogna
A core principle in statistical learning is that smoothness of target functions allows to break the curse of dimensionality. However, learning a smooth function seems to require enough samples close to one another to get meaningful estimate of high-order derivatives, which would be hard in machine learning problems where the ratio between number of data and input dimension is relatively small. By deriving new lower bounds on the generalization error, this paper formalizes such an intuition, before investigating the role of constants and transitory regimes which are usually not depicted beyond classical learning theory statements while they play a dominant role in practice.
****
Causal Imitability Under Context-Specific Independence Relations
Fateme Jamshidi, Sina Akbari, Negar Kiyavash
Drawbacks of ignoring the causal mechanisms when performing imitation learning have recently been acknowledged. Several approaches both to assess the feasibility of imitation and to circumvent causal confounding and causal misspecifications have been proposed in the literature.However, the potential benefits of the incorporation of additional information about the underlying causal structure are left unexplored.An example of such overlooked information is context-specific independence (CSI), i.e., independence that holds only in certain contexts.We consider the problem of causal imitation learning when CSI relations are known.We pro

ve that the decision problem pertaining to the feasibility of imitation in this setting is NP-hard.Further, we provide a necessary graphical criterion for imitation learning under CSI and show that under a structural assumption, this criterion is also sufficient.Finally, we propose a sound algorithmic approach for causal imitation learning which takes both CSI relations and data into account.
****

A Finite-Particle Convergence Rate for Stein Variational Gradient Descent
Jiaxin Shi, Lester Mackey
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Bringing regularized optimal transport to lightspeed: a splitting method adapted for GPUs
Jacob Lindbäck, Zesen Wang, Mikael Johansson
We present an efficient algorithm for regularized optimal transport. In contrast toprevious methods, we use the Douglas-Rachford splitting technique to developan efficient solver that can handle a broad class of regularizers. The algorithmhas strong global convergence guarantees, low per-iteration cost, and can exploit GPU parallelization, making it considerably faster than the state-of-the-art for many problems. We illustrate its competitiveness in several applications, includingdomain adaptation and learning of generative models.
****

Use perturbations when learning from explanations
Juyeon Heo, Vihari Piratla, Matthew Wicker, Adrian Weller
Machine learning from explanations (MLX) is an approach to learning that uses human-provided explanations of relevant or irrelevant features for each input to ensure that model predictions are right for the right reasons. Existing MLX approaches rely on local model interpretation methods and require strong model smoothing to align model and human explanations, leading to sub-optimal performance. We recast MLX as a robustness problem, where human explanations specify a lower dimensional manifold from which perturbations can be drawn, and show both theoretically and empirically how this approach alleviates the need for strong model smoothing. We consider various approaches to achieving robustness, leading to improved performance over prior MLX methods. Finally, we show how to combine robustness with an earlier MLX method, yielding state-of-the-art results on both synthetic and real-world benchmarks.
****

VisIT-Bench: A Dynamic Benchmark for Evaluating Instruction-Following Vision-and-Language Models
Yonatan Bitton, Hritik Bansal, Jack Hessel, Rulin Shao, Wanrong Zhu, Anas Awadalla, Josh Gardner, Rohan Taori, Ludwig Schmidt
We introduce VisIT-Bench (Visual InsTruction Benchmark), a benchmark for evaluating instruction-following vision-language models for real-world use. Our starting point is curating 70 "instruction families" that we envision instruction tuned vision-language models should be able to address. Extending beyond evaluations like VQAv2 and COCO, tasks range from basic recognition to game playing and creative generation. Following curation, our dataset comprises 592 test queries, each with a human-authored instruction-conditioned caption. These descriptions surface instruction-specific factors, e.g., for an instruction asking about the accessibility of a storefront for wheelchair users, the instruction-conditioned caption describes ramps/potential obstacles. These descriptions enable 1) collecting human-verified reference outputs for each instance; and 2) automatic evaluation of candidate multimodal generations using a text-only LLM, aligning with human judgment. We quantify quality gaps between models and references using both human and automatic evaluations; e.g., the top-performing instruction-following model wins against the GPT-4 reference in just 27% of the comparison. VisIT-Bench is dynamic to participate, practitioners simply submit their model's response on the project website; Data, code and leaderboard is available at https://visit-ben

ch.github.io/.
****
The Memory-Perturbation Equation: Understanding Model's Sensitivity to Data
Peter Nickl, Lu Xu, Dharmesh Tailor, Thomas Möllenhoff, Mohammad Emtiyaz E. Khan
Understanding model’s sensitivity to its training data is crucial but can also be challenging and costly, especially during training. To simplify such issues, we present the Memory-Perturbation Equation (MPE) which relates model's sensitivity to perturbation in its training data. Derived using Bayesian principles, the MPE unifies existing sensitivity measures, generalizes them to a wide-variety of models and algorithms, and unravels useful properties regarding sensitivities. Our empirical results show that sensitivity estimates obtained during training can be used to faithfully predict generalization on unseen test data. The proposed equation is expected to be useful for future research on robust and adaptive learning.
****
Contextual Gaussian Process Bandits with Neural Networks
Haoting Zhang, Jinghai He, Rhonda Righter, Zuo-Jun Shen, Zeyu Zheng
Contextual decision-making problems have witnessed extensive applications in various fields such as online content recommendation, personalized healthcare, and autonomous vehicles, where a core practical challenge is to select a suitable surrogate model for capturing unknown complicated reward functions. It is often the case that both high approximation accuracy and explicit uncertainty quantification are desired. In this work, we propose a neural network-accompanied Gaussian process (NN-AGP) model, which leverages neural networks to approximate the unknown and potentially complicated reward function regarding the contextual variable, and maintains a Gaussian process surrogate model with respect to the decision variable. Our model is shown to outperform existing approaches by offering better approximation accuracy thanks to the use of neural networks and possessing explicit uncertainty quantification from the Gaussian process. We also analyze the maximum information gain of the NN-AGP model and prove regret bounds for the corresponding algorithms. Moreover, we conduct experiments on both synthetic and practical problems, illustrating the effectiveness of our approach.
****
Auxiliary Losses for Learning Generalizable Concept-based Models
Ivaxi Sheth, Samira Ebrahimi Kahou
The increasing use of neural networks in various applications has lead to increasing apprehensions, underscoring the necessity to understand their operations beyond mere final predictions. As a solution to enhance model transparency, Concept Bottleneck Models (CBMs) have gained popularity since their introduction. CBMs essentially limit the latent space of a model to human-understandable high-level concepts. While beneficial, CBMs have been reported to often learn irrelevant concept representations that consecutively damage model performance. To overcome the performance trade-off, we propose a cooperative-Concept Bottleneck Model (coop-CBM). The concept representation of our model is particularly meaningful when fine-grained concept labels are absent. Furthermore, we introduce the concept orthogonal loss (COL) to encourage the separation between the concept representations and to reduce the intra-concept distance. This paper presents extensive experiments on real-world datasets for image classification tasks, namely CUB, AwA2, CelebA and TIL. We also study the performance of coop-CBM models under various distributional shift settings. We show that our proposed method achieves higher accuracy in all distributional shift settings even compared to the black-box models with the highest concept accuracy.
****
Make the U in UDA Matter: Invariant Consistency Learning for Unsupervised Domain Adaptation
Zhongqi Yue, QIANRU SUN, Hanwang Zhang
Domain Adaptation (DA) is always challenged by the spurious correlation between the domain-invariant features (e.g., class identity) and the domain-specific ones (e.g., environment) that does not generalize to the target domain. Unfortunately, even enriched with additional unsupervised target domains, existing Unsuperv

ised DA (UDA) methods still suffer from it. This is because the source domain supervision only considers the target domain samples as auxiliary data (e.g., by pseudo-labeling), yet the inherent distribution in the target domain---where the valuable de-correlation clues hide---is disregarded. We propose to make the U in UDA matter by giving equal status to the two domains. Specifically, we learn an invariant classifier whose prediction is simultaneously consistent with the labels in the source domain and clusters in the target domain, hence the spurious correlation inconsistent in the target domain is removed. We dub our approach "Invariant CONsistency learning" (ICON). Extensive experiments show that ICON achieves the state-of-the-art performance on the classic UDA benchmarks: Office-Home and VisDA-2017, and outperforms all the conventional methods on the challenging WILDS 2.0 benchmark. Codes are in https://github.com/yue-zhongqi/ICON.
****
Hyper-HMM: aligning human brains and semantic features in a common latent event space
Caroline Lee, Jane Han, Ma Feilong, Guo Jiahui, James Haxby, Christopher Baldassano
Naturalistic stimuli evoke complex neural responses with spatial and temporal properties that differ across individuals. Current alignment methods focus on either spatial hyperalignment (assuming exact temporal correspondence) or temporal alignment (assuming exact spatial correspondence). Here, we propose a hybrid model, the Hyper-HMM, that simultaneously aligns both temporal and spatial features across brains. The model learns to linearly project voxels to a reduced-dimension latent space, in which timecourses are segmented into corresponding temporal events. This approach allows tracking of each individual's mental trajectory through an event sequence, and also allows for alignment with other feature spaces such as stimulus content. Using an fMRI dataset in which students watch videos of class lectures, we demonstrate that the Hyper-HMM can be used to map all participants and the semantic content of the videos into a common low-dimensional space, and that these mappings generalize to held-out data. Our model provides a new window into individual cognitive dynamics evoked by complex naturalistic stimuli.
****
Active Learning for Semantic Segmentation with Multi-class Label Query
Sehyun Hwang, Sohyun Lee, Hoyoung Kim, Minhyeon Oh, Jungseul Ok, Suha Kwak
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Aleatoric and Epistemic Discrimination: Fundamental Limits of Fairness Interventions
Hao Wang, Luxi He, Rui Gao, Flavio Calmon
Machine learning (ML) models can underperform on certain population groups due to choices made during model development and bias inherent in the data. We categorize sources of discrimination in the ML pipeline into two classes: aleatoric discrimination, which is inherent in the data distribution, and epistemic discrimination, which is due to decisions made during model development. We quantify aleatoric discrimination by determining the performance limits of a model under fairness constraints, assuming perfect knowledge of the data distribution. We demonstrate how to characterize aleatoric discrimination by applying Blackwell's results on comparing statistical experiments. We then quantify epistemic discrimination as the gap between a model's accuracy when fairness constraints are applied and the limit posed by aleatoric discrimination. We apply this approach to benchmark existing fairness interventions and investigate fairness risks in data with missing values. Our results indicate that state-of-the-art fairness interventions are effective at removing epistemic discrimination on standard (overused) tabular datasets. However, when data has missing values, there is still significant room for improvement in handling aleatoric discrimination.
****

DISCOVER: Making Vision Networks Interpretable via Competition and Dissection
Konstantinos Panousis, Sotirios Chatzis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Adapting Neural Link Predictors for Data-Efficient Complex Query Answering
Erik Arakelyan, Pasquale Minervini, Daniel Daza, Michael Cochez, Isabelle Augenstein
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DataComp: In search of the next generation of multimodal datasets
Samir Yitzhak Gadre, Gabriel Ilharco, Alex Fang, Jonathan Hayase, Georgios Smyrnis, Thao Nguyen, Ryan Marten, Mitchell Wortsman, Dhruba Ghosh, Jieyu Zhang, Eyal Orgad, Rahim Entezari, Giannis Daras, Sarah Pratt, Vivek Ramanujan, Yonatan Bitton, Kalyani Marathe, Stephen Mussmann, Richard Vencu, Mehdi Cherti, Ranjay Krishna, Pang Wei W. Koh, Olga Saukh, Alexander J. Ratner, Shuran Song, Hannaneh Hajishirzi, Ali Farhadi, Romain Beaumont, Sewoong Oh, Alex Dimakis, Jenia Jitsev, Yair Carmon, Vaishaal Shankar, Ludwig Schmidt
Multimodal datasets are a critical component in recent breakthroughs such as CLIP, Stable Diffusion and GPT-4, yet their design does not receive the same research attention as model architectures or training algorithms. To address this shortcoming in the machine learning ecosystem, we introduce DataComp, a testbed for dataset experiments centered around a new candidate pool of 12.8 billion image-text pairs from Common Crawl. Participants in our benchmark design new filtering techniques or curate new data sources and then evaluate their new dataset by running our standardized CLIP training code and testing the resulting model on 38 downstream test sets. Our benchmark consists of multiple compute scales spanning four orders of magnitude, which enables the study of scaling trends and makes the benchmark accessible to researchers with varying resources. Our baseline experiments show that the DataComp workflow leads to better training sets. Our best baseline, DataComp-1B, enables training a CLIP ViT-L/14 from scratch to 79.2% zero-shot accuracy on ImageNet, outperforming OpenAI's CLIP ViT-L/14 by 3.7 percentage points while using the same training procedure and compute. We release \datanet and all accompanying code at www.datacomp.ai.
****
p-value Adjustment for Monotonous, Unbiased, and Fast Clustering Comparison
Kai Klede, Thomas Altstidl, Dario Zanca, Bjoern Eskofier
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Computing Pairwise Statistics with Local Differential Privacy
Badih Ghazi, Pritish Kamath, Ravi Kumar, Pasin Manurangsi, Adam Sealfon
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
STORM: Efficient Stochastic Transformer based World Models for Reinforcement Learning
Weipu Zhang, Gang Wang, Jian Sun, Yetian Yuan, Gao Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Is Heterogeneity Notorious? Taming Heterogeneity to Handle Test-Time Shift in Fe
derated Learning
Yue Tan, Chen Chen, Weiming Zhuang, Xin Dong, Lingjuan Lyu, Guodong Long
Federated learning (FL) is an effective machine learning paradigm where multiple
 clients can train models based on heterogeneous data in a decentralized manner
without accessing their private data. However, existing FL systems undergo perfo
rmance deterioration due to feature-level test-time shifts, which are well inves
tigated in centralized settings but rarely studied in FL. The common non-IID iss
ue in FL usually refers to inter-client heterogeneity during training phase, whi
le the test-time shift refers to the intra-client heterogeneity during test phas
e. Although the former is always deemed to be notorious for FL, there is still a
 wealth of useful information delivered by heterogeneous data sources, which may
 potentially help alleviate the latter issue. To explore the possibility of usin
g inter-client heterogeneity in handling intra-client heterogeneity, we firstly
propose a contrastive learning-based FL framework, namely FedICON, to capture in
variant knowledge among heterogeneous clients and consistently tune the model to
 adapt to test data. In FedICON, each client performs sample-wise supervised con
trastive learning during the local training phase, which enhances sample-wise in
variance encoding ability. Through global aggregation, the invariance extraction
 ability can be mutually boosted among inter-client heterogeneity. During the te
st phase, our test-time adaptation procedure leverages unsupervised contrastive
learning to guide the model to smoothly generalize to test data under intra-clie
nt heterogeneity. Extensive experiments validate the effectiveness of the propos
ed FedICON in taming heterogeneity to handle test-time shift problems.
****
Self-Predictive Universal AI
Elliot Catt, Jordi Grau-Moya, Marcus Hutter, Matthew Aitchison, Tim Genewein, Gr
égoire Delétang, Kevin Li, Joel Veness
Reinforcement Learning (RL) algorithms typically utilize learning and/or plannin
g techniques to derive effective policies. The integration of both approaches ha
s proven to be highly successful in addressing complex sequential decision-makin
g challenges, as evidenced by algorithms such as AlphaZero and MuZero, which con
solidate the planning process into a parametric search-policy. AIXI, the most po
tent theoretical universal agent, leverages planning through comprehensive searc
h as its primary means to find an optimal policy. Here we define an alternative
universal agent, which we call Self-AIXI, that on the contrary to AIXI, maximall
y exploits learning to obtain good policies. It does so by self-predicting its o
wn stream of action data, which is generated, similarly to other TD(0) agents, b
y taking an action maximization step over the current on-policy (universal mixtu
re-policy) Q-value estimates. We prove that Self-AIXI converges to AIXI, and inh
erits a series of properties like maximal Legg-Hutter intelligence and the self-
optimizing property.
****
Addressing Negative Transfer in Diffusion Models
Hyojun Go, Jin-Young Kim, Yunsung Lee, Seunghyun Lee, Shinhyeok Oh, Hyeongdon Mo
on, Seungtaek Choi
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
The Clock and the Pizza: Two Stories in Mechanistic Explanation of Neural Networ
ks
Ziqian Zhong, Ziming Liu, Max Tegmark, Jacob Andreas
Do neural networks, trained on well-understood algorithmic tasks, reliably redis
cover known algorithms? Several recent studies, on tasks ranging from group oper
ations to in-context linear regression, have suggested that the answer is yes. U
sing modular addition as a prototypical problem, we show that algorithm discover

y in neural networks is sometimes more complex: small changes to model hyperparameters and initializations can induce discovery of qualitatively different algorithms from a fixed training set, and even learning of multiple different solutions in parallel. In modular addition, we specifically show that models learn a known Clock algorithm, a previously undescribed, less intuitive, but comprehensible procedure we term the Pizza algorithm, and a variety of even more complex procedures. Our results show that even simple learning problems can admit a surprising diversity of solutions, motivating the development of new tools for mechanistically characterizing the behavior of neural networks across the algorithmic phase space.
****
Convergent Bregman Plug-and-Play Image Restoration for Poisson Inverse Problems
Samuel Hurault, Ulugbek Kamilov, Arthur Leclaire, Nicolas Papadakis
Plug-and-Play (PnP) methods are efficient iterative algorithms for solving ill-posed image inverse problems. PnP methods are obtained by using deep Gaussian denoisers instead of the proximal operator or the gradient-descent step within proximal algorithms. Current PnP schemes rely on data-fidelity terms that have either Lipschitz gradients or closed-form proximal operators, which is not applicable to Poisson inverse problems. Based on the observation that the Gaussian noise is not the adequate noise model in this setting, we propose to generalize PnP using the Bregman Proximal Gradient (BPG) method. BPG replaces the Euclidean distance with a Bregman divergence that can better capture the smoothness properties of the problem. We introduce the Bregman Score Denoiser specifically parametrized and trained for the new Bregman geometry and prove that it corresponds to the proximal operator of a nonconvex potential. We propose two PnP algorithms based on the Bregman Score Denoiser for solving Poisson inverse problems. Extending the convergence results of BPG in the nonconvex settings, we show that the proposed methods converge, targeting stationary points of an explicit global functional. Experimental evaluations conducted on various Poisson inverse problems validate the convergence results and showcase effective restoration performance.
****
SQ Lower Bounds for Learning Mixtures of Linear Classifiers
Ilias Diakonikolas, Daniel Kane, Yuxin Sun
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Canonical normalizing flows for manifold learning
Kyriakos Flouris, Ender Konukoglu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Regularizing Neural Networks with Meta-Learning Generative Models
Shin'ya Yamaguchi, Daiki Chijiwa, Sekitoshi Kanai, Atsutoshi Kumagai, Hisashi Kashima
This paper investigates methods for improving generative data augmentation for deep learning. Generative data augmentation leverages the synthetic samples produced by generative models as an additional dataset for classification with small dataset settings. A key challenge of generative data augmentation is that the synthetic data contain uninformative samples that degrade accuracy. This can be caused by the synthetic samples not perfectly representing class categories in real data and uniform sampling not necessarily providing useful samples for tasks. In this paper, we present a novel strategy for generative data augmentation called meta generative regularization (MGR). To avoid the degradation of generative data augmentation, MGR utilizes synthetic samples for regularizing feature extractors instead of training classifiers. These synthetic samples are dynamically determined to minimize the validation losses through meta-learning. We observed t

hat MGR can avoid the performance degradation of naive generative data augmentation and boost the baselines. Experiments on six datasets showed that MGR is effective particularly when datasets are smaller and stably outperforms baselines by up to 7 percentage points on test accuracy.
****
Landscape Surrogate: Learning Decision Losses for Mathematical Optimization Under Partial Information
Arman Zharmagambetov, Brandon Amos, Aaron Ferber, Taoan Huang, Bistra Dilkina, Yuandong Tian
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Quantifying & Modeling Multimodal Interactions: An Information Decomposition Framework
Paul Pu Liang, Yun Cheng, Xiang Fan, Chun Kai Ling, Suzanne Nie, Richard Chen, Zihao Deng, Nicholas Allen, Randy Auerbach, Faisal Mahmood, Russ R. Salakhutdinov, Louis-Philippe Morency
The recent explosion of interest in multimodal applications has resulted in a wide selection of datasets and methods for representing and integrating information from different modalities. Despite these empirical advances, there remain fundamental research questions: How can we quantify the interactions that are necessary to solve a multimodal task? Subsequently, what are the most suitable multimodal models to capture these interactions? To answer these questions, we propose an information-theoretic approach to quantify the degree of redundancy, uniqueness, and synergy relating input modalities with an output task. We term these three measures as the PID statistics of a multimodal distribution (or PID for short), and introduce two new estimators for these PID statistics that scale to high-dimensional distributions. To validate PID estimation, we conduct extensive experiments on both synthetic datasets where the PID is known and on large-scale multimodal benchmarks where PID estimations are compared with human annotations. Finally, we demonstrate their usefulness in (1) quantifying interactions within multimodal datasets, (2) quantifying interactions captured by multimodal models, (3) principled approaches for model selection, and (4) three real-world case studies engaging with domain experts in pathology, mood prediction, and robotic perception where our framework helps to recommend strong multimodal models for each application.
****
A Single 2D Pose with Context is Worth Hundreds for 3D Human Pose Estimation
Qitao Zhao, Ce Zheng, Mengyuan Liu, Chen Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On the Stability-Plasticity Dilemma in Continual Meta-Learning: Theory and Algorithm
Qi CHEN, Changjian Shui, Ligong Han, Mario Marchand
We focus on Continual Meta-Learning (CML), which targets accumulating and exploiting meta-knowledge on a sequence of non-i.i.d. tasks. The primary challenge is to strike a balance between stability and plasticity, where a model should be stable to avoid catastrophic forgetting in previous tasks and plastic to learn generalizable concepts from new tasks. To address this, we formulate the CML objective as controlling the average excess risk upper bound of the task sequence, which reflects the trade-off between forgetting and generalization. Based on the objective, we introduce a unified theoretical framework for CML in both static and shifting environments, providing guarantees for various task-specific learning algorithms. Moreover, we first present a rigorous analysis of a bi-level trade-off in shifting environments. To approach the optimal trade-off, we propose a nov

el algorithm that dynamically adjusts the meta-parameter and its learning rate w.r.t environment change. Empirical evaluations on synthetic and real datasets illustrate the effectiveness of the proposed theory and algorithm.
****

Paraphrasing evades detectors of AI-generated text, but retrieval is an effective defense
Kalpesh Krishna, Yixiao Song, Marzena Karpinska, John Wieting, Mohit Iyyer
The rise in malicious usage of large language models, such as fake content creation and academic plagiarism, has motivated the development of approaches that identify AI-generated text, including those based on watermarking or outlier detection. However, the robustness of these detection algorithms to paraphrases of AI-generated text remains unclear. To stress test these detectors, we build a 11B parameter paraphrase generation model (DIPPER) that can paraphrase paragraphs, condition on surrounding context, and control lexical diversity and content reordering. Paraphrasing text generated by three large language models (including GPT3.5-davinci-003) with DIPPER successfully evades several detectors, including watermarking, GPTZero, DetectGPT, and OpenAI's text classifier. For example, DIPPER drops detection accuracy of DetectGPT from 70.3% to 4.6% (at a constant false positive rate of 1%), without appreciably modifying the input semantics.To increase the robustness of AI-generated text detection to paraphrase attacks, we introduce a simple defense that relies on retrieving semantically-similar generations and must be maintained by a language model API provider. Given a candidate text, our algorithm searches a database of sequences previously generated by the API, looking for sequences that match the candidate text within a certain threshold. We empirically verify our defense using a database of 15M generations from a fine-tuned T5-XXL model and find that it can detect 80% to 97% of paraphrased generations across different settings while only classifying 1% of human-written sequences as AI-generated. We open-source our models, code and data.
****

ChimpACT: A Longitudinal Dataset for Understanding Chimpanzee Behaviors
Xiaoxuan Ma, Stephan Kaufhold, Jiajun Su, Wentao Zhu, Jack Terwilliger, Andres Meza, Yixin Zhu, Federico Rossano, Yizhou Wang
Understanding the behavior of non-human primates is crucial for improving animal welfare, modeling social behavior, and gaining insights into distinctively human and phylogenetically shared behaviors. However, the lack of datasets on non-human primate behavior hinders in-depth exploration of primate social interactions, posing challenges to research on our closest living relatives. To address these limitations, we present ChimpACT, a comprehensive dataset for quantifying the longitudinal behavior and social relations of chimpanzees within a social group. Spanning from 2015 to 2018, ChimpACT features videos of a group of over 20 chimpanzees residing at the Leipzig Zoo, Germany, with a particular focus on documenting the developmental trajectory of one young male, Azibo. ChimpACT is both comprehensive and challenging, consisting of 163 videos with a cumulative 160,500 frames, each richly annotated with detection, identification, pose estimation, and fine-grained spatiotemporal behavior labels. We benchmark representative methods of three tracks on ChimpACT: (i) tracking and identification, (ii) pose estimation, and (iii) spatiotemporal action detection of the chimpanzees. Our experiments reveal that ChimpACT offers ample opportunities for both devising new methods and adapting existing ones to solve fundamental computer vision tasks applied to chimpanzee groups, such as detection, pose estimation, and behavior analysis, ultimately deepening our comprehension of communication and sociality in non-human primates.
****

Energy Transformer
Benjamin Hoover, Yuchen Liang, Bao Pham, Rameswar Panda, Hendrik Strobelt, Duen Horng Chau, Mohammed Zaki, Dmitry Krotov
Our work combines aspects of three promising paradigms in machine learning, namely, attention mechanism, energy-based models, and associative memory. Attention is the power-house driving modern deep learning successes, but it lacks clear theoretical foundations. Energy-based models allow a principled approach to discri

minative and generative tasks, but the design of the energy functional is not st raightforward. At the same time, Dense Associative Memory models or Modern Hopfi eld Networks have a well-established theoretical foundation, and allow an intuit ive design of the energy function. We propose a novel architecture, called the E nergy Transformer (or ET for short), that uses a sequence of attention layers th at are purposely designed to minimize a specifically engineered energy function, which is responsible for representing the relationships between the tokens. In this work, we introduce the theoretical foundations of ET, explore its empirical capabilities using the image completion task, and obtain strong quantitative re sults on the graph anomaly detection and graph classification tasks.
****
Theoretical and Practical Perspectives on what Influence Functions Do
Andrea Schioppa, Katja Filippova, Ivan Titov, Polina Zablotskaia
Influence functions (IF) have been seen as a technique for explaining model pred ictions through the lens of the training data. Their utility is assumed to be in identifying training examples "responsible" for a prediction so that, for examp le, correcting a prediction is possible by intervening on those examples (removi ng or editing them) and retraining the model. However, recent empirical studies have shown that the existing methods of estimating IF predict the leave-one-out-and-retrain effect poorly. In order to understand the mismatch between the theor etical promise and the practical results, we analyse five assumptions made by IF methods which are problematic for modern-scale deep neural networks and which c oncern convexity, numeric stability, training trajectory and parameter divergenc e. This allows us to clarify what can be expected theoretically from IF. We show that while most assumptions can be addressed successfully, the parameter diverg ence poses a clear limitation on the predictive power of IF: influence fades ove r training time even with deterministic training. We illustrate this theoretical result with BERT and ResNet models.Another conclusion from the theoretical anal ysis is that IF are still useful for model debugging and correcting even though some of the assumptions made in prior work do not hold: using natural language p rocessing and computer vision tasks, we verify that mis-predictions can be succe ssfully corrected by taking only a few fine-tuning steps on influential examples .
****
trajdata: A Unified Interface to Multiple Human Trajectory Datasets
Boris Ivanovic, Guanyu Song, Igor Gilitschenski, Marco Pavone
The field of trajectory forecasting has grown significantly in recent years, par tially owing to the release of numerous large-scale, real-world human trajectory datasets for autonomous vehicles (AVs) and pedestrian motion tracking. While su ch datasets have been a boon for the community, they each use custom and unique data formats and APIs, making it cumbersome for researchers to train and evaluat e methods across multiple datasets. To remedy this, we present trajdata: a unifi ed interface to multiple human trajectory datasets. At its core, trajdata provid es a simple, uniform, and efficient representation and API for trajectory and ma p data. As a demonstration of its capabilities, in this work we conduct a compre hensive empirical evaluation of existing trajectory datasets, providing users wi th a rich understanding of the data underpinning much of current pedestrian and AV motion forecasting research, and proposing suggestions for future datasets fr om these insights. trajdata is permissively licensed (Apache 2.0) and can be acc essed online at https://github.com/NVlabs/trajdata.
****
On Sparse Modern Hopfield Model
Jerry Yao-Chieh Hu, Donglin Yang, Dennis Wu, Chenwei Xu, Bo-Yu Chen, Han Liu
We introduce the sparse modern Hopfield model as a sparse extension of the moder n Hopfield model.Like its dense counterpart, the sparse modern Hopfield model eq uips a memory-retrieval dynamics whose one-step approximation corresponds to the sparse attention mechanism. Theoretically, our key contribution is a principled derivation of a closed-form sparse Hopfield energy using the convex conjugate o f the sparse entropic regularizer.Building upon this, we derive the sparse memor y retrieval dynamics from the sparse energy function and show its one-step appro

ximation is equivalent to the sparse-structured attention.Importantly, we provide a sparsity-dependent memory retrieval error bound which is provably tighter than its dense analog.The conditions for the benefits of sparsity to arise are therefore identified and discussed.In addition, we show that the sparse modern Hopfield model maintains the robust theoretical properties of its dense counterpart, including rapid fixed point convergence and exponential memory capacity.Empirically, we use both synthetic and real-world datasets to demonstrate that the sparse Hopfield model outperforms its dense counterpart in many situations.
****
D-CIPHER: Discovery of Closed-form Partial Differential Equations
Krzysztof Kacprzyk, Zhaozhi Qian, Mihaela van der Schaar
Closed-form differential equations, including partial differential equations and higher-order ordinary differential equations, are one of the most important tools used by scientists to model and better understand natural phenomena. Discovering these equations directly from data is challenging because it requires modeling relationships between various derivatives that are not observed in the data (equation-data mismatch) and it involves searching across a huge space of possible equations. Current approaches make strong assumptions about the form of the equation and thus fail to discover many well-known phenomena. Moreover, many of them resolve the equation-data mismatch by estimating the derivatives, which makes them inadequate for noisy and infrequent observations. To this end, we propose D-CIPHER, which is robust to measurement artifacts and can uncover a new and very general class of differential equations. We further design a novel optimization procedure, CoLLie, to help D-CIPHER search through this class efficiently. Finally, we demonstrate empirically that it can discover many well-known equations that are beyond the capabilities of current methods.
****
Training neural operators to preserve invariant measures of chaotic attractors
Ruoxi Jiang, Peter Y. Lu, Elena Orlova, Rebecca Willett
Chaotic systems make long-horizon forecasts difficult because small perturbations in initial conditions cause trajectories to diverge at an exponential rate. In this setting, neural operators trained to minimize squared error losses, while capable of accurate short-term forecasts, often fail to reproduce statistical or structural properties of the dynamics over longer time horizons and can yield degenerate results. In this paper, we propose an alternative framework designed to preserve invariant measures of chaotic attractors that characterize the time-invariant statistical properties of the dynamics. Specifically, in the multi-environment setting (where each sample trajectory is governed by slightly different dynamics), we consider two novel approaches to training with noisy data. First, we propose a loss based on the optimal transport distance between the observed dynamics and the neural operator outputs. This approach requires expert knowledge of the underlying physics to determine what statistical features should be included in the optimal transport loss. Second, we show that a contrastive learning framework, which does not require any specialized prior knowledge, can preserve statistical properties of the dynamics nearly as well as the optimal transport approach. On a variety of chaotic systems, our method is shown empirically to preserve invariant measures of chaotic attractors.
****
Certification of Distributional Individual Fairness
Matthew Wicker, Vihari Piratla, Adrian Weller
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Leveraging sparse and shared feature activations for disentangled representation learning
Marco Fumero, Florian Wenzel, Luca Zancato, Alessandro Achille, Emanuele Rodolà, Stefano Soatto, Bernhard Schölkopf, Francesco Locatello
Recovering the latent factors of variation of high dimensional data has so far f

ocused on simple synthetic settings. Mostly building on unsupervised and weakly-supervised objectives, prior work missed out on the positive implications for representation learning on real world data. In this work, we propose to leverage knowledge extracted from a diversified set of supervised tasks to learn a common disentangled representation. Assuming each supervised task only depends on an unknown subset of the factors of variation, we disentangle the feature space of a supervised multi-task model, with features activating sparsely across different tasks and information being shared as appropriate. Importantly, we never directly observe the factors of variations but establish that access to multiple tasks is sufficient for identifiability under sufficiency and minimality assumptions.We validate our approach on six real world distribution shift benchmarks, and different data modalities (images, text), demonstrating how disentangled representations can be transferred to real settings.
****

Mathematical Capabilities of ChatGPT
Simon Frieder, Luca Pinchetti, Chevalier, Ryan-Rhys Griffiths, Tommaso Salvatori, Thomas Lukasiewicz, Philipp Petersen, Julius Berner
We investigate the mathematical capabilities of two iterations of ChatGPT (released 9-January-2023 and 30-January-2023) and of GPT-4 by testing them on publicly available datasets, as well as hand-crafted ones, using a novel methodology. In contrast to formal mathematics, where large databases of formal proofs are available (e.g., mathlib, the Lean Mathematical Library), current datasets of natural-language mathematics used to benchmark language models either cover only elementary mathematics or are very small. We address this by publicly releasing two new datasets: GHOSTS and miniGHOSTS. These are the first natural-language datasets curated by working researchers in mathematics that (1) aim to cover graduate-level mathematics, (2) provide a holistic overview of the mathematical capabilities of language models, and (3) distinguish multiple dimensions of mathematical reasoning. These datasets test on 1636 human expert evaluations whether ChatGPT and GPT-4 can be helpful assistants to professional mathematicians by emulating use cases that arise in the daily professional activities of mathematicians. We benchmark the models on a range of fine-grained performance metrics. For advanced mathematics, this is the most detailed evaluation effort to date. We find that ChatGPT and GPT-4 can be used most successfully as mathematical assistants for querying facts, acting as mathematical search engines and knowledge base interfaces. GPT-4 can additionally be used for undergraduate-level mathematics but fails on graduate-level difficulty. Contrary to many positive reports in the media about GPT-4 and ChatGPT's exam-solving abilities (a potential case of selection bias), their overall mathematical performance is well below the level of a graduate student. Hence, if you aim to use ChatGPT to pass a graduate-level math exam, you would be better off copying from your average peer!
****

A Unified Framework for U-Net Design and Analysis
Christopher Williams, Fabian Falck, George Deligiannidis, Chris C Holmes, Arnaud Doucet, Saifuddin Syed
U-Nets are a go-to neural architecture across numerous tasks for continuous signals on a square such as images and Partial Differential Equations (PDE), however their design and architecture is understudied. In this paper, we provide a framework for designing and analysing general U-Net architectures. We present theoretical results which characterise the role of the encoder and decoder in a U-Net, their high-resolution scaling limits and their conjugacy to ResNets via preconditioning. We propose Multi-ResNets, U-Nets with a simplified, wavelet-based encoder without learnable parameters. Further, we show how to design novel U-Net architectures which encode function constraints, natural bases, or the geometry of the data. In diffusion models, our framework enables us to identify that high-frequency information is dominated by noise exponentially faster, and show how U-Nets with average pooling exploit this. In our experiments, we demonstrate how Multi-ResNets achieve competitive and often superior performance compared to classical U-Nets in image segmentation, PDE surrogate modelling, and generative modelling with diffusion models. Our U-Net framework paves the way to study the theor

etical properties of U-Nets and design natural, scalable neural architectures fo
r a multitude of problems beyond the square.
****
On the Importance of Feature Separability in Predicting Out-Of-Distribution Erro
r
RENCHUNZI XIE, Hongxin Wei, Lei Feng, Yuzhou Cao, Bo An
Estimating the generalization performance is practically challenging on out-of-d
istribution (OOD) data without ground-truth labels. While previous methods empha
size the connection between distribution difference and OOD accuracy, we show th
at a large domain gap not necessarily leads to a low test accuracy. In this pape
r, we investigate this problem from the perspective of feature separability empi
rically and theoretically. Specifically, we propose a dataset-level score based
upon feature dispersion to estimate the test accuracy under distribution shift.
Our method is inspired by desirable properties of features in representation lea
rning: high inter-class dispersion and high intra-class compactness. Our analysi
s shows that inter-class dispersion is strongly correlated with the model accura
cy, while intra-class compactness does not reflect the generalization performanc
e on OOD data. Extensive experiments demonstrate the superiority of our method i
n both prediction performance and computational efficiency.
****
The Transient Nature of Emergent In-Context Learning in Transformers
Aaditya Singh, Stephanie Chan, Ted Moskovitz, Erin Grant, Andrew Saxe, Felix Hil
l
Transformer neural networks can exhibit a surprising capacity for in-context lea
rning (ICL) despite not being explicitly trained for it. Prior work has provide
d a deeper understanding of how ICL emerges in transformers, e.g. through the le
ns of mechanistic interpretability, Bayesian inference, or by examining the dist
ributional properties of training data. However, in each of these cases, ICL is
treated largely as a persistent phenomenon; namely, once ICL emerges, it is assu
med to persist asymptotically. Here, we show that the emergence of ICL during tr
ansformer training is, in fact, often transient. We train transformers on synthe
tic data designed so that both ICL and in-weights learning (IWL) strategies can
lead to correct predictions. We find that ICL first emerges, then disappears and
 gives way to IWL, all while the training loss decreases, indicating an asymptot
ic preference for IWL. The transient nature of ICL is observed in transformers a
cross a range of model sizes and datasets, raising the question of how much to `
`overtrain'' transformers when seeking compact, cheaper-to-run models. We find t
hat L2 regularization may offer a path to more persistent ICL that removes the n
eed for early stopping based on ICL-style validation tasks. Finally, we present
initial evidence that ICL transience may be caused by competition between ICL an
d IWL circuits.
****
When is Agnostic Reinforcement Learning Statistically Tractable?
Zeyu Jia, Gene Li, Alexander Rakhlin, Ayush Sekhari, Nati Srebro
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Imagine the Unseen World: A Benchmark for Systematic Generalization in Visual Wo
rld Models
Yeongbin Kim, Gautam Singh, Junyeong Park, Caglar Gulcehre, Sungjin Ahn
Systematic compositionality, or the ability to adapt to novel situations by crea
ting a mental model of the world using reusable pieces of knowledge, remains a s
ignificant challenge in machine learning. While there has been considerable prog
ress in the language domain, efforts towards systematic visual imagination, or e
nvisioning the dynamical implications of a visual observation, are in their infa
ncy. We introduce the Systematic Visual Imagination Benchmark (SVIB), the first
benchmark designed to address this problem head-on. SVIB offers a novel framewor
k for a minimal world modeling problem, where models are evaluated based on thei

r ability to generate one-step image-to-image transformations under a latent world dynamics. The framework provides benefits such as the possibility to jointly optimize for systematic perception and imagination, a range of difficulty levels, and the ability to control the fraction of possible factor combinations used during training. We provide a comprehensive evaluation of various baseline models on SVIB, offering insight into the current state-of-the-art in systematic visual imagination. We hope that this benchmark will help advance visual systematic compositionality.
****
Convolutional Visual Prompt for Robust Visual Perception
Yun-Yun Tsai, Chengzhi Mao, Junfeng Yang
Vision models are often vulnerable to out-of-distribution (OOD) samples without adapting. While visual prompts offer a lightweight method of input-space adaptation for large-scale vision models, they rely on a high-dimensional additive vector and labeled data. This leads to overfitting when adapting models in a self-supervised test-time setting without labels. We introduce convolutional visual prompts (CVP) for label-free test-time adaptation for robust visual perception. The structured nature of CVP demands fewer trainable parameters, less than 1\% compared to standard visual prompts, combating overfitting. Extensive experiments and analysis on a wide variety of OOD visual perception tasks show that our approach is effective, improving robustness by up to 5.87\% over several large-scale models.
****
LVM-Med: Learning Large-Scale Self-Supervised Vision Models for Medical Imaging via Second-order Graph Matching
Duy M. H. Nguyen, Hoang Nguyen, Nghiem Diep, Tan Ngoc Pham, Tri Cao, Binh Nguyen, Paul Swoboda, Nhat Ho, Shadi Albarqouni, Pengtao Xie, Daniel Sonntag, Mathias Niepert
Obtaining large pre-trained models that can be fine-tuned to new tasks with limited annotated samples has remained an open challenge for medical imaging data. While pre-trained networks on ImageNet and vision-language foundation models trained on web-scale data are the prevailing approaches, their effectiveness on medical tasks is limited due to the significant domain shift between natural and medical images. To bridge this gap, we introduce LVM-Med, the first family of deep networks trained on large-scale medical datasets. We have collected approximately 1.3 million medical images from 55 publicly available datasets, covering a large number of organs and modalities such as CT, MRI, X-ray, and Ultrasound. We benchmark several state-of-the-art self-supervised algorithms on this dataset and propose a novel self-supervised contrastive learning algorithm using a graph-matching formulation. The proposed approach makes three contributions: (i) it integrates prior pair-wise image similarity metrics based on local and global information; (ii) it captures the structural constraints of feature embeddings through a loss function constructed through a combinatorial graph-matching objective, and (iii) it can be trained efficiently end-to-end using modern gradient-estimation techniques for black-box solvers. We thoroughly evaluate the proposed LVM-Med on 15 downstream medical tasks ranging from segmentation and classification to object detection, and both for the in and out-of-distribution settings. LVM-Med empirically outperforms a number of state-of-the-art supervised, self-supervised, and foundation models. For challenging tasks such as Brain Tumor Classification or Diabetic Retinopathy Grading, LVM-Med improves previous vision-language models trained on 1 billion masks by 6-7% while using only a ResNet-50.
****
Lending Interaction Wings to Recommender Systems with Conversational Agents
Jiarui Jin, Xianyu Chen, Fanghua Ye, Mengyue Yang, Yue Feng, Weinan Zhang, Yong Yu, Jun Wang
An intelligent conversational agent (a.k.a., chat-bot) could embrace conversational technologies to obtain user preferences online, to overcome inherent limitations of recommender systems trained over the offline historical user behaviors. In this paper, we propose CORE, a new offline-training and online-checking framework to plug a COnversational agent into REcommender systems. Unlike most prior

conversational recommendation approaches that systemically combine conversationa
l and recommender parts through a reinforcement learning framework, CORE bridges
 the conversational agent and recommender system through a unified uncertainty m
inimization framework, which can be easily applied to any existing recommendatio
n approach. Concretely, CORE treats a recommender system as an offline estimator
 to produce an estimated relevance score for each item, while CORE regards a con
versational agent as an online checker that checks these estimated scores in eac
h online session. We define uncertainty as the sum of unchecked relevance scores
. In this regard, the conversational agent acts to minimize uncertainty via quer
ying either attributes or items. Towards uncertainty minimization, we derive the
 certainty gain of querying each attribute and item, and develop a novel online
decision tree algorithm to decide what to query at each turn. Our theoretical an
alysis reveals the bound of the expected number of turns of CORE in a cold-start
 setting. Experimental results demonstrate that CORE can be seamlessly employed
on a variety of recommendation approaches, and can consistently bring significan
t improvements in both hot-start and cold-start settings.
****
High-Fidelity Audio Compression with Improved RVQGAN
Rithesh Kumar, Prem Seetharaman, Alejandro Luebs, Ishaan Kumar, Kundan Kumar
Language models have been successfully used to model natural signals, such as im
ages, speech, and music. A key component of these models is a high quality neura
l compression model that can compress high-dimensional natural signals into lowe
r dimensional discrete tokens. To that end, we introduce a high-fidelity univers
al neural audio compression algorithm that achieves ~90x compression of 44.1 KHz
 audio into tokens at just 8kbps bandwidth. We achieve this by combining advance
s in high-fidelity audio generation with better vector quantization techniques f
rom the image domain, along with improved adversarial and reconstruction losses.
 We compress all domains (speech, environment, music, etc.) with a single univer
sal model, making it widely applicable to generative modeling of all audio. We c
ompare with competing audio compression algorithms, and find our method outperfo
rms them significantly. We provide thorough ablations for every design choice, a
s well as open-source code and trained model weights. We hope our work can lay t
he foundation for the next generation of high-fidelity audio modeling.
****
Comparing Apples to Oranges: Learning Similarity Functions for Data Produced by
Different Distributions
Leonidas Tsepenekas, Ivan Brugere, Freddy Lecue, Daniele Magazzeni
Similarity functions measure how comparable pairs of elements are, and play a ke
y role in a wide variety of applications, e.g., notions of Individual Fairness a
biding by the seminal paradigm of Dwork et al., as well as Clustering problems.
However, access to an accurate similarity function should not always be consider
ed guaranteed, and this point was even raised by Dwork et al. For instance, it i
s reasonable to assume that when the elements to be compared are produced by dif
ferent distributions, or in other words belong to different ``demographic'' grou
ps, knowledge of their true similarity might be very difficult to obtain. In thi
s work, we present an efficient sampling framework that learns these across-grou
ps similarity functions, using only a limited amount of experts' feedback. We sh
ow analytical results with rigorous theoretical bounds, and empirically validate
 our algorithms via a large suite of experiments.
****
Scalable Transformer for PDE Surrogate Modeling
Zijie Li, Dule Shu, Amir Barati Farimani
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
What is the Inductive Bias of Flatness Regularization? A Study of Deep Matrix Fa
ctorization Models
Khashayar Gatmiry, Zhiyuan Li, Tengyu Ma, Sashank Reddi, Stefanie Jegelka, Ching

-Yao Chuang

Recent works on over-parameterized neural networks have shown that the stochasticity in optimizers has the implicit regularization effect of minimizing the sharpness of the loss function (in particular, the trace of its Hessian) over the family zero-loss solutions. More explicit forms of flatness regularization also empirically improve the generalization performance. However, it remains unclear why and when flatness regularization leads to better generalization. This work takes the first step towards understanding the inductive bias of the minimum trace of the Hessian solutions in an important setting: learning deep linear networks from linear measurements, also known as \emph{deep matrix factorization}. We show that with the standard Restricted Isometry Property (RIP) on the measurements, minimizing the trace of Hessian is approximately equivalent to minimizing the Schatten 1-norm of the corresponding end-to-end matrix parameters (i.e., the product of all layer matrices), which in turn leads to better generalization.

****

Two Sides of The Same Coin: Bridging Deep Equilibrium Models and Neural ODEs via Homotopy Continuation

Shutong Ding, Tianyu Cui, Jingya Wang, Ye Shi

Deep Equilibrium Models (DEQs) and Neural Ordinary Differential Equations (Neural ODEs) are two branches of implicit models that have achieved remarkable success owing to their superior performance and low memory consumption. While both are implicit models, DEQs and Neural ODEs are derived from different mathematical formulations. Inspired by homotopy continuation, we establish a connection between these two models and illustrate that they are actually two sides of the same coin. Homotopy continuation is a classical method of solving nonlinear equations based on a corresponding ODE. Given this connection, we proposed a new implicit model called HomoODE that inherits the property of high accuracy from DEQs and the property of stability from Neural ODEs. Unlike DEQs, which explicitly solve an equilibrium-point-finding problem via Newton's methods in the forward pass, HomoODE solves the equilibrium-point-finding problem implicitly using a modified Neural ODE via homotopy continuation. Further, we developed an acceleration method for HomoODE with a shared learnable initial point. It is worth noting that our model also provides a better understanding of why Augmented Neural ODEs work as long as the augmented part is regarded as the equilibrium point to find. Comprehensive experiments with several image classification tasks demonstrate that HomoODE surpasses existing implicit models in terms of both accuracy and memory consumption.

****

Emergent and Predictable Memorization in Large Language Models

Stella Biderman, USVSN PRASHANTH, Lintang Sutawika, Hailey Schoelkopf, Quentin Anthony, Shivanshu Purohit, Edward Raff

Memorization, or the tendency of large language models (LLMs) to output entire sequences from their training data verbatim, is a key concern for deploying language models. In particular, it is vital to minimize a model's memorization of sensitive datapoints such as those containing personal identifiable information (PII). The prevalence of such undesirable memorization can pose issues for model trainers, and may even require discarding an otherwise functional model. We therefore seek to predict which sequences will be memorized before a large model's full train-time by extrapolating the memorization behavior of lower-compute trial runs. We measure memorization in the Pythia model suite and plot scaling laws for forecasting memorization, allowing us to provide equi-compute recommendations to maximize the reliability (recall) of such predictions. We additionally provide further novel discoveries on the distribution of memorization scores across models and data. We release all code and data necessary to reproduce the results in this paper at https://github.com/EleutherAI/pythia.

****

Mind2Web: Towards a Generalist Agent for the Web

Xiang Deng, Yu Gu, Boyuan Zheng, Shijie Chen, Sam Stevens, Boshi Wang, Huan Sun, Yu Su

We introduce Mind2Web, the first dataset for developing and evaluating generalis t agents for the web that can follow language instructions to complete complex tasks on any website. Existing datasets for web agents either use simulated websites or only cover a limited set of websites and tasks, thus not suitable for generalist web agents. With over 2,000 open-ended tasks collected from 137 websites spanning 31 domains and crowdsourced action sequences for the tasks, Mind2Web provides three necessary ingredients for building generalist web agents: 1) diverse domains, websites, and tasks, 2) use of real-world websites instead of simulated and simplified ones, and 3) a broad spectrum of user interaction patterns. Based on Mind2Web, we conduct an initial exploration of using large language models (LLMs) for building generalist web agents. While the raw HTML of real-world websites are often too large to be fed to LLMs, we show that first filtering it with a small LM significantly improves the effectiveness and efficiency of LLMs. Our solution demonstrates a decent level of performance, even on websites or entire domains the model has never seen before, but there is still a substantial room to improve towards truly generalizable agents. We open-source our dataset, model implementation, and trained models (https://osu-nlp-group.github.io/Mind2Web) to facilitate further research on building a generalist agent for the web.

****

MultiMoDN—Multimodal, Multi-Task, Interpretable Modular Networks

Vinitra Swamy, Malika Satayeva, Jibril Frej, Thierry Bossy, Thijs Vogels, Martin Jaggi, Tanja Käser, Mary-Anne Hartley

Predicting multiple real-world tasks in a single model often requires a particularly diverse feature space. Multimodal (MM) models aim to extract the synergistic predictive potential of multiple data types to create a shared feature space with aligned semantic meaning across inputs of drastically varying sizes (i.e. images, text, sound). Most current MM architectures fuse these representations in parallel, which not only limits their interpretability but also creates a dependency on modality availability. We present MultiModN, a multimodal, modular network that fuses latent representations in a sequence of any number, combination, or type of modality while providing granular real-time predictive feedback on any number or combination of predictive tasks. MultiModN's composable pipeline is interpretable-by-design, as well as innately multi-task and robust to the fundamental issue of biased missingness. We perform four experiments on several benchmark MM datasets across 10 real-world tasks (predicting medical diagnoses, academic performance, and weather), and show that MultiModN's sequential MM fusion does not compromise performance compared with a baseline of parallel fusion. By simulating the challenging bias of missing not-at-random (MNAR), this work shows that, contrary to MultiModN, parallel fusion baselines erroneously learn MNAR and suffer catastrophic failure when faced with different patterns of MNAR at inference. To the best of our knowledge, this is the first inherently MNAR-resistant approach to MM modeling. In conclusion, MultiModN provides granular insights, robustness, and flexibility without compromising performance.

****

Differentiable Neuro-Symbolic Reasoning on Large-Scale Knowledge Graphs

CHEN SHENGYUAN, Yunfeng Cai, Huang Fang, Xiao Huang, Mingming Sun

Knowledge graph (KG) reasoning utilizes two primary techniques, i.e., rule-based and KG-embedding based. The former provides precise inferences, but inferring via concrete rules is not scalable. The latter enables efficient reasoning at the cost of ambiguous inference accuracy. Neuro-symbolic reasoning seeks to amalgamate the advantages of both techniques. The crux of this approach is replacing the predicted existence of all possible triples (i.e., truth scores inferred from rules) with a suitable approximation grounded in embedding representations. However, constructing an effective approximation of all possible triples' truth scores is a challenging task, because it needs to balance the tradeoff between accuracy and efficiency, while compatible with both the rule-based and KG-embedding models. To this end, we proposed a differentiable framework - DiffLogic. Instead of directly approximating all possible triples, we design a tailored filter to adaptively select essential triples based on the dynamic rules and weights. The truth scores assessed by KG-embedding are continuous, so we employ a continuous Markov logic network named probabilistic soft logic (PSL). It employs the truth s

cores of essential triples to assess the overall agreement among rules, weights, and observed triples. PSL enables end-to-end differentiable optimization, so we can alternately update embedding and weighted rules. On benchmark datasets, we empirically show that DiffLogic surpasses baselines in both effectiveness and efficiency.
****
Topological Parallax: A Geometric Specification for Deep Perception Models
Abraham Smith, Michael Catanzaro, Gabrielle Angeloro, Nirav Patel, Paul Bendich
For safety and robustness of AI systems, we introduce topological parallax as a theoretical and computational tool that compares a trained model to a reference dataset to determine whether they have similar multiscale geometric structure. Our proofs and examples show that this geometric similarity between dataset and model is essential to trustworthy interpolation and perturbation, and we conjecture that this new concept will add value to the current debate regarding the unclear relationship between "overfitting"' and "generalization'' in applications of deep-learning. In typical deep-learning applications, an explicit geometric description of the model isimpossible, but parallax can estimate topological features (components, cycles, voids, etc.)in the model by examining the effect on the Rips complex of geodesic distortions using the reference dataset.Thus, parallax indicates whether the model shares similar multiscale geometric features with the dataset.Parallax presents theoretically via topological data analysis [TDA] as a bi-filtered persistence module,and the key properties of this module are stable under perturbation of the reference dataset.
****
Rewiring Neurons in Non-Stationary Environments
Zhicheng Sun, Yadong Mu
The human brain rewires itself for neuroplasticity in the presence of new tasks. We are inspired to harness this key process in continual reinforcement learning, prioritizing adaptation to non-stationary environments. In distinction to existing rewiring approaches that rely on pruning or dynamic routing, which may limit network capacity and plasticity, this work presents a novel rewiring scheme by permuting hidden neurons. Specifically, the neuron permutation is parameterized to be end-to-end learnable and can rearrange all available synapses to explore a large span of weight space, thereby promoting adaptivity. In addition, we introduce two main designs to steer the rewiring process in continual reinforcement learning: first, a multi-mode rewiring strategy is proposed which diversifies the policy and encourages exploration when encountering new environments. Secondly, to ensure stability on history tasks, the network is devised to cache each learned wiring while subtly updating its weights, allowing for retrospective recovery of any previous state appropriate for the task. Meanwhile, an alignment mechanism is curated to achieve better plasticity-stability tradeoff by jointly optimizing cached wirings and weights. Our proposed method is comprehensively evaluated on 18 continual reinforcement learning scenarios ranging from locomotion to manipulation, demonstrating its advantages over state-of-the-art competitors in performance-efficiency tradeoffs. Code is available at https://github.com/feifeiobama/RewireNeuron.
****
TransHP: Image Classification with Hierarchical Prompting
Wenhao Wang, Yifan Sun, Wei Li, Yi Yang
This paper explores a hierarchical prompting mechanism for the hierarchical image classification (HIC) task. Different from prior HIC methods, our hierarchical prompting is the first to explicitly inject ancestor-class information as a tokenized hint that benefits the descendant-class discrimination. We think it well imitates human visual recognition, i.e., humans may use the ancestor class as a prompt to draw focus on the subtle differences among descendant classes. We model this prompting mechanism into a Transformer with Hierarchical Prompting (TransHP). TransHP consists of three steps: 1) learning a set of prompt tokens to represent the coarse (ancestor) classes, 2) on-the-fly predicting the coarse class of the input image at an intermediate block, and 3) injecting the prompt token of the predicted coarse class into the intermediate feature. Though the parameters

of TransHP maintain the same for all input images, the injected coarse-class pro
mpt conditions (modifies) the subsequent feature extraction and encourages a dyn
amic focus on relatively subtle differences among the descendant classes. Extens
ive experiments show that TransHP improves image classification on accuracy (e.g
., improving ViT-B/16 by +2.83% ImageNet classification accuracy), training data
 efficiency (e.g., +12.69% improvement under 10% ImageNet training data), and mo
del explainability. Moreover, TransHP also performs favorably against prior HIC
methods, showing that TransHP well exploits the hierarchical information. The co
de is available at: https://github.com/WangWenhao0716/TransHP.
****
Practical Differentially Private Hyperparameter Tuning with Subsampling
Antti Koskela, Tejas D. Kulkarni
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Learning to Discover Skills through Guidance
HYUNSEUNG KIM, BYUNG KUN LEE, Hojoon Lee, Dongyoon Hwang, Sejik Park, Kyushik Mi
n, Jaegul Choo
In the field of unsupervised skill discovery (USD), a major challenge is limited
 exploration, primarily due to substantial penalties when skills deviate from th
eir initial trajectories. To enhance exploration, recent methodologies employ au
xiliary rewards to maximize the epistemic uncertainty or entropy of states. Howe
ver, we have identified that the effectiveness of these rewards declines as the
environmental complexity rises. Therefore, we present a novel USD algorithm, ski
ll discovery with guidance (DISCO-DANCE), which (1) selects the guide skill that
 possesses the highest potential to reach unexplored states, (2) guides other sk
ills to follow guide skill, then (3) the guided skills are dispersed to maximize
 their discriminability in unexplored states. Empirical evaluation demonstrates
that DISCO-DANCE outperforms other USD baselines in challenging environments, in
cluding two navigation benchmarks and a continuous control benchmark. Qualitativ
e visualizations and code of DISCO-DANCE are available at https://mynsng.github.
io/discodance/.
****
Polynomial-Time Linear-Swap Regret Minimization in Imperfect-Information Sequent
ial Games
Gabriele Farina, Charilaos Pipis
No-regret learners seek to minimize the difference between the loss they cumulat
ed through the actions they played, and the loss they would have cumulated in hi
ndsight had they consistently modified their behavior according to some strategy
 transformation function. The size of the set of transformations considered by t
he learner determines a natural notion of rationality. As the set of transformat
ions each learner considers grows, the strategies played by the learners recover
 more complex game-theoretic equilibria, including correlated equilibria in norm
al-form games and extensive-form correlated equilibria in extensive-form games.
At the extreme, a no-swap-regret agent is one that minimizes regret against the
set of all functions from the set of strategies to itself. While it is known tha
t the no-swap-regret condition can be attained efficiently in nonsequential (nor
mal-form) games, understanding what is the strongest notion of rationality that
can be attained efficiently in the worst case in sequential (extensive-form) gam
es is a longstanding open problem. In this paper we provide a positive result, b
y showing that it is possible, in any sequential game, to retain polynomial-time
 (in the game tree size) iterations while achieving sublinear regret with respec
t to all linear transformations of the mixed strategy space, a notion called no-
linear-swap regret. This notion of hindsight rationality is as strong as no-swap
-regret in nonsequential games, and stronger than no-trigger-regret in sequentia
l games—thereby proving the existence of a subset of extensive-form correlated e
quilibria robust to linear deviations, which we call linear-deviation correlated
 equilibria, that can be approached efficiently.

****
Counterfactually Comparing Abstaining Classifiers
Yo Joong Choe, Aditya Gangrade, Aaditya Ramdas
Abstaining classifiers have the option to abstain from making predictions on inputs that they are unsure about. These classifiers are becoming increasingly popular in high-stakes decision-making problems, as they can withhold uncertain predictions to improve their reliability and safety. When evaluating black-box abstaining classifier(s), however, we lack a principled approach that accounts for what the classifier would have predicted on its abstentions. These missing predictions matter when they can eventually be utilized, either directly or as a backup option in a failure mode. In this paper, we introduce a novel approach and perspective to the problem of evaluating and comparing abstaining classifiers by treating abstentions as missing data. Our evaluation approach is centered around defining the counterfactual score of an abstaining classifier, defined as the expected performance of the classifier had it not been allowed to abstain. We specify the conditions under which the counterfactual score is identifiable: if the abstentions are stochastic, and if the evaluation data is independent of the training data (ensuring that the predictions are missing at random), then the score is identifiable. Note that, if abstentions are deterministic, then the score is unidentifiable because the classifier can perform arbitrarily poorly on its abstentions. Leveraging tools from observational causal inference, we then develop nonparametric and doubly robust methods to efficiently estimate this quantity under identification. Our approach is examined in both simulated and real data experiments.
****
Video Timeline Modeling For News Story Understanding
Meng Liu, Mingda Zhang, Jialu Liu, Hanjun Dai, Ming-Hsuan Yang, Shuiwang Ji, Zheyun Feng, Boqing Gong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Understanding and Addressing the Pitfalls of Bisimulation-based Representations in Offline Reinforcement Learning
Hongyu Zang, Xin Li, Leiji Zhang, Yang Liu, Baigui Sun, Riashat Islam, Remi Tachet des Combes, Romain Laroche
While bisimulation-based approaches hold promise for learning robust state representations for Reinforcement Learning (RL) tasks, their efficacy in offline RL tasks has not been up to par. In some instances, their performance has even significantly underperformed alternative methods. We aim to understand why bisimulation methods succeed in online settings, but falter in offline tasks. Our analysis reveals that missing transitions in the dataset are particularly harmful to the bisimulation principle, leading to ineffective estimation. We also shed light on the critical role of reward scaling in bounding the scale of bisimulation measurements and of the value error they induce. Based on these findings, we propose to apply the expectile operator for representation learning to our offline RL setting, which helps to prevent overfitting to incomplete data. Meanwhile, by introducing an appropriate reward scaling strategy, we avoid the risk of feature collapse in representation space. We implement these recommendations on two state-of-the-art bisimulation-based algorithms, MICo and SimSR, and demonstrate performance gains on two benchmark suites: D4RL and Visual D4RL. Codes are provided at \url{https://github.com/zanghyu/Offline_Bisimulation}.
****
Predict, Refine, Synthesize: Self-Guiding Diffusion Models for Probabilistic Time Series Forecasting
Marcel Kollovieh, Abdul Fatir Ansari, Michael Bohlke-Schneider, Jasper Zschiegner, Hao Wang, Yuyang (Bernie) Wang
Diffusion models have achieved state-of-the-art performance in generative modeling tasks across various domains. Prior works on time series diffusion models hav

e primarily focused on developing conditional models tailored to specific foreca
sting or imputation tasks. In this work, we explore the potential of task-agnost
ic, unconditional diffusion models for several time series applications. We prop
ose TSDiff, an unconditionally-trained diffusion model for time series. Our prop
osed self-guidance mechanism enables conditioning TSDiff for downstream tasks du
ring inference, without requiring auxiliary networks or altering the training pr
ocedure. We demonstrate the effectiveness of our method on three different time
series tasks: forecasting, refinement, and synthetic data generation. First, we
show that TSDiff is competitive with several task-specific conditional forecasti
ng methods (predict). Second, we leverage the learned implicit probability densi
ty of TSDiff to iteratively refine the predictions of base forecasters with redu
ced computational overhead over reverse diffusion (refine). Notably, the generat
ive performance of the model remains intact — downstream forecasters trained on
synthetic samples from TSDiff outperform forecasters that are trained on samples
 from other state-of-the-art generative time series models, occasionally even ou
tperforming models trained on real data (synthesize).Our code is available at ht
tps://github.com/amazon-science/unconditional-time-series-diffusion
****
Learning Layer-wise Equivariances Automatically using Gradients
Tycho van der Ouderaa, Alexander Immer, Mark van der Wilk
Convolutions encode equivariance symmetries into neural networks leading to bett
er generalisation performance. However, symmetries provide fixed hard constraint
s on the functions a network can represent, need to be specified in advance, and
 can not be adapted. Our goal is to allow flexible symmetry constraints that can
 automatically be learned from data using gradients. Learning symmetry and assoc
iated weight connectivity structures from scratch is difficult for two reasons.
First, it requires efficient and flexible parameterisations of layer-wise equiva
riances. Secondly, symmetries act as constraints and are therefore not encourage
d by training losses measuring data fit. To overcome these challenges, we improv
e parameterisations of soft equivariance and learn the amount of equivariance in
 layers by optimising the marginal likelihood, estimated using differentiable La
place approximations. The objective balances data fit and model complexity enabl
ing layer-wise symmetry discovery in deep networks. We demonstrate the ability t
o automatically learn layer-wise equivariances on image classification tasks, ac
hieving equivalent or improved performance over baselines with hard-coded symmet
ry.
****
PRIOR: Personalized Prior for Reactivating the Information Overlooked in Federat
ed Learning.
Mingjia Shi, Yuhao Zhou, Kai Wang, Huaizheng Zhang, Shudong Huang, Qing Ye, Jian
cheng Lv
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Byzantine-Tolerant Methods for Distributed Variational Inequalities
Nazarii Tupitsa, Abdulla Jasem Almansoori, Yanlin Wu, Martin Takac, Karthik Nand
akumar, Samuel Horváth, Eduard Gorbunov
Robustness to Byzantine attacks is a necessity for various distributed training
scenarios. When the training reduces to the process of solving a minimization pr
oblem, Byzantine robustness is relatively well-understood. However, other proble
m formulations, such as min-max problems or, more generally, variational inequal
ities, arise in many modern machine learning and, in particular, distributed lea
rning tasks. These problems significantly differ from the standard minimization
ones and, therefore, require separate consideration. Nevertheless, only one work
 [Abidi et al., 2022] addresses this important question in the context of Byzant
ine robustness. Our work makes a further step in this direction by providing sev
eral (provably) Byzantine-robust methods for distributed variational inequality,
 thoroughly studying their theoretical convergence, removing the limitations of

the previous work, and providing numerical comparisons supporting the theoretica
l findings.
****
Asynchrony-Robust Collaborative Perception via Bird's Eye View Flow
Sizhe Wei, Yuxi Wei, Yue Hu, Yifan Lu, Yiqi Zhong, Siheng Chen, Ya Zhang
Collaborative perception can substantially boost each agent's perception ability by facilitating communication among multiple agents. However, temporal asynchrony among agents is inevitable in the real world due to communication delays, interruptions, and clock misalignments. This issue causes information mismatch during multi-agent fusion, seriously shaking the foundation of collaboration. To address this issue, we propose CoBEVFlow, an asynchrony-robust collaborative perception system based on bird's eye view (BEV) flow. The key intuition of CoBEVFlow is to compensate motions to align asynchronous collaboration messages sent by multiple agents. To model the motion in a scene, we propose BEV flow, which is a collection of the motion vector corresponding to each spatial location. Based on BEV flow, asynchronous perceptual features can be reassigned to appropriate positions, mitigating the impact of asynchrony. CoBEVFlow has two advantages: (i) CoBEVFlow can handle asynchronous collaboration messages sent at irregular, continuous time stamps without discretization; and (ii) with BEV flow, CoBEVFlow only transports the original perceptual features, instead of generating new perceptual features, avoiding additional noises. To validate CoBEVFlow's efficacy, we create IRregular V2V(IRV2V), the first synthetic collaborative perception dataset with various temporal asynchronies that simulate different real-world scenarios. Extensive experiments conducted on both IRV2V and the real-world dataset DAIR-V2X show that CoBEVFlow consistently outperforms other baselines and is robust in extremely asynchronous settings. The code is available at https://github.com/MediaBrain-SJTU/CoBEVFlow.
****
Train 'n Trade: Foundations of Parameter Markets
Tzu-Heng Huang, Harit Vishwakarma, Frederic Sala
Organizations typically train large models individually. This is costly and time-consuming, particularly for large-scale foundation models. Such vertical production is known to be suboptimal. Inspired by this economic insight, we ask whether it is possible to leverage others' expertise by trading the constituent parts in models, i.e., sets of weights, as if they were market commodities. While recent advances in aligning and interpolating models suggest that doing so may be possible, a number of fundamental questions must be answered to create viable parameter markets. In this work, we address these basic questions, propose a framework containing the infrastructure necessary for market operations to take place, study strategies for exchanging parameters, and offer means for agents to monetize parameters. Excitingly, compared to agents who train siloed models from scratch, we show that it is possible to mutually gain by using the market, even in competitive settings. This suggests that the notion of parameter markets may be a useful paradigm for improving large-scale model training in the future.
****
Relax, it doesn’t matter how you get there: A new self-supervised approach for multi-timescale behavior analysis
Mehdi Azabou, Michael Mendelson, Nauman Ahad, Maks Sorokin, Shantanu Thakoor, Carolina Urzay, Eva Dyer
Unconstrained and natural behavior consists of dynamics that are complex and unpredictable, especially when trying to predict what will happen multiple steps into the future. While some success has been found in building representations of animal behavior under constrained or simplified task-based conditions, many of these models cannot be applied to free and naturalistic settings where behavior becomes increasingly hard to model. In this work, we develop a multi-task representation learning model for animal behavior that combines two novel components: (i) an action-prediction objective that aims to predict the distribution of actions over future timesteps, and (ii) a multi-scale architecture that builds separate latent spaces to accommodate short- and long-term dynamics. After demonstrating the ability of the method to build representations of both local and glob

al dynamics in robots in varying environments and terrains, we apply our method to the MABe 2022 Multi-Agent Behavior challenge, where our model ranks first overall on both mice and fly benchmarks. In all of these cases, we show that our model can build representations that capture the many different factors that drive behavior and solve a wide range of downstream tasks.
****
A Measure-Theoretic Axiomatisation of Causality
Junhyung Park, Simon Buchholz, Bernhard Schölkopf, Krikamol Muandet
Causality is a central concept in a wide range of research areas, yet there is still no universally agreed axiomatisation of causality. We view causality both as an extension of probability theory and as a study of what happens when one intervenes on a system, and argue in favour of taking Kolmogorov's measure-theoretic axiomatisation of probability as the starting point towards an axiomatisation of causality. To that end, we propose the notion of a causal space, consisting of a probability space along with a collection of transition probability kernels, called causal kernels, that encode the causal information of the space. Our proposed framework is not only rigorously grounded in measure theory, but it also sheds light on long-standing limitations of existing frameworks including, for example, cycles, latent variables and stochastic processes.
****
LLaVA-Med: Training a Large Language-and-Vision Assistant for Biomedicine in One Day
Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, Jianfeng Gao
Conversational generative AI has demonstrated remarkable promise for empowering biomedical practitioners, but current investigations focus on unimodal text. Multimodal conversational AI has seen rapid progress by leveraging billions of image-text pairs from the public web, but such general-domain vision-language models still lack sophistication in understanding and conversing about biomedical images. In this paper, we propose a cost-efficient approach for training a vision-language conversational assistant that can answer open-ended research questions of biomedical images. The key idea is to leverage a large-scale, broad-coverage biomedical figure-caption dataset extracted from PubMed Central, use GPT-4 to self-instruct open-ended instruction-following data from the captions, and then fine-tune a large general-domain vision-language model using a novel curriculum learning method. Specifically, the model first learns to align biomedical vocabulary using the figure-caption pairs as is, then learns to master open-ended conversational semantics using GPT-4 generated instruction-following data, broadly mimicking how a layperson gradually acquires biomedical knowledge. This enables us to train a Large Language and Vision Assistant for BioMedicine (LLaVA-Med) in less than 15 hours (with eight A100s). LLaVA-Med exhibits excellent multimodal conversational capability and can follow open-ended instruction to assist with inquiries about a biomedical image. On three standard biomedical visual question answering datasets, LLaVA-Med outperforms previous supervised state-of-the-art on certain metrics. To facilitate biomedical multimodal research, we will release our instruction-following data and the LLaVA-Med model.
****
Networks are Slacking Off: Understanding Generalization Problem in Image Deraining
Jinjin Gu, Xianzheng Ma, Xiangtao Kong, Yu Qiao, Chao Dong
Deep deraining networks consistently encounter substantial generalization issues when deployed in real-world applications, although they are successful in laboratory benchmarks. A prevailing perspective in deep learning encourages using highly complex data for training, with the expectation that richer image background content will facilitate overcoming the generalization problem. However, through comprehensive and systematic experimentation, we discover that this strategy does not enhance the generalization capability of these networks. On the contrary, it exacerbates the tendency of networks to overfit specific degradations. Our experiments reveal that better generalization in a deraining network can be achieved by simplifying the complexity of the training background images. This is bec

ause that the networks are ``slacking off'' during training, that is, learning the least complex elements in the image background and degradation to minimize training loss. When the background images are less complex than the rain streaks, the network will prioritize the background reconstruction, thereby suppressing overfitting the rain patterns and leading to improved generalization performance. Our research offers a valuable perspective and methodology for better understanding the generalization problem in low-level vision tasks and displays promising potential for practical application.
****

Architecture Matters: Uncovering Implicit Mechanisms in Graph Contrastive Learning

Xiaojun Guo, Yifei Wang, Zeming Wei, Yisen Wang

With the prosperity of contrastive learning for visual representation learning (VCL), it is also adapted to the graph domain and yields promising performance. However, through a systematic study of various graph contrastive learning (GCL) methods, we observe that some common phenomena among existing GCL methods that are quite different from the original VCL methods, including 1) positive samples are not a must for GCL; 2) negative samples are not necessary for graph classification, neither for node classification when adopting specific normalization modules; 3) data augmentations have much less influence on GCL, as simple domain-agnostic augmentations (e.g., Gaussian noise) can also attain fairly good performance. By uncovering how the implicit inductive bias of GNNs works in contrastive learning, we theoretically provide insights into the above intriguing properties of GCL. Rather than directly porting existing VCL methods to GCL, we advocate for more attention toward the unique architecture of graph learning and consider its implicit influence when designing GCL methods. Code is available at https://github.com/PKU-ML/ArchitectureMattersGCL.
****

Text Promptable Surgical Instrument Segmentation with Vision-Language Models

Zijian Zhou, Oluwatosin Alabi, Meng Wei, Tom Vercauteren, Miaojing Shi

In this paper, we propose a novel text promptable surgical instrument segmentation approach to overcome challenges associated with diversity and differentiation of surgical instruments in minimally invasive surgeries. We redefine the task as text promptable, thereby enabling a more nuanced comprehension of surgical instruments and adaptability to new instrument types. Inspired by recent advancements in vision-language models, we leverage pretrained image and text encoders as our model backbone and design a text promptable mask decoder consisting of attention- and convolution-based prompting schemes for surgical instrument segmentation prediction. Our model leverages multiple text prompts for each surgical instrument through a new mixture of prompts mechanism, resulting in enhanced segmentation performance. Additionally, we introduce a hard instrument area reinforcement module to improve image feature comprehension and segmentation precision. Extensive experiments on several surgical instrument segmentation datasets demonstrate our model's superior performance and promising generalization capability. To our knowledge, this is the first implementation of a promptable approach to surgical instrument segmentation, offering significant potential for practical application in the field of robotic-assisted surgery. Code is available at https://github.com/franciszzj/TP-SIS.
****

OpenDataVal: a Unified Benchmark for Data Valuation

Kevin Jiang, Weixin Liang, James Y. Zou, Yongchan Kwon

Assessing the quality and impact of individual data points is critical for improving model performance and mitigating undesirable biases within the training dataset. Several data valuation algorithms have been proposed to quantify data quality, however, there lacks a systemic and standardized benchmarking system for data valuation. In this paper, we introduce OpenDataVal, an easy-to-use and unified benchmark framework that empowers researchers and practitioners to apply and compare various data valuation algorithms. OpenDataVal provides an integrated environment that includes (i) a diverse collection of image, natural language, and tabular datasets, (ii) implementations of eleven different state-of-the-art data

valuation algorithms, and (iii) a prediction model API that can import any models in scikit-learn. Furthermore, we propose four downstream machine learning tasks for evaluating the quality of data values. We perform benchmarking analysis using OpenDataVal, quantifying and comparing the efficacy of state-of-the-art data valuation approaches. We find that no single algorithm performs uniformly best across all tasks, and an appropriate algorithm should be employed for a user's downstream task. OpenDataVal is publicly available at https://opendataval.github.io with comprehensive documentation. Furthermore, we provide a leaderboard where researchers can evaluate the effectiveness of their own data valuation algorithms.
****
On the Consistency of Maximum Likelihood Estimation of Probabilistic Principal Component Analysis
Arghya Datta, Sayak Chakrabarty
Probabilistic principal component analysis (PPCA) is currently one of the most used statistical tools to reduce the ambient dimension of the data. From multidimensional scaling to the imputation of missing data, PPCA has a broad spectrum of applications ranging from science and engineering to quantitative finance.\Despite this wide applicability in various fields, hardly any theoretical guarantees exist to justify the soundness of the maximal likelihood (ML) solution for this model. In fact, it is well known that the maximum likelihood estimation (MLE) can only recover the true model parameters up to a rotation. The main obstruction is posed by the inherent identifiability nature of the PPCA model resulting from the rotational symmetry of the parameterization. To resolve this ambiguity, we propose a novel approach using quotient topological spaces and in particular, we show that the maximum likelihood solution is consistent in an appropriate quotient Euclidean space. Furthermore, our consistency results encompass a more general class of estimators beyond the MLE. Strong consistency of the ML estimate and consequently strong covariance estimation of the PPCA model have also been established under a compactness assumption.
****
Similarity, Compression and Local Steps: Three Pillars of Efficient Communications for Distributed Variational Inequalities
Aleksandr Beznosikov, Martin Takac, Alexander Gasnikov
Variational inequalities are a broad and flexible class of problems that includes minimization, saddle point, and fixed point problems as special cases. Therefore, variational inequalities are used in various applications ranging from equilibrium search to adversarial learning. With the increasing size of data and models, today's instances demand parallel and distributed computing for real-world machine learning problems, most of which can be represented as variational inequalities. Meanwhile, most distributed approaches have a significant bottleneck -- the cost of communications. The three main techniques to reduce the total number of communication rounds and the cost of one such round are the similarity of local functions, compression of transmitted information, and local updates. In this paper, we combine all these approaches. Such a triple synergy did not exist before for variational inequalities and saddle problems, nor even for minimization problems. The methods presented in this paper have the best theoretical guarantees of communication complexity and are significantly ahead of other methods for distributed variational inequalities. The theoretical results are confirmed by adversarial learning experiments on synthetic and real datasets.
****
Lookaround Optimizer: k steps around, 1 step average
Jiangtao Zhang, Shunyu Liu, Jie Song, Tongtian Zhu, Zhengqi Xu, Mingli Song
Weight Average (WA) is an active research topic due to its simplicity in ensembling deep networks and the effectiveness in promoting generalization. Existing weight average approaches, however, are often carried out along only one training trajectory in a post-hoc manner (i.e., the weights are averaged after the entire training process is finished), which significantly degrades the diversity between networks and thus impairs the effectiveness. In this paper, inspired by weight average, we propose Lookaround, a straightforward yet effective SGD-based opti

mizer leading to flatter minima with better generalization. Specifically, Lookaround iterates two steps during the whole training period: the around step and the average step. In each iteration, 1) the around step starts from a common point and trains multiple networks simultaneously, each on transformed data by a different data augmentation, and 2) the average step averages these trained networks to get the averaged network, which serves as the starting point for the next iteration. The around step improves the functionality diversity while the average step guarantees the weight locality of these networks during the whole training, which is essential for WA to work. We theoretically explain the superiority of Lookaround by convergence analysis, and make extensive experiments to evaluate Lookaround on popular benchmarks including CIFAR and ImageNet with both CNNs and ViTs, demonstrating clear superiority over state-of-the-arts. Our code is available at https://github.com/Ardcy/Lookaround.
****
The Quantization Model of Neural Scaling
Eric Michaud, Ziming Liu, Uzay Girit, Max Tegmark
We propose the Quantization Model of neural scaling laws, explaining both the observed power law dropoff of loss with model and data size, and also the sudden emergence of new capabilities with scale. We derive this model from what we call the Quantization Hypothesis, where network knowledge and skills are "quantized" into discrete chunks (quanta). We show that when quanta are learned in order of decreasing use frequency, then a power law in use frequencies explains observed power law scaling of loss. We validate this prediction on toy datasets, then study how scaling curves decompose for large language models. Using language model gradients, we automatically decompose model behavior into a diverse set of skills (quanta). We tentatively find that the frequency at which these quanta are used in the training distribution roughly follows a power law corresponding with the empirical scaling exponent for language models, a prediction of our theory.
****
ε-fractional core stability in Hedonic Games.
Simone Fioravanti, Michele Flammini, Bojana Kodric, Giovanna Varricchio
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Semi-Supervised Domain Generalization with Known and Unknown Classes
Lei Zhang, Ji-Fu Li, Wei Wang
Semi-Supervised Domain Generalization (SSDG) aims to learn a model that is generalizable to an unseen target domain with only a few labels, and most existing SSDG methods assume that unlabeled training and testing samples are all known classes. However, a more realistic scenario is that known classes may be mixed with some unknown classes in unlabeled training and testing data. To deal with such a scenario, we propose the Class-Wise Adaptive Exploration and Exploitation (CWAEE) method. In particular, we explore unlabeled training data by using one-vs-rest classifiers and class-wise adaptive thresholds to detect known and unknown classes, and exploit them by adopting consistency regularization on augmented samples based on Fourier Transformation to improve the unseen domain generalization. The experiments conducted on real-world datasets verify the effectiveness and superiority of our method.
****
When Do Graph Neural Networks Help with Node Classification? Investigating the Homophily Principle on Node Distinguishability
Sitao Luan, Chenqing Hua, Minkai Xu, Qincheng Lu, Jiaqi Zhu, Xiao-Wen Chang, Jie Fu, Jure Leskovec, Doina Precup
Homophily principle, i.e., nodes with the same labels are more likely to be connected, has been believed to be the main reason for the performance superiority of Graph Neural Networks (GNNs) over Neural Networks on node classification tasks. Recent research suggests that, even in the absence of homophily, the advantage of GNNs still exists as long as nodes from the same class share similar neighbo

rhood patterns. However, this argument only considers intra-class Node Distinguishability (ND) but neglects inter-class ND, which provides incomplete understanding of homophily on GNNs. In this paper, we first demonstrate such deficiency with examples and argue that an ideal situation for ND is to have smaller intra-class ND than inter-class ND. To formulate this idea and study ND deeply, we propose Contextual Stochastic Block Model for Homophily (CSBM-H) and define two metrics, Probabilistic Bayes Error (PBE) and negative generalized Jeffreys divergence, to quantify ND. With the metrics, we visualize and analyze how graph filters, node degree distributions and class variances influence ND, and investigate the combined effect of intra- and inter-class ND. Besides, we discovered the mid-homophily pitfall, which occurs widely in graph datasets. Furthermore, we verified that, in real-work tasks, the superiority of GNNs is indeed closely related to both intra- and inter-class ND regardless of homophily levels. Grounded in this observation, we propose a new hypothesis-testing based performance metric beyond homophily, which is non-linear, feature-based and can provide statistical threshold value for GNNs' the superiority. Experiments indicate that it is significantly more effective than the existing homophily metrics on revealing the advantage and disadvantage of graph-aware modes on both synthetic and benchmark real-world datasets.
****
(Almost) Provable Error Bounds Under Distribution Shift via Disagreement Discrepancy
Elan Rosenfeld, Saurabh Garg
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Schema-learning and rebinding as mechanisms of in-context learning and emergence
Sivaramakrishnan Swaminathan, Antoine Dedieu, Rajkumar Vasudeva Raju, Murray Shanahan, Miguel Lazaro-Gredilla, Dileep George
In-context learning (ICL) is one of the most powerful and most unexpected capabilities to emerge in recent transformer-based large language models (LLMs). Yet the mechanisms that underlie it are poorly understood. In this paper, we demonstrate that comparable ICL capabilities can be acquired by an alternative sequence prediction learning method using clone-structured causal graphs (CSCGs). Moreover, a key property of CSCGs is that, unlike transformer-based LLMs, they are {\em interpretable}, which considerably simplifies the task of explaining how ICL works. Specifically, we show that it uses a combination of (a) learning template (schema) circuits for pattern completion, (b) retrieving relevant templates in a context-sensitive manner, and (c) rebinding of novel tokens to appropriate slots in the templates. We go on to marshall evidence for the hypothesis that similar mechanisms underlie ICL in LLMs. For example, we find that, with CSCGs as with LLMs, different capabilities emerge at different levels of overparameterization, suggesting that overparameterization helps in learning more complex template (schema) circuits. By showing how ICL can be achieved with small models and datasets, we open up a path to novel architectures, and take a vital step towards a more general understanding of the mechanics behind this important capability.
****
CAP: Correlation-Aware Pruning for Highly-Accurate Sparse Vision Models
Denis Kuznedelev, Eldar Kurti■, Elias Frantar, Dan Alistarh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Your representations are in the network: composable and parallel adaptation for large scale models
Yonatan Dukler, Alessandro Achille, Hao Yang, Varsha Vivek, Luca Zancato, Benjamin Bowman, Avinash Ravichandran, Charless Fowlkes, Ashwin Swaminathan, Stefano S

oatto
We present a framework for transfer learning that efficiently adapts a large bas
e-model by learning lightweight cross-attention modules attached to its intermed
iate activations.We name our approach InCA (Introspective-Cross-Attention) and s
how that it can efficiently survey a network’s representations and identify stro
ng performing adapter models for a downstream task.During training, InCA enables
 training numerous adapters efficiently and in parallel, isolated from the froze
n base model. On the ViT-L/16 architecture, our experiments show that a single a
dapter, 1.3% of the full model, is able to reach full fine-tuning accuracy on av
erage across 11 challenging downstream classification tasks.Compared with other
forms of parameter-efficient adaptation, the isolated nature of the InCA adaptat
ion is computationally desirable for large-scale models. For instance, we adapt
ViT-G/14 (1.8B+ parameters) quickly with 20+ adapters in parallel on a single V1
00 GPU (76% GPU memory reduction) and exhaustively identify its most useful repr
esentations.We further demonstrate how the adapters learned by InCA can be incre
mentally modified or combined for flexible learning scenarios and our approach a
chieves state of the art performance on the ImageNet-to-Sketch multi-task benchm
ark.
****
Learning Energy-based Model via Dual-MCMC Teaching
Jiali Cui, Tian Han
This paper studies the fundamental learning problem of the energy-based model (E
BM). Learning the EBM can be achieved using the maximum likelihood estimation (M
LE), which typically involves the Markov Chain Monte Carlo (MCMC) sampling, such
 as the Langevin dynamics. However, the noise-initialized Langevin dynamics can
be challenging in practice and hard to mix. This motivates the exploration of jo
int training with the generator model where the generator model serves as a comp
lementary model to bypass MCMC sampling. However, such a method can be less accu
rate than the MCMC and result in biased EBM learning. While the generator can al
so serve as an initializer model for better MCMC sampling, its learning can be b
iased since it only matches the EBM and has no access to empirical training exam
ples. Such biased generator learning may limit the potential of learning the EBM
. To address this issue, we present a joint learning framework that interweaves
the maximum likelihood learning algorithm for both the EBM and the complementary
 generator model. In particular, the generator model is learned by MLE to match
both the EBM and the empirical data distribution, making it a more informative i
nitializer for MCMC sampling of EBM. Learning generator with observed examples t
ypically requires inference of the generator posterior. To ensure accurate and e
fficient inference, we adopt the MCMC posterior sampling and introduce a complem
entary inference model to initialize such latent MCMC sampling. We show that thr
ee separate models can be seamlessly integrated into our joint framework through
 two (dual-) MCMC teaching, enabling effective and efficient EBM learning.
****
Performance-optimized deep neural networks are evolving into worse models of inf
erotemporal visual cortex
Drew Linsley, Ivan F Rodriguez Rodriguez, Thomas FEL, Michael Arcaro, Saloni Sha
rma, Margaret Livingstone, Thomas Serre
One of the most impactful findings in computational neuroscience over the past d
ecade is that the object recognition accuracy of deep neural networks (DNNs) cor
relates with their ability to predict neural responses to natural images in the
inferotemporal (IT) cortex. This discovery supported the long-held theory that o
bject recognition is a core objective of the visual cortex, and suggested that m
ore accurate DNNs would serve as better models of IT neuron responses to images.
 Since then, deep learning has undergone a revolution of scale: billion paramete
r-scale DNNs trained on billions of images are rivaling or outperforming humans
at visual tasks including object recognition. Have today's DNNs become more accu
rate at predicting IT neuron responses to images as they have grown more accurat
e at object recognition?Surprisingly, across three independent experiments, we f
ind that this is not the case. DNNs have become progressively worse models of IT
 as their accuracy has increased on ImageNet. To understand why DNNs experience

this trade-off and evaluate if they are still an appropriate paradigm for modeling the visual system, we turn to recordings of IT that capture spatially resolved maps of neuronal activity elicited by natural images. These neuronal activity maps reveal that DNNs trained on ImageNet learn to rely on different visual features than those encoded by IT and that this problem worsens as their accuracy increases. We successfully resolved this issue with the neural harmonizer, a plug-and-play training routine for DNNs that aligns their learned representations with humans. Our results suggest that harmonized DNNs break the trade-off between ImageNet accuracy and neural prediction accuracy that assails current DNNs and offer a path to more accurate models of biological vision. Our work indicates that the standard approach for modeling IT with task-optimized DNNs needs revision, and other biological constraints, including human psychophysics data, are needed to accurately reverse-engineer the visual cortex.
****
Private Federated Frequency Estimation: Adapting to the Hardness of the Instance
Jingfeng Wu, Wennan Zhu, Peter Kairouz, Vladimir Braverman
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Class Distributions Induced by Nearest Neighbor Graphs for Node Classification of Tabular Data
Federico Errica
Researchers have used nearest neighbor graphs to transform classical machine learning problems on tabular data into node classification tasks to solve with graph representation learning methods. Such artificial structures often reflect the homophily assumption, believed to be a key factor in the performances of deep graph networks. In light of recent results demystifying these beliefs, we introduce a theoretical framework to understand the benefits of Nearest Neighbor (NN) graphs when a graph structure is missing. We formally analyze the Cross-Class Neighborhood Similarity (CCNS), used to empirically evaluate the usefulness of structures, in the context of nearest neighbor graphs. Moreover, we study the class separability induced by deep graph networks on a k-NN graph. Motivated by the theory, our quantitative experiments demonstrate that, under full supervision, employing a k-NN graph offers no benefits compared to a structure-agnostic baseline. Qualitative analyses suggest that our framework is good at estimating the CCNS and hint at k-NN graphs never being useful for such classification tasks under full supervision, thus advocating for the study of alternative graph construction techniques in combination with deep graph networks.
****
VRA: Variational Rectified Activation for Out-of-distribution Detection
Mingyu Xu, Zheng Lian, Bin Liu, Jianhua Tao
Out-of-distribution (OOD) detection is critical to building reliable machine learning systems in the open world. Researchers have proposed various strategies to reduce model overconfidence on OOD data. Among them, ReAct is a typical and effective technique to deal with model overconfidence, which truncates high activations to increase the gap between in-distribution and OOD. Despite its promising results, is this technique the best choice? To answer this question, we leverage the variational method to find the optimal operation and verify the necessity of suppressing abnormally low and high activations and amplifying intermediate activations in OOD detection, rather than focusing only on high activations like ReAct. This motivates us to propose a novel technique called ``Variational Rectified Activation (VRA)'', which simulates these suppression and amplification operations using piecewise functions. Experimental results on multiple benchmark datasets demonstrate that our method outperforms existing post-hoc strategies. Meanwhile, VRA is compatible with different scoring functions and network architectures. Our code is available at https://github.com/zeroQiaoba/VRA.
****
Variational Gaussian processes for linear inverse problems

Thibault RANDRIANARISOA, Botond Szabo

By now Bayesian methods are routinely used in practice for solving inverse problems. In inverse problems the parameter or signal of interest is observed only indirectly, as an image of a given map, and the observations are typically further corrupted with noise. Bayes offers a natural way to regularize these problems via the prior distribution and provides a probabilistic solution, quantifying the remaining uncertainty in the problem. However, the computational costs of standard, sampling based Bayesian approaches can be overly large in such complex models. Therefore, in practice variational Bayes is becoming increasingly popular. Nevertheless, the theoretical understanding of these methods is still relatively limited, especially in context of inverse problems.In our analysis we investigate variational Bayesian methods for Gaussian process priors to solve linear inverse problems. We consider both mildly and severely ill-posed inverse problems and work with the popular inducing variable variational Bayes approach proposed by Titsias [Titsias, 2009]. We derive posterior contraction rates for the variational posterior in general settings and show that the minimax estimation rate can be attained by correctly tunned procedures. As specific examples we consider a collection of inverse problems including the heat equation, Volterra operator and Radon transform and inducing variable methods based on population and empirical spectral features.

****

Self-Supervised Learning with Lie Symmetries for Partial Differential Equations

Grégoire Mialon, Quentin Garrido, Hannah Lawrence, Danyal Rehman, Yann LeCun, Bobak Kiani

Machine learning for differential equations paves the way for computationally efficient alternatives to numerical solvers, with potentially broad impacts in science and engineering. Though current algorithms typically require simulated training data tailored to a given setting, one may instead wish to learn useful information from heterogeneous sources, or from real dynamical systems observations that are messy or incomplete. In this work, we learn general-purpose representations of PDEs from heterogeneous data by implementing joint embedding methods for self-supervised learning (SSL), a framework for unsupervised representation learning that has had notable success in computer vision. Our representation outperforms baseline approaches to invariant tasks, such as regressing the coefficients of a PDE, while also improving the time-stepping performance of neural solvers. We hope that our proposed methodology will prove useful in the eventual development of general-purpose foundation models for PDEs.

****

TempME: Towards the Explainability of Temporal Graph Neural Networks via Motif Discovery

Jialin Chen, Rex Ying

Temporal graphs are widely used to model dynamic systems with time-varying interactions. In real-world scenarios, the underlying mechanisms of generating future interactions in dynamic systems are typically governed by a set of recurring substructures within the graph, known as temporal motifs. Despite the success and prevalence of current temporal graph neural networks (TGNN), it remains uncertain which temporal motifs are recognized as the significant indications that trigger a certain prediction from the model, which is a critical challenge for advancing the explainability and trustworthiness of current TGNNs. To address this challenge, we propose a novel approach, called Temporal Motifs Explainer (TempME), which uncovers the most pivotal temporal motifs guiding the prediction of TGNNs. Derived from the information bottleneck principle, TempME extracts the most interaction-related motifs while minimizing the amount of contained information to preserve the sparsity and succinctness of the explanation. Events in the explanations generated by TempME are verified to be more spatiotemporally correlated than those of existing approaches, providing more understandable insights. Extensive experiments validate the superiority of TempME, with up to 8.21% increase in terms of explanation accuracy across six real-world datasets and up to 22.96% increase in boosting the prediction Average Precision of current TGNNs.

****

YouTube-ASL: A Large-Scale, Open-Domain American Sign Language-English Parallel Corpus
Dave Uthus, Garrett Tanzer, Manfred Georg
Machine learning for sign languages is bottlenecked by data. In this paper, we present YouTube-ASL, a large-scale, open-domain corpus of American Sign Language (ASL) videos and accompanying English captions drawn from YouTube. With ~1000 hours of videos and >2500 unique signers, YouTube-ASL is ~3x as large and has ~10x as many unique signers as the largest prior ASL dataset. We train baseline models for ASL to English translation on YouTube-ASL and evaluate them on How2Sign, where we achieve a new fine-tuned state of the art of 12.397 BLEU and, for the first time, nontrivial zero-shot results.
****
Finite-Time Analysis of Whittle Index based Q-Learning for Restless Multi-Armed Bandits with Neural Network Function Approximation
GUOJUN XIONG, Jian Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Adapting to Continuous Covariate Shift via Online Density Ratio Estimation
Yu-Jie Zhang, Zhen-Yu Zhang, Peng Zhao, Masashi Sugiyama
Dealing with distribution shifts is one of the central challenges for modern machine learning. One fundamental situation is the covariate shift, where the input distributions of data change from the training to testing stages while the input-conditional output distribution remains unchanged. In this paper, we initiate the study of a more challenging scenario --- continuous covariate shift --- in which the test data appear sequentially, and their distributions can shift continuously. Our goal is to adaptively train the predictor such that its prediction risk accumulated over time can be minimized. Starting with the importance-weighted learning, we theoretically show the method works effectively if the time-varying density ratios of test and train inputs can be accurately estimated. However, existing density ratio estimation methods would fail due to data scarcity at each time step. To this end, we propose an online density ratio estimation method that can appropriately reuse historical information. Our method is proven to perform well by enjoying a dynamic regret bound, which finally leads to an excess risk guarantee for the predictor. Empirical results also validate the effectiveness.
****
Hierarchical Integration Diffusion Model for Realistic Image Deblurring
Zheng Chen, Yulun Zhang, Ding Liu, bin xia, Jinjin Gu, Linghe Kong, Xin Yuan
Diffusion models (DMs) have recently been introduced in image deblurring and exhibited promising performance, particularly in terms of details reconstruction. However, the diffusion model requires a large number of inference iterations to recover the clean image from pure Gaussian noise, which consumes massive computational resources. Moreover, the distribution synthesized by the diffusion model is often misaligned with the target results, leading to restrictions in distortion-based metrics. To address the above issues, we propose the Hierarchical Integration Diffusion Model (HI-Diff), for realistic image deblurring. Specifically, we perform the DM in a highly compacted latent space to generate the prior feature for the deblurring process. The deblurring process is implemented by a regression-based method to obtain better distortion accuracy. Meanwhile, the highly compact latent space ensures the efficiency of the DM. Furthermore, we design the hierarchical integration module to fuse the prior into the regression-based model from multiple scales, enabling better generalization in complex blurry scenarios. Comprehensive experiments on synthetic and real-world blur datasets demonstrate that our HI-Diff outperforms state-of-the-art methods. Code and trained models are available at https://github.com/zhengchen1999/HI-Diff.
****
Efficient Beam Tree Recursion

Jishnu Ray Chowdhury, Cornelia Caragea
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Holistic Transfer: Towards Non-Disruptive Fine-Tuning with Partial Target Data
Cheng-Hao Tu, Hong-You Chen, Zheda Mai, Jike Zhong, Vardaan Pahuja, Tanya Berger-Wolf, Song Gao, Charles Stewart, Yu Su, Wei-Lun (Harry) Chao
We propose a learning problem involving adapting a pre-trained source model to the target domain for classifying all classes that appeared in the source data, using target data that covers only a partial label space. This problem is practical, as it is unrealistic for the target end-users to collect data for all classes prior to adaptation. However, it has received limited attention in the literature. To shed light on this issue, we construct benchmark datasets and conduct extensive experiments to uncover the inherent challenges. We found a dilemma --- on the one hand, adapting to the new target domain is important to claim better performance; on the other hand, we observe that preserving the classification accuracy of classes missing in the target adaptation data is highly challenging, let alone improving them. To tackle this, we identify two key directions: 1) disentangling domain gradients from classification gradients, and 2) preserving class relationships. We present several effective solutions that maintain the accuracy of the missing classes and enhance the overall performance, establishing solid baselines for holistic transfer of pre-trained models with partial target data.
****
Rethinking Semi-Supervised Imbalanced Node Classification from Bias-Variance Decomposition
Divin Yan, Gengchen Wei, Chen Yang, Shengzhong Zhang, zengfeng Huang
This paper introduces a new approach to address the issue of class imbalance in graph neural networks (GNNs) for learning on graph-structured data. Our approach integrates imbalanced node classification and Bias-Variance Decomposition, establishing a theoretical framework that closely relates data imbalance to model variance. We also leverage graph augmentation technique to estimate the variance and design a regularization term to alleviate the impact of imbalance. Exhaustive tests are conducted on multiple benchmarks, including naturally imbalanced datasets and public-split class-imbalanced datasets, demonstrating that our approach outperforms state-of-the-art methods in various imbalanced scenarios. This work provides a novel theoretical perspective for addressing the problem of imbalanced node classification in GNNs.
****
Practical Equivariances via Relational Conditional Neural Processes
Daolang Huang, Manuel Haussmann, Ulpu Remes, ST John, Grégoire Clarté, Kevin Luck, Samuel Kaski, Luigi Acerbi
Conditional Neural Processes (CNPs) are a class of metalearning models popular for combining the runtime efficiency of amortized inference with reliable uncertainty quantification. Many relevant machine learning tasks, such as in spatio-temporal modeling, Bayesian Optimization and continuous control, inherently contain equivariances – for example to translation – which the model can exploit for maximal performance. However, prior attempts to include equivariances in CNPs do not scale effectively beyond two input dimensions. In this work, we propose Relational Conditional Neural Processes (RCNPs), an effective approach to incorporate equivariances into any neural process model. Our proposed method extends the applicability and impact of equivariant neural processes to higher dimensions. We empirically demonstrate the competitive performance of RCNPs on a large array of tasks naturally containing equivariances.
****
FourierHandFlow: Neural 4D Hand Representation Using Fourier Query Flow
Jihyun Lee, Junbong Jang, Donghwan Kim, Minhyuk Sung, Tae-Kyun Kim
Recent 4D shape representations model continuous temporal evolution of implicit shapes by (1) learning query flows without leveraging shape and articulation pri

ors or (2) decoding shape occupancies separately for each time value. Thus, they do not effectively capture implicit correspondences between articulated shapes or regularize jittery temporal deformations. In this work, we present FourierHandFlow, which is a spatio-temporally continuous representation for human hands that combines a 3D occupancy field with articulation-aware query flows represented as Fourier series. Given an input RGB sequence, we aim to learn a fixed number of Fourier coefficients for each query flow to guarantee smooth and continuous temporal shape dynamics. To effectively model spatio-temporal deformations of articulated hands, we compose our 4D representation based on two types of Fourier query flow: (1) pose flow that models query dynamics influenced by hand articulation changes via implicit linear blend skinning and (2) shape flow that models query-wise displacement flow. In the experiments, our method achieves state-of-the-art results on video-based 4D reconstruction while being computationally more efficient than the existing 3D/4D implicit shape representations. We additionally show our results on motion inter- and extrapolation and texture transfer using the learned correspondences of implicit shapes. To the best of our knowledge, FourierHandFlow is the first neural 4D continuous hand representation learned from RGB videos. The code will be publicly accessible.
****
Safe Exploration in Reinforcement Learning: A Generalized Formulation and Algorithms
Akifumi Wachi, Wataru Hashimoto, Xun Shen, Kazumune Hashimoto
Safe exploration is essential for the practical use of reinforcement learning (RL) in many real-world scenarios. In this paper, we present a generalized safe exploration (GSE) problem as a unified formulation of common safe exploration problems. We then propose a solution of the GSE problem in the form of a meta-algorithm for safe exploration, MASE, which combines an unconstrained RL algorithm with an uncertainty quantifier to guarantee safety in the current episode while properly penalizing unsafe explorations before actual safety violation to discourage them in future episodes. The advantage of MASE is that we can optimize a policy while guaranteeing with a high probability that no safety constraint will be violated under proper assumptions. Specifically, we present two variants of MASE with different constructions of the uncertainty quantifier: one based on generalized linear models with theoretical guarantees of safety and near-optimality, and another that combines a Gaussian process to ensure safety with a deep RL algorithm to maximize the reward. Finally, we demonstrate that our proposed algorithm achieves better performance than state-of-the-art algorithms on grid-world and Safety Gym benchmarks without violating any safety constraints, even during training.
****
RevColV2: Exploring Disentangled Representations in Masked Image Modeling
Qi Han, Yuxuan Cai, Xiangyu Zhang
Masked image modeling (MIM) has become a prevalent pre-training setup for vision foundation models and attains promising performance. Despite its success, existing MIM methods discard the decoder network during downstream applica- tions, resulting in inconsistent representations between pre-training and fine-tuning and can hamper downstream task performance. In this paper, we propose a new architecture, RevColV2, which tackles this issue by keeping the entire autoen- coder architecture during both pre-training and fine-tuning. The main body of RevColV2 contains bottom-up columns and top-down columns, between which information is reversibly propagated and gradually disentangled. Such design enables our architecture with the nice property: maintaining disentangled low-level and semantic information at the end of the network in MIM pre-training. Our experimental results suggest that a foundation model with decoupled features can achieve competitive performance across multiple downstream vision tasks such as image classification, semantic segmentation and object detection. For exam- ple, after intermediate fine-tuning on ImageNet-22K dataset, RevColV2-L attains 88.4\% top-1 accuracy on ImageNet-1K classification and 58.6 mIoU on ADE20K semantic segmentation. With extra teacher and large scale dataset, RevColv2-L achieves 62.1 APbox on COCO detection and 60.4 mIoU on ADE20K semantic segmentation.

****
Mass-Producing Failures of Multimodal Systems with Language Models
Shengbang Tong, Erik Jones, Jacob Steinhardt
Deployed multimodal models can fail in ways that evaluators did not anticipate. In order to find these failures before deployment, we introduce MultiMon, a system that automatically identifies systematic failures---generalizable, natural-language descriptions that describe categories of individual failures. To uncover systematic failures, MultiMon scrapes for examples of erroneous agreement: inputs that produce the same output, but should not. It then prompts a language model to identify common categories and describe them in natural language. We use MultiMon to find 14 systematic failures (e.g."ignores quantifiers'') of the CLIP text-encoder, each comprising hundreds of distinct inputs (e.g."a shelf with a few/many books''). Because CLIP is the backbone for most state-of-the-art multimodal models, these inputs produce failures in Midjourney 5.1, DALL-E, VideoFusion, and others. MultiMon can also steer towards failures relevant to specific use cases, such as self-driving cars. We see MultiMon as a step towards evaluation that autonomously explores the long-tail of potential system failures.
****
STXD: Structural and Temporal Cross-Modal Distillation for Multi-View 3D Object Detection
Sujin Jang, Dae Ung Jo, Sung Ju Hwang, Dongwook Lee, Daehyun Ji
3D object detection (3DOD) from multi-view images is an economically appealing alternative to expensive LiDAR-based detectors, but also an extremely challenging task due to the absence of precise spatial cues. Recent studies have leveraged the teacher-student paradigm for cross-modal distillation, where a strong LiDAR-modality teacher transfers useful knowledge to a multi-view-based image-modality student. However, prior approaches have only focused on minimizing global distances between cross-modal features, which may lead to suboptimal knowledge distillation results. Based on these insights, we propose a novel structural and temporal cross-modal knowledge distillation (STXD) framework for multi-view 3DOD. First, STXD reduces redundancy of the feature components of the student by regularizing the cross-correlation of cross-modal features, while maximizing their similarities. Second, to effectively transfer temporal knowledge, STXD encodes temporal relations of features across a sequence of frames via similarity maps. Lastly, STXD also adopts a response distillation method to further enhance the quality of knowledge distillation at the output-level. Our extensive experiments demonstrate that STXD significantly improves the NDS and mAP of the based student detectors by 2.8%~4.5% on the nuScenes testing dataset.
****
Battle of the Backbones: A Large-Scale Comparison of Pretrained Models across Computer Vision Tasks
Micah Goldblum, Hossein Souri, Renkun Ni, Manli Shu, Viraj Prabhu, Gowthami Somepalli, Prithvijit Chattopadhyay, Mark Ibrahim, Adrien Bardes, Judy Hoffman, Rama Chellappa, Andrew G. Wilson, Tom Goldstein
Neural network based computer vision systems are typically built on a backbone, a pretrained or randomly initialized feature extractor. Several years ago, the default option was an ImageNet-trained convolutional neural network. However, the recent past has seen the emergence of countless backbones pretrained using various algorithms and datasets. While this abundance of choice has led to performance increases for a range of systems, it is difficult for practitioners to make informed decisions about which backbone to choose. Battle of the Backbones (BoB) makes this choice easier by benchmarking a diverse suite of pretrained models, including vision-language models, those trained via self-supervised learning, and the Stable Diffusion backbone, across a diverse set of computer vision tasks ranging from classification to object detection to OOD generalization and more. Furthermore, BoB sheds light on promising directions for the research community to advance computer vision by illuminating strengths and weakness of existing approaches through a comprehensive analysis conducted on more than 1500 training runs. While vision transformers (ViTs) and self-supervised learning (SSL) are increasingly popular, we find that convolutional neural networks pretrained in a

supervised fashion on large training sets still perform best on most tasks among the models we consider. Moreover, in apples-to-apples comparisons on the same architectures and similarly sized pretraining datasets, we find that SSL backbones are highly competitive, indicating that future works should perform SSL pretraining with advanced architectures and larger pretraining datasets. We release the raw results of our experiments along with code that allows researchers to put their own backbones through the gauntlet here: https://github.com/hsouri/Battle-of-the-Backbones.
****
Beyond Deep Ensembles: A Large-Scale Evaluation of Bayesian Deep Learning under Distribution Shift
Florian Seligmann, Philipp Becker, Michael Volpp, Gerhard Neumann
Bayesian deep learning (BDL) is a promising approach to achieve well-calibrated predictions on distribution-shifted data. Nevertheless, there exists no large-scale survey that evaluates recent SOTA methods on diverse, realistic, and challenging benchmark tasks in a systematic manner. To provide a clear picture of the current state of BDL research, we evaluate modern BDL algorithms on real-world datasets from the WILDS collection containing challenging classification and regression tasks, with a focus on generalization capability and calibration under distribution shift. We compare the algorithms on a wide range of large, convolutional and transformer-based neural network architectures. In particular, we investigate a signed version of the expected calibration error that reveals whether the methods are over- or underconfident, providing further insight into the behavior of the methods. Further, we provide the first systematic evaluation of BDL for fine-tuning large pre-trained models, where training from scratch is prohibitively expensive. Finally, given the recent success of Deep Ensembles, we extend popular single-mode posterior approximations to multiple modes by the use of ensembles. While we find that ensembling single-mode approximations generally improves the generalization capability and calibration of the models by a significant margin, we also identify a failure mode of ensembles when finetuning large transformer-based language models. In this setting, variational inference based approaches such as last-layer Bayes By Backprop outperform other methods in terms of accuracy by a large margin, while modern approximate inference algorithms such as SWAG achieve the best calibration.
****
(S)GD over Diagonal Linear Networks: Implicit bias, Large Stepsizes and Edge of Stability
Mathieu Even, Scott Pesme, Suriya Gunasekar, Nicolas Flammarion
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Optimal Preconditioning and Fisher Adaptive Langevin Sampling
Michalis Titsias
We define an optimal preconditioning for the Langevin diffusion by analytically optimizing the expected squared jumped distance. This yields as the optimal preconditioning an inverse Fisher information covariance matrix, where the covariance matrix is computed as the outer product of log target gradients averaged under the target. We apply this result to the Metropolis adjusted Langevin algorithm (MALA) and derive a computationally efficient adaptive MCMC scheme that learns the preconditioning from the history of gradients produced as the algorithm runs. We show in several experiments that the proposed algorithm is very robust in high dimensions and significantly outperforms other methods, including a closely related adaptive MALA scheme that learns the preconditioning with standard adaptive MCMC as well as the position-dependent Riemannian manifold MALA sampler.
****
AdaptSSR: Pre-training User Model with Augmentation-Adaptive Self-Supervised Ranking
Yang Yu, Qi Liu, Kai Zhang, Yuren Zhang, Chao Song, Min Hou, Yuqing Yuan, Zhihao

Ye, ZAIXI ZHANG, Sanshi Lei Yu
User modeling, which aims to capture users' characteristics or interests, heavily relies on task-specific labeled data and suffers from the data sparsity issue. Several recent studies tackled this problem by pre-training the user model on massive user behavior sequences with a contrastive learning task. Generally, these methods assume different views of the same behavior sequence constructed via data augmentation are semantically consistent, i.e., reflecting similar characteristics or interests of the user, and thus maximizing their agreement in the feature space. However, due to the diverse interests and heavy noise in user behaviors, existing augmentation methods tend to lose certain characteristics of the user or introduce noisy behaviors. Thus, forcing the user model to directly maximize the similarity between the augmented views may result in a negative transfer. To this end, we propose to replace the contrastive learning task with a new pretext task: Augmentation-Adaptive SelfSupervised Ranking (AdaptSSR), which alleviates the requirement of semantic consistency between the augmented views while pre-training a discriminative user model. Specifically, we adopt a multiple pairwise ranking loss which trains the user model to capture the similarity orders between the implicitly augmented view, the explicitly augmented view, and views from other users. We further employ an in-batch hard negative sampling strategy to facilitate model training. Moreover, considering the distinct impacts of data augmentation on different behavior sequences, we design an augmentation-adaptive fusion mechanism to automatically adjust the similarity order constraint applied to each sample based on the estimated similarity between the augmented views. Extensive experiments on both public and industrial datasets with six downstream tasks verify the effectiveness of AdaptSSR.
****
Anytime Model Selection in Linear Bandits
Parnian Kassraie, Nicolas Emmenegger, Andreas Krause, Aldo Pacchiano
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Towards Personalized Federated Learning via Heterogeneous Model Reassembly
Jiaqi Wang, Xingyi Yang, Suhan Cui, Liwei Che, Lingjuan Lyu, Dongkuan (DK) Xu, Fenglong Ma
This paper focuses on addressing the practical yet challenging problem of model heterogeneity in federated learning, where clients possess models with different network structures. To track this problem, we propose a novel framework called pFedHR, which leverages heterogeneous model reassembly to achieve personalized federated learning. In particular, we approach the problem of heterogeneous model personalization as a model-matching optimization task on the server side. Moreover, pFedHR automatically and dynamically generates informative and diverse personalized candidates with minimal human intervention. Furthermore, our proposed heterogeneous model reassembly technique mitigates the adverse impact introduced by using public data with different distributions from the client data to a certain extent. Experimental results demonstrate that pFedHR outperforms baselines on three datasets under both IID and Non-IID settings. Additionally, pFedHR effectively reduces the adverse impact of using different public data and dynamically generates diverse personalized models in an automated manner.
****
Language Models Can Improve Event Prediction by Few-Shot Abductive Reasoning
Xiaoming Shi, Siqiao Xue, Kangrui Wang, Fan Zhou, James Zhang, Jun Zhou, Chenhao Tan, Hongyuan Mei
Large language models have shown astonishing performance on a wide range of reasoning tasks. In this paper, we investigate whether they could reason about real-world events and help improve the prediction performance of event sequence models. We design LAMP, a framework that integrates a large language model in event prediction. Particularly, the language model performs abductive reasoning to assist an event sequence model: the event model proposes predictions on future event

s given the past; instructed by a few expert-annotated demonstrations, the language model learns to suggest possible causes for each proposal; a search module finds out the previous events that match the causes; a scoring function learns to examine whether the retrieved events could actually cause the proposal. Through extensive experiments on several challenging real-world datasets, we demonstrate that our framework---thanks to the reasoning capabilities of large language models---could significantly outperform the state-of-the-art event sequence models.
****
Complexity Matters: Rethinking the Latent Space for Generative Modeling
Tianyang Hu, Fei Chen, Haonan Wang, Jiawei Li, Wenjia Wang, Jiacheng Sun, Zhenguo Li
In generative modeling, numerous successful approaches leverage a low-dimensional latent space, e.g., Stable Diffusion models the latent space induced by an encoder and generates images through a paired decoder. Although the selection of the latent space is empirically pivotal, determining the optimal choice and the process of identifying it remain unclear. In this study, we aim to shed light on this under-explored topic by rethinking the latent space from the perspective of model complexity. Our investigation starts with the classic generative adversarial networks (GANs). Inspired by the GAN training objective, we propose a novel "distance" between the latent and data distributions, whose minimization coincides with that of the generator complexity. The minimizer of this distance is characterized as the optimal data-dependent latent that most effectively capitalizes on the generator's capacity. Then, we consider parameterizing such a latent distribution by an encoder network and propose a two-stage training strategy called Decoupled Autoencoder (DAE), where the encoder is only updated in the first stage with an auxiliary decoder and then frozen in the second stage while the actual decoder is being trained. DAE can improve the latent distribution and as a result, improve the generative performance. Our theoretical analyses are corroborated by comprehensive experiments on various models such as VQGAN and Diffusion Transformer, where our modifications yield significant improvements in sample quality with decreased model complexity.
****
Riemannian stochastic optimization methods avoid strict saddle points
Ya-Ping Hsieh, Mohammad Reza Karimi Jaghargh, Andreas Krause, Panayotis Mertikopoulos
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Toward Better PAC-Bayes Bounds for Uniformly Stable Algorithms
Sijia Zhou, Yunwen Lei, Ata Kaban
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Cheap and Quick: Efficient Vision-Language Instruction Tuning for Large Language Models
Gen Luo, Yiyi Zhou, Tianhe Ren, Shengxin Chen, Xiaoshuai Sun, Rongrong Ji
Recently, growing interest has been aroused in extending the multimodal capability of large language models (LLMs), e.g., vision-language (VL) learning, which is regarded as the next milestone of artificial general intelligence. However, existing solutions are prohibitively expensive, which not only need to optimize excessive parameters, but also require another large-scale pre-training before VL instruction tuning. In this paper, we propose a novel and affordable solution for the effective VL adaption of LLMs, called Mixture-of-Modality Adaptation (MMA). Instead of using large neural networks to connect the image encoder and LLM, MMA adopts lightweight modules, i.e., adapters, to bridge the gap between L

LMs and VL tasks, which also enables the joint optimization of the image and lan guage models. Meanwhile, MMA is also equipped with a routing algorithm to help L LMs achieve an automatic shift between single- and multi-modal instructions wi thout compromising their ability of natural language understanding. To validate MMA, we apply it to a recent LLM called LLaMA and term this formed large vision -language instructed model as LaVIN. To validate MMA and LaVIN, we conduct exte nsive experiments under two setups, namely multimodal science question answeri ng and multimodal dialogue. The experimental results not only demonstrate the co mpetitive performance and the superior training efficiency of LaVIN than exist ing multimodal LLMs, but also confirm its great potential as a general-purpo se chatbot. More importantly, the actual expenditure of LaVIN is extremely cheap , e.g., only 1.4 training hours with 3.8M trainable parameters, greatly confirmi ng the effectiveness of MMA. Our code is anonymously released at: https://an onymous.4open.science/r/LaVIN--1067.
****

GADBench: Revisiting and Benchmarking Supervised Graph Anomaly Detection
Jianheng Tang, Fengrui Hua, Ziqi Gao, Peilin Zhao, Jia Li
With a long history of traditional Graph Anomaly Detection (GAD) algorithms and recently popular Graph Neural Networks (GNNs), it is still not clear (1) how the y perform under a standard comprehensive setting, (2) whether GNNs can outperfor m traditional algorithms such as tree ensembles, and (3) how about their efficie ncy on large-scale graphs. In response, we introduce GADBench---a benchmark tool dedicated to supervised anomalous node detection in static graphs. GADBench fac ilitates a detailed comparison across 29 distinct models on ten real-world GAD d atasets, encompassing thousands to millions (~6M) nodes. Our main finding is tha t tree ensembles with simple neighborhood aggregation can outperform the latest GNNs tailored for the GAD task. We shed light on the current progress of GAD, se tting a robust groundwork for subsequent investigations in this domain. GADBench is open-sourced at https://github.com/squareRoot3/GADBench.
****

Brain encoding models based on multimodal transformers can transfer across langu age and vision
Jerry Tang, Meng Du, Vy Vo, VASUDEV LAL, Alexander Huth
Encoding models have been used to assess how the human brain represents concepts in language and vision. While language and vision rely on similar concept repre sentations, current encoding models are typically trained and tested on brain re sponses to each modality in isolation. Recent advances in multimodal pretraining have produced transformers that can extract aligned representations of concepts in language and vision. In this work, we used representations from multimodal t ransformers to train encoding models that can transfer across fMRI responses to stories and movies. We found that encoding models trained on brain responses to one modality can successfully predict brain responses to the other modality, par ticularly in cortical regions that represent conceptual meaning. Further analysi s of these encoding models revealed shared semantic dimensions that underlie con cept representations in language and vision. Comparing encoding models trained u sing representations from multimodal and unimodal transformers, we found that mu ltimodal transformers learn more aligned representations of concepts in language and vision. Our results demonstrate how multimodal transformers can provide ins ights into the brain’s capacity for multimodal processing.
****

PointGPT: Auto-regressively Generative Pre-training from Point Clouds
Guangyan Chen, Meiling Wang, Yi Yang, Kai Yu, Li Yuan, Yufeng Yue
Large language models (LLMs) based on the generative pre-training transformer (G PT) have demonstrated remarkable effectiveness across a diverse range of downstr eam tasks. Inspired by the advancements of the GPT, we present PointGPT, a novel approach that extends the concept of GPT to point clouds, addressing the challe nges associated with disorder properties, low information density, and task gaps . Specifically, a point cloud auto-regressive generation task is proposed to pre -train transformer models. Our method partitions the input point cloud into mult iple point patches and arranges them in an ordered sequence based on their spati

al proximity. Then, an extractor-generator based transformer decode, with a dual masking strategy, learns latent representations conditioned on the preceding point patches, aiming to predict the next one in an auto-regressive manner. To explore scalability and enhance performance, a larger pre-training dataset is collected. Additionally, a subsequent post-pre-training stage is introduced, incorporating a labeled hybrid dataset. Our scalable approach allows for learning high-capacity models that generalize well, achieving state-of-the-art performance on various downstream tasks. In particular, our approach achieves classification accuracies of 94.9% on the ModelNet40 dataset and 93.4% on the ScanObjectNN dataset, outperforming all other transformer models. Furthermore, our method also attains new state-of-the-art accuracies on all four few-shot learning benchmarks. Codes are available at https://github.com/CGuangyan-BIT/PointGPT.
****
Symbol-LLM: Leverage Language Models for Symbolic System in Visual Human Activity Reasoning
Xiaoqian Wu, Yong-Lu Li, Jianhua Sun, Cewu Lu
Human reasoning can be understood as a cooperation between the intuitive, associative "System-1'' and the deliberative, logical "System-2''. For existing System-1-like methods in visual activity understanding, it is crucial to integrate System-2 processing to improve explainability, generalization, and data efficiency. One possible path of activity reasoning is building a symbolic system composed of symbols and rules, where one rule connects multiple symbols, implying human knowledge and reasoning abilities.Previous methods have made progress, but are defective with limited symbols from handcraft and limited rules from visual-based annotations, failing to cover the complex patterns of activities and lacking compositional generalization. To overcome the defects, we propose a new symbolic system with two ideal important properties: broad-coverage symbols and rational rules. Collecting massive human knowledge via manual annotations is expensive to instantiate this symbolic system. Instead, we leverage the recent advancement of LLMs (Large Language Models) as an approximation of the two ideal properties, i.e., Symbols from Large Language Models (Symbol-LLM). Then, given an image, visual contents from the images are extracted andchecked as symbols and activity semantics are reasoned out based on rules via fuzzy logic calculation.Our method shows superiority in extensive activity understanding tasks. Code and data are available at https://mvig-rhos.com/symbol_llm.
****
Non-Convex Bilevel Optimization with Time-Varying Objective Functions
Sen Lin, Daouda Sow, Kaiyi Ji, Yingbin Liang, Ness Shroff
Bilevel optimization has become a powerful tool in a wide variety of machine learning problems. However, the current nonconvex bilevel optimization considers an offline dataset and static functions, which may not work well in emerging online applications with streaming data and time-varying functions. In this work, we study online bilevel optimization (OBO) where the functions can be time-varying and the agent continuously updates the decisions with online streaming data. To deal with the function variations and the unavailability of the true hypergradients in OBO, we propose a single-loop online bilevel optimizer with window averaging (SOBOW), which updates the outer-level decision based on a window average of the most recent hypergradient estimations stored in the memory. Compared to existing algorithms, SOBOW is computationally efficient and does not need to know previous functions. To handle the unique technical difficulties rooted in single-loop update and function variations for OBO, we develop a novel analytical technique that disentangles the complex couplings between decision variables, and carefully controls the hypergradient estimation error. We show that SOBOW can achieve a sublinear bilevel local regret under mild conditions. Extensive experiments across multiple domains corroborate the effectiveness of SOBOW.
****
Online Pricing for Multi-User Multi-Item Markets
Yigit Efe Erginbas, Thomas Courtade, Kannan Ramchandran, Soham Phade
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Online (Multinomial) Logistic Bandit: Improved Regret and Constant Computation Cost
Yu-Jie Zhang, Masashi Sugiyama
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Transfer learning for atomistic simulations using GNNs and kernel mean embeddings
John Falk, Luigi Bonati, Pietro Novelli, Michele Parrinello, Massimiliano Pontil
Interatomic potentials learned using machine learning methods have been successfully applied to atomistic simulations. However, accurate models require large training datasets, while generating reference calculations is computationally demanding. To bypass this difficulty, we propose a transfer learning algorithm that leverages the ability of graph neural networks (GNNs) to represent chemical environments together with kernel mean embeddings. We extract a feature map from GNNs pre-trained on the OC20 dataset and use it to learn the potential energy surface from system-specific datasets of catalytic processes. Our method is further enhanced by incorporating into the kernel the chemical species information, resulting in improved performance and interpretability. We test our approach on a series of realistic datasets of increasing complexity, showing excellent generalization and transferability performance, and improving on methods that rely on GNNs or ridge regression alone, as well as similar fine-tuning approaches.
****
StressID: a Multimodal Dataset for Stress Identification
Hava Chaptoukaev, Valeriya Strizhkova, Michele Panariello, Bianca Dalpaos, Aglind Reka, Valeria Manera, Susanne Thümmler, Esma ISMAILOVA, Nicholas W., francois bremond, Massimiliano Todisco, Maria A Zuluaga, Laura M. Ferrari
StressID is a new dataset specifically designed for stress identification fromunimodal and multimodal data. It contains videos of facial expressions, audiorecordings, and physiological signals. The video and audio recordings are acquiredusing an RGB camera with an integrated microphone. The physiological datais composed of electrocardiography (ECG), electrodermal activity (EDA), andrespiration signals that are recorded and monitored using a wearable device. Thisexperimental setup ensures a synchronized and high-quality multimodal data col-lection. Different stress-inducing stimuli, such as emotional video clips, cognitivetasks including mathematical or comprehension exercises, and public speakingscenarios, are designed to trigger a diverse range of emotional responses. Thefinal dataset consists of recordings from 65 participants who performed 11 tasks,as well as their ratings of perceived relaxation, stress, arousal, and valence levels.StressID is one of the largest datasets for stress identification that features threedifferent sources of data and varied classes of stimuli, representing more than39 hours of annotated data in total. StressID offers baseline models for stressclassification including a cleaning, feature extraction, and classification phase foreach modality. Additionally, we provide multimodal predictive models combiningvideo, audio, and physiological inputs. The data and the code for the baselines areavailable at https://project.inria.fr/stressid/.
****
Statistical Knowledge Assessment for Large Language Models
Qingxiu Dong, Jingjing Xu, Lingpeng Kong, Zhifang Sui, Lei Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Color Equivariant Convolutional Networks

Attila Lengyel, Ombretta Strafforello, Robert-Jan Bruintjes, Alexander Gielisse, Jan van Gemert
Color is a crucial visual cue readily exploited by Convolutional Neural Networks (CNNs) for object recognition. However, CNNs struggle if there is data imbalance between color variations introduced by accidental recording conditions. Color invariance addresses this issue but does so at the cost of removing all color information, which sacrifices discriminative power. In this paper, we propose Color Equivariant Convolutions (CEConvs), a novel deep learning building block that enables shape feature sharing across the color spectrum while retaining important color information. We extend the notion of equivariance from geometric to photometric transformations by incorporating parameter sharing over hue-shifts in a neural network. We demonstrate the benefits of CEConvs in terms of downstream performance to various tasks and improved robustness to color changes, including train-test distribution shifts. Our approach can be seamlessly integrated into existing architectures, such as ResNets, and offers a promising solution for addressing color-based domain shifts in CNNs.
****
Realistic Synthetic Financial Transactions for Anti-Money Laundering Models
Erik Altman, Jovan Blanuša, Luc von Niederhäusern, Beni Egressy, Andreea Anghel, Kubilay Atasu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DiffVL: Scaling Up Soft Body Manipulation using Vision-Language Driven Differentiable Physics
Zhiao Huang, Feng Chen, Yewen Pu, Chunru Lin, Hao Su, Chuang Gan
Combining gradient-based trajectory optimization with differentiable physics simulation is an efficient technique for solving soft-body manipulation problems.Using a well-crafted optimization objective, the solver can quickly converge onto a valid trajectory.However, writing the appropriate objective functions requires expert knowledge, making it difficult to collect a large set of naturalistic problems from non-expert users.We introduce DiffVL, a method that enables non-expert users to communicate soft-body manipulation tasks -- a combination of vision and natural language, given in multiple stages -- that can be readily leveraged by a differential physics solver. We have developed GUI tools that enable non-expert users to specify 100 tasks inspired by real-life soft-body manipulations from online videos, which we'll make public.We leverage large language models to translate task descriptions into machine-interpretable optimization objectives. The optimization objectives can help differentiable physics solvers to solve these long-horizon multistage tasks that are challenging for previous baselines.
****
Label-efficient Segmentation via Affinity Propagation
Wentong Li, Yuqian Yuan, Song Wang, Wenyu Liu, Dongqi Tang, Jian liu, Jianke Zhu, Lei Zhang
Weakly-supervised segmentation with label-efficient sparse annotations has attracted increasing research attention to reduce the cost of laborious pixel-wise labeling process, while the pairwise affinity modeling techniques play an essential role in this task. Most of the existing approaches focus on using the local appearance kernel to model the neighboring pairwise potentials. However, such a local operation fails to capture the long-range dependencies and ignores the topology of objects. In this work, we formulate the affinity modeling as an affinity propagation process, and propose a local and a global pairwise affinity terms to generate accurate soft pseudo labels. An efficient algorithm is also developed to reduce significantly the computational cost. The proposed approach can be conveniently plugged into existing segmentation networks. Experiments on three typical label-efficient segmentation tasks, i.e. box-supervised instance segmentation, point/scribble-supervised semantic segmentation and CLIP-guided semantic segmentation, demonstrate the superior performance of the proposed approach.

****
Segment Anything in High Quality
Lei Ke, Mingqiao Ye, Martin Danelljan, Yifan liu, Yu-Wing Tai, Chi-Keung Tang, Fisher Yu
The recent Segment Anything Model (SAM) represents a big leap in scaling up segmentation models, allowing for powerful zero-shot capabilities and flexible prompting. Despite being trained with 1.1 billion masks, SAM's mask prediction quality falls short in many cases, particularly when dealing with objects that have intricate structures. We propose HQ-SAM, equipping SAM with the ability to accurately segment any object, while maintaining SAM's original promptable design, efficiency, and zero-shot generalizability. Our careful design reuses and preserves the pre-trained model weights of SAM, while only introducing minimal additional parameters and computation. We design a learnable High-Quality Output Token, which is injected into SAM's mask decoder and is responsible for predicting the high-quality mask. Instead of only applying it on mask-decoder features, we first fuse them with early and final ViT features for improved mask details. To train our introduced learnable parameters, we compose a dataset of 44K fine-grained masks from several sources. HQ-SAM is only trained on the introduced detaset of 44k masks, which takes only 4 hours on 8 GPUs. We show the efficacy of HQ-SAM in a suite of 10 diverse segmentation datasets across different downstream tasks, where 8 out of them are evaluated in a zero-shot transfer protocol. Our code and pretrained models are at https://github.com/SysCV/SAM-HQ.
****
Variational Inference with Gaussian Score Matching
Chirag Modi, Robert Gower, Charles Margossian, Yuling Yao, David Blei, Lawrence Saul
Variational inference (VI) is a method to approximate the computationally intractable posterior distributions that arise in Bayesian statistics. Typically, VI fits a simple parametric distribution to be close to the target posterior, optimizing an appropriate objective such as the evidence lower bound (ELBO). In this work, we present a new approach to VI. Our method is based on the principle of score matching---namely, that if two distributions are equal then their score functions (i.e., gradients of the log density) are equal at every point on their support. With this principle, we develop score-matching VI, an iterative algorithm that seeks to match the scores between the variational approximation and the exact posterior. At each iteration, score-matching VI solves an inner optimization, one that minimally adjusts the current variational estimate to match the scores at a newly sampled value of the latent variables. We show that when the variational family is a Gaussian, this inner optimization enjoys a closed-form solution, which we call Gaussian score matching VI (GSM-VI). GSM-VI is a ``black box'' variational algorithm in that it only requires a differentiable joint distribution, and as such it can be applied to a wide class of models. We compare GSM-VI to black box variational inference (BBVI), which has similar requirements but instead optimizes the ELBO. We first study how GSM-VI behaves as a function of the problem dimensionality, the condition number of the target covariance matrix (when the target is Gaussian), and the degree of mismatch between the approximating and exact posterior distribution. We then study GSM-VI on a collection of real-world Bayesian inference problems from the posteriorDB database of datasets and models. We find that GSM-VI is faster than BBVI and equally or more accurate. Specifically, over a wide range of target posteriors, GSM-VI requires 10-100x fewer gradient evaluations than BBVI to obtain a comparable quality of approximation.
****
Feature Adaptation for Sparse Linear Regression
Jonathan Kelner, Frederic Koehler, Raghu Meka, Dhruv Rohatgi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

SimMTM: A Simple Pre-Training Framework for Masked Time-Series Modeling
Jiaxiang Dong, Haixu Wu, Haoran Zhang, Li Zhang, Jianmin Wang, Mingsheng Long
Time series analysis is widely used in extensive areas. Recently, to reduce labe
ling expenses and benefit various tasks, self-supervised pre-training has attrac
ted immense interest. One mainstream paradigm is masked modeling, which successf
ully pre-trains deep models by learning to reconstruct the masked content based
on the unmasked part. However, since the semantic information of time series is
mainly contained in temporal variations, the standard way of randomly masking a
portion of time points will seriously ruin vital temporal variations of time ser
ies, making the reconstruction task too difficult to guide representation learni
ng. We thus present SimMTM, a Simple pre-training framework for Masked Time-seri
es Modeling. By relating masked modeling to manifold learning, SimMTM proposes t
o recover masked time points by the weighted aggregation of multiple neighbors o
utside the manifold, which eases the reconstruction task by assembling ruined bu
t complementary temporal variations from multiple masked series. SimMTM further
learns to uncover the local structure of the manifold, which is helpful for mask
ed modeling. Experimentally, SimMTM achieves state-of-the-art fine-tuning perfor
mance compared to the most advanced time series pre-training methods in two cano
nical time series analysis tasks: forecasting and classification, covering both
in- and cross-domain settings.
****
CommonScenes: Generating Commonsense 3D Indoor Scenes with Scene Graph Diffusion
Guangyao Zhai, Evin P■nar Örnek, Shun-Cheng Wu, Yan Di, Federico Tombari, Nassir
Navab, Benjamin Busam
Controllable scene synthesis aims to create interactive environments for numerou
s industrial use cases. Scene graphs provide a highly suitable interface to faci
litate these applications by abstracting the scene context in a compact manner.
Existing methods, reliant on retrieval from extensive databases or pre-trained s
hape embeddings, often overlook scene-object and object-object relationships, le
ading to inconsistent results due to their limited generation capacity. To addre
ss this issue, we present CommonScenes, a fully generative model that converts s
cene graphs into corresponding controllable 3D scenes, which are semantically re
alistic and conform to commonsense. Our pipeline consists of two branches, one p
redicting the overall scene layout via a variational auto-encoder and the other
generating compatible shapes via latent diffusion, capturing global scene-object
and local inter-object relationships in the scene graph while preserving shape
diversity. The generated scenes can be manipulated by editing the input scene gr
aph and sampling the noise in the diffusion model. Due to the lack of a scene gr
aph dataset offering high-quality object-level meshes with relations, we also co
nstruct SG-FRONT, enriching the off-the-shelf indoor dataset 3D-FRONT with addit
ional scene graph labels. Extensive experiments are conducted on SG-FRONT, where
CommonScenes shows clear advantages over other methods regarding generation con
sistency, quality, and diversity. Codes and the dataset are available on the web
site.
****
AlpacaFarm: A Simulation Framework for Methods that Learn from Human Feedback
Yann Dubois, Chen Xuechen Li, Rohan Taori, Tianyi Zhang, Ishaan Gulrajani, Jimmy
Ba, Carlos Guestrin, Percy S. Liang, Tatsunori B. Hashimoto
Large language models (LLMs) such as ChatGPT have seen widespread adoption due t
o their ability to follow user instructions well.Developing these LLMs involves
a complex yet poorly understood workflow requiring training with human feedback.
Replicating and understanding this instruction-following process faces three ma
jor challenges: the high cost of data collection, the lack of trustworthy evalua
tion, and the absence of reference method implementations. We address these bott
lenecks with AlpacaFarm, a simulator that enables research and development for l
earning from feedback at a low cost. First, we design LLM based simulator for hu
man feedback that is 45x cheaper than crowdworkers and displays high agreement w
ith humans. Second, we identify an evaluation dataset representative of real-wor
ld instructions and propose an automatic evaluation procedure. Third, we contrib
ute reference implementations for several methods (PPO, best-of-n, expert iterat

ion, among others) that learn from pairwise feedback. Finally, as an end-to-end validation of AlpacaFarm, we train and evaluate eleven models on 10k pairs of human feedback and show that rankings of models trained in AlpacaFarm match rankings of models trained on human data. As a demonstration of the research possible in AlpacaFarm, we find that methods that use a reward model can substantially improve over supervised fine-tuning and that our reference PPO implementation leads to a +10% win-rate improvement against Davinci003.
****
Beta Diffusion
Mingyuan Zhou, Tianqi Chen, Zhendong Wang, Huangjie Zheng
We introduce beta diffusion, a novel generative modeling method that integrates demasking and denoising to generate data within bounded ranges. Using scaled and shifted beta distributions, beta diffusion utilizes multiplicative transitions over time to create both forward and reverse diffusion processes, maintaining beta distributions in both the forward marginals and the reverse conditionals, given the data at any point in time. Unlike traditional diffusion-based generative models relying on additive Gaussian noise and reweighted evidence lower bounds (ELBOs), beta diffusion is multiplicative and optimized with KL-divergence upper bounds (KLUBs) derived from the convexity of the KL divergence. We demonstrate that the proposed KLUBs are more effective for optimizing beta diffusion compared to negative ELBOs, which can also be derived as the KLUBs of the same KL divergence with its two arguments swapped. The loss function of beta diffusion, expressed in terms of Bregman divergence, further supports the efficacy of KLUBs for optimization. Experimental results on both synthetic data and natural images demonstrate the unique capabilities of beta diffusion in generative modeling of range-bounded data and validate the effectiveness of KLUBs in optimizing diffusion models, thereby making them valuable additions to the family of diffusion-based generative models and the optimization techniques used to train them.
****
Minimax Optimal Rate for Parameter Estimation in Multivariate Deviated Models
Dat Do, Huy Nguyen, Khai Nguyen, Nhat Ho
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Partial Matrix Completion
Elad Hazan, Adam Tauman Kalai, Varun Kanade, Clara Mohri, Y. Jennifer Sun
The matrix completion problem involves reconstructing a low-rank matrix by using a given set of revealed (and potentially noisy) entries. Although existing methods address the completion of the entire matrix, the accuracy of the completed entries can vary significantly across the matrix, due to differences in the sampling distribution. For instance, users may rate movies primarily from their country or favorite genres, leading to inaccurate predictions for the majority of completed entries.We propose a novel formulation of the problem as Partial Matrix Completion, where the objective is to complete a substantial subset of the entries with high confidence. Our algorithm efficiently handles the unknown and arbitrarily complex nature of the sampling distribution, ensuring high accuracy for all completed entries and sufficient coverage across the matrix. Additionally, we introduce an online version of the problem and present a low-regret efficient algorithm based on iterative gradient updates. Finally, we conduct a preliminary empirical evaluation of our methods.
****
BLIP-Diffusion: Pre-trained Subject Representation for Controllable Text-to-Image Generation and Editing
DONGXU LI, Junnan Li, Steven Hoi
Subject-driven text-to-image generation models create novel renditions of an input subject based on text prompts. Existing models suffer from lengthy fine-tuning and difficulties preserving the subject fidelity. To overcome these limitations, we introduce BLIP-Diffusion, a new subject-driven image generation model that

supports multimodal control which consumes inputs of subject images and text prompts. Unlike other subject-driven generation models, BLIP-Diffusion introduces a new multimodal encoder which is pre-trained to provide subject representation. We first pre-train the multimodal encoder following BLIP-2 to produce visual representation aligned with the text.Then we design a subject representation learning task which enables a diffusion model to leverage such visual representation and generates new subject renditions. Compared with previous methods such as DreamBooth, our model enables zero-shot subject-driven generation, and efficient fine-tuning for customized subject with up to 20x speedup. We also demonstrate that BLIP-Diffusion can be flexibly combined with existing techniques such as ControlNet and prompt-to-prompt to enable novel subject-driven generation and editing applications. Implementations are available at: https://github.com/salesforce/LAVIS/tree/main/projects/blip-diffusion.
****
Implicit Bias of Gradient Descent for Two-layer ReLU and Leaky ReLU Networks on Nearly-orthogonal Data
Yiwen Kou, Zixiang Chen, Quanquan Gu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SpecTr: Fast Speculative Decoding via Optimal Transport
Ziteng Sun, Ananda Theertha Suresh, Jae Hun Ro, Ahmad Beirami, Himanshu Jain, Felix Yu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
LithoBench: Benchmarking AI Computational Lithography for Semiconductor Manufacturing
Su Zheng, Haoyu Yang, Binwu Zhu, Bei Yu, Martin Wong
Computational lithography provides algorithmic and mathematical support for resolution enhancement in optical lithography, which is the critical step in semiconductor manufacturing. The time-consuming lithography simulation and mask optimization processes limit the practical application of inverse lithography technology (ILT), a promising solution to the challenges of advanced-node lithography. Although various machine learning methods for ILT have shown promise for reducing the computational burden, this field is in lack of a dataset that can train the models thoroughly and evaluate the performance comprehensively. To boost the development of AI-driven computational lithography, we present the LithoBench dataset, a collection of circuit layout tiles for deep-learning-based lithography simulation and mask optimization. LithoBench consists of more than 120k tiles that are cropped from real circuit designs or synthesized according to the layout topologies of famous ILT testcases. The ground truths are generated by a famous lithography model in academia and an advanced ILT method. Based on the data, we provide a framework to design and evaluate deep neural networks (DNNs) with the data. The framework is used to benchmark state-of-the-art models on lithography simulation and mask optimization. We hope LithoBench can promote the research and development of computational lithography. LithoBench is available at https://anonymous.4open.science/r/lithobench-APPL.
****
Proportional Response: Contextual Bandits for Simple and Cumulative Regret Minimization
Sanath Kumar Krishnamurthy, Ruohan Zhan, Susan Athey, Emma Brunskill
In many applications, e.g. in healthcare and e-commerce, the goal of a contextual bandit may be to learn an optimal treatment assignment policy at the end of the experiment. That is, to minimize simple regret. However, this objective remains understudied. We propose a new family of computationally efficient bandit algo

rithms for the stochastic contextual bandit setting, where a tuning parameter de termines the weight placed on cumulative regret minimization (where we establish near-optimal minimax guarantees) versus simple regret minimization (where we es tablish state-of-the-art guarantees). Our algorithms work with any function clas s, are robust to model misspecification, and can be used in continuous arm setti ngs. This flexibility comes from constructing and relying on “conformal arm sets " (CASs). CASs provide a set of arms for every context, encompassing the context -specific optimal arm with a certain probability across the context distribution . Our positive results on simple and cumulative regret guarantees are contrasted with a negative result, which shows that no algorithm can achieve instance-depe ndent simple regret guarantees while simultaneously achieving minimax optimal cu mulative regret guarantees.
****
Higher-Order Uncoupled Dynamics Do Not Lead to Nash Equilibrium - Except When Th ey Do
Sarah Toonsi, Jeff Shamma
The framework of multi-agent learning explores the dynamics of how an agent's st rategies evolve in response to the evolving strategies of other agents. Of parti cular interest is whether or not agent strategies converge to well known solutio n concepts such as Nash Equilibrium (NE). In "higher order'' learning, agent dyn amics include auxiliary states that can capture phenomena such as path dependenc ies. We introduce higher-order gradient play dynamics that resemble projected gr adient ascent with auxiliary states. The dynamics are "payoff based'' and "uncou pled'' in that each agent's dynamics depend on its own evolving payoff and has n o explicit dependence on the utilities of other agents. We first show that for a ny specific game with an isolated completely mixed-strategy NE, there exist high er-order gradient play dynamics that lead (locally) to that NE, both for the spe cific game and nearby games with perturbed utility functions. Conversely, we sho w that for any higher-order gradient play dynamics, there exists a game with a u nique isolated completely mixed-strategy NE for which the dynamics do not lead t o NE. Finally, we show that convergence to the mixed-strategy equilibrium in coo rdination games, comes at the expense of the dynamics being inherently internall y unstable.
****
Subject-driven Text-to-Image Generation via Apprenticeship Learning
Wenhu Chen, Hexiang Hu, Yandong Li, Nataniel Ruiz, Xuhui Jia, Ming-Wei Chang, Wi lliam W. Cohen
Recent text-to-image generation models like DreamBooth have made remarkable prog ress in generating highly customized images of a target subject, by fine-tuning an ``expert model'' for a given subject from a few examples.However, this proces s is expensive, since a new expert model must be learned for each subject. In th is paper, we present SuTI, a Subject-driven Text-to-Image generator that replace s subject-specific fine tuning with {in-context} learning.Given a few demonstrat ions of a new subject, SuTI can instantly generate novel renditions of the subje ct in different scenes, without any subject-specific optimization.SuTI is powere d by {apprenticeship learning}, where a single apprentice model is learned from data generated by a massive number of subject-specific expert models. Specifical ly, we mine millions of image clusters from the Internet, each centered around a specific visual subject. We adopt these clusters to train a massive number of e xpert models, each specializing in a different subject. The apprentice model SuT I then learns to imitate the behavior of these fine-tuned experts. SuTI can gene rate high-quality and customized subject-specific images 20x faster than optimiz ation-based SoTA methods. On the challenging DreamBench and DreamBench-v2, our h uman evaluation shows that SuTI significantly outperforms existing models like I nstructPix2Pix, Textual Inversion, Imagic, Prompt2Prompt, Re-Imagen and DreamBoo th.
****
GSLB: The Graph Structure Learning Benchmark
Zhixun Li, Xin Sun, Yifan Luo, Yanqiao Zhu, Dingshuo Chen, Yingtao Luo, Xiangxin Zhou, Qiang Liu, Shu Wu, Liang Wang, Jeffrey Yu

Graph Structure Learning (GSL) has recently garnered considerable attention due to its ability to optimize both the parameters of Graph Neural Networks (GNNs) and the computation graph structure simultaneously. Despite the proliferation of GSL methods developed in recent years, there is no standard experimental setting or fair comparison for performance evaluation, which creates a great obstacle to understanding the progress in this field. To fill this gap, we systematically analyze the performance of GSL in different scenarios and develop a comprehensive Graph Structure Learning Benchmark (GSLB) curated from 20 diverse graph datasets and 16 distinct GSL algorithms. Specifically, GSLB systematically investigates the characteristics of GSL in terms of three dimensions: effectiveness, robustness, and complexity. We comprehensively evaluate state-of-the-art GSL algorithms in node- and graph-level tasks, and analyze their performance in robust learning and model complexity. Further, to facilitate reproducible research, we have developed an easy-to-use library for training, evaluating, and visualizing different GSL methods. Empirical results of our extensive experiments demonstrate the ability of GSL and reveal its potential benefits on various downstream tasks, offering insights and opportunities for future research. The code of GSLB is available at: https://github.com/GSL-Benchmark/GSLB.
****

Consensus and Subjectivity of Skin Tone Annotation for ML Fairness

Candice Schumann, Femi Olanubi, Auriel Wright, Ellis Monk, Courtney Heldreth, Susanna Ricco

Understanding different human attributes and how they affect model behavior may become a standard need for all model creation and usage, from traditional computer vision tasks to the newest multimodal generative AI systems. In computer vision specifically, we have relied on datasets augmented with perceived attribute signals (eg, gender presentation, skin tone, and age) and benchmarks enabled by these datasets. Typically labels for these tasks come from human annotators. However, annotating attribute signals, especially skin tone, is a difficult and subjective task. Perceived skin tone is affected by technical factors, like lighting conditions, and social factors that shape an annotator's lived experience.This paper examines the subjectivity of skin tone annotation through a series of annotation experiments using the Monk Skin Tone (MST) scale~\cite{Monk2022Monk}, a small pool of professional photographers, and a much larger pool of trained crowdsourced annotators. Along with this study we release the Monk Skin Tone Examples (MST-E) dataset, containing 1515 images and 31 videos spread across the full MST scale. MST-E is designed to help train human annotators to annotate MST effectively.Our study shows that annotators can reliably annotate skin tone in a way that aligns with an expert in the MST scale, even under challenging environmental conditions. We also find evidence that annotators from different geographic regions rely on different mental models of MST categories resulting in annotations that systematically vary across regions. Given this, we advise practitioners to use a diverse set of annotators and a higher replication count for each image when annotating skin tone for fairness research.
****

Large Language Models can Implement Policy Iteration

Ethan Brooks, Logan Walls, Richard L Lewis, Satinder Singh

In this work, we demonstrate a method for implementing policy iteration using a large language model. While the application of foundation models to RL has received considerable attention, most approaches rely on either (1) the curation of expert demonstrations (either through manual design or task-specific pretraining) or (2) adaptation to the task of interest using gradient methods (either fine-tuning or training of adapter layers). Both of these techniques have drawbacks. Collecting demonstrations is labor-intensive, and algorithms that rely on them do not outperform the experts from which the demonstrations were derived. All gradient techniques are inherently slow, sacrificing the “few-shot” quality that makes in-context learning attractive to begin with. Our method demonstrates that a large language model can be used to implement policy iteration using the machinery of in-context learning, enabling it to learn to perform RL tasks without expert demonstrations or gradients. Our approach iteratively updates the contents of

the prompt from which it derives its policy through trial-and-error interaction with an RL environment. In order to eliminate the role of in-weights learning (on which approaches like Decision Transformer rely heavily), we demonstrate our method using Codex (M. Chen et al. 2021b), a language model with no prior knowledge of the domains on which we evaluate it.
****
ForkMerge: Mitigating Negative Transfer in Auxiliary-Task Learning
Junguang Jiang, Baixu Chen, Junwei Pan, Ximei Wang, Dapeng Liu, Jie Jiang, Mingsheng Long
Auxiliary-Task Learning (ATL) aims to improve the performance of the target task by leveraging the knowledge obtained from related tasks. Occasionally, learning multiple tasks simultaneously results in lower accuracy than learning only the target task, which is known as negative transfer. This problem is often attributed to the gradient conflicts among tasks, and is frequently tackled by coordinating the task gradients in previous works. However, these optimization-based methods largely overlook the auxiliary-target generalization capability. To better understand the root cause of negative transfer, we experimentally investigate it from both optimization and generalization perspectives. Based on our findings, we introduce ForkMerge, a novel approach that periodically forks the model into multiple branches, automatically searches the varying task weights by minimizing target validation errors, and dynamically merges all branches to filter out detrimental task-parameter updates. On a series of auxiliary-task learning benchmarks, ForkMerge outperforms existing methods and effectively mitigates negative transfer.
****
Revisiting Adversarial Robustness Distillation from the Perspective of Robust Fairness
Xinli Yue, Mou Ningping, Qian Wang, Lingchen Zhao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Real3D-AD: A Dataset of Point Cloud Anomaly Detection
Jiaqi Liu, Guoyang Xie, Ruitao Chen, Xinpeng Li, Jinbao Wang, Yong Liu, Chengjie Wang, Feng Zheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Differentiable Sampling of Categorical Distributions Using the CatLog-Derivative Trick
Lennert De Smet, Emanuele Sansone, Pedro Zuidberg Dos Martires
Categorical random variables can faithfully represent the discrete and uncertain aspects of data as part of a discrete latent variable model. Learning in such models necessitates taking gradients with respect to the parameters of the categorical probability distributions, which is often intractable due to their combinatorial nature. A popular technique to estimate these otherwise intractable gradients is the Log-Derivative trick. This trick forms the basis of the well-known REINFORCE gradient estimator and its many extensions. While the Log-Derivative trick allows us to differentiate through samples drawn from categorical distributions, it does not take into account the discrete nature of the distribution itself. Our first contribution addresses this shortcoming by introducing the CatLog-Derivative trick -- a variation of the Log-Derivative trick tailored towards categorical distributions. Secondly, we use the CatLog-Derivative trick to introduce IndeCateR, a novel and unbiased gradient estimator for the important case of products of independent categorical distributions with provably lower variance than REINFORCE. Thirdly, we empirically show that IndeCateR can be efficiently implemented and that its gradient estimates have significantly lower bias and varian

ce for the same number of samples compared to the state of the art.
****
Variational Imbalanced Regression: Fair Uncertainty Quantification via Probabilistic Smoothing
Ziyan Wang, Hao Wang
Existing regression models tend to fall short in both accuracy and uncertainty estimation when the label distribution is imbalanced. In this paper, we propose a probabilistic deep learning model, dubbed variational imbalanced regression (VIR), which not only performs well in imbalanced regression but naturally produces reasonable uncertainty estimation as a byproduct. Different from typical variational autoencoders assuming I.I.D. representations (a data point's representation is not directly affected by other data points), our VIR borrows data with similar regression labels to compute the latent representation's variational distribution; furthermore, different from deterministic regression models producing point estimates, VIR predicts the entire normal-inverse-gamma distributions and modulates the associated conjugate distributions to impose probabilistic reweighting on the imbalanced data, thereby providing better uncertainty estimation. Experiments in several real-world datasets show that our VIR can outperform state-of-the-art imbalanced regression models in terms of both accuracy and uncertainty estimation. Code will soon be available at https://github.com/Wang-ML-Lab/variational-imbalanced-regression.
****
Towards A Richer 2D Understanding of Hands at Scale
Tianyi Cheng, Dandan Shan, Ayda Hassen, Richard Higgins, David Fouhey
As humans, we learn a lot about how to interact with the world by observing others interacting with their hands. To help AI systems obtain a better understanding of hand interactions, we introduce a new model that produces a rich understanding of hand interaction. Our system produces a richer output than past systems at a larger scale. Our outputs include boxes and segments for hands, in-contact objects, and second objects touched by tools as well as contact and grasp type. Supporting this method are annotations of 257K images, 401K hands, 288K objects, and 19K second objects spanning four datasets. We show that our method provides rich information and performs and generalizes well.
****
Effective Human-AI Teams via Learned Natural Language Rules and Onboarding
Hussein Mozannar, Jimin Lee, Dennis Wei, Prasanna Sattigeri, Subhro Das, David Sontag
People are relying on AI agents to assist them with various tasks. The human must know when to rely on the agent, collaborate with the agent, or ignore its suggestions. In this work, we propose to learn rules grounded in data regions and described in natural language that illustrate how the human should collaborate with the AI. Our novel region discovery algorithm finds local regions in the data as neighborhoods in an embedding space that corrects the human prior. Each region is then described using an iterative and contrastive procedure where a large language model describes the region. We then teach these rules to the human via an onboarding stage. Through user studies on object detection and question-answering tasks, we show that our method can lead to more accurate human-AI teams. We also evaluate our region discovery and description algorithms separately.
****
On Certified Generalization in Structured Prediction
Bastian Boll, Christoph Schnörr
In structured prediction, target objects have rich internal structure which does not factorize into independent components and violates common i.i.d. assumptions. This challenge becomes apparent through the exponentially large output space in applications such as image segmentation or scene graph generation.We present a novel PAC-Bayesian risk bound for structured prediction wherein the rate of generalization scales not only with the number of structured examples but also with their size.The underlying assumption, conforming to ongoing research on generative models, is that data are generated by the Knothe-Rosenblatt rearrangement of a factorizing reference measure. This allows to explicitly distill the structu

re between random output variables into a Wasserstein dependency matrix. Our work makes a preliminary step towards leveraging powerful generative models to establish generalization bounds for discriminative downstream tasks in the challenging setting of structured prediction.
****

SutraNets: Sub-series Autoregressive Networks for Long-Sequence, Probabilistic Forecasting
Shane Bergsma, Tim Zeyl, Lei Guo
We propose SutraNets, a novel method for neural probabilistic forecasting of long-sequence time series. SutraNets use an autoregressive generative model to factorize the likelihood of long sequences into products of conditional probabilities. When generating long sequences, most autoregressive approaches suffer from harmful error accumulation, as well as challenges in modeling long-distance dependencies. SutraNets treat long, univariate prediction as multivariate prediction over lower-frequency sub-series. Autoregression proceeds across time and across sub-series in order to ensure coherent multivariate (and, hence, high-frequency univariate) outputs. Since sub-series can be generated using fewer steps, SutraNets effectively reduce error accumulation and signal path distances. We find SutraNets to significantly improve forecasting accuracy over competitive alternatives on six real-world datasets, including when we vary the number of sub-series and scale up the depth and width of the underlying sequence models.
****

Complex Query Answering on Eventuality Knowledge Graph with Implicit Logical Constraints
Jiaxin Bai, Xin Liu, Weiqi Wang, Chen Luo, Yangqiu Song
Querying knowledge graphs (KGs) using deep learning approaches can naturally leverage the reasoning and generalization ability to learn to infer better answers. Traditional neural complex query answering (CQA) approaches mostly work on entity-centric KGs. However, in the real world, we also need to make logical inferences about events, states, and activities (i.e., eventualities or situations) to push learning systems from System I to System II, as proposed by Yoshua Bengio. Querying logically from an EVentuality-centric KG (EVKG) can naturally provide references to such kind of intuitive and logical inference. Thus, in this paper, we propose a new framework to leverage neural methods to answer complex logical queries based on an EVKG, which can satisfy not only traditional first-order logic constraints but also implicit logical constraints over eventualities concerning their occurrences and orders. For instance, if we know that Food is bad happens before PersonX adds soy sauce, then PersonX adds soy sauce is unlikely to be the cause of Food is bad due to implicit temporal constraint. To facilitate consistent reasoning on EVKGs, we propose Complex Eventuality Query Answering (CEQA), a more rigorous definition of CQA that considers the implicit logical constraints governing the temporal order and occurrence of eventualities. In this manner, we propose to leverage theorem provers for constructing benchmark datasets to ensure the answers satisfy implicit logical constraints. We also propose a Memory-Enhanced Query Encoding (MEQE) approach to significantly improve the performance of state-of-the-art neural query encoders on the CEQA task.
****

Fast Bellman Updates for Wasserstein Distributionally Robust MDPs
Zhuodong Yu, Ling Dai, Shaohang Xu, Siyang Gao, Chin Pang Ho
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

LoRA: A Logical Reasoning Augmented Dataset for Visual Question Answering
Jingying Gao, Qi Wu, Alan Blair, Maurice Pagnucco
The capacity to reason logically is a hallmark of human cognition. Humans excel at integrating multimodal information for locigal reasoning, as exemplified by the Visual Question Answering (VQA) task, which is a challenging multimodal task. VQA tasks and large vision-and-language models aim to tackle reasoning problems

, but the accuracy, consistency and fabrication of the generated answers is hard
to evaluate in the absence of a VQA dataset that can offer formal, comprehensiv
e and systematic complex logical reasoning questions. To address this gap, we pr
esent LoRA, a novel Logical Reasoning Augmented VQA dataset that requires formal
and complex description logic reasoning based on a food-and-kitchen knowledge b
ase. Our main objective in creating LoRA is to enhance the complex and formal lo
gical reasoning capabilities of VQA models, which are not adequately measured by
existing VQA datasets. We devise strong and flexible programs to automatically
generate 200,000 diverse description logic reasoning questions based on the SROI
Q Description Logic, along with realistic kitchen scenes and ground truth answer
s. We fine-tune the latest transformer VQA models and evaluate the zero-shot per
formance of the state-of-the-art large vision-and-language models on LoRA. The r
esults reveal that LoRA presents a unique challenge in logical reasoning, settin
g a systematic and comprehensive evaluation standard.
****
Practical Contextual Bandits with Feedback Graphs
Mengxiao Zhang, Yuheng Zhang, Olga Vrousgou, Haipeng Luo, Paul Mineiro
While contextual bandit has a mature theory, effectively leveraging different fe
edback patterns to enhance the pace of learning remains unclear. Bandits with fe
edback graphs, which interpolates between the full information and bandit regime
s, provides a promising framework to mitigate the statistical complexity of lear
ning. In this paper, we propose and analyze an approach to contextual bandits wi
th feedback graphs based upon reduction to regression. The resulting algorithms
are computationally practical and achieve established minimax rates, thereby re
ducing the statistical complexity in real-world applications.
****
Policy Optimization in a Noisy Neighborhood: On Return Landscapes in Continuous
Control
Nate Rahn, Pierluca D'Oro, Harley Wiltzer, Pierre-Luc Bacon, Marc Bellemare
Deep reinforcement learning agents for continuous control are known to exhibit s
ignificant instability in their performance over time. In this work, we provide
a fresh perspective on these behaviors by studying the return landscape: the map
ping between a policy and a return. We find that popular algorithms traverse noi
sy neighborhoods of this landscape, in which a single update to the policy param
eters leads to a wide range of returns. By taking a distributional view of these
returns, we map the landscape, characterizing failure-prone regions of policy s
pace and revealing a hidden dimension of policy quality. We show that the landsc
ape exhibits surprising structure by finding simple paths in parameter space whi
ch improve the stability of a policy. To conclude, we develop a distribution-awa
re procedure which finds such paths, navigating away from noisy neighborhoods in
order to improve the robustness of a policy. Taken together, our results provid
e new insight into the optimization, evaluation, and design of agents.
****
Real-World Image Variation by Aligning Diffusion Inversion Chain
Yuechen Zhang, Jinbo Xing, Eric Lo, Jiaya Jia
Recent diffusion model advancements have enabled high-fidelity images to be gene
rated using text prompts. However, a domain gap exists between generated images
and real-world images, which poses a challenge in generating high-quality variat
ions of real-world images. Our investigation uncovers that this domain gap origi
nates from a latents' distribution gap in different diffusion processes. To addr
ess this issue, we propose a novel inference pipeline called Real-world Image Va
riation by ALignment (RIVAL) that utilizes diffusion models to generate image va
riations from a single image exemplar. Our pipeline enhances the generation qual
ity of image variations by aligning the image generation process to the source i
mage's inversion chain. Specifically, we demonstrate that step-wise latent distr
ibution alignment is essential for generating high-quality variations. To attain
this, we design a cross-image self-attention injection for feature interaction
and a step-wise distribution normalization to align the latent features. Incorpo
rating these alignment processes into a diffusion model allows RIVAL to generate
high-quality image variations without further parameter optimization. Our exper

imental results demonstrate that our proposed approach outperforms existing meth ods concerning semantic similarity and perceptual quality. This generalized infe rence pipeline can be easily applied to other diffusion-based generation tasks, such as image-conditioned text-to-image generation and stylization. Project page : https://rival-diff.github.io
****
Vocabulary-free Image Classification
Alessandro Conti, Enrico Fini, Massimiliano Mancini, Paolo Rota, Yiming Wang, El isa Ricci
Recent advances in large vision-language models have revolutionized the image cl assification paradigm. Despite showing impressive zero-shot capabilities, a pre-defined set of categories, a.k.a. the vocabulary, is assumed at test time for co mposing the textual prompts. However, such assumption can be impractical when th e semantic context is unknown and evolving. We thus formalize a novel task, term ed as Vocabulary-free Image Classification (VIC), where we aim to assign to an i nput image a class that resides in an unconstrained language-induced semantic sp ace, without the prerequisite of a known vocabulary. VIC is a challenging task a s the semantic space is extremely large, containing millions of concepts, with h ard-to-discriminate fine-grained categories. In this work, we first empirically verify that representing this semantic space by means of an external vision-lang uage database is the most effective way to obtain semantically relevant content for classifying the image. We then propose Category Search from External Databas es (CaSED), a method that exploits a pre-trained vision-language model and an ex ternal vision-language database to address VIC in a training-free manner. CaSED first extracts a set of candidate categories from captions retrieved from the da tabase based on their semantic similarity to the image, and then assigns to the image the best matching candidate category according to the same vision-language model. Experiments on benchmark datasets validate that CaSED outperforms other complex vision-language frameworks, while being efficient with much fewer parame ters, paving the way for future research in this direction.
****
A Novel Framework for Policy Mirror Descent with General Parameterization and Li near Convergence
Carlo Alfano, Rui Yuan, Patrick Rebeschini
Modern policy optimization methods in reinforcement learning, such as TRPO and P PO, owe their success to the use of parameterized policies. However, while theor etical guarantees have been established for this class of algorithms, especially in the tabular setting, the use of general parameterization schemes remains mos tly unjustified. In this work, we introduce a novel framework for policy optimiz ation based on mirror descent that naturally accommodates general parameterizati ons. The policy class induced by our scheme recovers known classes, e.g., softma x, and generates new ones depending on the choice of mirror map. Using our frame work, we obtain the first result that guarantees linear convergence for a policy -gradient-based method involving general parameterization. To demonstrate the ab ility of our framework to accommodate general parameterization schemes, we provi de its sample complexity when using shallow neural networks, show that it repres ents an improvement upon the previous best results, and empirically validate the effectiveness of our theoretical claims on classic control tasks.
****
Weakly-Supervised Concealed Object Segmentation with SAM-based Pseudo Labeling a nd Multi-scale Feature Grouping
Chunming He, Kai Li, Yachao Zhang, Guoxia Xu, Longxiang Tang, Yulun Zhang, Zhenh ua Guo, Xiu Li
Weakly-Supervised Concealed Object Segmentation (WSCOS) aims to segment objects well blended with surrounding environments using sparsely-annotated data for mo del training. It remains a challenging task since (1) it is hard to distinguish concealed objects from the background due to the intrinsic similarity and (2) th e sparsely-annotated training data only provide weak supervision for model learn ing. In this paper, we propose a new WSCOS method to address these two challenge s. To tackle the intrinsic similarity challenge, we design a multi-scale feature

grouping module that first groups features at different granularities and then aggregates these grouping results. By grouping similar features together, it encourages segmentation coherence, helping obtain complete segmentation results for both single and multiple-object images. For the weak supervision challenge, we utilize the recently-proposed vision foundation model, ``Segment Anything Model (SAM)'', and use the provided sparse annotations as prompts to generate segmentation masks, which are used to train the model. To alleviate the impact of low-quality segmentation masks, we further propose a series of strategies, including multi-augmentation result ensemble, entropy-based pixel-level weighting, and entropy-based image-level selection. These strategies help provide more reliable supervision to train the segmentation model. We verify the effectiveness of our method on various WSCOS tasks, and experiments demonstrate that our method achieves state-of-the-art performance on these tasks.
****
Ordering-based Conditions for Global Convergence of Policy Gradient Methods
Jincheng Mei, Bo Dai, Alekh Agarwal, Mohammad Ghavamzadeh, Csaba Szepesvari, Dale Schuurmans
We prove that, for finite-arm bandits with linear function approximation, the global convergence of policy gradient (PG) methods depends on inter-related properties between the policy update and the representation. textcolor{blue}{First}, we establish a few key observations that frame the study: \textbf{(i)} Global convergence can be achieved under linear function approximation without policy or reward realizability, both for the standard Softmax PG and natural policy gradient (NPG). \textbf{(ii)} Approximation error is not a key quantity for characterizing global convergence in either algorithm. \textbf{(iii)} The conditions on the representation that imply global convergence are different between these two algorithms. Overall, these observations call into question approximation error as an appropriate quantity for characterizing the global convergence of PG methods under linear function approximation. \textcolor{blue}{Second}, motivated by these observations, we establish new general results: \textbf{(i)} NPG with linear function approximation achieves global convergence \emph{if and only if} the projection of the reward onto the representable space preserves the optimal action's rank, a quantity that is not strongly related to approximation error. \textbf{(ii)} The global convergence of Softmax PG occurs if the representation satisfies a non-domination condition and can preserve the ranking of rewards, which goes well beyond policy or reward realizability. We provide experimental results to support these theoretical findings.
****
Finding Order in Chaos: A Novel Data Augmentation Method for Time Series in Contrastive Learning
Berken Utku Demirel, Christian Holz
The success of contrastive learning is well known to be dependent on data augmentation.Although the degree of data augmentations has been well controlled by utilizing pre-defined techniques in some domains like vision, time-series data augmentation is less explored and remains a challenging problem due to the complexity of the data generation mechanism, such as the intricate mechanism involved in the cardiovascular system.Moreover, there is no widely recognized and general time-series augmentation method that can be applied across different tasks.In this paper, we propose a novel data augmentation method for time-series tasks that aims to connect intra-class samples together, and thereby find order in the latent space.Our method builds upon the well-known data augmentation technique of mixup by incorporating a novel approach that accounts for the non-stationary nature of time-series data.Also, by controlling the degree of chaos created by data augmentation, our method leads to improved feature representations and performance on downstream tasks.We evaluate our proposed method on three time-series tasks, including heart rate estimation, human activity recognition, and cardiovascular disease detection. Extensive experiments against the state-of-the-art methods show that the proposed method outperforms prior works on optimal data generation and known data augmentation techniques in three tasks, reflecting the effectiveness of the presented method. The source code is available at double-blind policy

.
****
List and Certificate Complexities in Replicable Learning
Peter Dixon, A. Pavan, Jason Vander Woude, N. V. Vinodchandran
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Flat Seeking Bayesian Neural Networks
Van-Anh Nguyen, Tung-Long Vuong, Hoang Phan, Thanh-Toan Do, Dinh Phung, Trung Le
Bayesian Neural Networks (BNNs) provide a probabilistic interpretation for deep learning models by imposing a prior distribution over model parameters and inferring a posterior distribution based on observed data. The model sampled from the posterior distribution can be used for providing ensemble predictions and quantifying prediction uncertainty. It is well-known that deep learning models with lower sharpness have better generalization ability. However, existing posterior inferences are not aware of sharpness/flatness in terms of formulation, possibly leading to high sharpness for the models sampled from them. In this paper, we develop theories, the Bayesian setting, and the variational inference approach for the sharpness-aware posterior. Specifically, the models sampled from our sharpness-aware posterior, and the optimal approximate posterior estimating this sharpness-aware posterior, have better flatness, hence possibly possessing higher generalization ability. We conduct experiments by leveraging the sharpness-aware posterior with state-of-the-art Bayesian Neural Networks, showing that the flat-seeking counterparts outperform their baselines in all metrics of interest.
****
Enhancing User Intent Capture in Session-Based Recommendation with Attribute Patterns
Xin Liu, Zheng Li, Yifan Gao, Jingfeng Yang, Tianyu Cao, Zhengyang Wang, Bing Yin, Yangqiu Song
The goal of session-based recommendation in E-commerce is to predict the next item that an anonymous user will purchase based on the browsing and purchase history. However, constructing global or local transition graphs to supplement session data can lead to noisy correlations and user intent vanishing. In this work, we propose the Frequent Attribute Pattern Augmented Transformer (FAPAT) that characterizes user intents by building attribute transition graphs and matching attribute patterns. Specifically, the frequent and compact attribute patterns are served as memory to augment session representations, followed by a gate and a transformer block to fuse the whole session information. Through extensive experiments on two public benchmarks and 100 million industrial data in three domains, we demonstrate that FAPAT consistently outperforms state-of-the-art methods by an average of 4.5% across various evaluation metrics (Hits, NDCG, MRR). Besides evaluating the next-item prediction, we estimate the models' capabilities to capture user intents via predicting items' attributes and period-item recommendations.
****
Can Language Models Solve Graph Problems in Natural Language?
Heng Wang, Shangbin Feng, Tianxing He, Zhaoxuan Tan, Xiaochuang Han, Yulia Tsvetkov
Large language models (LLMs) are increasingly adopted for a variety of tasks with implicit graphical structures, such as planning in robotics, multi-hop question answering or knowledge probing, structured commonsense reasoning, and more. While LLMs have advanced the state-of-the-art on these tasks with structure implications, whether LLMs could explicitly process textual descriptions of graphs and structures, map them to grounded conceptual spaces, and perform structured operations remains underexplored. To this end, we propose NLGraph (Natural Language Graph), a comprehensive benchmark of graph-based problem solving designed in natural language. NLGraph contains 29,370 problems, covering eight graph reasoning tasks with varying complexity from simple tasks such as connectivity and shortest path up to complex problems such as maximum flow and simulating graph neural n

etworks. We evaluate LLMs (GPT-3/4) with various prompting approaches on the NLGraph benchmark and find that 1) language models do demonstrate preliminary graph reasoning abilities, 2) the benefit of advanced prompting and in-context learning diminishes on more complex graph problems, while 3) LLMs are also (un)surprisingly brittle in the face of spurious correlations in graph and problem settings. We then propose Build-a-Graph Prompting and Algorithmic Prompting, two instruction-based approaches to enhance LLMs in solving natural language graph problems. Build-a-Graph and Algorithmic prompting improve the performance of LLMs on NLGraph by 3.07% to 16.85% across multiple tasks and settings, while how to solve the most complicated graph reasoning tasks in our setup with language models remains an open research question.
****

Language-driven Scene Synthesis using Multi-conditional Diffusion Model
An Dinh Vuong, Minh Nhat VU, Toan Nguyen, Baoru Huang, Dzung Nguyen, Thieu Vo, Anh Nguyen
Scene synthesis is a challenging problem with several industrial applications. Recently, substantial efforts have been directed to synthesize the scene using human motions, room layouts, or spatial graphs as the input. However, few studies have addressed this problem from multiple modalities, especially combining text prompts. In this paper, we propose a language-driven scene synthesis task, which is a new task that integrates text prompts, human motion, and existing objects for scene synthesis. Unlike other single-condition synthesis tasks, our problem involves multiple conditions and requires a strategy for processing and encoding them into a unified space. To address the challenge, we present a multi-conditional diffusion model, which differs from the implicit unification approach of other diffusion literature by explicitly predicting the guiding points for the original data distribution. We demonstrate that our approach is theoretically supportive. The intensive experiment results illustrate that our method outperforms state-of-the-art benchmarks and enables natural scene editing applications. The source code and dataset can be accessed at https://lang-scene-synth.github.io/.
****

Implicit Contrastive Representation Learning with Guided Stop-gradient
Byeongchan Lee, Sehyun Lee
In self-supervised representation learning, Siamese networks are a natural architecture for learning transformation-invariance by bringing representations of positive pairs closer together. But it is prone to collapse into a degenerate solution. To address the issue, in contrastive learning, a contrastive loss is used to prevent collapse by moving representations of negative pairs away from each other. But it is known that algorithms with negative sampling are not robust to a reduction in the number of negative samples. So, on the other hand, there are algorithms that do not use negative pairs. Many positive-only algorithms adopt asymmetric network architecture consisting of source and target encoders as a key factor in coping with collapse. By exploiting the asymmetric architecture, we introduce a methodology to implicitly incorporate the idea of contrastive learning. As its implementation, we present a novel method guided stop-gradient. We apply our method to benchmark algorithms SimSiam and BYOL and show that our method stabilizes training and boosts performance. We also show that the algorithms with our method work well with small batch sizes and do not collapse even when there is no predictor. The code is available in the supplementary material.
****

QATCH: Benchmarking SQL-centric tasks with Table Representation Learning Models on Your Data
Simone Papicchio, Paolo Papotti, Luca Cagliero
Table Representation Learning (TRL) models are commonly pre-trained on large open-domain datasets comprising millions of tables and then used to address downstream tasks. Choosing the right TRL model to use on proprietary data can be challenging, as the best results depend on the content domain, schema, and data quality. Our purpose is to support end-users in testing TRL models on proprietary data in two established SQL-centric tasks, i.e., Question Answering (QA) and Semantic Parsing (SP). We present QATCH (Query-Aided TRL Checklist), a toolbox to highl

ight TRL models' strengths and weaknesses on relational tables unseen at training time. For an input table, QATCH automatically generates a testing checklist tailored to QA and SP. Checklist generation is driven by a SQL query engine that crafts tests of different complexity. This design facilitates inherent portability, allowing the checks to be used by alternative models. We also introduce a set of cross-task performance metrics evaluating the TRL model's performance over its output. Finally, we show how QATCH automatically generates tests for proprietary datasets to evaluate various state-of-the-art models including TAPAS, TAPEX, and CHATGPT.
****
Improved Best-of-Both-Worlds Guarantees for Multi-Armed Bandits: FTRL with General Regularizers and Multiple Optimal Arms
Tiancheng Jin, Junyan Liu, Haipeng Luo
We study the problem of designing adaptive multi-armed bandit algorithms that perform optimally in both the stochastic setting and the adversarial setting simultaneously (often known as a best-of-both-world guarantee). A line of recent works shows that when configured and analyzed properly, the Follow-the-Regularized-Leader (FTRL) algorithm, originally designed for the adversarial setting, can in fact optimally adapt to the stochastic setting as well. Such results, however, critically rely on an assumption that there exists one unique optimal arm. Recently, Ito [2021] took the first step to remove such an undesirable uniqueness assumption for one particular FTRL algorithm withthe 1/2-Tsallis entropy regularizer. In this work, we significantly improve and generalize this result, showing that uniqueness is unnecessary for FTRL with a broad family of regularizers and a new learning rate schedule. For some regularizers, our regret bounds also improve upon prior results even when uniqueness holds. We further provide an application of our results to the decoupled exploration and exploitation problem, demonstrating that our techniques are broadly applicable.
****
AiluRus: A Scalable ViT Framework for Dense Prediction
Jin Li, Yaoming Wang, XIAOPENG ZHANG, Bowen Shi, Dongsheng Jiang, Chenglin Li, Wenrui Dai, Hongkai Xiong, Qi Tian
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CORL: Research-oriented Deep Offline Reinforcement Learning Library
Denis Tarasov, Alexander Nikulin, Dmitry Akimov, Vladislav Kurenkov, Sergey Kolesnikov
CORL is an open-source library that provides thoroughly benchmarked single-file implementations of both deep offline and offline-to-online reinforcement learning algorithms. It emphasizes a simple developing experience with a straightforward codebase and a modern analysis tracking tool. In CORL, we isolate methods implementation into separate single files, making performance-relevant details easier to recognize. Additionally, an experiment tracking feature is available to help log metrics, hyperparameters, dependencies, and more to the cloud. Finally, we have ensured the reliability of the implementations by benchmarking commonly employed D4RL datasets providing a transparent source of results that can be reused for robust evaluation tools such as performance profiles, probability of improvement, or expected online performance.
****
LightSpeed: Light and Fast Neural Light Fields on Mobile Devices
Aarush Gupta, Junli Cao, Chaoyang Wang, Ju Hu, Sergey Tulyakov, Jian Ren, László Jeni
Real-time novel-view image synthesis on mobile devices is prohibitive due to the limited computational power and storage. Using volumetric rendering methods, such as NeRF and its derivatives, on mobile devices is not suitable due to the high computational cost of volumetric rendering. On the other hand, recent advances in neural light field representations have shown promising real-time view synth

esis results on mobile devices. Neural light field methods learn a direct mappin
g from a ray representation to the pixel color. The current choice of ray repres
entation is either stratified ray sampling or Plücker coordinates, overlooking t
he classic light slab (two-plane) representation, the preferred representation t
o interpolate between light field views. In this work, we find that using the li
ght slab representation is an efficient representation for learning a neural lig
ht field. More importantly, it is a lower-dimensional ray representation enablin
g us to learn the 4D ray space using feature grids which are significantly faste
r to train and render. Although mostly designed for frontal views, we show that
the light-slab representation can be further extended to non-frontal scenes usin
g a divide-and-conquer strategy. Our method provides better rendering quality th
an prior light field methods and a significantly better trade-off between render
ing quality and speed than prior light field methods.
****
CLadder: Assessing Causal Reasoning in Language Models
Zhijing Jin, Yuen Chen, Felix Leeb, Luigi Gresele, Ojasv Kamal, Zhiheng LYU, Kev
in Blin, Fernando Gonzalez Adauto, Max Kleiman-Weiner, Mrinmaya Sachan, Bernhard
 Schölkopf
The ability to perform causal reasoning is widely considered a core feature of i
ntelligence. In this work, we investigate whether large language models (LLMs) c
an coherently reason about causality. Much of the existing work in natural langu
age processing (NLP) focuses on evaluating commonsense causal reasoning in LLMs,
 thus failing to assess whether a model can perform causal inference in accordan
ce with a set of well-defined formal rules. To address this, we propose a new NL
P task, causal inference in natural language, inspired by the "causal inference
engine" postulated by Judea Pearl et al. We compose a large dataset, CLadder, wi
th 10K samples: based on a collection of causal graphs and queries (associationa
l, interventional, and counterfactual), we obtain symbolic questions and ground-
truth answers, through an oracle causal inference engine. These are then transla
ted into natural language. We evaluate multiple LLMs on our dataset, and we intr
oduce and evaluate a bespoke chain-of-thought prompting strategy, CausalCoT. We
show that our task is highly challenging for LLMs, and we conduct an in-depth an
alysis to gain deeper insight into the causal reasoning abilities of LLMs. Our d
ata is open-sourced at https://huggingface.co/datasets/causalNLP/cladder, and ou
r code can be found at https://github.com/causalNLP/cladder.
****
Riemannian Laplace approximations for Bayesian neural networks
Federico Bergamin, Pablo Moreno-Muñoz, Søren Hauberg, Georgios Arvanitidis
Bayesian neural networks often approximate the weight-posterior with a Gaussian
distribution. However, practical posteriors are often, even locally, highly non-
Gaussian, and empirical performance deteriorates. We propose a simple parametric
 approximate posterior that adapts to the shape of the true posterior through a
Riemannian metric that is determined by the log-posterior gradient. We develop a
 Riemannian Laplace approximation where samples naturally fall into weight-regio
ns with low negative log-posterior. We show that these samples can be drawn by s
olving a system of ordinary differential equations, which can be done efficientl
y by leveraging the structure of the Riemannian metric and automatic differentia
tion. Empirically, we demonstrate that our approach consistently improves over t
he conventional Laplace approximation across tasks. We further show that, unlike
 the conventional Laplace approximation, our method is not overly sensitive to t
he choice of prior, which alleviates a practical pitfall of current approaches.
****
SugarCrepe: Fixing Hackable Benchmarks for Vision-Language Compositionality
Cheng-Yu Hsieh, Jieyu Zhang, Zixian Ma, Aniruddha Kembhavi, Ranjay Krishna
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Contrastive Retrospection: honing in on critical steps for rapid learning and ge

neralization in RL
Chen Sun, Wannan Yang, Thomas Jiralerspong, Dane Malenfant, Benjamin Alsbury-Nealy, Yoshua Bengio, Blake Richards
In real life, success is often contingent upon multiple critical steps that are distant in time from each other and from the final reward. These critical steps are challenging to identify with traditional reinforcement learning (RL) methods that rely on the Bellman equation for credit assignment. Here, we present a new RL algorithm that uses offline contrastive learning to hone in on these critical steps. This algorithm, which we call Contrastive Retrospection (ConSpec), can be added to any existing RL algorithm. ConSpec learns a set of prototypes for the critical steps in a task by a novel contrastive loss and delivers an intrinsic reward when the current state matches one of the prototypes. The prototypes in ConSpec provide two key benefits for credit assignment: (i) They enable rapid identification of all the critical steps. (ii) They do so in a readily interpretable manner, enabling out-of-distribution generalization when sensory features are altered. Distinct from other contemporary RL approaches to credit assignment, ConSpec takes advantage of the fact that it is easier to retrospectively identify the small set of steps that success is contingent upon (and ignoring other states) than it is to prospectively predict reward at every taken step. ConSpec greatly improves learning in a diverse set of RL tasks. The code is available at the link: https://github.com/sunchipsster1/ConSpec
****
A Hierarchical Training Paradigm for Antibody Structure-sequence Co-design
Fang Wu, Stan Z. Li
Therapeutic antibodies are an essential and rapidly flourishing drug modality. The binding specificity between antibodies and antigens is decided by complementarity-determining regions (CDRs) at the tips of these Y-shaped proteins. In this paper, we propose a \textbf{h}ierarchical \textbf{t}raining \textbf{p}aradigm (HTP) for the antibody sequence-structure co-design. HTP consists of four levels of training stages, each corresponding to a specific protein modality within a particular protein domain. Through carefully crafted tasks in different stages, HTP seamlessly and effectively integrates geometric graph neural networks (GNNs) with large-scale protein language models to excavate evolutionary information from not only geometric structures but also vast antibody and non-antibody sequence databases, which determines ligand binding pose and strength. Empirical experiments show HTP sets the new state-of-the-art performance in the co-design problem as well as the fix-backbone design. Our research offers a hopeful path to unleash the potential of deep generative architectures and seeks to illuminate the way forward for the antibody sequence and structure co-design challenge.
****
Differentiable Clustering with Perturbed Spanning Forests
Lawrence Stewart, Francis Bach, Felipe Llinares-Lopez, Quentin Berthet
We introduce a differentiable clustering method based on stochastic perturbations of minimum-weight spanning forests. This allows us to include clustering in end-to-end trainable pipelines, with efficient gradients. We show that our method performs well even in difficult settings, such as data sets with high noise and challenging geometries. We also formulate an ad hoc loss to efficiently learn from partial clustering data using this operation. We demonstrate its performance on several data sets for supervised and semi-supervised tasks.
****
Logarithmic-Regret Quantum Learning Algorithms for Zero-Sum Games
Minbo Gao, Zhengfeng Ji, Tongyang Li, Qisheng Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Joint Bayesian Inference of Graphical Structure and Parameters with a Single Generative Flow Network
Tristan Deleu, Mizu Nishikawa-Toomey, Jithendaraa Subramanian, Nikolay Malkin, L

aurent Charlin, Yoshua Bengio
Generative Flow Networks (GFlowNets), a class of generative models over discrete and structured sample spaces, have been previously applied to the problem of inferring the marginal posterior distribution over the directed acyclic graph (DAG) of a Bayesian Network, given a dataset of observations. Based on recent advances extending this framework to non-discrete sample spaces, we propose in this paper to approximate the joint posterior over not only the structure of a Bayesian Network, but also the parameters of its conditional probability distributions. We use a single GFlowNet whose sampling policy follows a two-phase process: the DAG is first generated sequentially one edge at a time, and then the corresponding parameters are picked once the full structure is known. Since the parameters are included in the posterior distribution, this leaves more flexibility for the local probability models of the Bayesian Network, making our approach applicable even to non-linear models parametrized by neural networks. We show that our method, called JSP-GFN, offers an accurate approximation of the joint posterior, while comparing favorably against existing methods on both simulated and real data.
****
DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models
Boxin Wang, Weixin Chen, Hengzhi Pei, Chulin Xie, Mintong Kang, Chenhui Zhang, Chejian Xu, Zidi Xiong, Ritik Dutta, Rylan Schaeffer, Sang Truong, Simran Arora, Mantas Mazeika, Dan Hendrycks, Zinan Lin, Yu Cheng, Sanmi Koyejo, Dawn Song, Bo Li
Generative Pre-trained Transformer (GPT) models have exhibited exciting progress in capabilities, capturing the interest of practitioners and the public alike. Yet, while the literature on the trustworthiness of GPT models remains limited, practitioners have proposed employing capable GPT models for sensitive applications to healthcare and finance – where mistakes can be costly. To this end, this work proposes a comprehensive trustworthiness evaluation for large language models with a focus on GPT-4 and GPT-3.5, considering diverse perspectives – including toxicity, stereotype bias, adversarial robustness, out-of-distribution robustness, robustness on adversarial demonstrations, privacy, machine ethics, and fairness. Based on our evaluations, we discover previously unpublished vulnerabilities to trustworthiness threats. For instance, we find that GPT models can be easily misled to generate toxic and biased outputs and leak private information in both training data and conversation history. We also find that although GPT-4 is usually more trustworthy than GPT-3.5 on standard benchmarks, GPT-4 is more vulnerable given jailbreaking system or user prompts, potentially due to the reason that GPT-4 follows the (misleading) instructions more precisely. Our work illustrates a comprehensive trustworthiness evaluation of GPT models and sheds light on the trustworthiness gaps. Our benchmark is publicly available at https://decodingtrust.github.io/.
****
Three Towers: Flexible Contrastive Learning with Pretrained Image Models
Jannik Kossen, Mark Collier, Basil Mustafa, Xiao Wang, Xiaohua Zhai, Lucas Beyer, Andreas Steiner, Jesse Berent, Rodolphe Jenatton, Effrosyni Kokiopoulou
We introduce Three Towers (3T), a flexible method to improve the contrastive learning of vision-language models by incorporating pretrained image classifiers. While contrastive models are usually trained from scratch, LiT (Zhai et al., 2022) has recently shown performance gains from using pretrained classifier embeddings. However, LiT directly replaces the image tower with the frozen embeddings, excluding any potential benefits from training the image tower contrastively. With 3T, we propose a more flexible strategy that allows the image tower to benefit from both pretrained embeddings and contrastive training. To achieve this, we introduce a third tower that contains the frozen pretrained embeddings, and we encourage alignment between this third tower and the main image-text towers. Empirically, 3T consistently improves over LiT and the CLIP-style from-scratch baseline for retrieval tasks. For classification, 3T reliably improves over the from-scratch baseline, and while it underperforms relative to LiT for JFT-pretrained models, it outperforms LiT for ImageNet-21k and Places365 pretraining.

****
Practical and Asymptotically Exact Conditional Sampling in Diffusion Models
Luhuan Wu, Brian Trippe, Christian Naesseth, David Blei, John P. Cunningham
Diffusion models have been successful on a range of conditional generation tasks including molecular design and text-to-image generation. However, these achievements have primarily depended on task-specific conditional training or error-prone heuristic approximations. Ideally, a conditional generation method should provide exact samples for a broad range of conditional distributions without requiring task-specific training. To this end, we introduce the Twisted Diffusion Sampler, or TDS. TDS is a sequential Monte Carlo (SMC) algorithm that targets the conditional distributions of diffusion models through simulating a set of weighted particles. The main idea is to use twisting, an SMC technique that enjoys good computational efficiency, to incorporate heuristic approximations without compromising asymptotic exactness. We first find in simulation and in conditional image generation tasks that TDS provides a computational statistical trade-off, yielding more accurate approximations with many particles but with empirical improvements over heuristics with as few as two particles. We then turn to motif-scaffolding, a core task in protein design, using a TDS extension to Riemannian diffusion models; on benchmark tasks, TDS allows flexible conditioning criteria and often outperforms the state-of-the-art, conditionally trained model. Code can be found in https://github.com/blt2114/twisteddiffusionsampler
****
Actively Testing Your Model While It Learns: Realizing Label-Efficient Learning in Practice
Dayou Yu, Weishi Shi, Qi Yu
In active learning (AL), we focus on reducing the data annotation cost from the model training perspective. However, "testing'', which often refers to the model evaluation process of using empirical risk to estimate the intractable true generalization risk, also requires data annotations. The annotation cost for "testing'' (model evaluation) is under-explored. Even in works that study active model evaluation or active testing (AT), the learning and testing ends are disconnected. In this paper, we propose a novel active testing while learning (ATL) framework that integrates active learning with active testing. ATL provides an unbiased sample-efficient estimation of the model risk during active learning. It leverages test samples annotated from different periods of a dynamic active learning process to achieve fair model evaluations based on a theoretically guaranteed optimal integration of different test samples. Periodic testing also enables effective early-stopping to further save the total annotation cost. ATL further integrates an "active feedback'' mechanism, which is inspired by human learning, where the teacher (active tester) provides immediate guidance given by the prior performance of the student (active learner). Our theoretical result reveals that active feedback maintains the label complexity of the integrated learning-testing objective, while improving the model's generalization capability. We study the realistic setting where we maximize the performance gain from choosing "testing'' samples for feedback without sacrificing the risk estimation accuracy. An agnostic-style analysis and empirical evaluations on real-world datasets demonstrate that the ATL framework can effectively improve the annotation efficiency of both active learning and evaluation tasks.
****
MagicBrush: A Manually Annotated Dataset for Instruction-Guided Image Editing
Kai Zhang, Lingbo Mo, Wenhu Chen, Huan Sun, Yu Su
Text-guided image editing is widely needed in daily life, ranging from personal use to professional applications such as Photoshop.However, existing methods are either zero-shot or trained on an automatically synthesized dataset, which contains a high volume of noise.Thus, they still require lots of manual tuning to produce desirable outcomes in practice.To address this issue, we introduce MagicBrush, the first large-scale, manually annotated dataset for instruction-guided real image editing that covers diverse scenarios: single-turn, multi-turn, mask-provided, and mask-free editing.MagicBrush comprises over 10K manually annotated triplets (source image, instruction, target image), which supports trainining lar

ge-scale text-guided image editing models.We fine-tune InstructPix2Pix on MagicB
rush and show that the new model can produce much better images according to hum
an evaluation.We further conduct extensive experiments to evaluate current image
 editing baselines from multiple dimensions including quantitative, qualitative,
 and human evaluations.The results reveal the challenging nature of our dataset
and the gap between current baselines and real-world editing needs.
****
Causal Interpretation of Self-Attention in Pre-Trained Transformers
Raanan Y. Rohekar, Yaniv Gurwicz, Shami Nisimov
We propose a causal interpretation of self-attention in the Transformer neural n
etwork architecture. We interpret self-attention as a mechanism that estimates a
 structural equation model for a given input sequence of symbols (tokens). The s
tructural equation model can be interpreted, in turn, as a causal structure over
 the input symbols under the specific context of the input sequence. Importantly
, this interpretation remains valid in the presence of latent confounders. Follo
wing this interpretation, we estimate conditional independence relations between
 input symbols by calculating partial correlations between their corresponding r
epresentations in the deepest attention layer. This enables learning the causal
structure over an input sequence using existing constraint-based algorithms. In
this sense, existing pre-trained Transformers can be utilized for zero-shot caus
al-discovery. We demonstrate this method by providing causal explanations for th
e outcomes of Transformers in two tasks: sentiment classification (NLP) and reco
mmendation.
****
Parsel■: Algorithmic Reasoning with Language Models by Composing Decompositions
Eric Zelikman, Qian Huang, Gabriel Poesia, Noah Goodman, Nick Haber
Despite recent success in large language model (LLM) reasoning, LLMs struggle wi
th hierarchical multi-step reasoning tasks like generating complex programs. For
 these tasks, humans often start with a high-level algorithmic design and implem
ent each part gradually. We introduce Parsel, a framework enabling automatic imp
lementation and validation of complex algorithms with code LLMs. With Parsel, we
 automatically decompose algorithmic tasks into hierarchical natural language fu
nction descriptions and then search over combinations of possible function imple
mentations using tests. We show that Parsel can be used across domains requiring
 hierarchical reasoning, including program synthesis and robotic planning. We fi
nd that, using Parsel, LLMs solve more competition-level problems in the APPS da
taset, resulting in pass rates over 75\% higher than prior results from directly
 sampling AlphaCode and Codex, while often using a smaller sample budget. Moreov
er, with automatically generated tests, we find that Parsel can improve the stat
e-of-the-art pass@1 performance on HumanEval from 67\% to 85\%. We also find tha
t LLM-generated robotic plans using Parsel are more than twice as likely to be c
onsidered accurate than directly generated plans. Lastly, we explore how Parsel
addresses LLM limitations and discuss how Parsel may be useful for human program
mers. We release our code at https://github.com/ezelikman/parsel.
****
Focus on Query: Adversarial Mining Transformer for Few-Shot Segmentation
Yuan Wang, Naisong Luo, Tianzhu Zhang
Few-shot segmentation (FSS) aims to segment objects of new categories given only
 a handful of annotated samples. Previous works focus their efforts on exploring
 the support information while paying less attention to the mining of the critic
al query branch. In this paper, we rethink the importance of support information
 and propose a new query-centric FSS model Adversarial Mining Transformer (AMFor
mer), which achieves accurate query image segmentation with only rough support g
uidance or even weak support labels. The proposed AMFormer enjoys several merits
. First, we design an object mining transformer (G) that can achieve the expansi
on of incomplete region activated by support clue, and a detail mining transform
er (D) to discriminate the detailed local difference between the expanded mask a
nd the ground truth. Second, we propose to train G and D via an adversarial proc
ess, where G is optimized to generate more accurate masks approaching ground tru
th to fool D. We conduct extensive experiments on commonly used Pascal-5i and CO

CO-20i benchmarks and achieve state-of-the-art results across all settings. In addition, the decent performance with weak support labels in our query-centric paradigm may inspire the development of more general FSS models.
****
Improving Language Plasticity via Pretraining with Active Forgetting
Yihong Chen, Kelly Marchisio, Roberta Raileanu, David Adelani, Pontus Lars Erik Saito Stenetorp, Sebastian Riedel, Mikel Artetxe
Pretrained language models (PLMs) are today the primary model for natural language processing. Despite their impressive downstream performance, it can be difficult to apply PLMs to new languages, a barrier to making their capabilities universally accessible. While prior work has shown it possible to address this issue by learning a new embedding layer for the new language, doing so is both data and compute inefficient. We propose to use an active forgetting mechanism during pretraining, as a simple way of creating PLMs that can quickly adapt to new languages. Concretely, by resetting the embedding layer every K updates during pretraining, we encourage the PLM to improve its ability of learning new embeddings within limited number of updates, similar to a meta-learning effect. Experiments with RoBERTa show that models pretrained with our forgetting mechanism not only demonstrate faster convergence during language adaptation, but also outperform standard ones in a low-data regime, particularly for languages that are distant from English. Code will be available at https://github.com/facebookresearch/language-model-plasticity.
****
Stability of Random Forests and Coverage of Random-Forest Prediction Intervals
Yan Wang, Huaiqing Wu, Dan Nettleton
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Multi-Fidelity Multi-Armed Bandits Revisited
Xuchuang Wang, Qingyun Wu, Wei Chen, John C.S. Lui
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Augmentation-Aware Self-Supervision for Data-Efficient GAN Training
Liang Hou, Qi Cao, Yige Yuan, Songtao Zhao, Chongyang Ma, Siyuan Pan, Pengfei Wan, Zhongyuan Wang, Huawei Shen, Xueqi Cheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Template-free Articulated Neural Point Clouds for Reposable View Synthesis
Lukas Uzolas, Elmar Eisemann, Petr Kellnhofer
Dynamic Neural Radiance Fields (NeRFs) achieve remarkable visual quality when synthesizing novel views of time-evolving 3D scenes. However, the common reliance on backward deformation fields makes reanimation of the captured object poses challenging. Moreover, the state of the art dynamic models are often limited by low visual fidelity, long reconstruction time or specificity to narrow application domains. In this paper, we present a novel method utilizing a point-based representation and Linear Blend Skinning (LBS) to jointly learn a Dynamic NeRF and an associated skeletal model from even sparse multi-view video. Our forward-warping approach achieves state-of-the-art visual fidelity when synthesizing novel views and poses while significantly reducing the necessary learning time when compared to existing work. We demonstrate the versatility of our representation on a variety of articulated objects from common datasets and obtain reposable 3D reconstructions without the need of object-specific skeletal templates.

****
Demystifying the Optimal Performance of Multi-Class Classification
Minoh Jeong, Martina Cardone, Alex Dytso
Classification is a fundamental task in science and engineering on which machine learning methods have shown outstanding performances. However, it is challenging to determine whether such methods have achieved the Bayes error rate, that is, the lowest error rate attained by any classifier. This is mainly due to the fact that the Bayes error rate is not known in general and hence, effectively estimating it is paramount. Inspired by the work by Ishida et al. (2023), we propose an estimator for the Bayes error rate of supervised multi-class classification problems. We analyze several theoretical aspects of such estimator, including its consistency, unbiasedness, convergence rate, variance, and robustness. We also propose a denoising method that reduces the noise that potentially corrupts the data labels, and we improve the robustness of the proposed estimator to outliers by incorporating the median-of-means estimator. Our analysis demonstrates the consistency, asymptotic unbiasedness, convergence rate, and robustness of the proposed estimators. Finally, we validate the effectiveness of our theoretical results via experiments both on synthetic data under various noise settings and on real data.
****
HyPoradise: An Open Baseline for Generative Speech Recognition with Large Language Models
CHEN CHEN, Yuchen Hu, Chao-Han Huck Yang, Sabato Marco Siniscalchi, Pin-Yu Chen, Eng-Siong Chng
Advancements in deep neural networks have allowed automatic speech recognition (ASR) systems to attain human parity on several publicly available clean speech datasets. However, even state-of-the-art ASR systems experience performance degradation when confronted with adverse conditions, as a well-trained acoustic model is sensitive to variations in the speech domain, e.g., background noise. Intuitively, humans address this issue by relying on their linguistic knowledge: the meaning of ambiguous spoken terms is usually inferred from contextual cues thereby reducing the dependency on the auditory system. Inspired by this observation, we introduce the first open-source benchmark to utilize external large language models (LLMs) for ASR error correction, where N-best decoding hypotheses provide informative elements for true transcription prediction. This approach is a paradigm shift from the traditional language model rescoring strategy that can only select one candidate hypothesis as output transcription. The proposed benchmark contains a novel dataset, "HyPoradise" (HP), encompassing more than 316,000 pairs of N-best hypotheses and corresponding accurate transcriptions across prevalent speech domains. Given this dataset, we examine three types of error correction techniques based on LLMs with varying amounts of labeled hypotheses-transcription pairs, which gains significant word error rate (WER) reduction. Experimental evidence demonstrates the proposed technique achieves a breakthrough by surpassing the upper bound of traditional re-ranking based methods. More surprisingly, LLM with reasonable prompt design can even correct those tokens that are missing in N-best list. We make our results publicly accessible for reproducible pipelines with released pre-trained models, thus providing a new paradigm for ASR error correction with LLMs.
****
Structured Voronoi Sampling
Afra Amini, Li Du, Ryan Cotterell
Gradient-based sampling algorithms have demonstrated their effectiveness in text generation, especially in the context of controlled text generation. However, there exists a lack of theoretically grounded and principled approaches for this task. In this paper, we take an important step toward building a principled approach for sampling from language models with gradient-based methods. We use discrete distributions given by language models to define densities and develop an algorithm based on Hamiltonian Monte Carlo to sample from them. We name our gradient-based technique Structured Voronoi Sampling (SVS). In an experimental setup where the reference distribution is known, we show that the empirical distributio

n of SVS samples is closer to the reference distribution compared to alternative sampling schemes. Furthermore, in a controlled generation task, SVS is able to generate fluent and diverse samples while following the control targets significantly better than other methods.
****

Stability and Generalization of the Decentralized Stochastic Gradient Descent Ascent Algorithm
Miaoxi Zhu, Li Shen, Bo Du, Dacheng Tao
The growing size of available data has attracted increasing interest in solving minimax problems in a decentralized manner for various machine learning tasks. Previous theoretical research has primarily focused on the convergence rate and communication complexity of decentralized minimax algorithms, with little attention given to their generalization. In this paper, we investigate the primal-dual generalization bound of the decentralized stochastic gradient descent ascent (D-SGDA) algorithm using the approach of algorithmic stability under both convex-concave and nonconvex-nonconcave settings. Our theory refines the algorithmic stability in a decentralized manner and demonstrates that the decentralized structure does not destroy the stability and generalization of D-SGDA, implying that it can generalize as well as the vanilla SGDA in certain situations. Our results analyze the impact of different topologies on the generalization bound of the D-SGDA algorithm beyond trivial factors such as sample sizes, learning rates, and iterations. We also evaluate the optimization error and balance it with the generalization gap to obtain the optimal population risk of D-SGDA in the convex-concave setting. Additionally, we perform several numerical experiments which validate our theoretical findings.
****

Hierarchical clustering with dot products recovers hidden tree structure
Annie Gray, Alexander Modell, Patrick Rubin-Delanchy, Nick Whiteley
In this paper we offer a new perspective on the well established agglomerative clustering algorithm, focusing on recovery of hierarchical structure. We recommend a simple variant of the standard algorithm, in which clusters are merged by maximum average dot product and not, for example, by minimum distance or within-cluster variance. We demonstrate that the tree output by this algorithm provides a bona fide estimate of generative hierarchical structure in data, under a generic probabilistic graphical model. The key technical innovations are to understand how hierarchical information in this model translates into tree geometry which can be recovered from data, and to characterise the benefits of simultaneously growing sample size and data dimension. We demonstrate superior tree recovery performance with real data over existing approaches such as UPGMA, Ward's method, and HDBSCAN.
****

Latent Field Discovery in Interacting Dynamical Systems with Neural Fields
Miltiadis (Miltos) Kofinas, Erik Bekkers, Naveen Nagaraja, Efstratios Gavves
Systems of interacting objects often evolve under the influence of underlying field effects that govern their dynamics, yet previous works have abstracted away from such effects, and assume that systems evolve in a vacuum. In this work, we focus on discovering these fields, and infer them from the observed dynamics alone, without directly observing them. We theorize the presence of latent force fields, and propose neural fields to learn them. Since the observed dynamics constitute the net effect of local object interactions and global field effects, recently popularized equivariant networks are inapplicable, as they fail to capture global information. To address this, we propose to disentangle local object interactions --which are SE(3) equivariant and depend on relative states-- from external global field effects --which depend on absolute states. We model the interactions with equivariant graph networks, and combine them with neural fields in a novel graph network that integrates field forces. Our experiments show that we can accurately discover the underlying fields in charged particles settings, traffic scenes, and gravitational n-body problems, and effectively use them to learn the system and forecast future trajectories.
****

Accelerating Reinforcement Learning with Value-Conditional State Entropy Exploration
Dongyoung Kim, Jinwoo Shin, Pieter Abbeel, Younggyo Seo
A promising technique for exploration is to maximize the entropy of visited state distribution, i.e., state entropy, by encouraging uniform coverage of visited state space. While it has been effective for an unsupervised setup, it tends to struggle in a supervised setup with a task reward, where an agent prefers to visit high-value states to exploit the task reward. Such a preference can cause an imbalance between the distributions of high-value states and low-value states, which biases exploration towards low-value state regions as a result of the state entropy increasing when the distribution becomes more uniform. This issue is exacerbated when high-value states are narrowly distributed within the state space, making it difficult for the agent to complete the tasks. In this paper, we present a novel exploration technique that maximizes the value-conditional state entropy, which separately estimates the state entropies that are conditioned on the value estimates of each state, then maximizes their average. By only considering the visited states with similar value estimates for computing the intrinsic bonus, our method prevents the distribution of low-value states from affecting exploration around high-value states, and vice versa. We demonstrate that the proposed alternative to the state entropy baseline significantly accelerates various reinforcement learning algorithms across a variety of tasks within MiniGrid, DeepMind Control Suite, and Meta-World benchmarks. Source code is available at https://sites.google.com/view/rl-vcse.
****
Learning World Models with Identifiable Factorization
Yuren Liu, Biwei Huang, Zhengmao Zhu, Honglong Tian, Mingming Gong, Yang Yu, Kun Zhang
Extracting a stable and compact representation of the environment is crucial for efficient reinforcement learning in high-dimensional, noisy, and non-stationary environments. Different categories of information coexist in such environments -- how to effectively extract and disentangle the information remains a challenging problem. In this paper, we propose IFactor, a general framework to model four distinct categories of latent state variables that capture various aspects of information within the RL system, based on their interactions with actions and rewards. Our analysis establishes block-wise identifiability of these latent variables, which not only provides a stable and compact representation but also discloses that all reward-relevant factors are significant for policy learning. We further present a practical approach to learning the world model with identifiable blocks, ensuring the removal of redundancies but retaining minimal and sufficient information for policy optimization. Experiments in synthetic worlds demonstrate that our method accurately identifies the ground-truth latent variables, substantiating our theoretical findings. Moreover, experiments in variants of the DeepMind Control Suite and RoboDesk showcase the superior performance of our approach over baselines.
****
Online Map Vectorization for Autonomous Driving: A Rasterization Perspective
Gongjie Zhang, Jiahao Lin, Shuang Wu, yilin song, Zhipeng Luo, Yang Xue, Shijian Lu, Zuoguan Wang
High-definition (HD) vectorized map is essential for autonomous driving, providing detailed and precise environmental information for advanced perception and planning. However, current map vectorization methods often exhibit deviations, and the existing evaluation metric for map vectorization lacks sufficient sensitivity to detect these deviations. To address these limitations, we propose integrating the philosophy of rasterization into map vectorization. Specifically, we introduce a new rasterization-based evaluation metric, which has superior sensitivity and is better suited to real-world autonomous driving scenarios. Furthermore, we propose MapVR (Map Vectorization via Rasterization), a novel framework that applies differentiable rasterization to vectorized outputs and then performs precise and geometry-aware supervision on rasterized HD maps. Notably, MapVR designs tailored rasterization strategies for various geometric shapes, enabling effec

tive adaptation to a wide range of map elements. Experiments show that incorporating rasterization into map vectorization greatly enhances performance with no extra computational cost during inference, leading to more accurate map perception and ultimately promoting safer autonomous driving. Codes are available at https://github.com/ZhangGongjie/MapVR. A standalone map vectorization evaluation toolkit is available at https://github.com/jiahaoLjh/MapVectorizationEvalToolkit.
****
NAP: Neural 3D Articulated Object Prior
Jiahui Lei, Congyue Deng, William B Shen, Leonidas J. Guibas, Kostas Daniilidis
We propose Neural 3D Articulated object Prior (NAP), the first 3D deep generative model to synthesize 3D articulated object models. Despite the extensive research on generating 3D static objects, compositions, or scenes, there are hardly any approaches on capturing the distribution of articulated objects, a common object category for human and robot interaction. To generate articulated objects, we first design a novel articulation tree/graph parameterization and then apply a diffusion-denoising probabilistic model over this representation where articulated objects can be generated via denoising from random complete graphs. In order to capture both the geometry and the motion structure whose distribution will affect each other, we design a graph denoising network for learning the reverse diffusion process. We propose a novel distance that adapts widely used 3D generation metrics to our novel task to evaluate generation quality. Experiments demonstrate our high performance in articulated object generation as well as its applications on conditioned generation, including Part2Motion, PartNet-Imagination, Motion2Part, and GAPart2Object.
****
Compressed Video Prompt Tuning
Bing Li, Jiaxin Chen, Xiuguo Bao, Di Huang
Compressed videos offer a compelling alternative to raw videos, showing the possibility to significantly reduce the on-line computational and storage cost. However, current approaches to compressed video processing generally follow the resource-consuming pre-training and fine-tuning paradigm, which does not fully take advantage of such properties, making them not favorable enough for widespread applications. Inspired by recent successes of prompt tuning techniques in computer vision, this paper presents the first attempt to build a prompt based representation learning framework, which enables effective and efficient adaptation of pre-trained raw video models to compressed video understanding tasks. To this end, we propose a novel prompt tuning approach, namely Compressed Video Prompt Tuning (CVPT), emphatically dealing with the challenging issue caused by the inconsistency between pre-training and downstream data modalities. Specifically, CVPT replaces the learnable prompts with compressed modalities (\emph{e.g.} Motion Vectors and Residuals) by re-parameterizing them into conditional prompts followed by layer-wise refinement. The conditional prompts exhibit improved adaptability and generalizability to instances compared to conventional individual learnable ones, and the Residual prompts enhance the noisy motion cues in the Motion Vector prompts for further fusion with the visual cues from I-frames. Additionally, we design Selective Cross-modal Complementary Prompt (SCCP) blocks. After inserting them into the backbone, SCCP blocks leverage semantic relations across diverse levels and modalities to improve cross-modal interactions between prompts and input flows. Extensive evaluations on HMDB-51, UCF-101 and Something-Something v2 demonstrate that CVPT remarkably outperforms the state-of-the-art counterparts, delivering a much better balance between accuracy and efficiency.
****
Sampling from Structured Log-Concave Distributions via a Soft-Threshold Dikin Walk
Oren Mangoubi, Nisheeth K. Vishnoi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Implicit Regularization in Over-Parameterized Support Vector Machine
Yang Sui, Xin HE, Yang Bai
In this paper, we design a regularization-free algorithm for high-dimensional support vector machines (SVMs) by integrating over-parameterization with Nesterov's smoothing method, and provide theoretical guarantees for the induced implicit regularization phenomenon. In particular, we construct an over-parameterized hinge loss function and estimate the true parameters by leveraging regularization-free gradient descent on this loss function. The utilization of Nesterov's method enhances the computational efficiency of our algorithm, especially in terms of determining the stopping criterion and reducing computational complexity. With appropriate choices of initialization, step size, and smoothness parameter, we demonstrate that unregularized gradient descent achieves a near-oracle statistical convergence rate. Additionally, we verify our theoretical findings through a variety of numerical experiments and compare the proposed method with explicit regularization. Our results illustrate the advantages of employing implicit regularization via gradient descent in conjunction with over-parameterization in sparse SVMs.
****
Large Language Models as Commonsense Knowledge for Large-Scale Task Planning
Zirui Zhao, Wee Sun Lee, David Hsu
Large-scale task planning is a major challenge. Recent work exploits large language models (LLMs) directly as a policy and shows surprisingly interesting results. This paper shows that LLMs provide a commonsense model of the world in addition to a policy that acts on it. The world model and the policy can be combined in a search algorithm, such as Monte Carlo Tree Search (MCTS), to scale up task planning. In our new LLM-MCTS algorithm, the LLM-induced world model provides a commonsense prior belief for MCTS to achieve effective reasoning; the LLM-induced policy acts as a heuristic to guide the search, vastly improving search efficiency. Experiments show that LLM-MCTS outperforms both MCTS alone and policies induced by LLMs (GPT2 and GPT3.5) by a wide margin, for complex, novel tasks.   Further experiments and analyses on multiple tasks -- multiplication, travel planning, object rearrangement -- suggest minimum description length (MDL) as a general guiding principle: if the description length of the world model is substantially smaller than that of the policy, using LLM as a world model for model-based planning is likely better than using LLM solely as a policy.
****
Uncovering Meanings of Embeddings via Partial Orthogonality
Yibo Jiang, Bryon Aragam, Victor Veitch
Machine learning tools often rely on embedding text as vectors of real numbers.In this paper, we study how the semantic structure of language is encoded in the algebraic structure of such embeddings.Specifically, we look at a notion of "semantic independence" capturing the idea that, e.g., "eggplant" and "tomato" are independent given "vegetable". Although such examples are intuitive, it is difficult to formalize such a notion of semantic independence. The key observation here is that any sensible formalization should obey a set of so-called independence axioms, and thus any algebraic encoding of this structure should also obey these axioms. This leads us naturally to use partial orthogonality as the relevant algebraic structure. We develop theory and methods that allow us to demonstrate that partial orthogonality does indeed capture semantic independence.Complementary to this, we also introduce the concept of independence preserving embeddings where embeddings preserve the conditional independence structures of a distribution, and we prove the existence of such embeddings and approximations to them.
****
Federated Learning with Bilateral Curation for Partially Class-Disjoint Data
Ziqing Fan, ruipeng zhang, Jiangchao Yao, Bo Han, Ya Zhang, Yanfeng Wang
Partially class-disjoint data (PCDD), a common yet under-explored data formation where each client contributes a part of classes (instead of all classes) of samples, severely challenges the performance of federated algorithms. Without full classes, the local objective will contradict the global objective, yielding the angle collapse problem for locally missing classes and the space waste problem f

or locally existing classes. As far as we know, none of the existing methods can intrinsically mitigate PCDD challenges to achieve holistic improvement in the bilateral views (both global view and local view) of federated learning. To address this dilemma, we are inspired by the strong generalization of simplex Equiangular Tight Frame (ETF) on the imbalanced data, and propose a novel approach called FedGELA where the classifier is globally fixed as a simplex ETF while locally adapted to the personal distributions. Globally, FedGELA provides fair and equal discrimination for all classes and avoids inaccurate updates of the classifier, while locally it utilizes the space of locally missing classes for locally existing classes. We conduct extensive experiments on a range of datasets to demonstrate that our FedGELA achieves promising performance (averaged improvement of 3.9% to FedAvg and 1.5% to best baselines) and provide both local and global convergence guarantees.
****
Partial Counterfactual Identification of Continuous Outcomes with a Curvature Sensitivity Model
Valentyn Melnychuk, Dennis Frauen, Stefan Feuerriegel
Counterfactual inference aims to answer retrospective "what if" questions and thus belongs to the most fine-grained type of inference in Pearl's causality ladder. Existing methods for counterfactual inference with continuous outcomes aim at point identification and thus make strong and unnatural assumptions about the underlying structural causal model. In this paper, we relax these assumptions and aim at partial counterfactual identification of continuous outcomes, i.e., when the counterfactual query resides in an ignorance interval with informative bounds. We prove that, in general, the ignorance interval of the counterfactual queries has non-informative bounds, already when functions of structural causal models are continuously differentiable. As a remedy, we propose a novel sensitivity model called Curvature Sensitivity Model. This allows us to obtain informative bounds by bounding the curvature of level sets of the functions. We further show that existing point counterfactual identification methods are special cases of our Curvature Sensitivity Model when the bound of the curvature is set to zero. We then propose an implementation of our Curvature Sensitivity Model in the form of a novel deep generative model, which we call Augmented Pseudo-Invertible Decoder. Our implementation employs (i) residual normalizing flows with (ii) variational augmentations. We empirically demonstrate the effectiveness of our Augmented Pseudo-Invertible Decoder. To the best of our knowledge, ours is the first partial identification model for Markovian structural causal models with continuous outcomes.
****
Training biologically plausible recurrent neural networks on cognitive tasks with long-term dependencies
Wayne Soo, Vishwa Goudar, Xiao-Jing Wang
Training recurrent neural networks (RNNs) has become a go-to approach for generating and evaluating mechanistic neural hypotheses for cognition. The ease and efficiency of training RNNs with backpropagation through time and the availability of robustly supported deep learning libraries has made RNN modeling more approachable and accessible to neuroscience. Yet, a major technical hindrance remains. Cognitive processes such as working memory and decision making involve neural population dynamics over a long period of time within a behavioral trial and across trials. It is difficult to train RNNs to accomplish tasks where neural representations and dynamics have long temporal dependencies without gating mechanisms such as LSTMs or GRUs which currently lack experimental support and prohibit direct comparison between RNNs and biological neural circuits. We tackled this problem based on the idea of specialized skip-connections through time to support the emergence of task-relevant dynamics, and subsequently reinstitute biological plausibility by reverting to the original architecture. We show that this approach enables RNNs to successfully learn cognitive tasks that prove impractical if not impossible to learn using conventional methods. Over numerous tasks considered here, we achieve less training steps and shorter wall-clock times, particularly in tasks that require learning long-term dependencies via temporal integratio

n over long timescales or maintaining a memory of past events in hidden-states. Our methods expand the range of experimental tasks that biologically plausible RNN models can learn, thereby supporting the development of theory for the emergent neural mechanisms of computations involving long-term dependencies.
****
Diffusion Probabilistic Models for Structured Node Classification
Hyosoon Jang, Seonghyun Park, Sangwoo Mo, Sungsoo Ahn
This paper studies structured node classification on graphs, where the predictions should consider dependencies between the node labels. In particular, we focus on solving the problem for partially labeled graphs where it is essential to incorporate the information in the known label for predicting the unknown labels. To address this issue, we propose a novel framework leveraging the diffusion probabilistic model for structured node classification (DPM-SNC). At the heart of our framework is the extraordinary capability of DPM-SNC to (a) learn a joint distribution over the labels with an expressive reverse diffusion process and (b) make predictions conditioned on the known labels utilizing manifold-constrained sampling. Since the DPMs lack training algorithms for partially labeled data, we design a novel training algorithm to apply DPMs, maximizing a new variational lower bound. We also theoretically analyze how DPMs benefit node classification by enhancing the expressive power of GNNs based on proposing AGG-WL, which is strictly more powerful than the classic 1-WL test. We extensively verify the superiority of our DPM-SNC in diverse scenarios, which include not only the transductive setting on partially labeled graphs but also the inductive setting and unlabeled graphs.
****
Non-stationary Experimental Design under Linear Trends
David Simchi-Levi, Chonghuan Wang, Zeyu Zheng
Experimentation has been critical and increasingly popular across various domains, such as clinical trials and online platforms, due to its widely recognized benefits. One of the primary objectives of classical experiments is to estimate the average treatment effect (ATE) to inform future decision-making. However, in healthcare and many other settings, treatment effects may be non-stationary, meaning that they can change over time, rendering the traditional experimental design inadequate and the classical static ATE uninformative. In this work, we address the problem of non-stationary experimental design under linear trends by considering two objectives: estimating the dynamic treatment effect and minimizing welfare loss within the experiment. We propose an efficient design that can be customized for optimal estimation error rate, optimal regret rate, or the Pareto optimal trade-off between the two objectives. We establish information-theoretical lower bounds that highlight the inherent challenge in estimating dynamic treatment effects and minimizing welfare loss, and also statistically reveal the fundamental trade-off between them.
****
EICIL: Joint Excitatory Inhibitory Cycle Iteration Learning for Deep Spiking Neural Networks
Zihang Shao, Xuanye Fang, Yaxin Li, Chaoran Feng, Jiangrong Shen, Qi Xu
Spiking neural networks (SNNs) have undergone continuous development and extensive study for decades, leading to increased biological plausibility and optimal energy efficiency. However, traditional training methods for deep SNNs have some limitations, as they rely on strategies such as pre-training and fine-tuning, indirect coding and reconstruction, and approximate gradients. These strategies lack a complete training model and require gradient approximation. To overcome these limitations, we propose a novel learning method named Joint Excitatory Inhibitory Cycle Iteration learning for Deep Spiking Neural Networks (EICIL) that integrates both excitatory and inhibitory behaviors inspired by the signal transmission of biological neurons.By organically embedding these two behavior patterns into one framework, the proposed EICIL significantly improves the bio-mimicry and adaptability of spiking neuron models, as well as expands the representation space of spiking neurons. Extensive experiments based on EICIL and traditional learning methods demonstrate that EICIL outperforms traditional methods on various

datasets, such as CIFAR10 and CIFAR100, revealing the crucial role of the learning approach that integrates both behaviors during training.
****
Encoding Time-Series Explanations through Self-Supervised Model Behavior Consistency
Owen Queen, Tom Hartvigsen, Teddy Koker, Huan He, Theodoros Tsiligkaridis, Marinka Zitnik
Interpreting time series models is uniquely challenging because it requires identifying both the location of time series signals that drive model predictions and their matching to an interpretable temporal pattern. While explainers from other modalities can be applied to time series, their inductive biases do not transfer well to the inherently challenging interpretation of time series. We present TimeX, a time series consistency model for training explainers. TimeX trains an interpretable surrogate to mimic the behavior of a pretrained time series model. It addresses the issue of model faithfulness by introducing model behavior consistency, a novel formulation that preserves relations in the latent space induced by the pretrained model with relations in the latent space induced by TimeX. TimeX provides discrete attribution maps and, unlike existing interpretability methods, it learns a latent space of explanations that can be used in various ways, such as to provide landmarks to visually aggregate similar explanations and easily recognize temporal patterns. We evaluate TimeX on eight synthetic and real-world datasets and compare its performance against state-of-the-art interpretability methods. We also conduct case studies using physiological time series. Quantitative evaluations demonstrate that TimeX achieves the highest or second-highest performance in every metric compared to baselines across all datasets. Through case studies, we show that the novel components of TimeX show potential for training faithful, interpretable models that capture the behavior of pretrained time series models.
****
NeuroEvoBench: Benchmarking Evolutionary Optimizers for Deep Learning Applications
Robert Lange, Yujin Tang, Yingtao Tian
Recently, the Deep Learning community has become interested in evolutionary optimization (EO) as a means to address hard optimization problems, e.g. meta-learning through long inner loop unrolls or optimizing non-differentiable operators. One core reason for this trend has been the recent innovation in hardware acceleration and compatible software -- making distributed population evaluations much easier than before. Unlike for gradient descent-based methods though, there is a lack of hyperparameter understanding and best practices for EO – arguably due to severely less `graduate student descent' and benchmarking being performed for EO methods. Additionally, classical benchmarks from the evolutionary community provide few practical insights for Deep Learning applications. This poses challenges for newcomers to hardware-accelerated EO and hinders significant adoption. Hence, we establish a new benchmark of EO methods (NEB) tailored toward Deep Learning applications and exhaustively evaluate traditional and meta-learned EO. We investigate core scientific questions including resource allocation, fitness shaping, normalization, regularization & scalability of EO. The benchmark is open-sourced at https://github.com/neuroevobench/neuroevobench under Apache-2.0 license.
****
HyTrel: Hypergraph-enhanced Tabular Data Representation Learning
Pei Chen, Soumajyoti Sarkar, Leonard Lausen, Balasubramaniam Srinivasan, Sheng Zha, Ruihong Huang, George Karypis
Language models pretrained on large collections of tabular data have demonstrated their effectiveness in several downstream tasks.However, many of these models do not take into account the row/column permutation invariances, hierarchical structure, etc. that exist in tabular data. To alleviate these limitations, we propose HyTrel, a tabular language model, that captures the permutation invariances and three more structural properties of tabular data by using hypergraphs--where the table cells make up the nodes and the cells occurring jointly together in

each row, column, and the entire table are used to form three different types of hyperedges. We show thatHyTrel is maximally invariant under certain conditions for tabular data, i.e., two tables obtain the same representations via HyTreliff the two tables are identical up to permutation. Our empirical results demonstrate that HyTrel consistently outperforms other competitive baselines on four downstream tasks with minimal pretraining, illustrating the advantages of incorporating inductive biases associated with tabular data into the representations. Finally, our qualitative analyses showcase that HyTrel can assimilate the table structure to generate robust representations for the cells, rows, columns, and the entire table.
****
PGDiff: Guiding Diffusion Models for Versatile Face Restoration via Partial Guidance
Peiqing Yang, Shangchen Zhou, Qingyi Tao, Chen Change Loy
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Convolutions Die Hard: Open-Vocabulary Segmentation with Single Frozen Convolutional CLIP
Qihang Yu, Ju He, Xueqing Deng, Xiaohui Shen, Liang-Chieh Chen
Open-vocabulary segmentation is a challenging task requiring segmenting and recognizing objects from an open set of categories in diverse environments. One way to address this challenge is to leverage multi-modal models, such as CLIP, to provide image and text features in a shared embedding space, which effectively bridges the gap between closed-vocabulary and open-vocabulary recognition.Hence, existing methods often adopt a two-stage framework to tackle the problem, where the inputs first go through a mask generator and then through the CLIP model along with the predicted masks. This process involves extracting features from raw images multiple times, which can be ineffective and inefficient. By contrast, we propose to build everything into a single-stage framework using a shared Frozen Convolutional CLIP backbone, which not only significantly simplifies the current two-stage pipeline, but also remarkably yields a better accuracy-cost trade-off. The resulting single-stage system, called FC-CLIP, benefits from the following observations: the frozen CLIP backbone maintains the ability of open-vocabulary classification and can also serve as a strong mask generator, and the convolutional CLIP generalizes well to a larger input resolution than the one used during contrastive image-text pretraining. Surprisingly, FC-CLIP advances state-of-the-art results on various benchmarks, while running practically fast. Specifically, when training on COCO panoptic data only and testing in a zero-shot manner, FC-CLIP achieve 26.8 PQ, 16.8 AP, and 34.1 mIoU on ADE20K, 18.2 PQ, 27.9 mIoU on Mapillary Vistas, 44.0 PQ, 26.8 AP, 56.2 mIoU on Cityscapes, outperforming the prior art under the same setting by +4.2 PQ, +2.4 AP, +4.2 mIoU on ADE20K, +4.0 PQ on Mapillary Vistas and +20.1 PQ on Cityscapes, respectively. Additionally, the training and testing time of FC-CLIP is 7.5x and 6.6x significantly faster than the same prior art, while using 5.9x fewer total model parameters. Meanwhile, FC-CLIP also sets a new state-of-the-art performance across various open-vocabulary semantic segmentation datasets. Code and models are available at https://github.com/bytedance/fc-clip
****
CLIP-OGD: An Experimental Design for Adaptive Neyman Allocation in Sequential Experiments
Jessica Dai, Paula Gradu, Christopher Harshaw
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Monitor-Guided Decoding of Code LMs with Static Analysis of Repository Context

Lakshya A Agrawal, Aditya Kanade, Navin Goyal, Shuvendu Lahiri, Sriram Rajamani
Language models of code (LMs) work well when the surrounding code provides sufficient context. This is not true when it becomes necessary to use types, functionality or APIs defined elsewhere in the repository or a linked library, especially those not seen during training. LMs suffer from limited awareness of such global context and end up hallucinating.Integrated development environments (IDEs) assist developers in understanding repository context using static analysis. We extend this assistance, enjoyed by developers, to LMs. We propose monitor-guided decoding (MGD) where a monitor uses static analysis to guide the decoding. We construct a repository-level dataset PragmaticCode for method-completion in Java and evaluate MGD on it. On models of varying parameter scale, by monitoring for type-consistent object dereferences, MGD consistently improves compilation rates and agreement with ground truth. Further, LMs with fewer parameters, when augmented with MGD, can outperform larger LMs. With MGD, SantaCoder-1.1B achieves better compilation rate and next-identifier match than the much larger text-davinci-003 model.We also conduct a generalizability study to evaluate the ability of MGD to generalize to multiple programming languages (Java, C# and Rust), coding scenarios (e.g., correct number of arguments to method calls), and to enforce richer semantic constraints (e.g., stateful API protocols). Our data and implementation are available at https://github.com/microsoft/monitors4codegen.
****
Towards Federated Foundation Models: Scalable Dataset Pipelines for Group-Structured Learning
Zachary Charles, Nicole Mitchell, Krishna Pillutla, Michael Reneer, Zachary Garrett
We introduce Dataset Grouper, a library to create large-scale group-structured (e.g., federated) datasets, enabling federated learning simulation at the scale of foundation models. This library facilitates the creation of group-structured versions of existing datasets based on user-specified partitions, and directly leads to a variety of useful heterogeneous datasets that can be plugged into existing software frameworks. Dataset Grouper offers three key advantages. First, it scales to settings where even a single group's dataset is too large to fit in memory. Second, it provides flexibility, both in choosing the base (non-partitioned) dataset and in defining partitions. Finally, it is framework-agnostic. We empirically demonstrate that Dataset Grouper enables large-scale federated language modeling simulations on datasets that are orders of magnitude larger than in previous work, allowing for federated training of language models with hundreds of millions, and even billions, of parameters. Our experimental results show that algorithms like FedAvg operate more as meta-learning methods than as empirical risk minimization methods at this scale, suggesting their utility in downstream personalization and task-specific adaptation. Dataset Grouper is available at https://github.com/google-research/dataset_grouper.
****
Sharp Spectral Rates for Koopman Operator Learning
Vladimir Kostic, Karim Lounici, Pietro Novelli, Massimiliano Pontil
Non-linear dynamical systems can be handily described by the associated Koopman operator, whose action evolves every observable of the system forward in time. Learning the Koopman operator and its spectral decomposition from data is enabled by a number of algorithms. In this work we present for the first time non-asymptotic learning bounds for the Koopman eigenvalues and eigenfunctions. We focus on time-reversal-invariant stochastic dynamical systems, including the important example of Langevin dynamics. We analyze two popular estimators: Extended Dynamic Mode Decomposition (EDMD) and Reduced Rank Regression (RRR). Our results critically hinge on novel {minimax} estimation bounds for the operator norm error, that may be of independent interest. Our spectral learning bounds are driven by the simultaneous control of the operator norm error and a novel metric distortion functional of the estimated eigenfunctions. The bounds indicates that both EDMD and RRR have similar variance, but EDMD suffers from a larger bias which might be detrimental to its learning rate. Our results shed new light on the emergence of spurious eigenvalues, an issue which is well known empirically. Numerical exp

eriments illustrate the implications of the bounds in practice.
****
Continual Learning for Instruction Following from Realtime Feedback
Alane Suhr, Yoav Artzi
We propose and deploy an approach to continually train an instruction-following agent from feedback provided by users during collaborative interactions. During interaction, human users instruct an agent using natural language, and provide realtime binary feedback as they observe the agent following their instructions. We design a contextual bandit learning approach, converting user feedback to immediate reward. We evaluate through thousands of human-agent interactions, demonstrating 15.4% absolute improvement in instruction execution accuracy over time. We also show our approach is robust to several design variations, and that the feedback signal is roughly equivalent to the learning signal of supervised demonstration data.
****
Embracing the chaos: analysis and diagnosis of numerical instability in variational flows
Zuheng Xu, Trevor Campbell
In this paper, we investigate the impact of numerical instability on the reliability of sampling, density evaluation, and evidence lower bound (ELBO) estimation in variational flows. We first empirically demonstrate that common flows can exhibit a catastrophic accumulation of error: the numerical flow map deviates significantly from the exact map---which affects sampling---and the numerical inverse flow map does not accurately recover the initial input---which affects density and ELBO computations. Surprisingly though, we find that results produced by flows are often accurate enough for applications despite the presence of serious numerical instability. In this work, we treat variational flows as chaotic dynamical systems, and leverage shadowing theory to elucidate this behavior via theoretical guarantees on the error of sampling, density evaluation, and ELBO estimation. Finally, we develop and empirically test a diagnostic procedure that can be used to validate results produced by numerically unstable flows in practice.
****
Masked Two-channel Decoupling Framework for Incomplete Multi-view Weak Multi-label Learning
Chengliang Liu, Jie Wen, Yabo Liu, Chao Huang, Zhihao Wu, Xiaoling Luo, Yong Xu
Multi-view learning has become a popular research topic in recent years, but research on the cross-application of classic multi-label classification and multi-view learning is still in its early stages. In this paper, we focus on the complex yet highly realistic task of incomplete multi-view weak multi-label learning and propose a masked two-channel decoupling framework based on deep neural networks to solve this problem. The core innovation of our method lies in decoupling the single-channel view-level representation, which is common in deep multi-view learning methods, into a shared representation and a view-proprietary representation. We also design a cross-channel contrastive loss to enhance the semantic property of the two channels. Additionally, we exploit supervised information to design a label-guided graph regularization loss, helping the extracted embedding features preserve the geometric structure among samples. Inspired by the success of masking mechanisms in image and text analysis, we develop a random fragment masking strategy for vector features to improve the learning ability of encoders. Finally, it is important to emphasize that our model is fully adaptable to arbitrary view and label absences while also performing well on the ideal full data. We have conducted sufficient and convincing experiments to confirm the effectiveness and advancement of our model.
****
Exponential Lower Bounds for Fictitious Play in Potential Games
Ioannis Panageas, Nikolas Patris, Stratis Skoulakis, Volkan Cevher
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****
Cocktail: Mixing Multi-Modality Control for Text-Conditional Image Generation
Minghui Hu, Jianbin Zheng, Daqing Liu, Chuanxia Zheng, Chaoyue Wang, Dacheng Tao
, Tat-Jen Cham
Text-conditional diffusion models are able to generate high-fidelity images with
 diverse contents.However, linguistic representations frequently exhibit ambiguo
us descriptions of the envisioned objective imagery, requiring the incorporation
 of additional control signals to bolster the efficacy of text-guided diffusion
models. In this work, we propose Cocktail, a pipeline to mix various modalities
into one embedding, amalgamated with a generalized ControlNet (gControlNet), a c
ontrollable normalisation (ControlNorm), and a spatial guidance sampling method,
 to actualize multi-modal and spatially-refined control for text-conditional dif
fusion models. Specifically, we introduce a hyper-network gControlNet, dedicated
 to the alignment and infusion of the control signals from disparate modalities
into the pre-trained diffusion model. gControlNet is capable of accepting flexib
le modality signals, encompassing the simultaneous reception of any combination
of modality signals, or the supplementary fusion of multiple modality signals. T
he control signals are then fused and injected into the backbone model according
 to our proposed ControlNorm.Furthermore, our advanced spatial guidance sampling
 methodology proficiently incorporates the control signal into the designated re
gion, thereby circumventing the manifestation of undesired objects within the ge
nerated image.We demonstrate the results of our method in controlling various mo
dalities, proving high-quality synthesis and fidelity to multiple external signa
ls.
****
RePo: Resilient Model-Based Reinforcement Learning by Regularizing Posterior Pre
dictability
Chuning Zhu, Max Simchowitz, Siri Gadipudi, Abhishek Gupta
Visual model-based RL methods typically encode image observations into low-dimen
sional representations in a manner that does not eliminate redundant information
. This leaves them susceptible to spurious variations -- changes in task-irrelev
ant components such as background distractors or lighting conditions. In this pa
per, we propose a visual model-based RL method that learns a latent representati
on resilient to such spurious variations. Our training objective encourages the
representation to be maximally predictive of dynamics and reward, while constrai
ning the information flow from the observation to the latent representation. We
demonstrate that this objective significantly bolsters the resilience of visual
model-based RL methods to visual distractors, allowing them to operate in dynami
c environments. We then show that while the learned encoder is able to operate i
n dynamic environments, it is not invariant under significant distribution shift
. To address this, we propose a simple reward-free alignment procedure that enab
les test time adaptation of the encoder. This allows for quick adaptation to wid
ely differing environments without having to relearn the dynamics and policy. Ou
r effort is a step towards making model-based RL a practical and useful tool for
 dynamic, diverse domains and we show its effectiveness in simulation tasks with
 significant spurious variations.
****
Circuit as Set of Points
Jialv Zou, Xinggang Wang, Jiahao Guo, Wenyu Liu, Qian Zhang, Chang Huang
As the size of circuit designs continues to grow rapidly, artificial intelligenc
e technologies are being extensively used in Electronic Design Automation (EDA)
to assist with circuit design.Placement and routing are the most time-consuming
parts of the physical design process, and how to quickly evaluate the placement
has become a hot research topic. Prior works either transformed circuit designs
into images using hand-crafted methods and then used Convolutional Neural Networ
ks (CNN) to extract features, which are limited by the quality of the hand-craft
ed methods and could not achieve end-to-end training, or treated the circuit des
ign as a graph structure and used Graph Neural Networks (GNN) to extract feature
s, which require time-consuming preprocessing.In our work, we propose a novel pe
rspective for circuit design by treating circuit components as point clouds and

using Transformer-based point cloud perception methods to extract features from
the circuit. This approach enables direct feature extraction from raw data witho
ut any preprocessing, allows for end-to-end training, and results in high perfor
mance.Experimental results show that our method achieves state-of-the-art perfor
mance in congestion prediction tasks on both the CircuitNet and ISPD2015 dataset
s, as well as in design rule check (DRC) violation prediction tasks on the Circu
itNet dataset.Our method establishes a bridge between the relatively mature poin
t cloud perception methods and the fast-developing EDA algorithms, enabling us t
o leverage more collective intelligence to solve this task. To facilitate the re
search of open EDA design, source codes and pre-trained models are released at h
ttps://github.com/hustvl/circuitformer.
****
Causal Component Analysis
Liang Wendong, Armin Keki■, Julius von Kügelgen, Simon Buchholz, Michel Besserve
, Luigi Gresele, Bernhard Schölkopf
Independent Component Analysis (ICA) aims to recover independent latent variable
s from observed mixtures thereof. Causal Representation Learning (CRL) aims inst
ead to infer causally related (thus often statistically dependent) latent variab
les, together with the unknown graph encoding their causal relationships. We int
roduce an intermediate problem termed Causal Component Analysis (CauCA). CauCA c
an be viewed as a generalization of ICA, modelling the causal dependence among t
he latent components, and as a special case of CRL. In contrast to CRL, it presu
pposes knowledge of the causal graph, focusing solely on learning the unmixing f
unction and the causal mechanisms. Any impossibility results regarding the recov
ery of the ground truth in CauCA also apply for CRL, while possibility results m
ay serve as a stepping stone for extensions to CRL. We characterize CauCA identi
fiability from multiple datasets generated through different types of interventi
ons on the latent causal variables. As a corollary, this interventional perspect
ive also leads to new identifiability results for nonlinear ICA—a special case o
f CauCA with an empty graph—requiring strictly fewer datasets than previous resu
lts. We introduce a likelihood-based approach using normalizing flows to estimat
e both the unmixing function and the causal mechanisms, and demonstrate its effe
ctiveness through extensive synthetic experiments in the CauCA and ICA setting.
****
Latent Graph Inference with Limited Supervision
Jianglin Lu, Yi Xu, Huan Wang, Yue Bai, Yun Fu
Latent graph inference (LGI) aims to jointly learn the underlying graph structur
e and node representations from data features. However, existing LGI methods com
monly suffer from the issue of supervision starvation, where massive edge weight
s are learned without semantic supervision and do not contribute to the training
 loss. Consequently, these supervision-starved weights, which determine the pred
ictions of testing samples, cannot be semantically optimal, resulting in poor ge
neralization. In this paper, we observe that this issue is actually caused by th
e graph sparsification operation, which severely destroys the important connecti
ons established between pivotal nodes and labeled ones. To address this, we prop
ose to restore the corrupted affinities and replenish the missed supervision for
 better LGI. The key challenge then lies in identifying the critical nodes and r
ecovering the corrupted affinities. We begin by defining the pivotal nodes as k-
hop starved nodes, which can be identified based on a given adjacency matrix. Co
nsidering the high computational burden, we further present a more efficient alt
ernative inspired by CUR matrix decomposition. Subsequently, we eliminate the st
arved nodes by reconstructing the destroyed connections. Extensive experiments o
n representative benchmarks demonstrate that reducing the starved nodes consiste
ntly improves the performance of state-of-the-art LGI methods, especially under
extremely limited supervision (6.12% improvement on Pubmed with a labeling rate
of only 0.3%).
****
Precision-Recall Divergence Optimization for Generative Modeling with GANs and N
ormalizing Flows
Alexandre Verine, Benjamin Negrevergne, Muni Sreenivas Pydi, Yann Chevaleyre

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Energy Guided Diffusion for Generating Neurally Exciting Images
Pawel Pierzchlewicz, Konstantin Willeke, Arne Nix, Pavithra Elumalai, Kelli Restivo, Tori Shinn, Cate Nealley, Gabrielle Rodriguez, Saumil Patel, Katrin Franke, Andreas Tolias, Fabian Sinz
In recent years, most exciting inputs (MEIs) synthesized from encoding models of neuronal activity have become an established method for studying tuning properties of biological and artificial visual systems. However, as we move up the visual hierarchy, the complexity of neuronal computations increases. Consequently, it becomes more challenging to model neuronal activity, requiring more complex models. In this study, we introduce a novel readout architecture inspired by the mechanism of visual attention. This new architecture, which we call attention readout, together with a data-driven convolutional core outperforms previous task-driven models in predicting the activity of neurons in macaque area V4. However, as our predictive network becomes deeper and more complex, synthesizing MEIs via straightforward gradient ascent (GA) can struggle to produce qualitatively good results and overfit to idiosyncrasies of a more complex model, potentially decreasing the MEI's model-to-brain transferability. To solve this problem, we propose a diffusion-based method for generating MEIs via Energy Guidance (EGG). We show that for models of macaque V4, EGG generates single neuron MEIs that generalize better across varying model architectures than the state-of-the-art GA, while at the same time reducing computational costs by a factor of 4.7x, facilitating experimentally challenging closed-loop experiments. Furthermore, EGG diffusion can be used to generate other neurally exciting images, like most exciting naturalistic images that are on par with a selection of highly activating natural images, or image reconstructions that generalize better across architectures. Finally, EGG is simple to implement, requires no retraining of the diffusion model, and can easily be generalized to provide other characterizations of the visual system, such as invariances. Thus, EGG provides a general and flexible framework to study the coding properties of the visual system in the context of natural images.
****

An active learning framework for multi-group mean estimation
Abdellah Aznag, Rachel Cummings, Adam N. Elmachtoub
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

CAT-Walk: Inductive Hypergraph Learning via Set Walks
Ali Behrouz, Farnoosh Hashemi, Sadaf Sadeghian, Margo Seltzer
Temporal hypergraphs provide a powerful paradigm for modeling time-dependent, higher-order interactions in complex systems. Representation learning for hypergraphs is essential for extracting patterns of the higher-order interactions that are critically important in real-world problems in social network analysis, neuroscience, finance, etc. However, existing methods are typically designed only for specific tasks or static hypergraphs. We present CAT-Walk, an inductive method that learns the underlying dynamic laws that govern the temporal and structural processes underlying a temporal hypergraph. CAT-Walk introduces a temporal, higher-order walk on hypergraphs, SetWalk, that extracts higher-order causal patterns. CAT-Walk uses a novel adaptive and permutation invariant pooling strategy, SetMixer, along with a set-based anonymization process that hides the identity of hyperedges. Finally, we present a simple yet effective neural network model to encode hyperedges. Our evaluation on 10 hypergraph benchmark datasets shows that CAT-Walk attains outstanding performance on temporal hyperedge prediction benchmarks in both inductive and transductive settings. It also shows competitive per

formance with state-of-the-art methods for node classification. (https://github.com/ubc-systopia/CATWalk)
****
Unbiased constrained sampling with Self-Concordant Barrier Hamiltonian Monte Carlo
Maxence Noble, Valentin De Bortoli, Alain Durmus
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Directional diffusion models for graph representation learning
Run Yang, Yuling Yang, Fan Zhou, Qiang Sun
Diffusion models have achieved remarkable success in diverse domains such as image synthesis, super-resolution, and 3D molecule generation. Surprisingly, the application of diffusion models in graph learning has garnered little attention. In this paper, we aim to bridge this gap by exploring the use of diffusion models for unsupervised graph representation learning. Our investigation commences with the identification of anisotropic structures within graphs and the recognition of a crucial limitation in the vanilla forward diffusion process when dealing with these anisotropic structures. The original forward diffusion process continually adds isotropic Gaussian noise to the data, which may excessively dilute anisotropic signals, leading to rapid signal-to-noise conversion. This rapid conversion poses challenges for training denoising neural networks and obstructs the acquisition of semantically meaningful representations during the reverse process. To overcome this challenge, we introduce a novel class of models termed {\it directional diffusion models}. These models adopt data-dependent, anisotropic, and directional noises in the forward diffusion process. In order to assess the effectiveness of our proposed models, we conduct extensive experiments on 12 publicly available datasets, with a particular focus on two distinct graph representation learning tasks. The experimental results unequivocally establish the superiority of our models over state-of-the-art baselines, underscoring their effectiveness in capturing meaningful graph representations. Our research not only sheds light on the intricacies of the forward process in diffusion models but also underscores the vast potential of these models in addressing a wide spectrum of graph-related tasks. Our code is available at \url{https://github.com/statsle/DDM}.
****
UniTSFace: Unified Threshold Integrated Sample-to-Sample Loss for Face Recognition
qiufu li, Xi Jia, Jiancan Zhou, Linlin Shen, Jinming Duan
Sample-to-class-based face recognition models can not fully explore the cross-sample relationship among large amounts of facial images, while sample-to-sample-based models require sophisticated pairing processes for training. Furthermore, neither method satisfies the requirements of real-world face verification applications, which expect a unified threshold separating positive from negative facial pairs. In this paper, we propose a unified threshold integrated sample-to-sample based loss (USS loss), which features an explicit unified threshold for distinguishing positive from negative pairs. Inspired by our USS loss, we also derive the sample-to-sample based softmax and BCE losses, and discuss their relationship. Extensive evaluation on multiple benchmark datasets, including MFR, IJB-C, LFW, CFP-FP, AgeDB, and MegaFace, demonstrates that the proposed USS loss is highly efficient and can work seamlessly with sample-to-class-based losses. The embedded loss (USS and sample-to-class Softmax loss) overcomes the pitfalls of previous approaches and the trained facial model UniTSFace exhibits exceptional performance, outperforming state-of-the-art methods, such as CosFace, ArcFace, VPL, AnchorFace, and UNPG. Our code is available at https://github.com/CVI-SZU/UniTSFace.
****
Defending Pre-trained Language Models as Few-shot Learners against Backdoor Atta

cks
Zhaohan Xi, Tianyu Du, Changjiang Li, Ren Pang, Shouling Ji, Jinghui Chen, Fenglong Ma, Ting Wang
Pre-trained language models (PLMs) have demonstrated remarkable performance as few-shot learners. However, their security risks under such settings are largely unexplored. In this work, we conduct a pilot study showing that PLMs as few-shot learners are highly vulnerable to backdoor attacks while existing defenses are inadequate due to the unique challenges of few-shot scenarios. To address such challenges, we advocate MDP, a novel lightweight, pluggable, and effective defense for PLMs as few-shot learners. Specifically, MDP leverages the gap between the masking-sensitivity of poisoned and clean samples: with reference to the limited few-shot data as distributional anchors, it compares the representations of given samples under varying masking and identifies poisoned samples as ones with significant variations. We show analytically that MDP creates an interesting dilemma for the attacker to choose between attack effectiveness and detection evasiveness. The empirical evaluation using benchmark datasets and representative attacks validates the efficacy of MDP. The code of MDP is publicly available.
****
On the Power of SVD in the Stochastic Block Model
Xinyu Mao, Jiapeng Zhang
A popular heuristic method for improving clustering results is to apply dimensionality reduction before running clustering algorithms.It has been observed that spectral-based dimensionality reduction tools, such as PCA or SVD, improve the performance of clustering algorithms in many applications. This phenomenon indicates that spectral method not only serves as a dimensionality reduction tool, but also contributes to the clustering procedure in some sense. It is an interesting question to understand the behavior of spectral steps in clustering problems.As an initial step in this direction, this paper studies the power of vanilla-SVD algorithm in the stochastic block model (SBM). We show that, in the symmetric setting, vanilla-SVD algorithm recovers all clusters correctly. This result answers an open question posed by Van Vu (Combinatorics Probability and Computing, 2018) in the symmetric setting.
****
Continuous-time Analysis of Anchor Acceleration
Jaewook Suh, Jisun Park, Ernest Ryu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
BEDD: The MineRL BASALT Evaluation and Demonstrations Dataset for Training and Benchmarking Agents that Solve Fuzzy Tasks
Stephanie Milani, Anssi Kanervisto, Karolis Ramanauskas, Sander Schulhoff, Brandon Houghton, Rohin Shah
The MineRL BASALT competition has served to catalyze advances in learning from human feedback through four hard-to-specify tasks in Minecraft, such as create and photograph a waterfall. Given the completion of two years of BASALT competitions, we offer to the community a formalized benchmark through the BASALT Evaluation and Demonstrations Dataset (BEDD), which serves as a resource for algorithm development and performance assessment. BEDD consists of a collection of 26 million image-action pairs from nearly 14,000 videos of human players completing the BASALT tasks in Minecraft. It also includes over 3,000 dense pairwise human evaluations of human and algorithmic agents. These comparisons serve as a fixed, preliminary leaderboard for evaluating newly-developed algorithms. To enable this comparison, we present a streamlined codebase for benchmarking new algorithms against the leaderboard. In addition to presenting these datasets, we conduct a detailed analysis of the data from both datasets to guide algorithm development and evaluation. The released code and data are available at https://github.com/minerllabs/basalt-benchmark.
****

Self-supervised Object-Centric Learning for Videos
Görkay Aydemir, Weidi Xie, Fatma Guney
Unsupervised multi-object segmentation has shown impressive results on images by utilizing powerful semantics learned from self-supervised pretraining. An additional modality such as depth or motion is often used to facilitate the segmentation in video sequences. However, the performance improvements observed in synthetic sequences, which rely on the robustness of an additional cue, do not translate to more challenging real-world scenarios. In this paper, we propose the first fully unsupervised method for segmenting multiple objects in real-world sequences. Our object-centric learning framework spatially binds objects to slots on each frame and then relates these slots across frames. From these temporally-aware slots, the training objective is to reconstruct the middle frame in a high-level semantic feature space. We propose a masking strategy by dropping a significant portion of tokens in the feature space for efficiency and regularization. Additionally, we address over-clustering by merging slots based on similarity. Our method can successfully segment multiple instances of complex and high-variety classes in YouTube videos.
****
Improving Adversarial Transferability via Intermediate-level Perturbation Decay
Qizhang Li, Yiwen Guo, Wangmeng Zuo, Hao Chen
Intermediate-level attacks that attempt to perturb feature representations following an adversarial direction drastically have shown favorable performance in crafting transferable adversarial examples. Existing methods in this category are normally formulated with two separate stages, where a directional guide is required to be determined at first and the scalar projection of the intermediate-level perturbation onto the directional guide is enlarged thereafter. The obtained perturbation deviates from the guide inevitably in the feature space, and it is revealed in this paper that such a deviation may lead to sub-optimal attack. To address this issue, we develop a novel intermediate-level method that crafts adversarial examples within a single stage of optimization. In particular, the proposed method, named intermediate-level perturbation decay (ILPD), encourages the intermediate-level perturbation to be in an effective adversarial direction and to possess a great magnitude simultaneously. In-depth discussion verifies the effectiveness of our method. Experimental results show that it outperforms state-of-the-arts by large margins in attacking various victim models on ImageNet (+10.07% on average) and CIFAR-10 (+3.88% on average). Our code is at https://github.com/qizhangli/ILPD-attack.
****
GUST: Combinatorial Generalization by Unsupervised Grouping with Neuronal Coherence
Hao Zheng, Hui Lin, Rong Zhao
Dynamically grouping sensory information into structured entities is essential for understanding the world of combinatorial nature. However, the grouping ability and therefore combinatorial generalization are still challenging artificial neural networks. Inspired by the evidence that successful grouping is indicated by neuronal coherence in the human brain, we introduce GUST (Grouping Unsupervisely by Spike Timing network), an iterative network architecture with biological constraints to bias the network towards a dynamical state of neuronal coherence that softly reflects the grouping information in the temporal structure of its spiking activity. We evaluate and analyze the model on synthetic datasets. Interestingly, the segregation ability is directly learned from superimposed stimuli with a succinct unsupervised objective. Two learning stages are present, from coarsely perceiving global features to additionally capturing local features. Further, the learned symbol-like building blocks can be systematically composed to represent novel scenes in a bio-plausible manner.
****
State Regularized Policy Optimization on Data with Dynamics Shift
Zhenghai Xue, Qingpeng Cai, Shuchang Liu, Dong Zheng, Peng Jiang, Kun Gai, Bo An
In many real-world scenarios, Reinforcement Learning (RL) algorithms are trained on data with dynamics shift, i.e., with different underlying environment dynami

cs. A majority of current methods address such issue by training context encoders to identify environment parameters. Data with dynamics shift are separated according to their environment parameters to train the corresponding policy.However, these methods can be sample inefficient as data are used \textit{ad hoc}, and policies trained for one dynamics cannot benefit from data collected in all other environments with different dynamics. In this paper, we find that in many environments with similar structures and different dynamics, optimal policies have similar stationary state distributions. We exploit such property and learn the stationary state distribution from data with dynamics shift for efficient data reuse. Such distribution is used to regularize the policy trained in a new environment, leading to the SRPO (\textbf{S}tate \textbf{R}egularized \textbf{P}olicy \textbf{O}ptimization) algorithm. To conduct theoretical analyses, the intuition of similar environment structures is characterized by the notion of homomorphous MDPs. We then demonstrate a lower-bound performance guarantee on policies regularized by the stationary state distribution. In practice, SRPO can be an add-on module to context-based algorithms in both online and offline RL settings.Experimental results show that SRPO can make several context-based algorithms far more data efficient and significantly improve their overall performance.
****
Distilling Out-of-Distribution Robustness from Vision-Language Foundation Models
Andy Zhou, Jindong Wang, Yu-Xiong Wang, Haohan Wang
We propose a conceptually simple and lightweight framework for improving the robustness of vision models through the combination of knowledge distillation and data augmentation. We address the conjecture that larger models do not make for better teachers by showing strong gains in out-of-distribution robustness when distilling from pretrained foundation models. Following this finding, we propose Discrete Adversarial Distillation (DAD), which leverages a robust teacher to generate adversarial examples and a VQGAN to discretize them, creating more informative samples than standard data augmentation techniques. We provide a theoretical framework for the use of a robust teacher in the knowledge distillation with data augmentation setting and demonstrate strong gains in out-of-distribution robustness and clean accuracy across different student architectures. Notably, our method adds minor computational overhead compared to similar techniques and can be easily combined with other data augmentations for further improvements.
****
XES3G5M: A Knowledge Tracing Benchmark Dataset with Auxiliary Information
Zitao Liu, Qiongqiong Liu, Teng Guo, Jiahao Chen, Shuyan Huang, Xiangyu Zhao, Jiliang Tang, Weiqi Luo, Jian Weng
Knowledge tracing (KT) is a task that predicts students' future performance based on their historical learning interactions. With the rapid development of deep learning techniques, existing KT approaches follow a data-driven paradigm that uses massive problem-solving records to model students' learning processes. However, although the educational contexts contain various factors that may have an influence on student learning outcomes, existing public KT datasets mainly consist of anonymized ID-like features, which may hinder the research advances towards this field. Therefore, in this work, we present, \emph{XES3G5M}, a large-scale dataset with rich auxiliary information about questions and their associated knowledge components (KCs)\footnote{\label{ft:kc}A KC is a generalization of everyday terms like concept, principle, fact, or skill.}. The XES3G5M dataset is collected from a real-world online math learning platform, which contains 7,652 questions, and 865 KCs with 5,549,635 interactions from 18,066 students. To the best of our knowledge, the XES3G5M dataset not only has the largest number of KCs in math domain but contains the richest contextual information including tree structured KC relations, question types, textual contents and analysis and student response timestamps. Furthermore, we build a comprehensive benchmark on 19 state-of-the-art deep learning based knowledge tracing (DLKT) models. Extensive experiments demonstrate the effectiveness of leveraging the auxiliary information in our XES3G5M with DLKT models. We hope the proposed dataset can effectively facilitate the KT research work.
****

Factorized Contrastive Learning: Going Beyond Multi-view Redundancy
Paul Pu Liang, Zihao Deng, Martin Q. Ma, James Y. Zou, Louis-Philippe Morency, Ruslan Salakhutdinov
In a wide range of multimodal tasks, contrastive learning has become a particularly appealing approach since it can successfully learn representations from abundant unlabeled data with only pairing information (e.g., image-caption or video-audio pairs). Underpinning these approaches is the assumption of multi-view redundancy - that shared information between modalities is necessary and sufficient for downstream tasks. However, in many real-world settings, task-relevant information is also contained in modality-unique regions: information that is only present in one modality but still relevant to the task. How can we learn self-supervised multimodal representations to capture both shared and unique information relevant to downstream tasks? This paper proposes FactorCL, a new multimodal representation learning method to go beyond multi-view redundancy. FactorCL is built from three new contributions: (1) factorizing task-relevant information into shared and unique representations, (2) capturing task-relevant information via maximizing MI lower bounds and removing task-irrelevant information via minimizing MI upper bounds, and (3) multimodal data augmentations to approximate task relevance without labels. On large-scale real-world datasets, FactorCL captures both shared and unique information and achieves state-of-the-art results on six benchmarks.
****
Semantic Image Synthesis with Unconditional Generator
JungWoo Chae, Hyunin Cho, Sooyeon Go, Kyungmook Choi, Youngjung Uh
Semantic image synthesis (SIS) aims to generate realistic images according to semantic masks given by a user. Although recent methods produce high quality results with fine spatial control, SIS requires expensive pixel-level annotation of the training images. On the other hand, manipulating intermediate feature maps in a pretrained unconditional generator such as StyleGAN supports coarse spatial control without heavy annotation. In this paper, we introduce a new approach, for reflecting user's detailed guiding masks on a pretrained unconditional generator. Our method converts a user's guiding mask to a proxy mask through a semantic mapper. Then the proxy mask conditions the resulting image through a rearranging network based on cross-attention mechanism. The proxy mask is simple clustering of intermediate feature maps in the generator. The semantic mapper and the rearranging network are easy to train (less than half an hour). Our method is useful for many tasks: semantic image synthesis, spatially editing real images, and unaligned local transplantation. Last but not least, it is generally applicable to various datasets such as human faces, animal faces, and churches.
****
Learning Neural Implicit through Volume Rendering with Attentive Depth Fusion Priors
Pengchong Hu, Zhizhong Han
Learning neural implicit representations has achieved remarkable performance in 3D reconstruction from multi-view images. Current methods use volume rendering to render implicit representations into either RGB or depth images that are supervised by the multi-view ground truth. However, rendering a view each time suffers from incomplete depth at holes and unawareness of occluded structures from the depth supervision, which severely affects the accuracy of geometry inference via a volume rendering. To resolve this issue, we propose to learn neural implicit representations from multi-view RGBD images through volume rendering with an attentive depth fusion prior. Our prior allows neural networks to sense coarse 3D structures from the Truncated Signed Distance Function (TSDF) fused from all available depth images for rendering. The TSDF enables accessing the missing depth at holes on one depth image and the occluded parts that are invisible from the current view. By introducing a novel attention mechanism, we allow neural networks to directly use the depth fusion prior with the inferred occupancy as the learned implicit function. Our attention mechanism works with either a one-time fused TSDF that represents a whole scene or an incrementally fused TSDF that represents a partial scene in the context of Simultaneous Localization and Mapping (SLAM)

. Our evaluations on widely used benchmarks including synthetic and real-world scans show our superiority over the latest neural implicit methods.
****
SimFBO: Towards Simple, Flexible and Communication-efficient Federated Bilevel Learning
Yifan Yang, Peiyao Xiao, Kaiyi Ji
Federated bilevel optimization (FBO) has shown great potential recently in machine learning and edge computing due to the emerging nested optimization structure in meta-learning, fine-tuning, hyperparameter tuning, etc. However, existing FBO algorithms often involve complicated computations and require multiple sub-loops per iteration, each of which contains a number of communication rounds. In this paper, we propose a simple and flexible FBO framework named SimFBO, which is easy to implement without sub-loops, and includes a generalized server-side aggregation and update for improving communication efficiency. We further propose System-level heterogeneity robust FBO (ShroFBO) as a variant of SimFBO with stronger resilience to heterogeneous local computation. We show that SimFBO and ShroFBO provably achieve a linear convergence speedup with partial client participation and client sampling without replacement, as well as improved sample and communication complexities. Experiments demonstrate the effectiveness of the proposed methods over existing FBO algorithms.
****
PICProp: Physics-Informed Confidence Propagation for Uncertainty Quantification
Qianli Shen, Wai Hoh Tang, Zhun Deng, Apostolos Psaros, Kenji Kawaguchi
Standard approaches for uncertainty quantification in deep learning and physics-informed learning have persistent limitations. Indicatively, strong assumptions regarding the data likelihood are required, the performance highly depends on the selection of priors, and the posterior can be sampled only approximately, which leads to poor approximations because of the associated computational cost.This paper introduces and studies confidence interval (CI) estimation for deterministic partial differential equations as a novel problem.That is, to propagate confidence, in the form of CIs, from data locations to the entire domain with probabilistic guarantees.We propose a method, termed Physics-Informed Confidence Propagation (PICProp), based on bi-level optimization to compute a valid CI without making heavy assumptions.We provide a theorem regarding the validity of our method, and computational experiments, where the focus is on physics-informed learning. Code is available at https://github.com/ShenQianli/PICProp.
****
Foundation Model is Efficient Multimodal Multitask Model Selector
Fanqing Meng, Wenqi Shao, zhanglin peng, Chonghe Jiang, Kaipeng Zhang, Yu Qiao, Ping Luo
This paper investigates an under-explored but important problem: given a collection of pre-trained neural networks, predicting their performance on each multi-modal task without fine-tuning them, such as image recognition, referring, captioning, visual question answering, and text question answering.A brute-force approach is to finetune all models on all target datasets, bringing high computational costs. Although recent-advanced approaches employed lightweight metrics to measure models’ transferability, they often depend heavily on the prior knowledge of a single task, making them inapplicable in a multi-modal multi-task scenario. To tackle this issue, we propose an efficient multi-task model selector (EMMS), which employs large-scale foundation models to transform diverse label formats such as categories, texts, and bounding boxes of different downstream tasks into a unified noisy label embedding. EMMS can estimate a model’s transferability through a simple weighted linear regression, which can be efficiently solved by an alternating minimization algorithm with a convergence guarantee. Extensive experiments on 5 downstream tasks with 24 datasets show that EMMS is fast, effective, and generic enough to assess the transferability of pre-trained models, making it the first model selection method in the multi-task scenario. For instance, compared with the state- of-the-art method LogME enhanced by our label embeddings, EMMS achieves 9.0%, 26.3%, 20.1%, 54.8%, 12.2% performance gain on image recognition, referring, captioning, visual question answering, and text question answe

ring, while bringing 5.13×, 6.29×, 3.59×, 6.19×, and 5.66× speedup in wall-clock time, respectively. The code is available at https://github.com/OpenGVLab/Multitask-Model-Selector.
****
Feature Likelihood Divergence: Evaluating the Generalization of Generative Models Using Samples
Marco Jiralerspong, Joey Bose, Ian Gemp, Chongli Qin, Yoram Bachrach, Gauthier Gidel
The past few years have seen impressive progress in the development of deep generative models capable of producing high-dimensional, complex, and photo-realistic data. However, current methods for evaluating such models remain incomplete: standard likelihood-based metrics do not always apply and rarely correlate with perceptual fidelity, while sample-based metrics, such as FID, are insensitive to overfitting, i.e., inability to generalize beyond the training set. To address these limitations, we propose a new metric called the Feature Likelihood Divergence (FLD), a parametric sample-based score that uses density estimation to provide a comprehensive trichotomic evaluation accounting for novelty (i.e., different from the training samples), fidelity, and diversity of generated samples. We empirically demonstrate the ability of FLD to identify specific overfitting problem cases, where previously proposed metrics fail. We also extensively evaluate FLD on various image datasets and model classes, demonstrating its ability to match intuitions of previous metrics like FID while offering a more comprehensive evaluation of generative models.
****
LOVM: Language-Only Vision Model Selection
Orr Zohar, Shih-Cheng Huang, Kuan-Chieh Wang, Serena Yeung
Pre-trained multi-modal vision-language models (VLMs) are becoming increasingly popular due to their exceptional performance on downstream vision applications, particularly in the few- and zero-shot settings. However, selecting the best-performing VLM for some downstream applications is non-trivial, as it is dataset and task-dependent. Meanwhile, the exhaustive evaluation of all available VLMs on a novel application is not only time and computationally demanding but also necessitates the collection of a labeled dataset for evaluation. As the number of open-source VLM variants increases, there is a need for an efficient model selection strategy that does not require access to a curated evaluation dataset. This paper proposes a novel task and benchmark for efficiently evaluating VLMs' zero-shot performance on downstream applications without access to the downstream task dataset. Specifically, we introduce a new task LOVM: Language-Only Vision Model Selection , where methods are expected to perform both model selection and performance prediction based solely on a text description of the desired downstream application. We then introduced an extensive LOVM benchmark consisting of ground-truth evaluations of 35 pre-trained VLMs and 23 datasets, where methods are expected to rank the pre-trained VLMs and predict their zero-shot performance.
****
Statistical Analysis of Quantum State Learning Process in Quantum Neural Networks
Hao-Kai Zhang, Chenghong Zhu, Mingrui Jing, Xin Wang
Quantum neural networks (QNNs) have been a promising framework in pursuing near-term quantum advantage in various fields, where many applications can be viewed as learning a quantum state that encodes useful data. As a quantum analog of probability distribution learning, quantum state learning is theoretically and practically essential in quantum machine learning. In this paper, we develop a no-go theorem for learning an unknown quantum state with QNNs even starting from a high-fidelity initial state. We prove that when the loss value is lower than a critical threshold, the probability of avoiding local minima vanishes exponentially with the qubit count, while only grows polynomially with the circuit depth. The curvature of local minima is concentrated to the quantum Fisher information times a loss-dependent constant, which characterizes the sensibility of the output state with respect to parameters in QNNs. These results hold for any circuit structures, initialization strategies, and work for both fixed ansatzes and adaptiv

e methods. Extensive numerical simulations are performed to validate our theoretical results. Our findings place generic limits on good initial guesses and adaptive methods for improving the learnability and scalability of QNNs, and deepen the understanding of prior information's role in QNNs.
****
SOL: Sampling-based Optimal Linear bounding of arbitrary scalar functions
Yuriy Biktairov, Jyotirmoy Deshmukh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Opening the Vocabulary of Egocentric Actions
Dibyadip Chatterjee, Fadime Sener, Shugao Ma, Angela Yao
Human actions in egocentric videos often feature hand-object interactions composed of a verb (performed by the hand) applied to an object. Despite their extensive scaling up, egocentric datasets still face two limitations — sparsity of action compositions and a closed set of interacting objects. This paper proposes a novel open vocabulary action recognition task. Given a set of verbs and objects observed during training, the goal is to generalize the verbs to an open vocabulary of actions with seen and novel objects. To this end, we decouple the verb and object predictions via an object-agnostic verb encoder and a prompt-based object encoder. The prompting leverages CLIP representations to predict an open vocabulary of interacting objects. We create open vocabulary benchmarks on the EPIC-KITCHENS-100 and Assembly101 datasets; whereas closed-action methods fail to generalize, our proposed method is effective. In addition, our object encoder significantly outperforms existing open-vocabulary visual recognition methods in recognizing novel interacting objects.
****
On the Pareto Front of Multilingual Neural Machine Translation
Liang Chen, Shuming Ma, Dongdong Zhang, Furu Wei, Baobao Chang
In this work, we study how the performance of a given direction changes with its sampling ratio in Multilingual Neural Machine Translation (MNMT). By training over 200 multilingual models with various model sizes, data sizes, and language directions, we find it interesting that the performance of certain translation direction does not always improve with the increase of its weight in the multi-task optimization objective. Accordingly, scalarization method leads to a multitask trade-off front that deviates from the traditional Pareto front when there exists data imbalance in the training corpus, which poses a great challenge to improve the overall performance of all directions. Based on our observations, we propose the Double Power Law to predict the unique performance trade-off front in MNMT, which is robust across various languages, data adequacy, and the number of tasks. Finally, we formulate the sample ratio selection problem in MNMT as an optimization problem based on the Double Power Law. Extensive experiments show that it achieves better performance than temperature searching and gradient manipulation methods with only 1/5 to 1/2 of the total training budget. We release the code at https://github.com/pkunlp-icler/ParetoMNMT for reproduction.
****
Hierarchically Gated Recurrent Neural Network for Sequence Modeling
Zhen Qin, Songlin Yang, Yiran Zhong
Transformers have surpassed RNNs in popularity due to their superior abilities in parallel training and long-term dependency modeling.Recently, there has been a renewed interest in using linear RNNs for efficient sequence modeling.These linear RNNs often employ gating mechanisms in the output of the linear recurrence layer while ignoring the significance of using forget gates within the recurrence. In this paper, we propose a gated linear RNN model dubbed Hierarchically Gated Recurrent Neural Network (HGRN), which includes forget gates that are lower bounded by a learnable value. The lower bound increases monotonically when moving up layers. This allows the upper layers to model long-term dependencies and the lower layers to model more local, short-term dependencies. Experiments on languag

e modeling, image classification, and long-range arena benchmarks showcase the efficiency and effectiveness of our proposed model. The source code is available at https://github.com/OpenNLPLab/HGRN.
****

Why Did This Model Forecast This Future? Information-Theoretic Saliency for Counterfactual Explanations of Probabilistic Regression Models

Chirag Raman, Alec Nonnemaker, Amelia Villegas-Morcillo, Hayley Hung, Marco Loog

We propose a post hoc saliency-based explanation framework for counterfactual reasoning in probabilistic multivariate time-series forecasting (regression) settings. Building upon Miller's framework of explanations derived from research in multiple social science disciplines, we establish a conceptual link between counterfactual reasoning and saliency-based explanation techniques. To address the lack of a principled notion of saliency, we leverage a unifying definition of information-theoretic saliency grounded in preattentive human visual cognition and extend it to forecasting settings. Specifically, we obtain a closed-form expression for commonly used density functions to identify which observed timesteps appear salient to an underlying model in making its probabilistic forecasts. We empirically validate our framework in a principled manner using synthetic data to establish ground-truth saliency that is unavailable for real-world data. Finally, using real-world data and forecasting models, we demonstrate how our framework can assist domain experts in forming new data-driven hypotheses about the causal relationships between features in the wild.
****

Category-Extensible Out-of-Distribution Detection via Hierarchical Context Descriptions

Kai Liu, Zhihang Fu, Chao Chen, Sheng Jin, Ze Chen, Mingyuan Tao, Rongxin Jiang, Jieping Ye

The key to OOD detection has two aspects: generalized feature representation and precise category description. Recently, vision-language models such as CLIP provide significant advances in both two issues, but constructing precise category descriptions is still in its infancy due to the absence of unseen categories. This work introduces two hierarchical contexts, namely perceptual context and spurious context, to carefully describe the precise category boundary through automatic prompt tuning. Specifically, perceptual contexts perceive the inter-category difference (e.g., cats vs apples) for current classification tasks, while spurious contexts further identify spurious (similar but exactly not) OOD samples for every single category (e.g., cats vs panthers, apples vs peaches). The two contexts hierarchically construct the precise description for a certain category, which is, first roughly classifying a sample to the predicted category and then delicately identifying whether it is truly an ID sample or actually OOD. Moreover, the precise descriptions for those categories within the vision-language framework present a novel application: CATegory-EXtensible OOD detection (CATEX). One can efficiently extend the set of recognizable categories by simply merging the hierarchical contexts learned under different sub-task settings. And extensive experiments are conducted to demonstrate CATEX’s effectiveness, robustness, and category-extensibility. For instance, CATEX consistently surpasses the rivals by a large margin with several protocols on the challenging ImageNet-1K dataset. In addition, we offer new insights on how to efficiently scale up the prompt engineering in vision-language models to recognize thousands of object categories, as well as how to incorporate large language models (like GPT-3) to boost zero-shot applications.
****

Online Corrupted User Detection and Regret Minimization

Zhiyong Wang, Jize Xie, Tong Yu, Shuai Li, John C.S. Lui

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Nash Regret Guarantees for Linear Bandits

Ayush Sawarni, Soumyabrata Pal, Siddharth Barman
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ShiftAddViT: Mixture of Multiplication Primitives Towards Efficient Vision Transformer
Haoran You, Huihong Shi, Yipin Guo, Yingyan Lin
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Optimal Extragradient-Based Algorithms for Stochastic Variational Inequalities with Separable Structure
Angela Yuan, Chris Junchi Li, Gauthier Gidel, Michael Jordan, Quanquan Gu, Simon S. Du
We consider the problem of solving stochastic monotone variational inequalities with a separable structure using a stochastic first-order oracle. Building on standard extragradient for variational inequalities we propose a novel algorithm---stochastic \emph{accelerated gradient-extragradient} (AG-EG)---for strongly monotone variational inequalities (VIs). Our approach combines the strengths of extragradient and Nesterov acceleration. By showing that its iterates remain in a bounded domain and applying scheduled restarting, we prove that AG-EG has an optimal convergence rate for strongly monotone VIs. Furthermore, when specializing to the particular case of bilinearly coupled strongly-convex-strongly-concave saddle-point problems, including bilinear games, our algorithm achieves fine-grained convergence rates that match the respective lower bounds, with the stochasticity being characterized by an additive statistical error term that is optimal up to a constant prefactor.
****
Combinatorial Group Testing with Selfish Agents
Georgios Chionas, Dariusz Kowalski, Piotr Krysta
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Hierarchical Spatial Transformer for Massive Point Samples in Continuous Space
Wenchong He, Zhe Jiang, Tingsong Xiao, Zelin Xu, Shigang Chen, Ronald Fick, MILES MEDINA, Christine Angelini
Transformers are widely used deep learning architectures. Existing transformers are mostly designed for sequences (texts or time series), images or videos, and graphs. This paper proposes a novel transformer model for massive (up to a million) point samples in continuous space. Such data are ubiquitous in environment sciences (e.g., sensor observations), numerical simulations (e.g., particle-laden flow, astrophysics), and location-based services (e.g., POIs and trajectories). However, designing a transformer for massive spatial points is non-trivial due to several challenges, including implicit long-range and multi-scale dependency on irregular points in continuous space, a non-uniform point distribution, the potential high computational costs of calculating all-pair attention across massive points, and the risks of over-confident predictions due to varying point density. To address these challenges, we propose a new hierarchical spatial transformer model, which includes multi-resolution representation learning within a quad-tree hierarchy and efficient spatial attention via coarse approximation. We also design an uncertainty quantification branch to estimate prediction confidence related to input feature noise and point sparsity. We provide a theoretical analysis of computational time complexity and memory costs. Extensive experiments on

both real-world and synthetic datasets show that our method outperforms multiple baselines in prediction accuracy and our model can scale up to one million points on one NVIDIA A100 GPU. The code is available at https://github.com/spatialdatasciencegroup/HST
****

Rethinking Gauss-Newton for learning over-parameterized models
Michael Arbel, Romain Menegaux, Pierre Wolinski
This work studies the global convergence and implicit bias of Gauss Newton's (GN) when optimizing over-parameterized one-hidden layer networks in the mean-field regime. We first establish a global convergence result for GN in the continuous-time limit exhibiting a faster convergence rate compared to GD due to improved conditioning. We then perform an empirical study on a synthetic regression task to investigate the implicit bias of GN's method.While GN is consistently faster than GD in finding a global optimum, the learned model generalizes well on test data when starting from random initial weights with a small variance and using a small step size to slow down convergence. Specifically, our study shows that such a setting results in a hidden learning phenomenon, where the dynamics are able to recover features with good generalization properties despite the model having sub-optimal training and test performances due to an under-optimized linear layer. This study exhibits a trade-off between the convergence speed of GN and the generalization ability of the learned solution.
****

Knowledge Distillation for High Dimensional Search Index
Zepu Lu, Jin Chen, Defu Lian, ZAIXI ZHANG, Yong Ge, Enhong Chen
Lightweight compressed models are prevalent in Approximate Nearest Neighbor Search (ANNS) and Maximum Inner Product Search (MIPS) owing to their superiority of retrieval efficiency in large-scale datasets. However, results given by compressed methods are less accurate due to the curse of dimension and the limitations of optimization objectives (e.g., lacking interactions between queries and documents). Thus, we are encouraged to design a new learning algorithm for the compressed search index on high dimensions to improve retrieval performance. In this paper, we propose a novel KnowledgeDistillation for high dimensional search index framework (KDindex), with the aim of efficiently learning lightweight indexes by distilling knowledge from high-precision ANNS and MIPS models such as graph-based indexes. Specifically, the student is guided to keep the same ranking order of the top-k relevant results yielded by the teacher model, which acts as the additional supervision signals between queries and documents to learn the similarities between documents. Furthermore, to avoid the trivial solutions that all candidates are partitioned to the same centroid, the reconstruction loss that minimizes the compressed error, and the posting list balance strategy that equally allocates the candidates, are integrated into the learning objective. Experiment results demonstrate that KDindex outperforms existing learnable quantization-based indexes and is 40× lighter than the state-of-the-art non-exhaustive methods while achieving comparable recall quality.
****

Exploring the Optimal Choice for Generative Processes in Diffusion Models: Ordinary vs Stochastic Differential Equations
Yu Cao, Jingrun Chen, Yixin Luo, Xiang ZHOU
The diffusion model has shown remarkable success in computer vision, but it remains unclear whether the ODE-based probability flow or the SDE-based diffusion model is more superior and under what circumstances. Comparing the two is challenging due to dependencies on data distributions, score training, and other numerical issues. In this paper, we study the problem mathematically for two limiting scenarios: the zero diffusion (ODE) case and the large diffusion case. We first introduce a pulse-shape error to perturb the score function and analyze error accumulation of sampling quality, followed by a thorough analysis for generalization to arbitrary error. Our findings indicate that when the perturbation occurs at the end of the generative process, the ODE model outperforms the SDE model with a large diffusion coefficient. However, when the perturbation occurs earlier, the SDE model outperforms the ODE model, and we demonstrate that the error of sam

ple generation due to such a pulse-shape perturbation is exponentially suppresse d as the diffusion term's magnitude increases to infinity. Numerical validation of this phenomenon is provided using Gaussian, Gaussian mixture, and Swiss roll distribution, as well as realistic datasets like MNIST and CIFAR-10.
****
PIXIU: A Comprehensive Benchmark, Instruction Dataset and Large Language Model f or Finance
Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-L ira, Jimin Huang
Although large language models (LLMs) have shown great performance in natural la nguage processing (NLP) in the financial domain, there are no publicly available financially tailored LLMs, instruction tuning datasets, and evaluation benchmar ks, which is critical for continually pushing forward the open-source developmen t of financial artificial intelligence (AI). This paper introduces PIXIU, a comp rehensive framework including the first financial LLM based on fine-tuning LLaMA with instruction data, the first instruction data with 128K data samples to sup port the fine-tuning, and an evaluation benchmark with 8 tasks and 15 datasets. We first construct the large-scale multi-task instruction data considering a var iety of financial tasks, financial document types, and financial data modalities . We then propose a financial LLM called FinMA by fine-tuning LLaMA with the con structed dataset to be able to follow instructions for various financial tasks. To support the evaluation of financial LLMs, we propose a standardized benchmark that covers a set of critical financial tasks, including six financial NLP task s and two financial prediction tasks. With this benchmark, we conduct a detailed analysis of FinMA and several existing LLMs, uncovering their strengths and wea knesses in handling critical financial tasks. The model, datasets, benchmark, an d experimental results are open-sourced to facilitate future research in financi al AI.
****
EgoDistill: Egocentric Head Motion Distillation for Efficient Video Understandin g
Shuhan Tan, Tushar Nagarajan, Kristen Grauman
Recent advances in egocentric video understanding models are promising, but thei r heavy computational expense is a barrier for many real-world applications. To address this challenge, we propose EgoDistill, a distillation-based approach tha t learns to reconstruct heavy ego-centric video clip features by combining the s emantics from a sparse set of video frames with head motion from lightweight IMU readings. We further devise a novel IMU-based self-supervised pretraining strat egy. Our method leads to significant improvements in efficiency, requiring 200× fewer GFLOPs than equivalent video models. We demonstrate its effectiveness on t he Ego4D and EPIC- Kitchens datasets, where our method outperforms state-of-the-art efficient video understanding methods.
****
Does Continual Learning Meet Compositionality? New Benchmarks and An Evaluation Framework
Weiduo Liao, Ying Wei, Mingchen Jiang, Qingfu Zhang, Hisao Ishibuchi
Compositionality facilitates the comprehension of novel objects using acquired c oncepts and the maintenance of a knowledge pool. This is particularly crucial fo r continual learners to prevent catastrophic forgetting and enable compositional ly forward transfer of knowledge. However, the existing state-of-the-art benchma rks inadequately evaluate the capability of compositional generalization, leavin g an intriguing question unanswered. To comprehensively assess this capability, we introduce two vision benchmarks, namely Compositional GQA (CGQA) and Composit ional OBJects365 (COBJ), along with a novel evaluation framework called Composit ional Few-Shot Testing (CFST). These benchmarks evaluate the systematicity, prod uctivity, and substitutivity aspects of compositional generalization. Experiment al results on five baselines and two modularity-based methods demonstrate that c urrent continual learning techniques do exhibit somewhat favorable compositional ity in their learned feature extractors. Nonetheless, further efforts are requir ed in developing modularity-based approaches to enhance compositional generaliza

tion. We anticipate that our proposed benchmarks and evaluation protocol will fo
ster research on continual learning and compositionality.
****
Unsupervised Protein-Ligand Binding Energy Prediction via Neural Euler's Rotatio
n Equation
Wengong Jin, Siranush Sarkizova, Xun Chen, Nir HaCohen, Caroline Uhler
Protein-ligand binding prediction is a fundamental problem in AI-driven drug dis
covery. Previous work focused on supervised learning methods for small molecules
 where binding affinity data is abundant, but it is hard to apply the same strat
egy to other ligand classes like antibodies where labelled data is limited. In t
his paper, we explore unsupervised approaches and reformulate binding energy pre
diction as a generative modeling task. Specifically, we train an energy-based mo
del on a set of unlabelled protein-ligand complexes using SE(3) denoising score
matching (DSM) and interpret its log-likelihood as binding affinity. Our key con
tribution is a new equivariant rotation prediction network called Neural Euler's
 Rotation Equations (NERE) for SE(3) DSM. It predicts a rotation by modeling the
 force and torque between protein and ligand atoms, where the force is defined a
s the gradient of an energy function with respect to atom coordinates. Using two
 protein-ligand and antibody-antigen binding affinity prediction benchmarks, we
show that NERE outperforms all unsupervised baselines (physics-based potentials
and protein language models) in both cases and surpasses supervised baselines in
 the antibody case.
****
ProteinNPT: Improving Protein Property Prediction and Design with Non-Parametric
 Transformers
Pascal Notin, Ruben Weitzman, Debora Marks, Yarin Gal
Protein design holds immense potential for optimizing naturally occurring protei
ns, with broad applications in drug discovery, material design, and sustainabili
ty. However, computational methods for protein engineering are confronted with
significant challenges, such as an expansive design space, sparse functional reg
ions, and a scarcity of available labels. These issues are further exacerbated i
n practice by the fact most real-life design scenarios necessitate the simultane
ous optimization of multiple properties. In this work, we introduce ProteinNPT,
a non-parametric transformer variant tailored to protein sequences and particula
rly suited to label-scarce and multi-task learning settings. We first focus on t
he supervised fitness prediction setting and develop several cross-validation sc
hemes which support robust performance assessment. We subsequently reimplement p
rior top-performing baselines, introduce several extensions of these baselines b
y integrating diverse branches of the protein engineering literature, and demons
trate that ProteinNPT consistently outperforms all of them across a diverse set
of protein property prediction tasks. Finally, we demonstrate the value of our a
pproach for iterative protein design across extensive in silico Bayesian optimiz
ation and conditional sampling experiments.
****
Mitigating Source Bias for Fairer Weak Supervision
Changho Shin, Sonia Cromp, Dyah Adila, Frederic Sala
Weak supervision enables efficient development of training sets by reducing the
need for ground truth labels. However, the techniques that make weak supervision
 attractive---such as integrating any source of signal to estimate unknown label
s---also entail the danger that the produced pseudolabels are highly biased. Sur
prisingly, given everyday use and the potential for increased bias, weak supervi
sion has not been studied from the point of view of fairness. We begin such a st
udy, starting with the observation that even when a fair model can be built from
 a dataset with access to ground-truth labels, the corresponding dataset labeled
 via weak supervision can be arbitrarily unfair. To address this, we propose and
 empirically validate a model for source unfairness in weak supervision, then in
troduce a simple counterfactual fairness-based technique that can mitigate these
 biases. Theoretically, we show that it is possible for our approach to simultan
eously improve both accuracy and fairness---in contrast to standard fairness app
roaches that suffer from tradeoffs. Empirically, we show that our technique impr

oves accuracy on weak supervision baselines by as much as 32\% while reducing demographic parity gap by 82.5\%. A simple extension of our method aimed at maximizing performance produces state-of-the-art performance in five out of ten datasets in the WRENCH benchmark.
****
GNNEvaluator: Evaluating GNN Performance On Unseen Graphs Without Labels
Xin Zheng, Miao Zhang, Chunyang Chen, Soheila Molaei, Chuan Zhou, Shirui Pan
Evaluating the performance of graph neural networks (GNNs) is an essential task for practical GNN model deployment and serving, as deployed GNNs face significant performance uncertainty when inferring on unseen and unlabeled test graphs, due to mismatched training-test graph distributions. In this paper, we study a new problem, GNN model evaluation, that aims to assess the performance of a specific GNN model trained on labeled and observed graphs, by precisely estimating its performance (e.g., node classification accuracy) on unseen graphs without labels. Concretely, we propose a two-stage GNN model evaluation framework, including (1) DiscGraph set construction and (2) GNNEvaluator training and inference. The DiscGraph set captures wide-range and diverse graph data distribution discrepancies through a discrepancy measurement function, which exploits the GNN outputs of latent node embeddings and node class predictions. Under the effective training supervision from the DiscGraph set, GNNEvaluator learns to precisely estimate node classification accuracy of the to-be-evaluated GNN model and makes an accurate inference for evaluating GNN model performance. Extensive experiments on real-world unseen and unlabeled test graphs demonstrate the effectiveness of our proposed method for GNN model evaluation.
****
Kronecker-Factored Approximate Curvature for Modern Neural Network Architectures
Runa Eschenhagen, Alexander Immer, Richard Turner, Frank Schneider, Philipp Hennig
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Tanimoto Random Features for Scalable Molecular Machine Learning
Austin Tripp, Sergio Bacallado, Sukriti Singh, José Miguel Hernández-Lobato
The Tanimoto coefficient is commonly used to measure the similarity between molecules represented as discrete fingerprints,either as a distance metric or a positive definite kernel. While many kernel methods can be accelerated using random feature approximations, at present there is a lack of such approximations for the Tanimoto kernel. In this paper we propose two kinds of novel random features to allow this kernel to scale to large datasets, and in the process discover a novel extension of the kernel to real-valued vectors. We theoretically characterize these random features, and provide error bounds on the spectral norm of the Gram matrix. Experimentally, we show that these random features are effective at approximating the Tanimoto coefficient of real-world datasetsand are useful for molecular property prediction and optimization tasks. Future updates to this work will be available at http://arxiv.org/abs/2306.14809.
****
Probabilistic Inference in Reinforcement Learning Done Right
Jean Tarbouriech, Tor Lattimore, Brendan O'Donoghue
A popular perspective in Reinforcement learning (RL) casts the problem as probabilistic inference on a graphical model of the Markov decision process (MDP). The core object of study is the probability of each state-action pair being visited under the optimal policy. Previous approaches to approximate this quantity can be arbitrarily poor, leading to algorithms that do not implement genuine statistical inference and consequently do not perform well in challenging problems. In this work, we undertake a rigorous Bayesian treatment of the posterior probability of state-action optimality and clarify how it flows through the MDP. We first reveal that this quantity can indeed be used to generate a policy that explores efficiently, as measured by regret. Unfortunately, computing it is intractable,

so we derive a new variational Bayesian approximation yielding a tractable conv
ex optimization problem and establish that the resulting policy also explores ef
ficiently. We call our approach VAPOR and show that it has strong connections to
 Thompson sampling, K-learning, and maximum entropy exploration. We conclude wit
h some experiments demonstrating the performance advantage of a deep RL version
of VAPOR.
****
Scale-teaching: Robust Multi-scale Training for Time Series Classification with
Noisy Labels
Zhen Liu, ma peitian, Dongliang Chen, Wenbin Pei, Qianli Ma
Deep Neural Networks (DNNs) have been criticized because they easily overfit noi
sy (incorrect) labels. To improve the robustness of DNNs, existing methods for i
mage data regard samples with small training losses as correctly labeled data (s
mall-loss criterion). Nevertheless, time series' discriminative patterns are eas
ily distorted by external noises (i.e., frequency perturbations) during the reco
rding process. This results in training losses of some time series samples that
do not meet the small-loss criterion. Therefore, this paper proposes a deep lear
ning paradigm called Scale-teaching to cope with time series noisy labels. Speci
fically, we design a fine-to-coarse cross-scale fusion mechanism for learning di
scriminative patterns by utilizing time series at different scales to train mult
iple DNNs simultaneously. Meanwhile, each network is trained in a cross-teaching
 manner by using complementary information from different scales to select small
-loss samples as clean labels. For unselected large-loss samples, we introduce m
ulti-scale embedding graph learning via label propagation to correct their label
s by using selected clean samples. Experiments on multiple benchmark time series
 datasets demonstrate the superiority of the proposed Scale-teaching paradigm ov
er state-of-the-art methods in terms of effectiveness and robustness.
****
VOCE: Variational Optimization with Conservative Estimation for Offline Safe Rei
nforcement Learning
Jiayi Guan, Guang Chen, Jiaming Ji, Long Yang, ao zhou, Zhijun Li, changjun jian
g
Offline safe reinforcement learning (RL) algorithms promise to learn policies th
at satisfy safety constraints directly in offline datasets without interacting w
ith the environment. This arrangement is particularly important in scenarios wit
h high sampling costs and potential dangers, such as autonomous driving and robo
tics. However, the influence of safety constraints and out-of-distribution (OOD)
 actions have made it challenging for previous methods to achieve high reward re
turns while ensuring safety. In this work, we propose a Variational Optimization
 with Conservative Eestimation algorithm (VOCE) to solve the problem of optimizi
ng safety policies in the offline dataset. Concretely, we reframe the problem of
 offline safe RL using probabilistic inference, which introduces variational dis
tributions to make the optimization of policies more flexible. Subsequently, we
utilize pessimistic estimation methods to estimate the Q-value of cost and rewar
d, which mitigates the extrapolation errors induced by OOD actions. Finally, ext
ensive experiments demonstrate that the VOCE algorithm achieves competitive perf
ormance across multiple experimental tasks, particularly outperforming state-of-
the-art algorithms in terms of safety.
****
Reimagining Synthetic Tabular Data Generation through Data-Centric AI: A Compreh
ensive Benchmark
Lasse Hansen, Nabeel Seedat, Mihaela van der Schaar, Andrija Petrovic
Synthetic data serves as an alternative in training machine learning models, par
ticularly when real-world data is limited or inaccessible. However, ensuring tha
t synthetic data mirrors the complex nuances of real-world data is a challenging
 task. This paper addresses this issue by exploring the potential of integrating
 data-centric AI techniques which profile the data to guide the synthetic data g
eneration process. Moreover, we shed light on the often ignored consequences of
neglecting these data profiles during synthetic data generation --- despite seem
ingly high statistical fidelity. Subsequently, we propose a novel framework to e

valuate the integration of data profiles to guide the creation of more representative synthetic data. In an empirical study, we evaluate the performance of five state-of-the-art models for tabular data generation on eleven distinct tabular datasets. The findings offer critical insights into the successes and limitations of current synthetic data generation techniques. Finally, we provide practical recommendations for integrating data-centric insights into the synthetic data generation process, with a specific focus on classification performance, model selection, and feature selection. This study aims to reevaluate conventional approaches to synthetic data generation and promote the application of data-centric AI techniques in improving the quality and effectiveness of synthetic data.
****

Beyond Geometry: Comparing the Temporal Structure of Computation in Neural Circuits with Dynamical Similarity Analysis

Mitchell Ostrow, Adam Eisen, Leo Kozachkov, Ila Fiete

How can we tell whether two neural networks utilize the same internal processes for a particular computation? This question is pertinent for multiple subfields of neuroscience and machine learning, including neuroAI, mechanistic interpretability, and brain-machine interfaces. Standard approaches for comparing neural networks focus on the spatial geometry of latent states. Yet in recurrent networks, computations are implemented at the level of dynamics, and two networks performing the same computation with equivalent dynamics need not exhibit the same geometry. To bridge this gap, we introduce a novel similarity metric that compares two systems at the level of their dynamics, called Dynamical Similarity Analysis (DSA). Our method incorporates two components: Using recent advances in data-driven dynamical systems theory, we learn a high-dimensional linear system that accurately captures core features of the original nonlinear dynamics. Next, we compare different systems passed through this embedding using a novel extension of Procrustes Analysis that accounts for how vector fields change under orthogonal transformation. In four case studies, we demonstrate that our method disentangles conjugate and non-conjugate recurrent neural networks (RNNs), while geometric methods fall short. We additionally show that our method can distinguish learning rules in an unsupervised manner. Our method opens the door to comparative analyses of the essential temporal structure of computation in neural circuits.
****

H-nobs: Achieving Certified Fairness and Robustness in Distributed Learning on Heterogeneous Datasets

Guanqiang Zhou, Ping Xu, Yue Wang, Zhi Tian

Fairness and robustness are two important goals in the design of modern distributed learning systems. Despite a few prior works attempting to achieve both fairness and robustness, some key aspects of this direction remain underexplored. In this paper, we try to answer three largely unnoticed and unaddressed questions that are of paramount significance to this topic: (i) What makes jointly satisfying fairness and robustness difficult? (ii) Is it possible to establish theoretical guarantee for the dual property of fairness and robustness? (iii) How much does fairness have to sacrifice at the expense of robustness being incorporated into the system? To address these questions, we first identify data heterogeneity as the key difficulty of combining fairness and robustness. Accordingly, we propose a fair and robust framework called H-nobs which can offer certified fairness and robustness through the adoption of two key components, a fairness-promoting objective function and a simple robust aggregation scheme called norm-based screening (NBS). We explain in detail why NBS is the suitable scheme in our algorithm in contrast to other robust aggregation measures. In addition, we derive three convergence theorems for H-nobs in cases of the learning model being nonconvex, convex, and strongly convex respectively, which provide theoretical guarantees for both fairness and robustness. Further, we empirically investigate the influence of the robust mechanism (NBS) on the fairness performance of H-nobs, the very first attempt of such exploration.
****

A Randomized Approach to Tight Privacy Accounting

Jiachen (Tianhao) Wang, Saeed Mahloujifar, Tong Wu, Ruoxi Jia, Prateek Mittal

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Triple Eagle: Simple, Fast and Practical Budget-Feasible Mechanisms
Kai Han, You Wu, He Huang, Shuang Cui
We revisit the classical problem of designing Budget-Feasible Mechanisms (BFMs) for submodular valuation functions, which has been extensively studied since the seminal paper of Singer [FOCS'10] due to its wide applications in crowdsourcing and social marketing. We propose TripleEagle, a novel algorithmic framework for designing BFMs, based on which we present several simple yet effective BFMs thatachieve better approximation ratios than the state-of-the-art work for both monotone and non-monotone submodular valuation functions. Moreover, our BFMs are the first in the literature to achieve linear complexities while ensuring obvious strategyproofness, making them more practical than the previous BFMs. We conduct extensive experiments to evaluate the empirical performance of our BFMs, and the experimental results strongly demonstrate the efficiency and effectiveness of our approach.
****
VillanDiffusion: A Unified Backdoor Attack Framework for Diffusion Models
Sheng-Yen Chou, Pin-Yu Chen, Tsung-Yi Ho
Diffusion Models (DMs) are state-of-the-art generative models that learn a reversible corruption process from iterative noise addition and denoising. They are the backbone of many generative AI applications, such as text-to-image conditional generation. However, recent studies have shown that basic unconditional DMs (e.g., DDPM and DDIM) are vulnerable to backdoor injection, a type of output manipulation attack triggered by a maliciously embedded pattern at model input. This paper presents a unified backdoor attack framework (VillanDiffusion) to expand the current scope of backdoor analysis for DMs. Our framework covers mainstream unconditional and conditional DMs (denoising-based and score-based) and various training-free samplers for holistic evaluations. Experiments show that our unified framework facilitates the backdoor analysis of different DM configurations and provides new insights into caption-based backdoor attacks on DMs.
****
An Information Theory Perspective on Variance-Invariance-Covariance Regularization
Ravid Shwartz-Ziv, Randall Balestriero, Kenji Kawaguchi, Tim G. J. Rudner, Yann LeCun
Variance-Invariance-Covariance Regularization (VICReg) is a self-supervised learning (SSL) method that has shown promising results on a variety of tasks. However, the fundamental mechanisms underlying VICReg remain unexplored. In this paper, we present an information-theoretic perspective on the VICReg objective. We begin by deriving information-theoretic quantities for deterministic networks as an alternative to unrealistic stochastic network assumptions. We then relate the optimization of the VICReg objective to mutual information optimization, highlighting underlying assumptions and facilitating a constructive comparison with other SSL algorithms and derive a generalization bound for VICReg, revealing its inherent advantages for downstream tasks. Building on these results, we introduce a family of SSL methods derived from information-theoretic principles that outperform existing SSL techniques.
****
Learning and processing the ordinal information of temporal sequences in recurrent neural circuits
xiaolong zou, Zhikun Chu, Qinghai Guo, Jie Cheng, Bo Ho, Si Wu, Yuanyuan Mi
Temporal sequence processing is fundamental in brain cognitive functions. Experimental data has indicated that the representations of ordinal information and contents of temporal sequences are disentangled in the brain, but the neural mechanism underlying this disentanglement remains largely unclear. Here, we investigate how recurrent neural circuits learn to represent the abstract order structure

of temporal sequences, and how this disentangled representation of order structure from that of contents facilitates the processing of temporal sequences. We show that with an appropriate learn protocol, a recurrent neural circuit can learn a set of tree-structured attractor states to encode the corresponding tree-structured orders of given temporal sequences. This abstract temporal order template can then be bound with different contents, allowing for flexible and robust temporal sequence processing. Using a transfer learning task, we demonstrate that the reuse of a temporal order template facilitates the acquisition of new temporal sequences of the same or similar ordinal structure. Using a key-word spotting task, we demonstrate that the attractor representation of order structure improves the robustness of temporal sequence discrimination, if the ordinal information is the key to differentiate different sequences. We hope this study gives us insights into the neural mechanism of representing the ordinal information of temporal sequences in the brain, and helps us to develop brain-inspired temporal sequence processing algorithms.
****
UNSSOR: Unsupervised Neural Speech Separation by Leveraging Over-determined Training Mixtures
Zhong-Qiu Wang, Shinji Watanabe
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Improving Self-supervised Molecular Representation Learning using Persistent Homology
Yuankai Luo, Lei Shi, Veronika Thost
Self-supervised learning (SSL) has great potential for molecular representation learning given the complexity of molecular graphs, the large amounts of unlabelled data available, the considerable cost of obtaining labels experimentally, and the hence often only small training datasets. The importance of the topic is reflected in the variety of paradigms and architectures that have been investigated recently, most focus on designing views for contrastive learning.In this paper, we study SSL based on persistent homology (PH), a mathematical tool for modeling topological features of data that persist across multiple scales. It has several unique features which particularly suit SSL, naturally offering: different views of the data, stability in terms of distance preservation, and the opportunity to flexibly incorporate domain knowledge.We (1) investigate an autoencoder, which shows the general representational power of PH, and (2) propose a contrastive loss that complements existing approaches. We rigorously evaluate our approach for molecular property prediction and demonstrate its particular features in improving the embedding space:after SSL, the representations are better and offer considerably more predictive power than the baselines over different probing tasks; our loss increases baseline performance, sometimes largely; and we often obtain substantial improvements over very small datasets, a common scenario in practice.
****
Characteristic Circuits
Zhongjie Yu, Martin Trapp, Kristian Kersting
In many real-world scenarios it is crucial to be able to reliably and efficiently reason under uncertainty while capturing complex relationships in data. Probabilistic circuits (PCs), a prominent family of tractable probabilistic models, offer a remedy to this challenge by composing simple, tractable distributions into a high-dimensional probability distribution.   However, learning PCs on heterogeneous data is challenging and densities of some parametric distributions are not available in closed form, limiting their potential use.   We introduce characteristic circuits (CCs), a family of tractable probabilistic models providing a unified formalization of distributions over heterogeneous data in the spectral domain. The one-to-one relationship between characteristic functions and probability measures enables us to learn high-dimensional distributions on heterogeneou

s data domains and facilitates efficient probabilistic inference even when no cl osed-form density function is available.   We show that the structure and parame ters of CCs can be learned efficiently from the data and find that CCs outperfor m state-of-the-art density estimators for heterogeneous data domains on common b enchmark data sets.
****
Posterior Contraction Rates for Matérn Gaussian Processes on Riemannian Manifold s
Paul Rosa, Slava Borovitskiy, Alexander Terenin, Judith Rousseau
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Causal Context Connects Counterfactual Fairness to Robust Prediction and Group F airness
Jacy Anthis, Victor Veitch
Counterfactual fairness requires that a person would have been classified in the same way by an AI or other algorithmic system if they had a different protected class, such as a different race or gender. This is an intuitive standard, as re flected in the U.S. legal system, but its use is limited because counterfactuals cannot be directly observed in real-world data. On the other hand, group fairne ss metrics (e.g., demographic parity or equalized odds) are less intuitive but m ore readily observed. In this paper, we use \textit{causal context} to bridge th e gaps between counterfactual fairness, robust prediction, and group fairness. F irst, we motivate counterfactual fairness by showing that there is not necessari ly a fundamental trade-off between fairness and accuracy because, under plausibl e conditions, the counterfactually fair predictor is in fact accuracy-optimal in an unbiased target distribution. Second, we develop a correspondence between th e causal graph of the data-generating process and which, if any, group fairness metrics are equivalent to counterfactual fairness. Third, we show that in three common fairness contexts—measurement error, selection on label, and selection on predictors—counterfactual fairness is equivalent to demographic parity, equaliz ed odds, and calibration, respectively. Counterfactual fairness can sometimes be tested by measuring relatively simple group fairness metrics.
****
Banana: Banach Fixed-Point Network for Pointcloud Segmentation with Inter-Part E quivariance
Congyue Deng, Jiahui Lei, William B Shen, Kostas Daniilidis, Leonidas J. Guibas
Equivariance has gained strong interest as a desirable network property that inh erently ensures robust generalization. However, when dealing with complex system s such as articulated objects or multi-object scenes, effectively capturing inte r-part transformations poses a challenge, as it becomes entangled with the overa ll structure and local transformations. The interdependence of part assignment a nd per-part group action necessitates a novel equivariance formulation that allo ws for their co-evolution. In this paper, we present Banana, a Banach fixed-poin t network for equivariant segmentation with inter-part equivariance by construct ion. Our key insight is to iteratively solve a fixed-point problem, where point-part assignment labels and per-part SE(3)-equivariance co-evolve simultaneously. We provide theoretical derivations of both per-step equivariance and global con vergence, which induces an equivariant final convergent state. Our formulation n aturally provides a strict definition of inter-part equivariance that generalize s to unseen inter-part configurations. Through experiments conducted on both art iculated objects and multi-object scans, we demonstrate the efficacy of our appr oach in achieving strong generalization under inter-part transformations, even w hen confronted with substantial changes in pointcloud geometry and topology.
****
Describe, Explain, Plan and Select: Interactive Planning with LLMs Enables Open-World Multi-Task Agents
Zihao Wang, Shaofei Cai, Guanzhou Chen, Anji Liu, Xiaojian (Shawn) Ma, Yitao Lia

ng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Partial Label Learning with Dissimilarity Propagation guided Candidate Label Shrinkage
Yuheng Jia, Fuchao Yang, Yongqiang Dong
In partial label learning (PLL), each sample is associated with a group of candidate labels, among which only one label is correct. The key of PLL is to disambiguate the candidate label set to find the ground-truth label. To this end, we first construct a constrained regression model to capture the confidence of the candidate labels, and multiply the label confidence matrix by its transpose to build a second-order similarity matrix, whose elements indicate the pairwise similarity relationships of samples globally. Then we develop a semantic dissimilarity matrix by considering the complement of the intersection of the candidate label set, and further propagate the initial dissimilarity relationships to the whole data set by leveraging the local geometric structure of samples. The similarity and dissimilarity matrices form an adversarial relationship, which is further utilized to shrink the solution space of the label confidence matrix and promote the dissimilarity matrix. We finally extend the proposed model to a kernel version to exploit the non-linear structure of samples and solve the proposed model by the inexact augmented Lagrange multiplier method. By exploiting the adversarial prior, the proposed method can significantly outperformstate-of-the-art PLL algorithms when evaluated on 10 artificial and 7 real-world partial label data sets. We also prove the effectiveness of our method with some theoretical guarantees. The code is publicly available at https://github.com/Yangfc-ML/DPCLS.
****
Data Selection for Language Models via Importance Resampling
Sang Michael Xie, Shibani Santurkar, Tengyu Ma, Percy S. Liang
Selecting a suitable pretraining dataset is crucial for both general-domain (e.g., GPT-3) and domain-specific (e.g., Codex) language models (LMs). We formalize this problem as selecting a subset of a large raw unlabeled dataset to match a desired target distribution given unlabeled target samples. Due to the scale and dimensionality of the raw text data, existing methods use simple heuristics or require human experts to manually curate data. Instead, we extend the classic importance resampling approach used in low-dimensions for LM data selection. We propose Data Selection with Importance Resampling (DSIR), an efficient and scalable framework that estimates importance weights in a reduced feature space for tractability and selects data with importance resampling according to these weights. We instantiate the DSIR framework with hashed n-gram features for efficiency, enabling the selection of 100M documents from the full Pile dataset in 4.5 hours. To measure whether hashed n-gram features preserve the aspects of the data that are relevant to the target, we define KL reduction, a data metric that measures the proximity between the selected pretraining data and the target on some feature space. Across 8 data selection methods (including expert selection), KL reduction on hashed n-gram features highly correlates with average downstream accuracy (r=0.82). When selecting data for continued pretraining on a specific domain, DSIR performs comparably to expert curation across 8 target distributions. When pretraining general-domain models (target is Wikipedia and books), DSIR improves over random selection and heuristic filtering baselines by 2--2.5% on the GLUE benchmark.
****
Video Dynamics Prior: An Internal Learning Approach for Robust Video Enhancements
Gaurav Shrivastava, Ser Nam Lim, Abhinav Shrivastava
In this paper, we present a novel robust framework for low-level vision tasks, including denoising, object removal, frame interpolation, and super-resolution, that does not require any external training data corpus. Our proposed approach di

rectly learns the weights of neural modules by optimizing over the corrupted test sequence, leveraging the spatio-temporal coherence and internal statistics of videos. Furthermore, we introduce a novel spatial pyramid loss that leverages the property of spatio-temporal patch recurrence in a video across the different scales of the video. This loss enhances robustness to unstructured noise in both the spatial and temporal domains. This further results in our framework being highly robust to degradation in input frames and yields state-of-the-art results on downstream tasks such as denoising, object removal, and frame interpolation. To validate the effectiveness of our approach, we conduct qualitative and quantitative evaluations on standard video datasets such as DAVIS, UCF-101, and VIMEO90K-T.
****
Glance and Focus: Memory Prompting for Multi-Event Video Question Answering
Ziyi Bai, Ruiping Wang, Xilin Chen
Video Question Answering (VideoQA) has emerged as a vital tool to evaluate agents’ ability to understand human daily behaviors. Despite the recent success of large vision language models in many multi-modal tasks, complex situation reasoning over videos involving multiple human-object interaction events still remains challenging. In contrast, humans can easily tackle it by using a series of episode memories as anchors to quickly locate question-related key moments for reasoning. To mimic this effective reasoning strategy, we propose the Glance- Focus model. One simple way is to apply an action detection model to predict a set of actions as key memories. However, these actions within a closed set vocabulary are hard to generalize to various video domains. Instead of that, we train an Encoder-Decoder to generate a set of dynamic event memories at the glancing stage. Apart from using supervised bipartite matching to obtain the event memories, we further design an unsupervised memory generation method to get rid of dependence on event annotations. Next, at the focusing stage, these event memories act as a bridge to establish the correlation between the questions with high-level event concepts and low-level lengthy video content. Given the question, the model first focuses on the generated key event memory, then focuses on the most relevant moment for reasoning through our designed multi-level cross- attention mechanism. We conduct extensive experiments on four Multi-Event VideoQA benchmarks including STAR, EgoTaskQA, AGQA, and NExT-QA. Our proposed model achieves state-of-the-art results, surpassing current large models in various challenging reasoning tasks. The code and models are available at https://github.com/ByZ0e/Glance-Focus.
****
Learning To Dive In Branch And Bound
Max Paulus, Andreas Krause
Primal heuristics are important for solving mixed integer linear programs, because they find feasible solutions that facilitate branch and bound search. A prominent group of primal heuristics are diving heuristics. They iteratively modify and resolve linear programs to conduct a depth-first search from any node in the search tree. Existing divers rely on generic decision rules that fail to exploit structural commonality between similar problem instances that often arise in practice. Therefore, we propose L2Dive to learn specific diving heuristics with graph neural networks: We train generative models to predict variable assignments and leverage the duality of linear programs to make diving decisions based on the model's predictions. L2Dive is fully integrated into the open-source solver SCIP. We find that L2Dive outperforms standard divers to find better feasible solutions on a range of combinatorial optimization problems. For real-world applications from server load balancing and neural network verification, L2Dive improves the primal-dual integral by up to 7% (35%) on average over a tuned (default) solver baseline and reduces average solving time by 20% (29%).
****
Intriguing Properties of Quantization at Scale
Arash Ahmadian, Saurabh Dash, Hongyu Chen, Bharat Venkitesh, Zhen Stephen Gou, Phil Blunsom, Ahmet Üstün, Sara Hooker
Emergent properties have been widely adopted as a term to describe behavior not present in smaller models but observed in larger models (Wei et al., 2022a). Re

cent work suggests that the trade-off incurred by quantization is also an emergent property, with sharp drops in performance in models over 6B parameters. In this work, we ask are quantization cliffs in performance solely a factor of scale? Against a backdrop of increased research focus on why certain emergent properties surface at scale, this work provides a useful counter-example. We posit that it is possible to optimize for a quantization friendly training recipe that suppresses large activation magnitude outliers. Here, we find that outlier dimensions are not an inherent product of scale, but rather sensitive to the optimization conditions present during pre-training. This both opens up directions for more efficient quantization, and poses the question of whether other emergent properties are inherent or can be altered and conditioned by optimization and architecture design choices. We successfully quantize models ranging in size from 410M to 52B with minimal degradation in performance.
****
Self-supervised Graph Neural Networks via Low-Rank Decomposition
Liang Yang, Runjie Shi, Qiuliang Zhang, bingxin niu, Zhen Wang, Xiaochun Cao, Chuan Wang
Self-supervised learning is introduced to train graph neural networks (GNNs) by employing propagation-based GNNs designed for semi-supervised learning tasks. Unfortunately, this common choice tends to cause two serious issues. Firstly, global parameters cause the model lack the ability to capture the local property. Secondly, it is difficult to handle networks beyond homophily without label information.This paper tends to break through the common choice of employing propagation-based GNNs, which aggregate representations of nodes belonging to different classes and tend to lose discriminative information. If the propagation in each ego-network is just between the nodes from the same class, the obtained representation matrix should follow the low-rank characteristic. To meet this requirement, this paper proposes the Low-Rank Decomposition-based GNNs (LRD-GNN-Matrix) by employing Low-Rank Decomposition to the attribute matrix. Furthermore, to incorporate long-distance information, Low-Rank Tensor Decomposition-based GNN (LRD-GNN-Tensor) is proposed by constructing the node attribute tensor from selected similar ego-networks and performing Low-Rank Tensor Decomposition. The employed tensor nuclear norm facilitates the capture of the long-distance relationship between original and selected similar ego-networks. Extensive experiments demonstrate the superior performance and the robustness of LRD-GNNs.
****
Cookie Consent Has Disparate Impact on Estimation Accuracy
Erik Miehling, Rahul Nair, Elizabeth Daly, Karthikeyan Natesan Ramamurthy, Robert Redmond
Cookies are designed to enable more accurate identification and tracking of user behavior, in turn allowing for more personalized ads and better performing ad campaigns. Given the additional information that is recorded, questions related to privacy and fairness naturally arise. How does a user's consent decision influence how much the system can learn about their demographic and tastes? Is the impact of a user's consent decision on the recommender system's ability to learn about their latent attributes uniform across demographics? We investigate these questions in the context of an engagement-driven recommender system using simulation. We empirically demonstrate that when consent rates exhibit demographic-dependence, user consent has a disparate impact on the recommender agent's ability to estimate users' latent attributes. In particular, we find that when consent rates are demographic-dependent, a user disagreeing to share their cookie may counter-intuitively cause the recommender agent to know more about the user than if the user agreed to share their cookie. Furthermore, the gap in base consent rates across demographics serves as an amplifier: users from the lower consent rate demographic who agree to cookie sharing generally experience higher estimation errors than the same users from the higher consent rate demographic, and conversely for users who choose to disagree to cookie sharing, with these differences increasing in consent rate gap. We discuss the need for new notions of fairness that encourage consistency between a user's privacy decisions and the system's ability to estimate their latent attributes.

****
NPCL: Neural Processes for Uncertainty-Aware Continual Learning
Saurav Jha, Dong Gong, He Zhao, Lina Yao
Continual learning (CL) aims to train deep neural networks efficiently on streaming data while limiting the forgetting caused by new tasks. However, learning transferable knowledge with less interference between tasks is difficult, and real-world deployment of CL models is limited by their inability to measure predictive uncertainties. To address these issues, we propose handling CL tasks with neural processes (NPs), a class of meta-learners that encode different tasks into probabilistic distributions over functions all while providing reliable uncertainty estimates. Specifically, we propose an NP-based CL approach (NPCL) with task-specific modules arranged in a hierarchical latent variable model. We tailor regularizers on the learned latent distributions to alleviate forgetting. The uncertainty estimation capabilities of the NPCL can also be used to handle the task head/module inference challenge in CL. Our experiments show that the NPCL outperforms previous CL approaches. We validate the effectiveness of uncertainty estimation in the NPCL for identifying novel data and evaluating instance-level model confidence. Code is available at https://github.com/srvCodes/NPCL.
****
From Pixels to UI Actions: Learning to Follow Instructions via Graphical User Interfaces
Peter Shaw, Mandar Joshi, James Cohan, Jonathan Berant, Panupong Pasupat, Hexiang Hu, Urvashi Khandelwal, Kenton Lee, Kristina N Toutanova
Much of the previous work towards digital agents for graphical user interfaces (GUIs) has relied on text-based representations (derived from HTML or other structured data sources), which are not always readily available. These input representations have been often coupled with custom, task-specific action spaces. This paper focuses on creating agents that interact with the digital world using the same conceptual interface that humans commonly use — via pixel-based screenshots and a generic action space corresponding to keyboard and mouse actions. Building upon recent progress in pixel-based pretraining, we show, for the first time, that it is possible for such agents to outperform human crowdworkers on the MiniWob++ benchmark of GUI-based instruction following tasks.
****
Robust Mean Estimation Without Moments for Symmetric Distributions
Gleb Novikov, David Steurer, Stefan Tiegel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Fast and Simple Spectral Clustering in Theory and Practice
Peter Macgregor
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CL-NeRF: Continual Learning of Neural Radiance Fields for Evolving Scene Representation
Xiuzhe Wu, Peng Dai, Weipeng DENG, Handi Chen, Yang Wu, Yan-Pei Cao, Ying Shan, Xiaojuan Qi
Existing methods for adapting Neural Radiance Fields (NeRFs) to scene changes require extensive data capture and model retraining, which is both time-consuming and labor-intensive. In this paper, we tackle the challenge of efficiently adapting NeRFs to real-world scene changes over time using a few new images while retaining the memory of unaltered areas, focusing on the continual learning aspect of NeRFs. To this end, we propose CL-NeRF, which consists of two key components: a lightweight expert adaptor for adapting to new changes and evolving scene representations and a conflict-aware knowledge distillation learning objective for

memorizing unchanged parts. We also present a new benchmark for evaluating Continual Learning of NeRFs with comprehensive metrics. Our extensive experiments demonstrate that CL-NeRF can synthesize high-quality novel views of both changed and unchanged regions with high training efficiency, surpassing existing methods in terms of reducing forgetting and adapting to changes. Code and benchmark will be made available.
****
Generalised f-Mean Aggregation for Graph Neural Networks
Ryan Kortvelesy, Steven Morad, Amanda Prorok
Graph Neural Network (GNN) architectures are defined by their implementations of update and aggregation modules. While many works focus on new ways to parametrise the update modules, the aggregation modules receive comparatively little attention. Because it is difficult to parametrise aggregation functions, currently most methods select a ``standard aggregator'' such as mean, sum, or max. While this selection is often made without any reasoning, it has been shown that the choice in aggregator has a significant impact on performance, and the best choice in aggregator is problem-dependent. Since aggregation is a lossy operation, it is crucial to select the most appropriate aggregator in order to minimise information loss. In this paper, we present GenAgg, a generalised aggregation operator, which parametrises a function space that includes all standard aggregators. In our experiments, we show that GenAgg is able to represent the standard aggregators with much higher accuracy than baseline methods. We also show that using GenAgg as a drop-in replacement for an existing aggregator in a GNN often leads to a significant boost in performance across various tasks.
****
Certified Robustness via Dynamic Margin Maximization and Improved Lipschitz Regularization
Mahyar Fazlyab, Taha Entesari, Aniket Roy, Rama Chellappa
To improve the robustness of deep classifiers against adversarial perturbations, many approaches have been proposed, such as designing new architectures with better robustness properties (e.g., Lipschitz-capped networks), or modifying the training process itself (e.g., min-max optimization, constrained learning, or regularization). These approaches, however, might not be effective at increasing the margin in the input (feature) space. In this paper, we propose a differentiable regularizer that is a lower bound on the distance of the data points to the classification boundary. The proposed regularizer requires knowledge of the model's Lipschitz constant along certain directions. To this end, we develop a scalable method for calculating guaranteed differentiable upper bounds on the Lipschitz constant of neural networks accurately and efficiently. The relative accuracy of the bounds prevents excessive regularization and allows for more direct manipulation of the decision boundary. Furthermore, our Lipschitz bounding algorithm exploits the monotonicity and Lipschitz continuity of the activation layers, and the resulting bounds can be used to design new layers with controllable bounds on their Lipschitz constant. Experiments on the MNIST, CIFAR-10, and Tiny-ImageNet data sets verify that our proposed algorithm obtains competitively improved results compared to the state-of-the-art.
****
Unpaired Multi-Domain Causal Representation Learning
Nils Sturma, Chandler Squires, Mathias Drton, Caroline Uhler
The goal of causal representation learning is to find a representation of data that consists of causally related latent variables. We consider a setup where one has access to data from multiple domains that potentially share a causal representation. Crucially, observations in different domains are assumed to be unpaired, that is, we only observe the marginal distribution in each domain but not their joint distribution. In this paper, we give sufficient conditions for identifiability of the joint distribution and the shared causal graph in a linear setup. Identifiability holds if we can uniquely recover the joint distribution and the shared causal representation from the marginal distributions in each domain. We transform our results into a practical method to recover the shared latent causal graph.

****
Flow-Based Feature Fusion for Vehicle-Infrastructure Cooperative 3D Object Detection
Haibao Yu, Yingjuan Tang, Enze Xie, Jilei Mao, Ping Luo, Zaiqing Nie
Cooperatively utilizing both ego-vehicle and infrastructure sensor data can significantly enhance autonomous driving perception abilities. However, the uncertain temporal asynchrony and limited communication conditions that are present in traffic environments can lead to fusion misalignment and constrain the exploitation of infrastructure data. To address these issues in vehicle-infrastructure cooperative 3D (VIC3D) object detection, we propose the Feature Flow Net (FFNet), a novel cooperative detection framework. FFNet is a flow-based feature fusion framework that uses a feature flow prediction module to predict future features and compensate for asynchrony. Instead of transmitting feature maps extracted from still-images, FFNet transmits feature flow, leveraging the temporal coherence of sequential infrastructure frames. Furthermore, we introduce a self-supervised training approach that enables FFNet to generate feature flow with feature prediction ability from raw infrastructure sequences. Experimental results demonstrate that our proposed method outperforms existing cooperative detection methods while only requiring about 1/100 of the transmission cost of raw data and covers all latency in one model on the DAIR-V2X dataset. The code is available https://github.com/haibao-yu/FFNet-VIC3D.
****
Subspace Identification for Multi-Source Domain Adaptation
Zijian Li, Ruichu Cai, Guangyi Chen, Boyang Sun, Zhifeng Hao, Kun Zhang
Multi-source domain adaptation (MSDA) methods aim to transfer knowledge from multiple labeled source domains to an unlabeled target domain. Although current methods achieve target joint distribution identifiability by enforcing minimal changes across domains, they often necessitate stringent conditions, such as an adequate number of domains, monotonic transformation of latent variables, and invariant label distributions. These requirements are challenging to satisfy in real-world applications. To mitigate the need for these strict assumptions, we propose a subspace identification theory that guarantees the disentanglement of domain-invariant and domain-specific variables under less restrictive constraints regarding domain numbers and transformation properties and thereby facilitating domain adaptation by minimizing the impact of domain shifts on invariant variables. Based on this theory, we develop a Subspace Identification Guarantee (SIG) model that leverages variational inference. Furthermore, the SIG model incorporates class-aware conditional alignment to accommodate target shifts where label distributions change with the domain. Experimental results demonstrate that our SIG model outperforms existing MSDA techniques on various benchmark datasets, highlighting its effectiveness in real-world applications.
****
Optimistic Exploration in Reinforcement Learning Using Symbolic Model Estimates
Sarath Sreedharan, Michael Katz
There has been an increasing interest in using symbolic models along with reinforcement learning (RL) problems, where these coarser abstract models are used as a way to provide RL agents with higher level guidance. However, most of these works are inherently limited by their assumption of having an access to a symbolic approximation of the underlying problem. To address this issue, we introduce a new method for learning optimistic symbolic approximations of the underlying world model. We will see how these representations, coupled with fast diverse planners developed by the automated planning community, provide us with a new paradigm for optimistic exploration in sparse reward settings. We investigate the possibility of speeding up the learning process by generalizing learned model dynamics across similar actions with minimal human input. Finally, we evaluate the method, by testing it on multiple benchmark domains and compare it with other RL strategies.
****
Feature learning via mean-field Langevin dynamics: classifying sparse parities and beyond

Taiji Suzuki, Denny Wu, Kazusato Oko, Atsushi Nitanda
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Improving Graph Matching with Positional Reconstruction Encoder-Decoder Network
Yixiao Zhou, Ruiqi Jia, Hongxiang Lin, Hefeng Quan, Yumeng Zhao, Xiaoqing Lyu
Deriving from image matching and understanding, semantic keypoint matching aims at establishing correspondence between keypoint sets in images. As graphs are powerful tools to represent points and their complex relationships, graph matching provides an effective way to find desired semantic keypoint correspondences. Recent deep graph matching methods have shown excellent performance, but there is still a lack of exploration and utilization of spatial information of keypoints as nodes in graphs. More specifically, existing methods are insufficient to capture the relative spatial relations through current graph construction approaches from the locations of semantic keypoints. To address these issues, we introduce a positional reconstruction encoder-decoder (PR-EnDec) to model intrinsic graph spatial structure, and present an end-to-end graph matching network PREGM based on PR-EnDec. Our PR-EnDec consists of a positional encoder that learns effective node spatial embedding with the affine transformation invariance, and a spatial relation decoder that further utilizes the high-order spatial information by reconstructing the locational structure of graphs contained in the node coordinates. Extensive experimental results on three public keypoint matching datasets demonstrate the effectiveness of our proposed PREGM.
****
A Causal Framework for Decomposing Spurious Variations
Drago Plecko, Elias Bareinboim
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Revisiting Logistic-softmax Likelihood in Bayesian Meta-Learning for Few-Shot Classification
Tianjun Ke, Haoqun Cao, Zenan Ling, Feng Zhou
Meta-learning has demonstrated promising results in few-shot classification (FSC) by learning to solve new problems using prior knowledge. Bayesian methods are effective at characterizing uncertainty in FSC, which is crucial in high-risk fields. In this context, the logistic-softmax likelihood is often employed as an alternative to the softmax likelihood in multi-class Gaussian process classification due to its conditional conjugacy property. However, the theoretical property of logistic-softmax is not clear and previous research indicated that the inherent uncertainty of logistic-softmax leads to suboptimal performance. To mitigate these issues, we revisit and redesign the logistic-softmax likelihood, which enables control of the \textit{a priori} confidence level through a temperature parameter. Furthermore, we theoretically and empirically show that softmax can be viewed as a special case of logistic-softmax and logistic-softmax induces a larger family of data distribution than softmax. Utilizing modified logistic-softmax, we integrate the data augmentation technique into the deep kernel based Gaussian process meta-learning framework, and derive an analytical mean-field approximation for task-specific updates. Our approach yields well-calibrated uncertainty estimates and achieves comparable or superior results on standard benchmark datasets. Code is publicly available at \url{https://github.com/keanson/revisit-logistic-softmax}.
****
Functional-Group-Based Diffusion for Pocket-Specific Molecule Generation and Elaboration
Haitao Lin, Yufei Huang, Odin Zhang, Yunfan Liu, Lirong Wu, Siyuan Li, Zhiyuan Chen, Stan Z. Li

In recent years, AI-assisted drug design methods have been proposed to generate molecules given the pockets' structures of target proteins. Most of them are {\em atom-level-based} methods, which consider atoms as basic components and generate atom positions and types. In this way, however, it is hard to generate realistic fragments with complicated structures. To solve this, we propose \textsc{D3FG}, a {\em functional-group-based} diffusion model for pocket-specific molecule generation and elaboration. \textsc{D3FG} decomposes molecules into two categories of components: functional groups defined as rigid bodies and linkers as mass points. And the two kinds of components can together form complicated fragments that enhance ligand-protein interactions. To be specific, in the diffusion process, \textsc{D3FG} diffuses the data distribution of the positions, orientations, and types of the components into a prior distribution; In the generative process, the noise is gradually removed from the three variables by denoisers parameterized with designed equivariant graph neural networks. In the experiments, our method can generate molecules with more realistic 3D structures, competitive affinities toward the protein targets, and better drug properties. Besides, \textsc{D3FG} as a solution to a new task of molecule elaboration, could generate molecules with high affinities based on existing ligands and the hotspots of target proteins.

****

Approximately Equivariant Graph Networks

Ningyuan Huang, Ron Levie, Soledad Villar

Graph neural networks (GNNs) are commonly described as being permutation equivariant with respect to node relabeling in the graph. This symmetry of GNNs is often compared to the translation equivariance of Euclidean convolution neural networks (CNNs). However, these two symmetries are fundamentally different: The translation equivariance of CNNs corresponds to symmetries of the fixed domain acting on the image signals (sometimes known as active symmetries), whereas in GNNs any permutation acts on both the graph signals and the graph domain (sometimes described as passive symmetries). In this work, we focus on the active symmetries of GNNs, by considering a learning setting where signals are supported on a fixed graph. In this case, the natural symmetries of GNNs are the automorphisms of the graph. Since real-world graphs tend to be asymmetric, we relax the notion of symmetries by formalizing approximate symmetries via graph coarsening. We present a bias-variance formula that quantifies the tradeoff between the loss in expressivity and the gain in the regularity of the learned estimator, depending on the chosen symmetry group. To illustrate our approach, we conduct extensive experiments on image inpainting, traffic flow prediction, and human pose estimation with different choices of symmetries. We show theoretically and empirically that the best generalization performance can be achieved by choosing a suitably larger group than the graph automorphism, but smaller than the permutation group.

****

H2O: Heavy-Hitter Oracle for Efficient Generative Inference of Large Language Models

Zhenyu Zhang, Ying Sheng, Tianyi Zhou, Tianlong Chen, Lianmin Zheng, Ruisi Cai, Zhao Song, Yuandong Tian, Christopher Ré, Clark Barrett, Zhangyang "Atlas" Wang, Beidi Chen

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

Uncovering motifs of concurrent signaling across multiple neuronal populations

Evren Gokcen, Anna Jasper, Alison Xu, Adam Kohn, Christian K. Machens, Byron M Yu

Modern recording techniques now allow us to record from distinct neuronal populations in different brain networks. However, especially as we consider multiple (more than two) populations, new conceptual and statistical frameworks are needed to characterize the multi-dimensional, concurrent flow of signals among these populations. Here, we develop a dimensionality reduction framework that determine

s (1) the subset of populations described by each latent dimension, (2) the direction of signal flow among those populations, and (3) how those signals evolve over time within and across experimental trials. We illustrate these features in simulation, and further validate the method by applying it to previously studied recordings from neuronal populations in macaque visual areas V1 and V2. Then we study interactions across select laminar compartments of areas V1, V2, and V3d, recorded simultaneously with multiple Neuropixels probes. Our approach uncovered signatures of selective communication across these three areas that related to their retinotopic alignment. This work advances the study of concurrent signaling across multiple neuronal populations.
****

NVFi: Neural Velocity Fields for 3D Physics Learning from Dynamic Videos
Jinxi Li, Ziyang Song, Bo Yang
In this paper, we aim to model 3D scene dynamics from multi-view videos. Unlike the majority of existing works which usually focus on the common task of novel view synthesis within the training time period, we propose to simultaneously learn the geometry, appearance, and physical velocity of 3D scenes only from video frames, such that multiple desirable applications can be supported, including future frame extrapolation, unsupervised 3D semantic scene decomposition, and dynamic motion transfer. Our method consists of three major components, 1) the keyframe dynamic radiance field, 2) the interframe velocity field, and 3) a joint keyframe and interframe optimization module which is the core of our framework to effectively train both networks. To validate our method, we further introduce two dynamic 3D datasets: 1) Dynamic Object dataset, and 2) Dynamic Indoor Scene dataset. We conduct extensive experiments on multiple datasets, demonstrating the superior performance of our method over all baselines, particularly in the critical tasks of future frame extrapolation and unsupervised 3D semantic scene decomposition.
****

Don't be so Monotone: Relaxing Stochastic Line Search in Over-Parameterized Models
Leonardo Galli, Holger Rauhut, Mark Schmidt
Recent works have shown that line search methods can speed up Stochastic Gradient Descent (SGD) and Adam in modern over-parameterized settings. However, existing line searches may take steps that are smaller than necessary since they require a monotone decrease of the (mini-)batch objective function. We explore nonmonotone line search methods to relax this condition and possibly accept larger step sizes. Despite the lack of a monotonic decrease, we prove the same fast rates of convergence as in the monotone case. Our experiments show that nonmonotone methods improve the speed of convergence and generalization properties of SGD/Adam even beyond the previous monotone line searches. We propose a POlyak NOnmonotone Stochastic (PoNoS) method, obtained by combining a nonmonotone line search with a Polyak initial step size. Furthermore, we develop a new resetting technique that in the majority of the iterations reduces the amount of backtracks to zero while still maintaining a large initial step size. To the best of our knowledge, a first runtime comparison shows that the epoch-wise advantage of line-search-based methods gets reflected in the overall computational time.
****

OFCOURSE: A Multi-Agent Reinforcement Learning Environment for Order Fulfillment
Yiheng Zhu, Yang Zhan, Xuankun Huang, Yuwei Chen, yujie Chen, Jiangwen Wei, Wei Feng, Yinzhi Zhou, Haoyuan Hu, Jieping Ye
The dramatic growth of global e-commerce has led to a surge in demand for efficient and cost-effective order fulfillment which can increase customers' service levels and sellers' competitiveness. However, managing order fulfillment is challenging due to a series of interdependent online sequential decision-making problems. To clear this hurdle, rather than solving the problems separately as attempted in some recent researches, this paper proposes a method based on multi-agent reinforcement learning to integratively solve the series of interconnected problems, encompassing order handling, packing and pickup, storage, order consolidation, and last-mile delivery. In particular, we model the integrated problem as a

Markov game, wherein a team of agents learns a joint policy via interacting wit
h a simulated environment. Since no simulated environment supporting the complet
e order fulfillment problem exists, we devise Order Fulfillment COoperative mUlt
i-agent Reinforcement learning Scalable Environment (OFCOURSE) in the OpenAI Gym
 style, which allows reproduction and re-utilization to build customized applica
tions. By constructing the fulfillment system in OFCOURSE, we optimize a joint p
olicy that solves the integrated problem, facilitating sequential order-wise ope
rations across all fulfillment units and minimizing the total cost of fulfilling
 all orders within the promised time. With OFCOURSE, we also demonstrate that th
e joint policy learned by multi-agent reinforcement learning outperforms the com
bination of locally optimal policies. The source code of OFCOURSE is available a
t: https://github.com/GitYiheng/ofcourse.
****
On Private and Robust Bandits
Yulian Wu, Xingyu Zhou, Youming Tao, Di Wang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
RiskQ: Risk-sensitive Multi-Agent Reinforcement Learning Value Factorization
Siqi Shen, Chennan Ma, Chao Li, Weiquan Liu, Yongquan Fu, Songzhu Mei, Xinwang L
iu, Cheng Wang
Multi-agent systems are characterized by environmental uncertainty, varying poli
cies of agents, and partial observability, which result in significant risks. In
 the context of Multi-Agent Reinforcement Learning (MARL), learning coordinated
and decentralized policies that are sensitive to risk is challenging. To formula
te the coordination requirements in risk-sensitive MARL, we introduce the Risk-s
ensitive Individual-Global-Max (RIGM) principle as a generalization of the Indiv
idual-Global-Max (IGM) and Distributional IGM (DIGM) principles. This principle
requires that the collection of risk-sensitive action selections of each agent s
hould be equivalent to the risk-sensitive action selection of the central policy
. Current MARL value factorization methods do not satisfy the RIGM principle for
 common risk metrics such as the Value at Risk (VaR) metric or distorted risk me
asurements. Therefore, we propose RiskQ to address this limitation, which models
 the joint return distribution by modeling quantiles of it as weighted quantile
mixtures of per-agent return distribution utilities. RiskQ satisfies the RIGM pr
inciple for the VaR and distorted risk metrics. We show that RiskQ can obtain pr
omising performance through extensive experiments. The source code of RiskQ is a
vailable in https://github.com/xmu-rl-3dv/RiskQ.
****
Learning Exponential Families from Truncated Samples
Jane Lee, Andre Wibisono, Emmanouil Zampetakis
Missing data problems have many manifestations across many scientific fields. A
fundamental type of missing data problem arises when samples are \textit{truncat
ed}, i.e., samples that lie in a subset of the support are not observed. Statist
ical estimation from truncated samples is a classical problem in statistics whic
h dates back to Galton, Pearson, and Fisher. A recent line of work provides the
first efficient estimation algorithms for the parameters of a Gaussian distribut
ion and for linear regression with Gaussian noise.In this paper we generalize th
ese results to log-concave exponential families. We provide an estimation algori
thm that shows that \textit{extrapolation} is possible for a much larger class o
f distributions while it maintains a polynomial sample and time complexity on av
erage. Our algorithm is based on Projected Stochastic Gradient Descent and is no
t only applicable in a more general setting but is also simpler and more efficie
nt than recent algorithms. Our work also has interesting implications for learni
ng general log-concave distributions and sampling given only access to truncated
 data.
****
XAGen: 3D Expressive Human Avatars Generation

Zhongcong XU, Jianfeng Zhang, Jun Hao Liew, Jiashi Feng, Mike Zheng Shou
Recent advances in 3D-aware GAN models have enabled the generation of realistic and controllable human body images. However, existing methods focus on the control of major body joints, neglecting the manipulation of expressive attributes, such as facial expressions, jaw poses, hand poses, and so on. In this work, we present XAGen, the first 3D generative model for human avatars capable of expressive control over body, face, and hands. To enhance the fidelity of small-scale regions like face and hands, we devise a multi-scale and multi-part 3D representation that models fine details. Based on this representation, we propose a multi-part rendering technique that disentangles the synthesis of body, face, and hands to ease model training and enhance geometric quality. Furthermore, we design multi-part discriminators that evaluate the quality of the generated avatars with respect to their appearance and fine-grained control capabilities. Experiments show that XAGen surpasses state-of-the-art methods in terms of realism, diversity, and expressive control abilities. Code and data will be made available at https://showlab.github.io/xagen.
****
HIQL: Offline Goal-Conditioned RL with Latent States as Actions
Seohong Park, Dibya Ghosh, Benjamin Eysenbach, Sergey Levine
Unsupervised pre-training has recently become the bedrock for computer vision and natural language processing. In reinforcement learning (RL), goal-conditioned RL can potentially provide an analogous self-supervised approach for making use of large quantities of unlabeled (reward-free) data. However, building effective algorithms for goal-conditioned RL that can learn directly from diverse offline data is challenging, because it is hard to accurately estimate the exact value function for faraway goals. Nonetheless, goal-reaching problems exhibit structure, such that reaching distant goals entails first passing through closer subgoals. This structure can be very useful, as assessing the quality of actions for nearby goals is typically easier than for more distant goals. Based on this idea, we propose a hierarchical algorithm for goal-conditioned RL from offline data. Using one action-free value function, we learn two policies that allow us to exploit this structure: a high-level policy that treats states as actions and predicts (a latent representation of) a subgoal and a low-level policy that predicts the action for reaching this subgoal. Through analysis and didactic examples, we show how this hierarchical decomposition makes our method robust to noise in the estimated value function. We then apply our method to offline goal-reaching benchmarks, showing that our method can solve long-horizon tasks that stymie prior methods, can scale to high-dimensional image observations, and can readily make use of action-free data. Our code is available at https://seohong.me/projects/hiql/
****
Visual Instruction Tuning
Haotian Liu, Chunyuan Li, Qingyang Wu, Yong Jae Lee
Instruction tuning large language models (LLMs) using machine-generated instruction-following data has been shown to improve zero-shot capabilities on new tasks, but the idea is less explored in the multimodal field. We present the first attempt to use language-only GPT-4 to generate multimodal language-image instruction-following data. By instruction tuning on such generated data, we introduce LLaVA: Large Language and Vision Assistant, an end-to-end trained large multimodal model that connects a vision encoder and an LLM for general-purpose visual and language understanding. To facilitate future research on visual instruction following, we construct two evaluation benchmarks with diverse and challenging application-oriented tasks. Our experiments show that LLaVA demonstrates impressive multimodal chat abilities, sometimes exhibiting the behaviors of multimodal GPT-4 on unseen images/instructions, and yields a 85.1% relative score compared with GPT-4 on a synthetic multimodal instruction-following dataset. When fine-tuned on Science QA, the synergy of LLaVA and GPT-4 achieves a new state-of-the-art accuracy of 92.53%. We make GPT-4 generated visual instruction tuning data, our model, and code publicly available.
****

A Fast and Accurate Estimator for Large Scale Linear Model via Data Averaging
Rui Wang, Yanyan Ouyang, Yu Panpan, Wangli Xu
This work is concerned with the estimation problem of linear model when thesampl e size is extremely large and the data dimension can vary with the samplesize. I n this setting, the least square estimator based on the full data is not feasibl ewith limited computational resources. Many existing methods for this problem ar ebased on the sketching technique which uses the sketched data to perform leasts quare estimation. We derive fine-grained lower bounds of the conditional meansqu ared error for sketching methods. For sampling methods, our lower boundprovides an attainable optimal convergence rate. Our result implies that when thedimensio n is large, there is hardly a sampling method can have a faster convergencerate than the uniform sampling method. To achieve a better statistical performance,we propose a new sketching method based on data averaging. The proposedmethod redu ces the original data to a few averaged observations. These averagedobservations still satisfy the linear model and are used to estimate the regressioncoefficie nts. The asymptotic behavior of the proposed estimation procedure isstudied. Our theoretical results show that the proposed method can achieve afaster convergen ce rate than the optimal convergence rate for sampling methods.Theoretical and n umerical results show that the proposed estimator has goodstatistical performanc e as well as low computational cost.
****
Correlative Information Maximization: A Biologically Plausible Approach to Super vised Deep Neural Networks without Weight Symmetry
Bariscan Bozkurt, Cengiz Pehlevan, Alper Erdogan
The backpropagation algorithm has experienced remarkable success in training lar ge-scale artificial neural networks; however, its biological plausibility has be en strongly criticized, and it remains an open question whether the brain employ s supervised learning mechanisms akin to it. Here, we propose correlative inform ation maximization between layer activations as an alternative normative approac h to describe the signal propagation in biological neural networks in both forwa rd and backward directions. This new framework addresses many concerns about the biological-plausibility of conventional artificial neural networks and the back propagation algorithm. The coordinate descent-based optimization of the correspo nding objective, combined with the mean square error loss function for fitting l abeled supervision data, gives rise to a neural network structure that emulates a more biologically realistic network of multi-compartment pyramidal neurons wit h dendritic processing and lateral inhibitory neurons. Furthermore, our approach provides a natural resolution to the weight symmetry problem between forward an d backward signal propagation paths, a significant critique against the plausibi lity of the conventional backpropagation algorithm. This is achieved by leveragi ng two alternative, yet equivalent forms of the correlative mutual information o bjective. These alternatives intrinsically lead to forward and backward predicti on networks without weight symmetry issues, providing a compelling solution to t his long-standing challenge.
****
What Distributions are Robust to Indiscriminate Poisoning Attacks for Linear Lea rners?
Fnu Suya, Xiao Zhang, Yuan Tian, David Evans
We study indiscriminate poisoning for linear learners where an adversary injects a few crafted examples into the training data with the goal of forcing the ind uced model to incur higher test error. Inspired by the observation that linear l earners on some datasets are able to resist the best known attacks even without any defenses, we further investigate whether datasets can be inherently robust t o indiscriminate poisoning attacks for linear learners. For theoretical Gaussian distributions, we rigorously characterize the behavior of an optimal poisoning attack, defined as the poisoning strategy that attains the maximum risk of the i nduced model at a given poisoning budget. Our results prove that linear learners can indeed be robust to indiscriminate poisoning if the class-wise data distrib utions are well-separated with low variance and the size of the constraint set c ontaining all permissible poisoning points is also small. These findings largely

explain the drastic variation in empirical attack performance of the state-of-the-art poisoning attacks on linear learners across benchmark datasets, making an important initial step towards understanding the underlying reasons some learning tasks are vulnerable to data poisoning attacks.
****
Expressive probabilistic sampling in recurrent neural networks
Shirui Chen, Linxing Jiang, Rajesh PN Rao, Eric Shea-Brown
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Counting Distinct Elements Under Person-Level Differential Privacy
Thomas Steinke, Alexander Knop
We study the problem of counting the number of distinct elements in a dataset subject to the constraint of differential privacy. We consider the challenging setting of person-level DP (a.k.a. user-level DP) where each person may contribute an unbounded number of items and hence the sensitivity is unbounded.Our approach is to compute a bounded-sensitivity version of this query, which reduces to solving a max-flow problem. The sensitivity bound is optimized to balance the noise we must add to privatize the answer against the error of the approximation of the bounded-sensitivity query to the true number of unique elements.
****
Stochastic Collapse: How Gradient Noise Attracts SGD Dynamics Towards Simpler Subnetworks
Feng Chen, Daniel Kunin, Atsushi Yamamura, Surya Ganguli
In this work, we reveal a strong implicit bias of stochastic gradient descent (SGD) that drives overly expressive networks to much simpler subnetworks, thereby dramatically reducing the number of independent parameters, and improving generalization. To reveal this bias, we identify invariant sets, or subsets of parameter space that remain unmodified by SGD. We focus on two classes of invariant sets that correspond to simpler (sparse or low-rank) subnetworks and commonly appear in modern architectures. Our analysis uncovers that SGD exhibits a property of stochastic attractivity towards these simpler invariant sets. We establish a sufficient condition for stochastic attractivity based on a competition between the loss landscape's curvature around the invariant set and the noise introduced by stochastic gradients. Remarkably, we find that an increased level of noise strengthens attractivity, leading to the emergence of attractive invariant sets associated with saddle-points or local maxima of the train loss. We observe empirically the existence of attractive invariant sets in trained deep neural networks, implying that SGD dynamics often collapses to simple subnetworks with either vanishing or redundant neurons. We further demonstrate how this simplifying process of stochastic collapse benefits generalization in a linear teacher-student framework. Finally, through this analysis, we mechanistically explain why early training with large learning rates for extended periods benefits subsequent generalization.
****
Conservative State Value Estimation for Offline Reinforcement Learning
Liting Chen, Jie Yan, Zhengdao Shao, Lu Wang, Qingwei Lin, Saravanakumar Rajmohan, Thomas Moscibroda, Dongmei Zhang
Offline reinforcement learning faces a significant challenge of value over-estimation due to the distributional drift between the dataset and the current learned policy, leading to learning failure in practice. The common approach is to incorporate a penalty term to reward or value estimation in the Bellman iterations. Meanwhile, to avoid extrapolation on out-of-distribution (OOD) states and actions, existing methods focus on conservative Q-function estimation. In this paper, we propose Conservative State Value Estimation (CSVE), a new approach that learns conservative V-function via directly imposing penalty on OOD states. Compared to prior work, CSVE allows more effective state value estimation with conservative guarantees and further better policy optimization. Further, we apply CSVE an

d develop a practical actor-critic algorithm in which the critic does the conser vative value estimation by additionally sampling and penalizing the states aroun d the dataset, and the actor applies advantage weighted updates extended with st ate exploration to improve the policy. We evaluate in classic continual control tasks of D4RL, showing that our method performs better than the conservative Q-f unction learning methods and is strongly competitive among recent SOTA methods.
****
Demystifying Oversmoothing in Attention-Based Graph Neural Networks
Xinyi Wu, Amir Ajorlou, Zihui Wu, Ali Jadbabaie
Oversmoothing in Graph Neural Networks (GNNs) refers to the phenomenon where inc reasing network depth leads to homogeneous node representations. While previous work has established that Graph Convolutional Networks (GCNs) exponentially lose expressive power, it remains controversial whether the graph attention mechanis m can mitigate oversmoothing. In this work, we provide a definitive answer to th is question through a rigorous mathematical analysis, by viewing attention-based GNNs as nonlinear time-varying dynamical systems and incorporating tools and te chniques from the theory of products of inhomogeneous matrices and the joint spe ctral radius. We establish that, contrary to popular belief, the graph attention mechanism cannot prevent oversmoothing and loses expressive power exponentially . The proposed framework extends the existing results on oversmoothing for symme tric GCNs to a significantly broader class of GNN models, including random walk GCNs, Graph Attention Networks (GATs) and (graph) transformers. In particular, o ur analysis accounts for asymmetric, state-dependent and time-varying aggregatio n operators and a wide range of common nonlinear activation functions, such as R eLU, LeakyReLU, GELU and SiLU.
****
A Comprehensive Benchmark for Neural Human Radiance Fields
Kenkun Liu, Derong Jin, Ailing Zeng, Xiaoguang Han, Lei Zhang
The past two years have witnessed a significant increase in interest concerning NeRF-based human body rendering. While this surge has propelled considerable adv ancements, it has also led to an influx of methods and datasets. This explosion complicates experimental settings and makes fair comparisons challenging. In thi s work, we design and execute thorough studies into unified evaluation settings and metrics to establish a fair and reasonable benchmark for human NeRF models. To reveal the effects of extant models, we benchmark them against diverse and ha rd scenes. Additionally, we construct a cross-subject benchmark pre-trained on l arge-scale datasets to assess generalizable methods. Finally, we analyze the ess ential components for animatability and generalizability, and make HumanNeRF fro m monocular videos generalizable, as the inaugural baseline. We hope these bench marks and analyses could serve the community.
****
Sample Complexity for Quadratic Bandits: Hessian Dependent Bounds and Optimal Al gorithms
Qian Yu, Yining Wang, Baihe Huang, Qi Lei, Jason D. Lee
In stochastic zeroth-order optimization, a problem of practical relevance is und erstanding how to fully exploit the local geometry of the underlying objective f unction. We consider a fundamental setting in which the objective function is qu adratic, and provide the first tight characterization of the optimal Hessian-dep endent sample complexity. Our contribution is twofold. First, from an informatio n-theoretic point of view, we prove tight lower bounds on Hessian-dependent comp lexities by introducing a concept called \emph{energy allocation}, which capture s the interaction between the searching algorithm and the geometry of objective functions. A matching upper bound is obtained by solving the optimal energy spec trum. Then, algorithmically, we show the existence of a Hessian-independent algo rithm that universally achieves the asymptotic optimal sample complexities for a ll Hessian instances. The optimal sample complexities achieved by our algorithm remain valid for heavy-tailed noise distributions, which are enabled by a trunca tion method.
****
Training shallow ReLU networks on noisy data using hinge loss: when do we overfi

t and is it benign?
Erin George, Michael Murray, William Swartworth, Deanna Needell
We study benign overfitting in two-layer ReLU networks trained using gradient de
scent and hinge loss on noisy data for binary classification. In particular, we
consider linearly separable data for which a relatively small proportion of labe
ls are corrupted or flipped. We identify conditions on the margin of the clean d
ata that give rise to three distinct training outcomes: benign overfitting, in w
hich zero loss is achieved and with high probability test data is classified cor
rectly; overfitting, in which zero loss is achieved but test data is misclassifi
ed with probability lower bounded by a constant; and non-overfitting, in which c
lean points, but not corrupt points, achieve zero loss and again with high proba
bility test data is classified correctly. Our analysis provides a fine-grained d
escription of the dynamics of neurons throughout training and reveals two distin
ct phases: in the first phase clean points achieve close to zero loss, in the se
cond phase clean points oscillate on the boundary of zero loss while corrupt poi
nts either converge towards zero loss or are eventually zeroed by the network. W
e prove these results using a combinatorial approach that involves bounding the
number of clean versus corrupt updates during these phases of training.
****
Adaptive Algorithms for Relaxed Pareto Set Identification
Cyrille KONE, Emilie Kaufmann, Laura Richert
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
PHOTOSWAP: Personalized Subject Swapping in Images
Jing Gu, Yilin Wang, Nanxuan Zhao, Tsu-Jui Fu, Wei Xiong, Qing Liu, Zhifei Zhang
, HE Zhang, Jianming Zhang, HyunJoon Jung, Xin Eric Wang
In an era where images and visual content dominate our digital landscape, the ab
ility to manipulate and personalize these images has become a necessity.Envision
 seamlessly substituting a tabby cat lounging on a sunlit window sill in a photo
graph with your own playful puppy, all while preserving the original charm and c
omposition of the image. We present \emph{Photoswap}, a novel approach that enab
les this immersive image editing experience through personalized subject swappin
g in existing images.\emph{Photoswap} first learns the visual concept of the sub
ject from reference images and then swaps it into the target image using pre-tra
ined diffusion models in a training-free manner. We establish that a well-concep
tualized visual subject can be seamlessly transferred to any image with appropri
ate self-attention and cross-attention manipulation, maintaining the pose of the
 swapped subject and the overall coherence of the image. Comprehensive experimen
ts underscore the efficacy and controllability of \emph{Photoswap} in personaliz
ed subject swapping. Furthermore, \emph{Photoswap} significantly outperforms bas
eline methods in human ratings across subject swapping, background preservation,
 and overall quality, revealing its vast application potential, from entertainme
nt to professional editing.
****
Simplifying Neural Network Training Under Class Imbalance
Ravid Shwartz-Ziv, Micah Goldblum, Yucen Li, C. Bayan Bruss, Andrew G. Wilson
Real-world datasets are often highly class-imbalanced, which can adversely impac
t the performance of deep learning models. The majority of research on training
neural networks under class imbalance has focused on specialized loss functions
and sampling techniques. Notably, we demonstrate that simply tuning existing com
ponents of standard deep learning pipelines, such as the batch size, data augmen
tation, architecture size, pre-training, optimizer, and label smoothing, can ach
ieve state-of-the-art performance without any specialized loss functions or samp
lers. We also provide key prescriptions and considerations for training under cl
ass imbalance, and an understanding of why imbalance methods succeed or fail.
****
Regret Minimization via Saddle Point Optimization

Johannes Kirschner, Alireza Bakhtiari, Kushagra Chandak, Volodymyr Tkachuk, Csaba Szepesvari
A long line of works characterizes the sample complexity of regret minimization in sequential decision-making by min-max programs. In the corresponding saddle-point game, the min-player optimizes the sampling distribution against an adversarial max-player that chooses confusing models leading to large regret. The most recent instantiation of this idea is the decision-estimation coefficient (DEC), which was shown to provide nearly tight lower and upper bounds on the worst-case expected regret in structured bandits and reinforcement learning. By re-parametrizing the offset DEC with the confidence radius and solving the corresponding min-max program, we derive an anytime variant of the Estimation-To-Decisions algorithm (Anytime-E2D). Importantly, the algorithm optimizes the exploration-exploitation trade-off online instead of via the analysis. Our formulation leads to a practical algorithm for finite model classes and linear feedback models. We further point out connections to the information ratio, decoupling coefficient and PAC-DEC, and numerically evaluate the performance of E2D on simple examples.
****
On the Sublinear Regret of GP-UCB
Justin Whitehouse, Aaditya Ramdas, Steven Z. Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Train Once and Explain Everywhere: Pre-training Interpretable Graph Neural Networks
Jun Yin, Chaozhuo Li, Hao Yan, Jianxun Lian, Senzhang Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Quantum speedups for stochastic optimization
Aaron Sidford, Chenyi Zhang
We consider the problem of minimizing a continuous function given given access to a natural quantum generalization of a stochastic gradient oracle. We provide two new methods for the special case of minimizing a Lipschitz convex function. Each method obtains a dimension versus accuracy trade-off which is provably unachievable classically and we prove that one method is asymptotically optimal in low-dimensional settings. Additionally, we provide quantum algorithms for computing a critical point of a smooth non-convex function at rates not known to be achievable classically. To obtain these results we build upon the quantum multivariate mean estimation result of Cornelissen et al. and provide a general quantum variance reduction technique of independent interest.
****
Concept Algebra for (Score-Based) Text-Controlled Generative Models
Zihao Wang, Lin Gui, Jeffrey Negrea, Victor Veitch
This paper concerns the structure of learned representations in text-guided generative models, focusing on score-based models. A key property of such models is that they can compose disparate concepts in a 'disentangled' manner.This suggests these models have internal representations that encode concepts in a 'disentangled' manner. Here, we focus on the idea that concepts are encoded as subspaces of some representation space. We formalize what this means, show there's a natural choice for the representation, and develop a simple method for identifying the part of the representation corresponding to a given concept. In particular, this allows us to manipulate the concepts expressed by the model through algebraic manipulation of the representation. We demonstrate the idea with examples using Stable Diffusion.
****
Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics

Guillaume Mahey, Laetitia Chapel, Gilles Gasso, Clément Bonet, Nicolas Courty
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Aggregating Capacity in FL through Successive Layer Training for Computationally-Constrained Devices
Kilian Pfeiffer, Ramin Khalili, Joerg Henkel
Federated learning (FL) is usually performed on resource-constrained edge devices, e.g., with limited memory for the computation. If the required memory to train a model exceeds this limit, the device will be excluded from the training. This can lead to a lower accuracy as valuable data and computation resources are excluded from training, also causing bias and unfairness. The FL training process should be adjusted to such constraints. The state-of-the-art techniques propose training subsets of the FL model at constrained devices, reducing their resource requirements for training. However, these techniques largely limit the co-adaptation among parameters of the model and are highly inefficient, as we show: it is actually better to train a smaller (less accurate) model by the system where all the devices can train the model end-to-end than applying such techniques. We propose a new method that enables successive freezing and training of the parameters of the FL model at devices, reducing the training's resource requirements at the devices while still allowing enough co-adaptation between parameters. We show through extensive experimental evaluation that our technique greatly improves the accuracy of the trained model (by 52.4 p.p.) compared with the state of the art, efficiently aggregating the computation capacity available on distributed devices.
****
FiGURe: Simple and Efficient Unsupervised Node Representations with Filter Augmentations
Chanakya Ekbote, Ajinkya Deshpande, Arun Iyer, SUNDARARAJAN SELLAMANICKAM, Ramakrishna Bairi
Unsupervised node representations learnt using contrastive learning-based methods have shown good performance on downstream tasks. However, these methods rely on augmentations that mimic low-pass filters, limiting their performance on tasks requiring different eigen-spectrum parts. This paper presents a simple filter-based augmentation method to capture different parts of the eigen-spectrum. We show significant improvements using these augmentations. Further, we show that sharing the same weights across these different filter augmentations is possible, reducing the computational load. In addition, previous works have shown that good performance on downstream tasks requires high dimensional representations. Working with high dimensions increases the computations, especially when multiple augmentations are involved. We mitigate this problem and recover good performance through lower dimensional embeddings using simple random Fourier feature projections. Our method, FiGURe, achieves an average gain of up to 4.4\%, compared to the state-of-the-art unsupervised models, across all datasets in consideration, both homophilic and heterophilic. Our code can be found at: https://github.com/Microsoft/figure.
****
Mixed-Initiative Multiagent Apprenticeship Learning for Human Training of Robot Teams
Esmaeil Seraj, Jerry Xiong, Mariah Schrum, Matthew Gombolay
Extending recent advances in Learning from Demonstration (LfD) frameworks to multi-robot settings poses critical challenges such as environment non-stationarity due to partial observability which is detrimental to the applicability of existing methods. Although prior work has shown that enabling communication among agents of a robot team can alleviate such issues, creating inter-agent communication under existing Multi-Agent LfD (MA-LfD) frameworks requires the human expert to provide demonstrations for both environment actions and communication actions, which necessitates an efficient communication strategy on a known message space

s. To address this problem, we propose Mixed-Initiative Multi-Agent Apprenticesh ip Learning (MixTURE). MixTURE enables robot teams to learn from a human expert-generated data a preferred policy to accomplish a collaborative task, while simu ltaneously learning emergent inter-agent communication to enhance team coordinat ion. The key ingredient to MixTURE's success is automatically learning a communi cation policy, enhanced by a mutual-information maximizing reverse model that ra tionalizes the underlying expert demonstrations without the need for human gener ated data or an auxiliary reward function. MixTURE outperforms a variety of rele vant baselines on diverse data generated by human experts in complex heterogeneo us domains. MixTURE is the first MA-LfD framework to enable learning multi-robot collaborative policies directly from real human data, resulting in ~44% less hu man workload, and ~46% higher usability score.
****

Meta-Learning Adversarial Bandit Algorithms

Misha Khodak, Ilya Osadchiy, Keegan Harris, Maria-Florina F. Balcan, Kfir Y. Lev y, Ron Meir, Steven Z. Wu

We study online meta-learning with bandit feedback, with the goal of improving p erformance across multiple tasks if they are similar according to some natural s imilarity measure. As the first to target the adversarial online-within-online partial-information setting, we design meta-algorithms that combine outer learne rs to simultaneously tune the initialization and other hyperparameters of an inn er learner for two important cases: multi-armed bandits (MAB) and bandit linear optimization (BLO). For MAB, the meta-learners initialize and set hyperparamet ers of the Tsallis-entropy generalization of Exp3, with the task-averaged regret improving if the entropy of the optima-in-hindsight is small. For BLO, we lear n to initialize and tune online mirror descent (OMD) with self-concordant barrie r regularizers, showing that task-averaged regret varies directly with an action space-dependent measure they induce. Our guarantees rely on proving that unreg ularized follow-the-leader combined with two levels of low-dimensional hyperpara meter tuning is enough to learn a sequence of affine functions of non-Lipschitz and sometimes non-convex Bregman divergences bounding the regret of OMD.
****

Geometric Algebra Transformer

Johann Brehmer, Pim de Haan, Sönke Behrends, Taco S. Cohen

Problems involving geometric data arise in physics, chemistry, robotics, compute r vision, and many other fields. Such data can take numerous forms, for instance points, direction vectors, translations, or rotations, but to date there is no single architecture that can be applied to such a wide variety of geometric type s while respecting their symmetries. In this paper we introduce the Geometric Al gebra Transformer (GATr), a general-purpose architecture for geometric data. GAT r represents inputs, outputs, and hidden states in the projective geometric (or Clifford) algebra, which offers an efficient 16-dimensional vector-space represe ntation of common geometric objects as well as operators acting on them. GATr is equivariant with respect to E(3), the symmetry group of 3D Euclidean space. As a Transformer, GATr is versatile, efficient, and scalable. We demonstrate GATr i n problems from n-body modeling to wall-shear-stress estimation on large arteria l meshes to robotic motion planning. GATr consistently outperforms both non-geom etric and equivariant baselines in terms of error, data efficiency, and scalabil ity.
****

Top-Ambiguity Samples Matter: Understanding Why Deep Ensemble Works in Selective Classification

Qiang Ding, Yixuan Cao, Ping Luo

Selective classification allows a machine learning model to reject some hard inp uts and thus improve the reliability of its predictions. In this area, the ensem ble method is powerful in practice, but there has been no solid analysis on why the ensemble method works. Inspired by an interesting empirical result that the improvement of the ensemble largely comes from top-ambiguity samples where its m ember models diverge, we prove that, based on some assumptions, the ensemble has a lower selective risk than the member model for any coverage within a range. T

he proof is nontrivial since the selective risk is a non-convex function of the model prediction. The assumptions and the theoretical results are supported by systematic experiments on both computer vision and natural language processing tasks.
****

Unlimiformer: Long-Range Transformers with Unlimited Length Input
Amanda Bertsch, Uri Alon, Graham Neubig, Matthew Gormley
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Improving CLIP Training with Language Rewrites
Lijie Fan, Dilip Krishnan, Phillip Isola, Dina Katabi, Yonglong Tian
Contrastive Language-Image Pre-training (CLIP) stands as one of the most effective and scalable methods for training transferable vision models using paired image and text data. CLIP models are trained using contrastive loss, which typically relies on data augmentations to prevent overfitting and shortcuts. However, in the CLIP training paradigm, data augmentations are exclusively applied to image inputs, while language inputs remain unchanged throughout the entire training process, limiting the exposure of diverse texts to the same image. In this paper, we introduce Language augmented CLIP (LaCLIP), a simple yet highly effective approach to enhance CLIP training through language rewrites. Leveraging the in-context learning capability of large language models, we rewrite the text descriptions associated with each image. These rewritten texts exhibit diversity in sentence structure and vocabulary while preserving the original key concepts and meanings. During training, LaCLIP randomly selects either the original texts or the rewritten versions as text augmentations for each image. Extensive experiments on CC3M, CC12M, RedCaps and LAION-400M datasets show that CLIP pre-training with language rewrites significantly improves the transfer performance without computation or memory overhead during training. Specifically for ImageNet zero-shot accuracy, LaCLIP outperforms CLIP by 8.2% on CC12M and 2.4% on LAION-400M.
****

Extensible Prompts for Language Models on Zero-shot Language Style Customization
Tao Ge, Hu Jing, Li Dong, Shaoguang Mao, Yan Xia, Xun Wang, Si-Qing Chen, Furu Wei
We propose eXtensible Prompt (X-Prompt) for prompting a large language model (LLM) beyond natural language (NL). X-Prompt instructs an LLM with not only NL but also an extensible vocabulary of imaginary words. Registering new imaginary words allows us to instruct the LLM to comprehend concepts that are difficult to describe with NL words, thereby making a prompt more descriptive. Also, these imaginary words are designed to be out-of-distribution (OOD) robust so that they can be (re)used like NL words in various prompts, distinguishing X-Prompt from soft prompt that is for fitting in-distribution data. We propose context-augmented learning (CAL) to learn imaginary words for general usability, enabling them to work properly in OOD (unseen) prompts. We experiment X-Prompt for zero-shot language style customization as a case study. The promising results of X-Prompt demonstrate its potential to facilitate advanced interaction beyond the natural language interface, bridging the communication gap between humans and LLMs.
****

MIMEx: Intrinsic Rewards from Masked Input Modeling
Toru Lin, Allan Jabri
Exploring in environments with high-dimensional observations is hard. One promising approach for exploration is to use intrinsic rewards, which often boils down to estimating "novelty" of states, transitions, or trajectories with deep networks. Prior works have shown that conditional prediction objectives such as masked autoencoding can be seen as stochastic estimation of pseudo-likelihood. We show how this perspective naturally leads to a unified view on existing intrinsic reward approaches: they are special cases of conditional prediction, where the estimation of novelty can be seen as pseudo-likelihood estimation with different m

ask distributions. From this view, we propose a general framework for deriving intrinsic rewards -- Masked Input Modeling for Exploration (MIMEx) -- where the mask distribution can be flexibly tuned to control the difficulty of the underlying conditional prediction task. We demonstrate that MIMEx can achieve superior results when compared against competitive baselines on a suite of challenging sparse-reward visuomotor tasks.
****
RGMIL: Guide Your Multiple-Instance Learning Model with Regressor
Zhaolong Du, Shasha Mao, Yimeng Zhang, Shuiping Gou, Licheng Jiao, Lin Xiong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Stochastic Multi-armed Bandits: Optimal Trade-off among Optimality, Consistency, and Tail Risk
David Simchi-Levi, Zeyu Zheng, Feng Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Mask-aware CLIP Representations for Zero-Shot Segmentation
Siyu Jiao, Yunchao Wei, Yaowei Wang, Yao Zhao, Humphrey Shi
Recently, pre-trained vision-language models have been increasingly used to tackle the challenging zero-shot segmentation task. Typical solutions follow the paradigm of first generating mask proposals and then adopting CLIP to classify them. To maintain the CLIP's zero-shot transferability, previous practices favour to freeze CLIP during training. However, in the paper, we reveal that CLIP is insensitive to different mask proposals and tends to produce similar predictions for various mask proposals of the same image. This insensitivity results in numerous false positives when classifying mask proposals. This issue mainly relates to the fact that CLIP is trained with image-level supervision. To alleviate this issue, we propose a simple yet effective method, named Mask-aware Fine-tuning (MAFT). Specifically, Image-Proposals CLIP Encoder (IP-CLIP Encoder) is proposed to handle arbitrary numbers of image and mask proposals simultaneously. Then, mask-aware loss and self-distillation loss are designed to fine-tune IP-CLIP Encoder, ensuring CLIP is responsive to different mask proposals while not sacrificing transferability. In this way, mask-aware representations can be easily learned to make the true positives stand out. Notably, our solution can seamlessly plug into most existing methods without introducing any new parameters during the fine-tuning process. We conduct extensive experiments on the popular zero-shot benchmarks. With MAFT, the performance of the state-of-the-art methods is promoted by a large margin: 50.4\% (+ 8.2\%) on COCO, 81.8\% (+ 3.2\%) on Pascal-VOC, and 8.7\% (+4.3\%) on ADE20K in terms of mIoU for unseen classes. Codes will be provided for reproducibility. Code is available at https://github.com/jiaosiyu1999/MAFT.git .
****
Understanding Multi-phase Optimization Dynamics and Rich Nonlinear Behaviors of ReLU Networks
Mingze Wang, Chao Ma
The training process of ReLU neural networks often exhibits complicated nonlinear phenomena. The nonlinearity of models and non-convexity of loss pose significant challenges for theoretical analysis. Therefore, most previous theoretical works on the optimization dynamics of neural networks focus either on local analysis (like the end of training) or approximate linear models (like Neural Tangent Kernel). In this work, we conduct a complete theoretical characterization of the training process of a two-layer ReLU network trained by Gradient Flow on a linearly separable data. In this specific setting, our analysis captures the whole optimization process starting from random initialization to final convergence. Des

pite the relatively simple model and data that we studied, we reveal four different phases from the whole training process showing a general simplifying-to-complicating learning trend.Specific nonlinear behaviors can also be precisely identified and captured theoretically, such asinitial condensation, saddle-to-plateau dynamics, plateau escape, changes of activation patterns, learning with increasing complexity, etc.
****
RD-Suite: A Benchmark for Ranking Distillation
Zhen Qin, Rolf Jagerman, Rama Kumar Pasumarthi, Honglei Zhuang, He Zhang, Aijun Bai, Kai Hui, Le Yan, Xuanhui Wang
The distillation of ranking models has become an important topic in both academia and industry. In recent years, several advanced methods have been proposed to tackle this problem, often leveraging ranking information from teacher rankers that is absent in traditional classification settings. To date, there is no well-established consensus on how to evaluate this class of models. Moreover, inconsistent benchmarking on a wide range of tasks and datasets make it difficult to assess or invigorate advances in this field. This paper first examines representative prior arts on ranking distillation, and raises three questions to be answered around methodology and reproducibility. To that end, we propose a systematic and unified benchmark, Ranking Distillation Suite (RD-Suite), which is a suite of tasks with 4 large real-world datasets, encompassing two major modalities (textual and numeric) and two applications (standard distillation and distillation transfer). RD-Suite consists of benchmark results that challenge some of the common wisdom in the field, and the release of datasets with teacher scores and evaluation scripts for future research. RD-Suite paves the way towards better understanding of ranking distillation, facilities more research in this direction, and presents new challenges.
****
Gradient Descent with Linearly Correlated Noise: Theory and Applications to Differential Privacy
Anastasiia Koloskova, Ryan McKenna, Zachary Charles, John Rush, H. Brendan McMahan
We study gradient descent under linearly correlated noise. Our work is motivated by recent practical methods for optimization with differential privacy (DP), such as DP-FTRL, which achieve strong performance in settings where privacy amplification techniques are infeasible (such as in federated learning). These methods inject privacy noise through a matrix factorization mechanism, making the noise linearly correlated over iterations. We propose a simplified setting that distills key facets of these methods and isolates the impact of linearly correlated noise. We analyze the behavior of gradient descent in this setting, for both convex and non-convex functions. Our analysis is demonstrably tighter than prior work and recovers multiple important special cases exactly (including anticorrelated perturbed gradient descent). We use our results to develop new, effective matrix factorizations for differentially private optimization, and highlight the benefits of these factorizations theoretically and empirically.
****
A Framework for Fast and Stable Representations of Multiparameter Persistent Homology Decompositions
David Loiseaux, Mathieu Carrière, Andrew Blumberg
Topological data analysis (TDA) is an area of data science that focuses on using invariants from algebraic topology to provide multiscale shape descriptors for geometric data sets such as point clouds. One of the most important such descriptors is persistent homology, which encodes the change in shape as a filtration parameter changes; a typical parameter is the feature scale. For many data sets, it is useful to simultaneously vary multiple filtration parameters, for example feature scale and density. While the theoretical properties of single parameter persistent homology are well understood, less is known about the multiparameter case. A central question is the problem of representing multiparameter persistent homology by elements of a vector space for integration with standard machine learning algorithms. Existing approaches to this problem either ignore most of t

he multiparameter information to reduce to the one-parameter case or are heurist
ic and potentially unstable in the face of noise. In this article, we introduce
a new general representation framework that leverages recent results on decompos
itions of multiparameter persistent homology. This framework is rich in informat
ion, fast to compute, and encompasses previous approaches. Moreover, we establis
h theoretical stability guarantees under this framework as well as efficient alg
orithms for practical computation, making this framework an applicable and versa
tile tool for analyzing geometric and point cloud data. We validate our stabilit
y results and algorithms with numerical experiments that demonstrate statistical
 convergence, prediction accuracy, and fast running times on several real data s
ets.
****
Objaverse-XL: A Universe of 10M+ 3D Objects
Matt Deitke, Ruoshi Liu, Matthew Wallingford, Huong Ngo, Oscar Michel, Aditya Ku
supati, Alan Fan, Christian Laforte, Vikram Voleti, Samir Yitzhak Gadre, Eli Van
derBilt, Aniruddha Kembhavi, Carl Vondrick, Georgia Gkioxari, Kiana Ehsani, Ludw
ig Schmidt, Ali Farhadi
Natural language processing and 2D vision models have attained remarkable profic
iency on many tasks primarily by escalating the scale of training data. However,
 3D vision tasks have not seen the same progress, in part due to the challenges
of acquiring high-quality 3D data. In this work, we present Objaverse-XL, a data
set of over 10 million 3D objects. Our compilation comprises deduplicated 3D obj
ects from a diverse set of sources, including manually designed objects, photogr
ammetry scans of landmarks and everyday items, and professional scans of histori
c and antique artifacts. Representing the largest scale and diversity in the rea
lm of 3D datasets, Objaverse-XL enables significant new possibilities for 3D vis
ion. Our experiments demonstrate the vast improvements enabled with the scale pr
ovided by Objaverse-XL. We show that by training Zero123 on novel view synthesis
, utilizing over 100 million multi-view rendered images, we achieve strong zero-
shot generalization abilities. We hope that releasing Objaverse-XL will enable f
urther innovations in the field of 3D vision at scale.
****
Should We Learn Most Likely Functions or Parameters?
Shikai Qiu, Tim G. J. Rudner, Sanyam Kapoor, Andrew G. Wilson
Standard regularized training procedures correspond to maximizing a posterior di
stribution over parameters, known as maximum a posteriori (MAP) estimation. Howe
ver, model parameters are of interest only insomuch as they combine with the fun
ctional form of a model to provide a function that can make good predictions. Mo
reover, the most likely parameters under the parameter posterior do not generall
y correspond to the most likely function induced by the parameter posterior. In
fact, we can re-parametrize a model such that any setting of parameters can maxi
mize the parameter posterior. As an alternative, we investigate the benefits and
 drawbacks of directly estimating the most likely function implied by the model
and the data. We show that this procedure leads to pathological solutions when u
sing neural networks and prove conditions under which the procedure is well-beha
ved, as well as a scalable approximation. Under these conditions, we find that f
unction-space MAP estimation can lead to flatter minima, better generalization,
and improved robustness to overfitting.
****
Geometry-Informed Neural Operator for Large-Scale 3D PDEs
Zongyi Li, Nikola Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Otta, Mo
hammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli,
 Animashree Anandkumar
We propose the geometry-informed neural operator (GINO), a highly efficient appr
oach for learning the solution operator of large-scale partial differential equa
tions with varying geometries. GINO uses a signed distance function (SDF) repres
entation of the input shape and neural operators based on graph and Fourier arch
itectures to learn the solution operator. The graph neural operator handles irre
gular grids and transforms them into and from regular latent grids on which Four
ier neural operator can be efficiently applied. We provide an efficient implemen

tation of GINO using an optimized hashing approach, which allows efficient learn ing in a shared, compressed latent space with reduced computation and memory co sts. GINO is discretization-invariant, meaning the trained model can be applied to arbitrary discretizations of the continuous domain and applies to any shape o r resolution. To empirically validate the performance of our method on large-sc ale simulation, we generate the industry-standard aerodynamics dataset of 3D veh icle geometries with Reynolds numbers as high as five million. For this large-sc ale 3D fluid simulation, numerical methods are expensive to compute surface pres sure. We successfully trained GINO to predict the pressure on car surfaces using only five hundred data points. The cost-accuracy experiments show a 26,000x spe ed-up compared to optimized GPU-based computational fluid dynamics (CFD) simulat ors on computing the drag coefficient. When tested on new combinations of geomet ries and boundary conditions (inlet velocities), GINO obtains a one-fourth reduc tion in error rate compared to deep neural network approaches.
****
Differentially Private Image Classification by Learning Priors from Random Proce sses
Xinyu Tang, Ashwinee Panda, Vikash Sehwag, Prateek Mittal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
ImageNet-Hard: The Hardest Images Remaining from a Study of the Power of Zoom an d Spatial Biases in Image Classification
Mohammad Reza Taesiri, Giang Nguyen, Sarra Habchi, Cor-Paul Bezemer, Anh Nguyen
Image classifiers are information-discarding machines, by design. Yet, how these models discard information remains mysterious. We hypothesize that one way for image classifiers to reach high accuracy is to first zoom to the most discrimina tive region in the image and then extract features from there to predict image l abels, discarding the rest of the image. Studying six popular networks ranging f rom AlexNet to CLIP, we find that proper framing of the input image can lead to the correct classification of 98.91% of ImageNet images. Furthermore, we uncove r positional biases in various datasets, especially a strong center bias in two popular datasets: ImageNet-A and ObjectNet. Finally, leveraging our insights int o the potential of zooming, we propose a test-time augmentation (TTA) technique that improves classification accuracy by forcing models to explicitly perform zo om-in operations before making predictions.Our method is more interpretable, acc urate, and faster than MEMO, a state-of-the-art (SOTA) TTA method. We introduce ImageNet-Hard, a new benchmark that challenges SOTA classifiers including large vision-language models even when optimal zooming is allowed.
****
NuTrea: Neural Tree Search for Context-guided Multi-hop KGQA
Hyeong Kyu Choi, Seunghun Lee, Jaewon Chu, Hyunwoo J. Kim
Multi-hop Knowledge Graph Question Answering (KGQA) is a task that involves retr ieving nodes from a knowledge graph (KG) to answer natural language questions. Recent GNN-based approaches formulate this task as a KG path searching problem, where messages are sequentially propagated from the seed node towards the answer nodes. However, these messages are past-oriented, and they do not consider the full KG context. To make matters worse, KG nodes often represent pronoun entitie s and are sometimes encrypted, being uninformative in selecting between paths. T o address these problems, we propose Neural Tree Search (NuTrea), a tree search-based GNN model that incorporates the broader KG context. Our model adopts a mes sage-passing scheme that probes the unreached subtree regions to boost the past-oriented embeddings. In addition, we introduce the Relation Frequency-Inverse En tity Frequency (RF-IEF) node embedding that considers the global KG context to b etter characterize ambiguous KG nodes. The general effectiveness of our approach is demonstrated through experiments on three major multi-hop KGQA benchmark dat asets, and our extensive analyses further validate its expressiveness and robust ness. Overall, NuTrea provides a powerful means to query the KG with complex nat

ural language questions. Code is available at https://github.com/mlvlab/NuTrea.
****
Anonymous Learning via Look-Alike Clustering: A Precise Analysis of Model Generalization
Adel Javanmard, Vahab Mirrokni
While personalized recommendations systems have become increasingly popular, ensuring user data protection remains a top concern in the development of these learning systems. A common approach to enhancing privacy involves training models using anonymous data rather than individual data. In this paper, we explore a natural technique called "look-alike clustering", which involves replacing sensitive features of individuals with the cluster's average values. We provide a precise analysis of how training models using anonymous cluster centers affects their generalization capabilities. We focus on an asymptotic regime where the size of the training set grows in proportion to the features dimension. Our analysis is based on the Convex Gaussian Minimax Theorem (CGMT) and allows us to theoretically understand the role of different model components on the generalization error. In addition, we demonstrate that in certain high-dimensional regimes, training over anonymous cluster centers acts as a regularization and improves generalization error of the trained models. Finally, we corroborate our asymptotic theory with finite-sample numerical experiments where we observe a perfect match when the sample size is only of order of a few hundreds.
****
Saving 100x Storage: Prototype Replay for Reconstructing Training Sample Distribution in Class-Incremental Semantic Segmentation
Jinpeng Chen, Runmin Cong, Yuxuan LUO, Horace Ip, Sam Kwong
Existing class-incremental semantic segmentation (CISS) methods mainly tackle catastrophic forgetting and background shift, but often overlook another crucial issue. In CISS, each step focuses on different foreground classes, and the training set for a single step only includes images containing pixels of the current foreground classes, excluding images without them. This leads to an overrepresentation of these foreground classes in the single-step training set, causing the classification biased towards these classes. To address this issue, we present STAR, which preserves the main characteristics of each past class by storing a compact prototype and necessary statistical data, and aligns the class distribution of single-step training samples with the complete dataset by replaying these prototypes and repeating background pixels with appropriate frequency. Compared to the previous works that replay raw images, our method saves over 100 times the storage while achieving better performance. Moreover, STAR incorporates an old-class features maintaining (OCFM) loss, keeping old-class features unchanged while preserving sufficient plasticity for learning new classes. Furthermore, a similarity-aware discriminative (SAD) loss is employed to specifically enhance the feature diversity between similar old-new class pairs. Experiments on two public datasets, Pascal VOC 2012 and ADE20K, reveal that our model surpasses all previous state-of-the-art methods.
****
Skill-it! A data-driven skills framework for understanding and training language models
Mayee Chen, Nicholas Roberts, Kush Bhatia, Jue WANG, Ce Zhang, Frederic Sala, Christopher Ré
The quality of training data impacts the performance of pre-trained large language models (LMs). Given a fixed budget of tokens, we study how to best select data that leads to good downstream model performance across tasks. We develop a new framework based on a simple hypothesis: just as humans acquire interdependent skills in a deliberate order, language models also follow a natural order when learning a set of skills from their training data. If such an order exists, it can be utilized for improved understanding of LMs and for data-efficient training. Using this intuition, our framework formalizes the notion of a skill and of an ordered set of skills in terms of the associated data. First, using both synthetic and real data, we demonstrate that these ordered skill sets exist, and that their existence enables more advanced skills to be learned with less data when we

train on their prerequisite skills. Second, using our proposed framework, we introduce an online data sampling algorithm, Skill-It, over mixtures of skills for both continual pre-training and fine-tuning regimes, where the objective is to efficiently learn multiple skills in the former and an individual skill in the latter. On the LEGO synthetic in the continual pre-training setting, Skill-It obtains 37.5 points higher accuracy than random sampling. On the Natural Instructions dataset in the fine-tuning setting, Skill-It reduces the validation loss on the target skill by 13.6% versus training on data associated with the target skill itself. We apply our skills framework on the RedPajama dataset to continually pre-train a 3B-parameter LM, achieving higher accuracy on the LM Evaluation Harness with 1B tokens than the baseline approach of sampling uniformly over data sources with 3B tokens.
****
Strategic Behavior in Two-sided Matching Markets with Prediction-enhanced Preference-formation
Stefania Ionescu, Yuhao Du, Kenneth Joseph, Ancsa Hannak
Two-sided matching markets have long existed to pair agents in the absence of regulated exchanges. A common example is school choice, where a matching mechanism uses student and school preferences to assign students to schools. In such settings, forming preferences is both difficult and critical. Prior work has suggested various prediction mechanisms that help agents make decisions about their preferences. Although often deployed together, these matching and prediction mechanisms are almost always analyzed separately. The present work shows that at the intersection of the two lies a previously unexplored type of strategic behavior: agents returning to the market (e.g., schools) can attack future predictions by interacting short-term non-optimally with their matches. Here, we first introduce this type of strategic behavior, which we call an adversarial interaction attack. Next, we construct a formal economic model that captures the feedback loop between prediction mechanisms designed to assist agents and the matching mechanism used to pair them. Finally, in a simplified setting, we prove that returning agents can benefit from using adversarial interaction attacks and gain progressively more as the trust in and accuracy of predictions increases. We also show that this attack increases inequality in the student population.
****
Sample Complexity of Forecast Aggregation
Tao Lin, Yiling Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning to Influence Human Behavior with Offline Reinforcement Learning
Joey Hong, Sergey Levine, Anca Dragan
When interacting with people, AI agents do not just influence the state of the world -- they also influence the actions people take in response to the agent, and even their underlying intentions and strategies. Accounting for and leveraging this influence has mostly been studied in settings where it is sufficient to assume that human behavior is near-optimal: competitive games, or general-sum settings like autonomous driving alongside human drivers. Instead, we focus on influence in settings where there is a need to capture human suboptimality. For instance, imagine a collaborative task in which, due either to cognitive biases or lack of information, people do not perform very well -- how could an agent influence them towards more optimal behavior? Assuming near-optimal human behavior will not work here, and so the agent needs to learn from real human data. But experimenting online with humans is potentially unsafe, and creating a high-fidelity simulator of the environment is often impractical. Hence, we focus on learning from an offline dataset of human-human interactions. Our observation is that offline reinforcement learning (RL) can learn to effectively influence suboptimal humans by extending and combining elements of observed human-human behavior. We demonstrate that offline RL can solve two challenges with effective influence. Fir

st, we show that by learning from a dataset of suboptimal human-human interaction on a variety of tasks -- none of which contains examples of successful influence -- an agent can learn influence strategies to steer humans towards better performance even on new tasks. Second, we show that by also modeling and conditioning on human behavior, offline RL can learn to affect not just the human's actions but also their underlying strategy, and adapt to changes in their strategy.
****
Discriminative Calibration: Check Bayesian Computation from Simulations and Flexible Classifier
Yuling Yao, Justin Domke
To check the accuracy of Bayesian computations, it is common to use rank-based simulation-based calibration (SBC). However, SBC has drawbacks: The test statistic is somewhat ad-hoc, interactions are difficult to examine, multiple testing is a challenge, and the resulting p-value is not a divergence metric. We propose to replace the marginal rank test with a flexible classification approach that learns test statistics from data. This measure typically has a higher statistical power than the SBC test and returns an interpretable divergence measure of miscalibration, computed from classification accuracy. This approach can be used with different data generating processes to address simulation-based inference or traditional inference methods like Markov chain Monte Carlo or variational inference. We illustrate an automated implementation using neural networks and statistically-inspired features, and validate the method with numerical and real data experiments.
****
Epidemic Learning: Boosting Decentralized Learning with Randomized Communication
Martijn De Vos, Sadegh Farhadkhani, Rachid Guerraoui, Anne-marie Kermarrec, Rafael Pires, Rishi Sharma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Global Identifiability of ℓ_1-based Dictionary Learning via Matrix Volume Optimization
Jingzhou Hu, Kejun Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Memory-Efficient Fine-Tuning of Compressed Large Language Models via sub-4-bit Integer Quantization
Jeonghoon Kim, Jung Hyun Lee, Sungdong Kim, Joonsuk Park, Kang Min Yoo, Se Jung Kwon, Dongsoo Lee
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Corruption-Robust Offline Reinforcement Learning with General Function Approximation
Chenlu Ye, Rui Yang, Quanquan Gu, Tong Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Training Fully Connected Neural Networks is $\exists\mathbb{R}$-Complete
Daniel Bertschinger, Christoph Hertrich, Paul Jungeblut, Tillmann Miltzow, Simon Weber

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Uncertainty Estimation for Safety-critical Scene Segmentation via Fine-grained Reward Maximization
Hongzheng Yang, Cheng Chen, Yueyao CHEN, Scheppach, Hon Chi Yip, DOU QI
Uncertainty estimation plays an important role for future reliable deployment of deep segmentation models in safety-critical scenarios such as medical applications. However, existing methods for uncertainty estimation have been limited by the lack of explicit guidance for calibrating the prediction risk and model confidence. In this work, we propose a novel fine-grained reward maximization (FGRM) framework, to address uncertainty estimation by directly utilizing an uncertainty metric related reward function with a reinforcement learning based model tuning algorithm. This would benefit the model uncertainty estimation with direct optimization guidance for model calibration. Specifically, our method designs a new uncertainty estimation reward function using the calibration metric, which is maximized to fine-tune an evidential learning pre-trained segmentation model for calibrating prediction risk. Importantly, we innovate an effective fine-grained parameter update scheme, which imposes fine-grained reward-weighting of each network parameter according to the parameter importance quantified by the fisher information matrix. To the best of our knowledge, this is the first work exploring reward optimization for model uncertainty estimation in safety-critical vision tasks. The effectiveness of our method is demonstrated on two large safety-critical surgical scene segmentation datasets under two different uncertainty estimation settings. With real-time one forward pass at inference, our method outperforms state-of-the-art methods by a clear margin on all the calibration metrics of uncertainty estimation, while maintaining a high task accuracy for the segmentation results. Code is available at https://github.com/med-air/FGRM.
****
GLIME: General, Stable and Local LIME Explanation
Zeren Tan, Yang Tian, Jian Li
As black-box machine learning models become more complex and are applied in high-stakes settings, the need for providing explanations for their predictions becomes crucial. Although Local Interpretable Model-agnostic Explanations (LIME) \cite{ribeiro2016should} is a widely adopted method for understanding model behavior, it suffers from instability with respect to random seeds \cite{zafar2019dlime, shankaranarayana2019alime, bansal2020sam} and exhibits low local fidelity (i.e., how the explanation explains model's local behaviors) \cite{rahnama2019study, laugel2018defining}. Our study demonstrates that this instability is caused by small sample weights, resulting in the dominance of regularization and slow convergence. Additionally, LIME's sampling approach is non-local and biased towards the reference, leading to diminished local fidelity and instability to references. To address these challenges, we propose \textsc{Glime}, an enhanced framework that extends LIME and unifies several previous methods. Within the \textsc{Glime} framework, we derive an equivalent formulation of LIME that achieves significantly faster convergence and improved stability. By employing a local and unbiased sampling distribution, \textsc{Glime} generates explanations with higher local fidelity compared to LIME, while being independent of the reference choice. Moreover, \textsc{Glime} offers users the flexibility to choose sampling distribution based on their specific scenarios.
****
Efficient Symbolic Policy Learning with Differentiable Symbolic Expression
Jiaming Guo, Rui Zhang, Shaohui Peng, Qi Yi, Xing Hu, Ruizhi Chen, Zidong Du, xishan zhang, Ling Li, Qi Guo, Yunji Chen
Deep reinforcement learning (DRL) has led to a wide range of advances in sequential decision-making tasks. However, the complexity of neural network policies makes it difficult to understand and deploy with limited computational resources. Currently, employing compact symbolic expressions as symbolic policies is a prom

ising strategy to obtain simple and interpretable policies. Previous symbolic po
licy methods usually involve complex training processes and pre-trained neural n
etwork policies, which are inefficient and limit the application of symbolic pol
icies. In this paper, we propose an efficient gradient-based learning method nam
ed Efficient Symbolic Policy Learning (ESPL) that learns the symbolic policy fro
m scratch in an end-to-end way. We introduce a symbolic network as the search sp
ace and employ a path selector to find the compact symbolic policy. By doing so
we represent the policy with a differentiable symbolic expression and train it i
n an off-policy manner which further improves the efficiency. In addition, in co
ntrast with previous symbolic policies which only work in single-task RL because
 of complexity, we expand ESPL on meta-RL to generate symbolic policies for unse
en tasks. Experimentally, we show that our approach generates symbolic policies
with higher performance and greatly improves data efficiency for single-task RL.
 In meta-RL, we demonstrate that compared with neural network policies the propo
sed symbolic policy achieves higher performance and efficiency and shows the pot
ential to be interpretable.
****
PAC-Bayesian Spectrally-Normalized Bounds for Adversarially Robust Generalizatio
n
Jiancong Xiao, Ruoyu Sun, Zhi-Quan Luo
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Neural Graph Generation from Graph Statistics
Kiarash Zahirnia, Yaochen Hu, Mark Coates, Oliver Schulte
We describe a new setting for learning a deep graph generative model (GGM) from
aggregate graph statistics, rather than from the graph adjacency matrix. Matchin
g the statistics of observed training graphs is the main approach for learning t
raditional GGMs (e.g, BTER, Chung-Lu, and Erdos-Renyi models). Privacy research
ers have proposed learning from graph statistics as a way to protect privacy. We
 develop an architecture for training a deep GGM to match statistics while pres
erving local differential privacy guarantees. Empirical evaluation on 8 datasets
 indicates that our deep GGM model generates more realistic graphs than the trad
itional GGMs when both are learned from graph statistics only. We also benchmark
 our deep GGM trained on statistics only, against state-of-the-art deep GGM mode
ls that are trained on the entire adjacency matrix. The results show that graph
statistics are often sufficient to build a competitive deep GGM that generates r
ealistic graphs while protecting local privacy.
****
DIN-SQL: Decomposed In-Context Learning of Text-to-SQL with Self-Correction
Mohammadreza Pourreza, Davood Rafiei
There is currently a significant gap between the performance of fine-tuned model
s and prompting approaches using Large Language Models (LLMs) on the challenging
 task of text-to-SQL, as evaluated on datasets such as Spider. To improve the pe
rformance of LLMs in the reasoning process, we study how decomposing the task in
to smaller sub-tasks can be effective. In particular, we show that breaking down
 the generation problem into sub-problems and feeding the solutions of those sub
-problems into LLMs can be an effective approach for significantly improving the
ir performance. Our experiments with three LLMs show that this approach consist
ently improves their simple few-shot performance by roughly 10%, pushing the acc
uracy of LLMs towards SOTA or surpassing it. On the holdout test set of Spider,
the SOTA, in terms of execution accuracy, was 79.9 and the new SOTA at the time
of this writing using our approach is 85.3. Our approach with in-context learni
ng beats many heavily fine-tuned models by at least 5%. Additionally, when evalu
ated on the BIRD benchmark, our approach achieved an execution accuracy of 55.9%
, setting a new SOTA on its holdout test set.
****
Scaling Up Differentially Private LASSO Regularized Logistic Regression via Fast

er Frank-Wolfe Iterations
Edward Raff, Amol Khanna, Fred Lu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Uncoupled and Convergent Learning in Two-Player Zero-Sum Markov Games with Bandit Feedback
Yang Cai, Haipeng Luo, Chen-Yu Wei, Weiqiang Zheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Deductive Verification of Chain-of-Thought Reasoning
Zhan Ling, Yunhao Fang, Xuanlin Li, Zhiao Huang, Mingu Lee, Roland Memisevic, Hao Su
Large Language Models (LLMs) significantly benefit from Chain-of-thought (CoT) prompting in performing various reasoning tasks. While CoT allows models to produce more comprehensive reasoning processes, its emphasis on intermediate reasoning steps can inadvertently introduce hallucinations and accumulated errors, thereby limiting models’ ability to solve complex reasoning tasks. Inspired by how humans engage in careful and meticulous deductive logical reasoning processes to solve tasks, we seek to enable language models to perform explicit and rigorous deductive reasoning, and also ensure the trustworthiness of their reasoning process through self-verification. However, directly verifying the validity of an entire deductive reasoning process is challenging, even with advanced models like ChatGPT. In light of this, we propose to decompose a reasoning verification process into a series of step-by-step subprocesses, each only receiving their necessary context and premises. To facilitate this procedure, we propose Natural Program, a natural language-based deductive reasoning format. Our approach enables models to generate precise reasoning steps where subsequent steps are more rigorously grounded on prior steps. It also empowers language models to carry out reasoning self-verification in a step-by-step manner. By integrating this verification process into each deductive reasoning stage, we significantly enhance the rigor and trustfulness of generated reasoning steps. Along this process, we also improve the answer correctness on complex reasoning tasks.
****
Rigorous Runtime Analysis of MOEA/D for Solving Multi-Objective Minimum Weight Base Problems
Anh Viet Do, Aneta Neumann, Frank Neumann, Andrew Sutton
We study the multi-objective minimum weight base problem, an abstraction of classical NP-hard combinatorial problems such as the multi-objective minimum spanning tree problem. We prove some important properties of the convex hull of the non-dominated front, such as its approximation quality and an upper bound on the number of extreme points. Using these properties, we give the first run-time analysis of the MOEA/D algorithm for this problem, an evolutionary algorithm that effectively optimizes by decomposing the objectives into single-objective components. We show that the MOEA/D, given an appropriate decomposition setting, finds all extreme points within expected fixed-parameter polynomial time, in the oracle model. Experiments are conducted on random bi-objective minimum spanning tree instances, and the results agree with our theoretical findings. Furthermore, compared with a previously studied evolutionary algorithm for the problem GSEMO, MOEA/D finds all extreme points much faster across all instances.
****
Implicit Transfer Operator Learning: Multiple Time-Resolution Models for Molecular Dynamics
Mathias Schreiner, Ole Winther, Simon Olsson
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Causal de Finetti: On the Identification of Invariant Causal Structure in Exchangeable Data
Siyuan Guo, Viktor Toth, Bernhard Schölkopf, Ferenc Huszar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Batch Bayesian Optimization For Replicable Experimental Design
Zhongxiang Dai, Quoc Phong Nguyen, Sebastian Tay, Daisuke Urano, Richalynn Leong, Bryan Kian Hsiang Low, Patrick Jaillet
Many real-world experimental design problems (a) evaluate multiple experimental conditions in parallel and (b) replicate each condition multiple times due to large and heteroscedastic observation noise. Given a fixed total budget, this naturally induces a trade-off between evaluating more unique conditions while replicating each of them fewer times vs. evaluating fewer unique conditions and replicating each more times. Moreover, in these problems, practitioners may be risk-averse and hence prefer an input with both good average performance and small variability. To tackle both challenges, we propose the Batch Thompson Sampling for Replicable Experimental Design (BTS-RED) framework, which encompasses three algorithms. Our BTS-RED-Known and BTS-RED-Unknown algorithms, for, respectively, known and unknown noise variance, choose the number of replications adaptively rather than deterministically such that an input with a larger noise variance is replicated more times. As a result, despite the noise heteroscedasticity, both algorithms enjoy a theoretical guarantee and are asymptotically no-regret. Our Mean-Var-BTS-RED algorithm aims at risk-averse optimization and is also asymptotically no-regret. We also show the effectiveness of our algorithms in two practical real-world applications: precision agriculture and AutoML.
****
Contrastive Modules with Temporal Attention for Multi-Task Reinforcement Learning
Siming Lan, Rui Zhang, Qi Yi, Jiaming Guo, Shaohui Peng, Yunkai Gao, Fan Wu, Ruizhi Chen, Zidong Du, Xing Hu, xishan zhang, Ling Li, Yunji Chen
In the field of multi-task reinforcement learning, the modular principle, which involves specializing functionalities into different modules and combining them appropriately, has been widely adopted as a promising approach to prevent the negative transfer problem that performance degradation due to conflicts between tasks. However, most of the existing multi-task RL methods only combine shared modules at the task level, ignoring that there may be conflicts within the task. In addition, these methods do not take into account that without constraints, some modules may learn similar functions, resulting in restricting the model's expressiveness and generalization capability of modular methods.In this paper, we propose the Contrastive Modules with Temporal Attention(CMTA) method to address these limitations. CMTA constrains the modules to be different from each other by contrastive learning and combining shared modules at a finer granularity than the task level with temporal attention, alleviating the negative transfer within the task and improving the generalization ability and the performance for multi-task RL.We conducted the experiment on Meta-World, a multi-task RL benchmark containing various robotics manipulation tasks. Experimental results show that CMTA outperforms learning each task individually for the first time and achieves substantial performance improvements over the baselines.
****
Scalable Primal-Dual Actor-Critic Method for Safe Multi-Agent RL with General Utilities
Donghao Ying, Yunkai Zhang, Yuhao Ding, Alec Koppel, Javad Lavaei
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Optimal Transport-Guided Conditional Score-Based Diffusion Model
Xiang Gu, Liwei Yang, Jian Sun, Zongben Xu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
GNeSF: Generalizable Neural Semantic Fields
Hanlin Chen, Chen Li, Mengqi Guo, Zhiwen Yan, Gim Hee Lee
3D scene segmentation based on neural implicit representation has emerged recently with the advantage of training only on 2D supervision. However, existing approaches still requires expensive per-scene optimization that prohibits generalization to novel scenes during inference. To circumvent this problem, we introduce a \textit{generalizable} 3D segmentation framework based on implicit representation. Specifically, our framework takes in multi-view image features and semantic maps as the inputs instead of only spatial information to avoid overfitting to scene-specific geometric and semantic information. We propose a novel soft voting mechanism to aggregate the 2D semantic information from different views for each 3D point. In addition to the image features, view difference information is also encoded in our framework to predict the voting scores. Intuitively, this allows the semantic information from nearby views to contribute more compared to distant ones. Furthermore, a visibility module is also designed to detect and filter out detrimental information from occluded views. Due to the generalizability of our proposed method, we can synthesize semantic maps or conduct 3D semantic segmentation for novel scenes with solely 2D semantic supervision. Experimental results show that our approach achieves comparable performance with scene-specific approaches. More importantly, our approach can even outperform existing strong supervision-based approaches with only 2D annotations.
****
When can Regression-Adjusted Control Variate Help? Rare Events, Sobolev Embedding and Minimax Optimality
Jose Blanchet, Haoxuan Chen, Yiping Lu, Lexing Ying
This paper studies the use of a machine learning-based estimator as a control variate for mitigating the variance of Monte Carlo sampling. Specifically, we seek to uncover the key factors that influence the efficiency of control variates in reducing variance. We examine a prototype estimation problem that involves simulating the moments of a Sobolev function based on observations obtained from (random) quadrature nodes. Firstly, we establish an information-theoretic lower bound for the problem. We then study a specific quadrature rule that employs a nonparametric regression-adjusted control variate to reduce the variance of the Monte Carlo simulation. We demonstrate that this kind of quadrature rule can improve the Monte Carlo rate and achieve the minimax optimal rate under a sufficient smoothness assumption. Due to the Sobolev Embedding Theorem, the sufficient smoothness assumption eliminates the existence of rare and extreme events. Finally, we show that, in the presence of rare and extreme events, a truncated version of the Monte Carlo algorithm can achieve the minimax optimal rate while the control variate cannot improve the convergence rate.
****
Sharp Calibrated Gaussian Processes
Alexandre Capone, Sandra Hirche, Geoff Pleiss
While Gaussian processes are a mainstay for various engineering and scientific applications, the uncertainty estimates don't satisfy frequentist guarantees and can be miscalibrated in practice. State-of-the-art approaches for designing calibrated models rely on inflating the Gaussian process posterior variance, which yields confidence intervals that are potentially too coarse. To remedy this, we present a calibration approach that generates predictive quantiles using a comput

ation inspired by the vanilla Gaussian process posterior variance but using a different set of hyperparameters chosen to satisfy an empirical calibration constraint. This results in a calibration approach that is considerably more flexible than existing approaches, which we optimize to yield tight predictive quantiles. Our approach is shown to yield a calibrated model under reasonable assumptions. Furthermore, it outperforms existing approaches in sharpness when employed for calibrated regression.
****
GeoPhy: Differentiable Phylogenetic Inference via Geometric Gradients of Tree Topologies
Takahiro Mimori, Michiaki Hamada
Phylogenetic inference, grounded in molecular evolution models, is essential for understanding the evolutionary relationships in biological data. Accounting for the uncertainty of phylogenetic tree variables, which include tree topologies and evolutionary distances on branches, is crucial for accurately inferring species relationships from molecular data and tasks requiring variable marginalization. Variational Bayesian methods are key to developing scalable, practical models; however, it remains challenging to conduct phylogenetic inference without restricting the combinatorially vast number of possible tree topologies. In this work, we introduce a novel, fully differentiable formulation of phylogenetic inference that leverages a unique representation of topological distributions in continuous geometric spaces. Through practical considerations on design spaces and control variates for gradient estimations, our approach, GeoPhy, enables variational inference without limiting the topological candidates. In experiments using real benchmark datasets, GeoPhy significantly outperformed other approximate Bayesian methods that considered whole topologies.
****
AbdomenAtlas-8K: Annotating 8,000 CT Volumes for Multi-Organ Segmentation in Three Weeks
Chongyu Qu, Tiezheng Zhang, Hualin Qiao, jie liu, Yucheng Tang, Alan L. Yuille, Zongwei Zhou
Annotating medical images, particularly for organ segmentation, is laborious and time-consuming. For example, annotating an abdominal organ requires an estimated rate of 30-60 minutes per CT volume based on the expertise of an annotator and the size, visibility, and complexity of the organ. Therefore, publicly available datasets for multi-organ segmentation are often limited in data size and organ diversity. This paper proposes an active learning procedure to expedite the annotation process for organ segmentation and creates the largest multi-organ dataset (by far) with the spleen, liver, kidneys, stomach, gallbladder, pancreas, aorta, and IVC annotated in 8,448 CT volumes, equating to 3.2 million slices. The conventional annotation methods would take an experienced annotator up to 1,600 weeks (or roughly 30.8 years) to complete this task. In contrast, our annotation procedure has accomplished this task in three weeks (based on an 8-hour workday, five days a week) while maintaining a similar or even better annotation quality. This achievement is attributed to three unique properties of our method: (1) label bias reduction using multiple pre-trained segmentation models, (2) effective error detection in the model predictions, and (3) attention guidance for annotators to make corrections on the most salient errors. Furthermore, we summarize the taxonomy of common errors made by AI algorithms and annotators. This allows for continuous improvement of AI and annotations, significantly reducing the annotation costs required to create large-scale datasets for a wider variety of medical imaging tasks. Code and dataset are available at https://github.com/MrGiovanni/AbdomenAtlas
****
The Learnability of In-Context Learning
Noam Wies, Yoav Levine, Amnon Shashua
In-context learning is a surprising and important phenomenon that emerged when modern language models were scaled to billions of learned parameters. Without modifying a large language model's weights, it can be tuned to perform various downstream natural language tasks simply by including concatenated training exampl

es of these tasks in its input. Though disruptive for many practical applicatio
ns of large language models, this emergent learning paradigm is not well underst
ood from a theoretical perspective. In this paper, we propose a first-of-its-kin
d PAC based framework for in-context learnability, and use it to provide the fir
st finite sample complexity results for the in-context learning setup. Our fram
ework includes an initial pretraining phase, which fits a function to the pretra
ining distribution, and then a second in-context learning phase, which keeps thi
s function constant and concatenates training examples of the downstream task in
 its input. We use our framework in order to prove that, under mild assumptions
, when the pretraining distribution is a mixture of latent tasks (a model often
considered for natural language pretraining), these tasks can be efficiently lea
rned via in-context learning, even though the model's weights are unchanged and
the input significantly diverges from the pretraining distribution. Our theoret
ical analysis reveals that in this setting, in-context learning is more about id
entifying the task than about learning it, a result which is in line with a seri
es of recent empirical findings.   We hope that the in-context learnability fram
ework presented in this paper will facilitate future progress towards a deeper u
nderstanding of this important new learning paradigm.
****
Pick-a-Pic: An Open Dataset of User Preferences for Text-to-Image Generation
Yuval Kirstain, Adam Polyak, Uriel Singer, Shahbuland Matiana, Joe Penna, Omer L
evy
The ability to collect a large dataset of human preferences from text-to-image u
sers is usually limited to companies, making such datasets inaccessible to the p
ublic. To address this issue, we create a web app that enables text-to-image use
rs to generate images and specify their preferences. Using this web app we build
 Pick-a-Pic, a large, open dataset of text-to-image prompts and real users’ pref
erences over generated images. We leverage this dataset to train a CLIP-based sc
oring function, PickScore, which exhibits superhuman performance on the task of
predicting human preferences. Then, we test PickScore’s ability to perform model
 evaluation and observe that it correlates better with human rankings than other
 automatic evaluation metrics. Therefore, we recommend using PickScore for evalu
ating future text-to-image generation models, and using Pick-a-Pic prompts as a
more relevant dataset than MS-COCO. Finally, we demonstrate how PickScore can en
hance existing text-to-image models via ranking.
****
Decorate3D: Text-Driven High-Quality Texture Generation for Mesh Decoration in t
he Wild
Yanhui Guo, Xinxin Zuo, Peng Dai, Juwei Lu, Xiaolin Wu, Li cheng, Youliang Yan,
Songcen Xu, Xiaofei Wu
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Representational Strengths and Limitations of Transformers
Clayton Sanford, Daniel J. Hsu, Matus Telgarsky
Attention layers, as commonly used in transformers, form the backbone of modern
deep learning, yet there is no mathematical description of their benefits and de
ficiencies as compared with other architectures. In this work we establish both
positive and negative results on the representation power of attention layers, w
ith a focus on intrinsic complexity parameters such as width, depth, and embeddi
ng dimension. On the positive side, we present a sparse averaging task, where re
current networks and feedforward networks all have complexity scaling polynomial
ly in the input size, whereas transformers scale merely logarithmically in the i
nput size; furthermore, we use the same construction to show the necessity and r
ole of a large embedding dimension in a transformer. On the negative side, we pr
esent a triple detection task, where attention layers in turn have complexity sc
aling linearly in the input size; as this scenario seems rare in practice, we al
so present natural variants that can be efficiently solved by attention layers.

The proof techniques emphasize the value of communication complexity in the analysis of transformers and related models, and the role of sparse averaging as a prototypical attention task, which even finds use in the analysis of triple detection.
****
On the Relationship Between Relevance and Conflict in Online Social Link Recommendations
Yanbang Wang, Jon Kleinberg
In an online social network, link recommendations are a way for users to discover relevant links to people they may know, thereby potentially increasing their engagement on the platform. However, the addition of links to a social network can also have an effect on the level of conflict in the network --- expressed in terms of polarization and disagreement. To date, however, we have very little understanding of how these two implications of link formation relate to each other: are the goals of high relevance and conflict reduction aligned, or are the links that users are most likely to accept fundamentally different from the ones with the greatest potential for reducing conflict? Here we provide the first analysis of this question, using the recently popular Friedkin-Johnsen model of opinion dynamics. We first present a surprising result on how link additions shift the level of opinion conflict, followed by explanation work that relates the amount of shift to structural features of the added links. We then characterize the gap in conflict reduction between the set of links achieving the largest reduction and the set of links achieving the highest relevance. The gap is measured on real-world data, based on instantiations of relevance defined by 13 link recommendation algorithms. We find that some, but not all, of the more accurate algorithms actually lead to better reduction of conflict. Our work suggests that social links recommended for increasing user engagement may not be as conflict-provoking as people might have thought.
****
Mobilizing Personalized Federated Learning in Infrastructure-Less and Heterogeneous Environments via Random Walk Stochastic ADMM
Ziba Parsons, Fei Dou, Houyi Du, Zheng Song, Jin Lu
This paper explores the challenges of implementing Federated Learning (FL) in practical scenarios featuring isolated nodes with data heterogeneity, which can only be connected to the server through wireless links in an infrastructure-less environment. To overcome these challenges, we propose a novel mobilizing personalized FL approach, which aims to facilitate mobility and resilience. Specifically, we develop a novel optimization algorithm called Random Walk Stochastic Alternating Direction Method of Multipliers (RWSADMM). RWSADMM capitalizes on the server's random movement toward clients and formulates local proximity among their adjacent clients based on hard inequality constraints rather than requiring consensus updates or introducing bias via regularization methods. To mitigate the computational burden on the clients, an efficient stochastic solver of the approximated optimization problem is designed in RWSADMM, which provably converges to the stationary point almost surely in expectation. Our theoretical and empirical results demonstrate the provable fast convergence and substantial accuracy improvements achieved by RWSADMM compared to baseline methods, along with its benefits of reduced communication costs and enhanced scalability.
****
An Optimal Structured Zeroth-order Algorithm for Non-smooth Optimization
Marco Rando, Cesare Molinari, Lorenzo Rosasco, Silvia Villa
Finite-difference methods are a class of algorithms designed to solve black-box optimization problems by approximating a gradient of the target function on a set of directions. In black-box optimization, the non-smooth setting is particularly relevant since, in practice, differentiability and smoothness assumptions cannot be verified. To cope with nonsmoothness, several authors use a smooth approximation of the target function and show that finite difference methods approximate its gradient. Recently, it has been proved that imposing a structure in the directions allows improving performance. However, only the smooth setting was considered. To close this gap, we introduce and analyze O-ZD, the first structured

finite-difference algorithm for non-smooth black-box optimization. Our method exploits a smooth approximation of the target function and we prove that it approximates its gradient on a subset of random {\em orthogonal} directions. We analyze the convergence of O-ZD under different assumptions. For non-smooth convex functions, we obtain the optimal complexity. In the non-smooth non-convex setting, we characterize the number of iterations needed to bound the expected norm of the smoothed gradient. For smooth functions, our analysis recovers existing results for structured zeroth-order methods for the convex case and extends them to the non-convex setting. We conclude with numerical simulations where assumptions are satisfied, observing that our algorithm has very good practical performances.
****
Online Control for Meta-optimization
Xinyi Chen, Elad Hazan
Choosing the optimal hyperparameters, including learning rate and momentum, for specific optimization instances is a significant yet non-convex challenge. This makes conventional iterative techniques such as hypergradient descent \cite{baydin2017online} insufficient in obtaining global optimality guarantees.We consider the more general task of meta-optimization -- online learning of the best optimization algorithm given problem instances, and introduce a novel approach based on control theory. We show how meta-optimization can be formulated as an optimal control problem, departing from existing literature that use stability-based methods to study optimization. Our approach leverages convex relaxation techniques in the recently-proposed nonstochastic control framework to overcome the challenge of nonconvexity, and obtains regret guarantees vs. the best offline solution. This guarantees that in meta-optimization, we can learn a method that attains convergence comparable to that of the best optimization method in hindsight from a class of methods.
****
Emergent Communication in Interactive Sketch Question Answering
Zixing Lei, Yiming Zhang, Yuxin Xiong, Siheng Chen
Vision-based emergent communication (EC) aims to learn to communicate through sketches and demystify the evolution of human communication. Ironically, previous works neglect multi-round interaction, which is indispensable in human communication. To fill this gap, we first introduce a novel Interactive Sketch Question Answering (ISQA) task, where two collaborative players are interacting through sketches to answer a question about an image. To accomplish this task, we design a new and efficient interactive EC system, which can achieve an effective balance among three evaluation factors, including the question answering accuracy, drawing complexity and human interpretability. Our experimental results demonstrate that multi-round interactive mechanism facilitates tar- geted and efficient communication between intelligent agents. The code will be released.
****
Computing Approximate ℓ_p Sensitivities
Swati Padmanabhan, David Woodruff, Richard Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Automated Classification of Model Errors on ImageNet
Momchil Peychev, Mark Müller, Marc Fischer, Martin Vechev
While the ImageNet dataset has been driving computer vision research over the past decade, significant label noise and ambiguity have made top-1 accuracy an insufficient measure of further progress. To address this, new label-sets and evaluation protocols have been proposed for ImageNet showing that state-of-the-art models already achieve over 95% accuracy and shifting the focus on investigating why the remaining errors persist.Recent work in this direction employed a panel of experts to manually categorize all remaining classification errors for two selected models. However, this process is time-consuming, prone to inconsistencies,

and requires trained experts, making it unsuitable for regular model evaluation thus limiting its utility. To overcome these limitations, we propose the first automated error classification framework, a valuable tool to study how modeling choices affect error distributions. We use our framework to comprehensively evaluate the error distribution of over 900 models. Perhaps surprisingly, we find that across model architectures, scales, and pre-training corpora, top-1 accuracy is a strong predictor for the portion of all error types. In particular, we observe that the portion of severe errors drops significantly with top-1 accuracy indicating that, while it underreports a model's true performance, it remains a valuable performance metric.We release all our code at https://github.com/eth-sri/automated-error-analysis.
****
Sampling from Gaussian Process Posteriors using Stochastic Gradient Descent
Jihao Andreas Lin, Javier Antorán, Shreyas Padhy, David Janz, José Miguel Hernández-Lobato, Alexander Terenin
Gaussian processes are a powerful framework for quantifying uncertainty and for sequential decision-making but are limited by the requirement of solving linear systems. In general, this has a cubic cost in dataset size and is sensitive to conditioning. We explore stochastic gradient algorithms as a computationally efficient method of approximately solving these linear systems: we develop low-variance optimization objectives for sampling from the posterior and extend these to inducing points. Counterintuitively, stochastic gradient descent often produces accurate predictions, even in cases where it does not converge quickly to the optimum. We explain this through a spectral characterization of the implicit bias from non-convergence. We show that stochastic gradient descent produces predictive distributions close to the true posterior both in regions with sufficient data coverage, and in regions sufficiently far away from the data. Experimentally, stochastic gradient descent achieves state-of-the-art performance on sufficiently large-scale or ill-conditioned regression tasks. Its uncertainty estimates match the performance of significantly more expensive baselines on a large-scale Bayesian~optimization~task.
****
Selectivity Drives Productivity: Efficient Dataset Pruning for Enhanced Transfer Learning
Yihua Zhang, Yimeng Zhang, Aochuan Chen, jinghan jia, Jiancheng Liu, Gaowen Liu, Mingyi Hong, Shiyu Chang, Sijia Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Slicing Optimality for Mutual Information
Ammar Fayad, Majd Ibrahim
Measuring dependence between two random variables is of great importance in various domains but is difficult to compute in today's complex environments with high-dimensional data. Recently, slicing methods have shown to be a scalable approach to measuring mutual information (MI) between high-dimensional variables by projecting these variables into one-dimensional spaces. Unfortunately, these methods use uniform distributions of slicing directions, which generally discard informative features between variables and thus lead to inaccurate quantification of dependence. In this paper, we propose a principled framework that searches for an \textit{optimal} distribution of slices for MI. Importantly, we answer theoretical questions about finding the optimal slicing distribution in the context of MI and develop corresponding theoretical analyses. We also develop a practical algorithm, connecting our theoretical results with modern machine learning frameworks. Through comprehensive experiments in benchmark domains, we demonstrate significant gains in our information measure than state-of-the-art baselines.
****
OpenIllumination: A Multi-Illumination Dataset for Inverse Rendering Evaluation on Real Objects

Isabella Liu, Linghao Chen, Ziyang Fu, Liwen Wu, Haian Jin, Zhong Li, Chin Ming Ryan Wong, Yi Xu, Ravi Ramamoorthi, Zexiang Xu, Hao Su
We introduce OpenIllumination, a real-world dataset containing over 108K images of 64 objects with diverse materials, captured under 72 camera views and a large number of different illuminations. For each image in the dataset, we provide accurate camera parameters, illumination ground truth, and foreground segmentation masks. Our dataset enables the quantitative evaluation of most inverse rendering and material decomposition methods for real objects. We examine several state-of-the-art inverse rendering methods on our dataset and compare their performances. The dataset and code can be found on the project page: https://oppo-us-research.github.io/OpenIllumination.
****
Language Model Tokenizers Introduce Unfairness Between Languages
Aleksandar Petrov, Emanuele La Malfa, Philip Torr, Adel Bibi
Recent language models have shown impressive multilingual performance, even when not explicitly trained for it.Despite this, there are concerns about the quality of their outputs across different languages.In this paper, we show how disparity in the treatment of different languages arises at the tokenization stage, well before a model is even invoked.The same text translated into different languages can have drastically different tokenization lengths, with differences up to 15 times in some cases.These disparities persist even for tokenizers that are intentionally trained for multilingual support.Character-level and byte-level models also exhibit over 4 times the difference in the encoding length for some language pairs.This induces unfair treatment for some language communities in regard to the cost of accessing commercial language services, the processing time and latency, as well as the amount of content that can be provided as context to the models.Therefore, we make the case that we should train future language models using multilingually fair subword tokenizers.
****
Interaction Measures, Partition Lattices and Kernel Tests for High-Order Interactions
Zhaolu Liu, Robert Peach, Pedro A.M Mediano, Mauricio Barahona
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Demystifying Structural Disparity in Graph Neural Networks: Can One Size Fit All?
Haitao Mao, Zhikai Chen, Wei Jin, Haoyu Han, Yao Ma, Tong Zhao, Neil Shah, Jiliang Tang
Recent studies on Graph Neural Networks(GNNs) provide both empirical and theoretical evidence supporting their effectiveness in capturing structural patterns on both homophilic and certain heterophilic graphs. Notably, most real-world homophilic and heterophilic graphs are comprised of a mixture of nodes in both homophilic and heterophilic structural patterns, exhibiting a structural disparity. However, the analysis of GNN performance with respect to nodes exhibiting different structural patterns, e.g., homophilic nodes in heterophilic graphs, remains rather limited. In the present study, we provide evidence that Graph Neural Networks(GNNs) on node classification typically perform admirably on homophilic nodes within homophilic graphs and heterophilic nodes within heterophilic graphs while struggling on the opposite node set, exhibiting a performance disparity. We theoretically and empirically identify effects of GNNs on testing nodes exhibiting distinct structural patterns. We then propose a rigorous, non-i.i.d PAC-Bayesian generalization bound for GNNs, revealing reasons for the performance disparity, namely the aggregated feature distance and homophily ratio difference between training and testing nodes. Furthermore, we demonstrate the practical implications of our new findings via (1) elucidating the effectiveness of deeper GNNs; and (2) revealing an over-looked distribution shift factor on graph out-of-distribution problem and proposing a new scenario accordingly.

****
Deciphering Spatio-Temporal Graph Forecasting: A Causal Lens and Treatment
Yutong Xia, Yuxuan Liang, Haomin Wen, Xu Liu, Kun Wang, Zhengyang Zhou, Roger Zimmermann
Spatio-Temporal Graph (STG) forecasting is a fundamental task in many real-world applications. Spatio-Temporal Graph Neural Networks have emerged as the most popular method for STG forecasting, but they often struggle with temporal out-of-distribution (OoD) issues and dynamic spatial causation. In this paper, we propose a novel framework called CaST to tackle these two challenges via causal treatments. Concretely, leveraging a causal lens, we first build a structural causal model to decipher the data generation process of STGs. To handle the temporal OoD issue, we employ the back-door adjustment by a novel disentanglement block to separate the temporal environments from input data. Moreover, we utilize the front-door adjustment and adopt edge-level convolution to model the ripple effect of causation. Experiments results on three real-world datasets demonstrate the effectiveness of CaST, which consistently outperforms existing methods with good interpretability. Our source code is available at https://github.com/yutong-xia/CaST.
****
Greedy Poisson Rejection Sampling
Gergely Flamich
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Uncertainty Quantification via Neural Posterior Principal Components
Elias Nehme, Omer Yair, Tomer Michaeli
Uncertainty quantification is crucial for the deployment of image restoration models in safety-critical domains, like autonomous driving and biological imaging. To date, methods for uncertainty visualization have mainly focused on per-pixel estimates. Yet, a heatmap of per-pixel variances is typically of little practical use, as it does not capture the strong correlations between pixels. A more natural measure of uncertainty corresponds to the variances along the principal components (PCs) of the posterior distribution. Theoretically, the PCs can be computed by applying PCA on samples generated from a conditional generative model for the input image. However, this requires generating a very large number of samples at test time, which is painfully slow with the current state-of-the-art (diffusion) models. In this work, we present a method for predicting the PCs of the posterior distribution for any input image, in a single forward pass of a neural network. Our method can either wrap around a pre-trained model that was trained to minimize the mean square error (MSE), or can be trained from scratch to output both a predicted image and the posterior PCs. We showcase our method on multiple inverse problems in imaging, including denoising, inpainting, super-resolution, and biological image-to-image translation. Our method reliably conveys instance-adaptive uncertainty directions, achieving uncertainty quantification comparable with posterior samplers while being orders of magnitude faster. Code and examples are available on our webpage.
****
Deep Reinforcement Learning with Plasticity Injection
Evgenii Nikishin, Junhyuk Oh, Georg Ostrovski, Clare Lyle, Razvan Pascanu, Will Dabney, Andre Barreto
A growing body of evidence suggests that neural networks employed in deep reinforcement learning (RL) gradually lose their plasticity, the ability to learn from new data; however, the analysis and mitigation of this phenomenon is hampered by the complex relationship between plasticity, exploration, and performance in RL. This paper introduces plasticity injection, a minimalistic intervention that increases the network plasticity without changing the number of trainable parameters or biasing the predictions. The applications of this intervention are two-fold: first, as a diagnostic tool — if injection increases the performance, we ma

y conclude that an agent's network was losing its plasticity. This tool allows us to identify a subset of Atari environments where the lack of plasticity causes performance plateaus, motivating future studies on understanding and combating plasticity loss. Second, plasticity injection can be used to improve the computational efficiency of RL training if the agent has to re-learn from scratch due to exhausted plasticity or by growing the agent's network dynamically without compromising performance. The results on Atari show that plasticity injection attains stronger performance compared to alternative methods while being computationally efficient.
****
StreamNet: Memory-Efficient Streaming Tiny Deep Learning Inference on the Microcontroller
Hong-Sheng Zheng, Yu-Yuan Liu, Chen-Fong Hsu, Tsung Tai Yeh
With the emerging Tiny Machine Learning (TinyML) inference applications, there is a growing interest when deploying TinyML models on the low-power Microcontroller Unit (MCU). However, deploying TinyML models on MCUs reveals several challenges due to the MCU’s resource constraints, such as small flash memory, tight SRAM memory budget, and slow CPU performance. Unlike typical layer-wise inference, patch-based inference reduces the peak usage of SRAM memory on MCUs by saving small patches rather than the entire tensor in the SRAM memory. However, the processing of patch-based inference tremendously increases the amount of MACs against the layer-wise method. Thus, this notoriously computational overhead makes patch-based inference undesirable on MCUs. This work designs StreamNet that employs the stream buffer to eliminate the redundant computation of patch-based inference. StreamNet uses 1D and 2D streaming processing and provides an parameter selection algorithm that automatically improve the performance of patch-based inference with minimal requirements on the MCU’s SRAM memory space. In 10 TinyML models, StreamNet-2D achieves a geometric mean of 7.3X speedup and saves 81\% of MACs over the state-of-the-art patch-based inference.
****
Estimating and Controlling for Equalized Odds via Sensitive Attribute Predictors
Beepul Bharti, Paul Yi, Jeremias Sulam
As the use of machine learning models in real world high-stakes decision settings continues to grow, it is highly important that we are able to audit and control for any potential fairness violations these models may exhibit towards certain groups. To do so, one naturally requires access to sensitive attributes, such as demographics, biological sex, or other potentially sensitive features that determine group membership. Unfortunately, in many settings, this information is often unavailable. In this work we study the well known equalized odds (EOD) definition of fairness. In a setting without sensitive attributes, we first provide tight and computable upper bounds for the EOD violation of a predictor. These bounds precisely reflect the worst possible EOD violation. Second, we demonstrate how one can provably control the worst-case EOD by a new post-processing correction method. Our results characterize when directly controlling for EOD with respect to the predicted sensitive attributes is -- and when is not -- optimal when it comes to controlling worst-case EOD. Our results hold under assumptions that are milder than previous works, and we illustrate these results with experiments on synthetic and real datasets.
****
Segment Any Point Cloud Sequences by Distilling Vision Foundation Models
Youquan Liu, Lingdong Kong, Jun CEN, Runnan Chen, Wenwei Zhang, Liang Pan, Kai Chen, Ziwei Liu
Recent advancements in vision foundation models (VFMs) have opened up new possibilities for versatile and efficient visual perception. In this work, we introduce Seal, a novel framework that harnesses VFMs for segmenting diverse automotive point cloud sequences. Seal exhibits three appealing properties: i) Scalability: VFMs are directly distilled into point clouds, obviating the need for annotations in either 2D or 3D during pretraining. ii) Consistency: Spatial and temporal relationships are enforced at both the camera-to-LiDAR and point-to-segment regularization stages, facilitating cross-modal representation learning. iii) Genera

lizability: Seal enables knowledge transfer in an off-the-shelf manner to downstream tasks involving diverse point clouds, including those from real/synthetic, low/high-resolution, large/small-scale, and clean/corrupted datasets. Extensive experiments conducted on eleven different point cloud datasets showcase the effectiveness and superiority of Seal. Notably, Seal achieves a remarkable 45.0% mIoU on nuScenes after linear probing, surpassing random initialization by 36.9% mIoU and outperforming prior arts by 6.1% mIoU. Moreover, Seal demonstrates significant performance gains over existing methods across 20 different few-shot fine-tuning tasks on all eleven tested point cloud datasets. The code is available at this link.
****
Time Series Kernels based on Nonlinear Vector AutoRegressive Delay Embeddings
Giovanni De Felice, John Goulermas, Vladimir Gusev
Kernel design is a pivotal but challenging aspect of time series analysis, especially in the context of small datasets. In recent years, Reservoir Computing (RC) has emerged as a powerful tool to compare time series based on the underlying dynamics of the generating process rather than the observed data. However, the performance of RC highly depends on the hyperparameter setting, which is hard to interpret and costly to optimize because of the recurrent nature of RC. Here, we present a new kernel for time series based on the recently established equivalence between reservoir dynamics and Nonlinear Vector AutoRegressive (NVAR) processes. The kernel is non-recurrent and depends on a small set of meaningful hyperparameters, for which we suggest an effective heuristic. We demonstrate excellent performance on a wide range of real-world classification tasks, both in terms of accuracy and speed. This further advances the understanding of RC representation learning models and extends the typical use of the NVAR framework to kernel design and representation of real-world time series data.
****
SPA: A Graph Spectral Alignment Perspective for Domain Adaptation
Zhiqing Xiao, Haobo Wang, Ying Jin, Lei Feng, Gang Chen, Fei Huang, Junbo Zhao
Unsupervised domain adaptation (UDA) is a pivotal form in machine learning to extend the in-domain model to the distinctive target domains where the data distributions differ. Most prior works focus on capturing the inter-domain transferability but largely overlook rich intra-domain structures, which empirically results in even worse discriminability. In this work, we introduce a novel graph SPectral Alignment (SPA) framework to tackle the tradeoff. The core of our method is briefly condensed as follows: (i)-by casting the DA problem to graph primitives, SPA composes a coarse graph alignment mechanism with a novel spectral regularizer towards aligning the domain graphs in eigenspaces; (ii)-we further develop a fine-grained message propagation module --- upon a novel neighbor-aware self-training mechanism --- in order for enhanced discriminability in the target domain. On standardized benchmarks, the extensive experiments of SPA demonstrate that its performance has surpassed the existing cutting-edge DA methods. Coupled with dense model analysis, we conclude that our approach indeed possesses superior efficacy, robustness, discriminability, and transferability. Code and data are available at: https://github.com/CrownX/SPA.
****
CosNet: A Generalized Spectral Kernel Network
Yanfang Xue, Pengfei Fang, Jinyue Tian, Shipeng Zhu, hui xue
Complex-valued representation exists inherently in the time-sequential data that can be derived from the integration of harmonic waves. The non-stationary spectral kernel, realizing a complex-valued feature mapping, has shown its potential to analyze the time-varying statistical characteristics of the time-sequential data, as a result of the modeling frequency parameters. However, most existing spectral kernel-based methods eliminate the imaginary part, thereby limiting the representation power of the spectral kernel. To tackle this issue, we propose a generalized spectral kernel network, namely, \underline{Co}mplex-valued \underline{s}pectral kernel \underline{Net}work (CosNet), which includes spectral kernel mapping generalization (SKMG) module and complex-valued spectral kernel embedding (CSKE) module. Concretely, the SKMG module is devised to generalize the spectr

al kernel mapping in the real number domain to the complex number domain, recovering the inherent complex-valued representation for the real-valued data. Then a following CSKE module is further developed to combine the complex-valued spectral kernels and neural networks to effectively capture long-range or periodic relations of the data. Along with the CosNet, we study the effect of the complex-valued spectral kernel mapping via theoretically analyzing the bound of covering number and generalization error. Extensive experiments demonstrate that CosNet performs better than the mainstream kernel methods and complex-valued neural networks.
****
A Theory of Unsupervised Translation Motivated by Understanding Animal Communication
Shafi Goldwasser, David Gruber, Adam Tauman Kalai, Orr Paradise
Neural networks are capable of translating between languages—in some cases even between two languages where there is little or no access to parallel translations, in what is known as Unsupervised Machine Translation (UMT). Given this progress, it is intriguing to ask whether machine learning tools can ultimately enable understanding animal communication, particularly that of highly intelligentanimals. We propose a theoretical framework for analyzing UMT when no parallel translations are available and when it cannot be assumed that the source and target corpora address related subject domains or posses similar linguistic structure. Weexemplify this theory with two stylized models of language, for which our framework provides bounds on necessary sample complexity; the bounds are formally proven and experimentally verified on synthetic data. These bounds show that the error rates are inversely related to the language complexity and amount of common ground. This suggests that unsupervised translation of animal communication may be feasible if the communication system is sufficiently complex.
****
Adversarial Self-Training Improves Robustness and Generalization for Gradual Domain Adaptation
Lianghe Shi, Weiwei Liu
Gradual Domain Adaptation (GDA), in which the learner is provided with additional intermediate domains, has been theoretically and empirically studied in many contexts. Despite its vital role in security-critical scenarios, the adversarial robustness of the GDA model remains unexplored. In this paper, we adopt the effective gradual self-training method and replace vanilla self-training with adversarial self-training (AST). AST first predicts labels on the unlabeled data and then adversarially trains the model on the pseudo-labeled distribution. Intriguingly, we find that gradual AST improves not only adversarial accuracy but also clean accuracy on the target domain. We reveal that this is because adversarial training (AT) performs better than standard training when the pseudo-labels contain a portion of incorrect labels. Accordingly, we first present the generalization error bounds for gradual AST in a multiclass classification setting. We then use the optimal value of the Subset Sum Problem to bridge the standard error on a real distribution and the adversarial error on a pseudo-labeled distribution. The result indicates that AT may obtain a tighter bound than standard training on data with incorrect pseudo-labels. We further present an example of a conditional Gaussian distribution to provide more insights into why gradual AST can improve the clean accuracy for GDA.
****
TensorNet: Cartesian Tensor Representations for Efficient Learning of Molecular Potentials
Guillem Simeon, Gianni De Fabritiis
The development of efficient machine learning models for molecular systems representation is becoming crucial in scientific research. We introduce TensorNet, an innovative O(3)-equivariant message-passing neural network architecture that leverages Cartesian tensor representations. By using Cartesian tensor atomic embeddings, feature mixing is simplified through matrix product operations. Furthermore, the cost-effective decomposition of these tensors into rotation group irreducible representations allows for the separate processing of scalars, vectors, an

d tensors when necessary. Compared to higher-rank spherical tensor models, TensorNet demonstrates state-of-the-art performance with significantly fewer parameters. For small molecule potential energies, this can be achieved even with a single interaction layer. As a result of all these properties, the model's computational cost is substantially decreased. Moreover, the accurate prediction of vector and tensor molecular quantities on top of potential energies and forces is possible. In summary, TensorNet's framework opens up a new space for the design of state-of-the-art equivariant models.
****

Multi-Player Zero-Sum Markov Games with Networked Separable Interactions
Chanwoo Park, Kaiqing Zhang, Asuman Ozdaglar
We study a new class of Markov games, \textit{(multi-player) zero-sum Markov Games} with {\it Networked separable interactions} (zero-sum NMGs), to model the local interaction structure in non-cooperative multi-agent sequential decision-making. We define a zero-sum NMG as a model where {the payoffs of the auxiliary games associated with each state are zero-sum and} have some separable (i.e., polymatrix) structure across the neighbors over some interaction network. We first identify the necessary and sufficient conditions under which an MG can be presented as a zero-sum NMG, and show that the set of Markov coarse correlated equilibrium (CCE) collapses to the set of Markov Nash equilibrium (NE) in these games, in that the {product of} per-state marginalization of the former for all players yields the latter. Furthermore, we show that finding approximate Markov \emph{stationary} CCE in infinite-horizon discounted zero-sum NMGs is \texttt{PPAD}-hard, unless the underlying network has a ``star topology''. Then, we propose fictitious-play-type dynamics, the classical learning dynamics in normal-form games, for zero-sum NMGs, and establish convergence guarantees to Markov stationary NE under a star-shaped network structure. Finally, in light of the hardness result, we focus on computing a Markov \emph{non-stationary} NE and provide finite-iteration guarantees for a series of value-iteration-based algorithms. We also provide numerical experiments to corroborate our theoretical results.
****

Continuous-Time Functional Diffusion Processes
Giulio Franzese, Giulio Corallo, Simone Rossi, Markus Heinonen, Maurizio Filippone, Pietro Michiardi
We introduce Functional Diffusion Processes (FDPs), which generalize score-based diffusion models to infinite-dimensional function spaces. FDPs require a new mathematical framework to describe the forward and backward dynamics, and several extensions to derive practical training objectives. These include infinite-dimensional versions of Girsanov theorem, in order to be able to compute an ELBO, and of the sampling theorem, in order to guarantee that functional evaluations in a countable set of points are equivalent to infinite-dimensional functions. We use FDPs to build a new breed of generative models in function spaces, which do not require specialized network architectures, and that can work with any kind of continuous data.Our results on real data show that FDPs achieve high-quality image generation, using a simple MLP architecture with orders of magnitude fewer parameters than existing diffusion models.
****

Knowledge-based in silico models and dataset for the comparative evaluation of mammography AI for a range of breast characteristics, lesion conspicuities and doses
Elena Sizikova, Niloufar Saharkhiz, Diksha Sharma, Miguel Lago, Berkman Sahiner, Jana Delfino, Aldo Badano
To generate evidence regarding the safety and efficacy of artificial intelligence (AI) enabled medical devices, AI models need to be evaluated on a diverse population of patient cases, some of which may not be readily available. We propose an evaluation approach for testing medical imaging AI models that relies on in silico imaging pipelines in which stochastic digital models of human anatomy (in object space) with and without pathology are imaged using a digital replica imaging acquisition system to generate realistic synthetic image datasets. Here, we release M-SYNTH, a dataset of cohorts with four breast fibroglandular density di

stributions imaged at different exposure levels using Monte Carlo x-ray simulations with the publicly available Virtual Imaging Clinical Trial for Regulatory Evaluation (VICTRE) toolkit. We utilize the synthetic dataset to analyze AI model performance and find that model performance decreases with increasing breast density and increases with higher mass density, as expected. As exposure levels decrease, AI model performance drops with the highest performance achieved at exposure levels lower than the nominal recommended dose for the breast type.
****
Is Distance Matrix Enough for Geometric Deep Learning?
Zian Li, Xiyuan Wang, Yinan Huang, Muhan Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Optimized Covariance Design for AB Test on Social Network under Interference
Qianyi Chen, Bo Li, Lu Deng, Yong Wang
Online A/B tests have become increasingly popular and important for social platforms. However, accurately estimating the global average treatment effect (GATE) has proven to be challenging due to network interference, which violates the Stable Unit Treatment Value Assumption (SUTVA) and poses great challenge to experimental design. Existing network experimental design research was mostly based on the unbiased Horvitz-Thompson (HT) estimator with substantial data trimming to ensure unbiasedness at the price of high resultant estimation variance. In this paper, we strive to balance the bias and variance in designing randomized network experiments. Under a potential outcome model with 1-hop interference, we derive the bias and variance of the standard HT estimator and reveal their relation to the network topological structure and the covariance of the treatment assignment vector. We then propose to formulate the experimental design problem as to optimize the covariance matrix of the treatment assignment vector to achieve the bias and variance balance by minimizing the mean squared error (MSE) of the estimator. An efficient projected gradient descent algorithm is presented to the implement of the desired randomization scheme. Finally, we carry out extensive simulation studies to demonstrate the advantages of our proposed method over other existing methods in many settings, with different levels of model misspecification.
****
AV-NeRF: Learning Neural Fields for Real-World Audio-Visual Scene Synthesis
Susan Liang, Chao Huang, Yapeng Tian, Anurag Kumar, Chenliang Xu
Can machines recording an audio-visual scene produce realistic, matching audio-visual experiences at novel positions and novel view directions? We answer it by studying a new task---real-world audio-visual scene synthesis---and a first-of-its-kind NeRF-based approach for multimodal learning. Concretely, given a video recording of an audio-visual scene, the task is to synthesize new videos with spatial audios along arbitrary novel camera trajectories in that scene. We propose an acoustic-aware audio generation module that integrates prior knowledge of audio propagation into NeRF, in which we implicitly associate audio generation with the 3D geometry and material properties of a visual environment. Furthermore, we present a coordinate transformation module that expresses a view direction relative to the sound source, enabling the model to learn sound source-centric acoustic fields. To facilitate the study of this new task, we collect a high-quality Real-World Audio-Visual Scene (RWAVS) dataset. We demonstrate the advantages of our method on this real-world dataset and the simulation-based SoundSpaces dataset. Notably, we refer readers to view our demo videos for convincing comparisons.
****
Is This Loss Informative? Faster Text-to-Image Customization by Tracking Objective Dynamics
Anton Voronov, Mikhail Khoroshikh, Artem Babenko, Max Ryabinin
Text-to-image generation models represent the next step of evolution in image sy

nthesis, offering a natural way to achieve flexible yet fine-grained control over the result.One emerging area of research is the fast adaptation of large text-to-image models to smaller datasets or new visual concepts.However, many efficient methods of adaptation have a long training time, which limits their practical applications, slows down experiments, and spends excessive GPU resources.In this work, we study the training dynamics of popular text-to-image personalization methods (such as Textual Inversion or DreamBooth), aiming to speed them up.We observe that most concepts are learned at early stages and do not improve in quality later, but standard training convergence metrics fail to indicate that.Instead, we propose a simple drop-in early stopping criterion that only requires computing the regular training objective on a fixed set of inputs for all training iterations.Our experiments on Stable Diffusion for 48 different concepts and three personalization methods demonstrate the competitive performance of our approach, which makes adaptation up to 8 times faster with no significant drops in quality.
****
DesCo: Learning Object Recognition with Rich Language Descriptions
Liunian Li, Zi-Yi Dou, Nanyun Peng, Kai-Wei Chang
Recent development in vision-language approaches has instigated a paradigm shift in learning visual recognition models from language supervision. These approaches align objects with language queries (e.g. "a photo of a cat") and thus improve the models' adaptability to novel objects and domains. Recent studies have attempted to query these models with complex language expressions that include specifications of fine-grained details, such as colors, shapes, and relations. However, simply incorporating language descriptions into queries does not guarantee accurate interpretation by the models. In fact, our experiments show that GLIP, a state-of-the-art vision-language model for object detection, often disregards contextual information in the language descriptions and instead relies heavily on detecting objects solely by their names. To tackle the challenge, we propose a new description-conditioned (DesCo) paradigm of learning object recognition models with rich language descriptions consisting of two innovations: 1) we employ a large language model as a commonsense knowledge engine to generate rich language descriptions of objects; 2) we design context-sensitive queries to improve the model's ability in deciphering intricate nuances embedded within descriptions and enforce the model to focus on context rather than object names alone. On two novel object detection benchmarks, LVIS and OminiLabel, under the zero-shot detection setting, our approach achieves 34.8 APr minival (+9.1) and 29.3 AP (+3.6), respectively, surpassing the prior state-of-the-art models, GLIP and FIBER, by a large margin.
****
On the Variance, Admissibility, and Stability of Empirical Risk Minimization
Gil Kur, Eli Putterman, Alexander Rakhlin
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
NetHack is Hard to Hack
Ulyana Piterbarg, Lerrel Pinto, Rob Fergus
Neural policy learning methods have achieved remarkable results in various control problems, ranging from Atari games to simulated locomotion. However, these methods struggle in long-horizon tasks, especially in open-ended environments with multi-modal observations, such as the popular dungeon-crawler game, NetHack. Intriguingly, the NeurIPS 2021 NetHack Challenge revealed that symbolic agents outperformed neural approaches by over four times in median game score. In this paper, we delve into the reasons behind this performance gap and present an extensive study on neural policy learning for NetHack. To conduct this study, we analyze the winning symbolic agent, extending its codebase to track internal strategy selection in order to generate one of the largest available demonstration datasets. Utilizing this dataset, we examine (i) the advantages of an action hierarchy

; (ii) enhancements in neural architecture; and (iii) the integration of reinforcement learning with imitation learning. Our investigations produce a state-of-the-art neural agent that surpasses previous fully neural policies by 127% in offline settings and 25% in online settings on median game score. However, we also demonstrate that mere scaling is insufficient to bridge the performance gap with the best symbolic models or even the top human players.
****
SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning
Benjamin Ellis, Jonathan Cook, Skander Moalla, Mikayel Samvelyan, Mingfei Sun, Anuj Mahajan, Jakob Foerster, Shimon Whiteson
The availability of challenging benchmarks has played a key role in the recent progress of machine learning. In cooperative multi-agent reinforcement learning, the StarCraft Multi-Agent Challenge (SMAC) has become a popular testbed for centralised training with decentralised execution. However, after years of sustained improvement on SMAC, algorithms now achieve near-perfect performance. In this work, we conduct new analysis demonstrating that SMAC lacks the stochasticity and partial observability to require complex closed-loop policies. In particular, we show that an open-loop policy conditioned only on the timestep can achieve non-trivial win rates for many SMAC scenarios. To address this limitation, we introduce SMACv2, a new version of the benchmark where scenarios are procedurally generated and require agents to generalise to previously unseen settings (from the same distribution) during evaluation. We also introduce the extended partial observability challenge (EPO), which augments SMACv2 to ensure meaningful partial observability. We show that these changes ensure the benchmarkrequires the use of closed-loop policies. We evaluate state-of-the-art algorithms on SMACv2 and show that it presents significant challenges not present in the original benchmark. Our analysis illustrates that SMACv2 addresses the discovered deficiencies of SMAC and can help benchmark the next generation of MARL methods. Videos of training are available on our website.
****
LightZero: A Unified Benchmark for Monte Carlo Tree Search in General Sequential Decision Scenarios
Yazhe Niu, YUAN PU, Zhenjie Yang, Xueyan Li, Tong Zhou, Jiyuan Ren, Shuai Hu, Hongsheng Li, Yu Liu
Building agents based on tree-search planning capabilities with learned models has achieved remarkable success in classic decision-making problems, such as Go and Atari.However, it has been deemed challenging or even infeasible to extend Monte Carlo Tree Search (MCTS) based algorithms to diverse real-world applications, especially when these environments involve complex action spaces and significant simulation costs, or inherent stochasticity.In this work, we introduce LightZero, the first unified benchmark for deploying MCTS/MuZero in general sequential decision scenarios. Specificially, we summarize the most critical challenges in designing a general MCTS-style decision-making solver, then decompose the tightly-coupled algorithm and system design of tree-search RL methods into distinct sub-modules.By incorporating more appropriate exploration and optimization strategies, we can significantly enhance these sub-modules and construct powerful LightZero agents to tackle tasks across a wide range of domains, such as board games, Atari, MuJoCo, MiniGrid and GoBigger.Detailed benchmark results reveal the significant potential of such methods in building scalable and efficient decision intelligence.The code is available as part of OpenDILab at https://github.com/opendilab/LightZero.
****
Improving Diffusion-Based Image Synthesis with Context Prediction
Ling Yang, Jingwei Liu, Shenda Hong, Zhilong Zhang, Zhilin Huang, Zheming Cai, Wentao Zhang, Bin CUI
Diffusion models are a new class of generative models, and have dramatically promoted image generation with unprecedented quality and diversity. Existing diffusion models mainly try to reconstruct input image from a corrupted one with a pixel-wise or feature-wise constraint along spatial axes. However, such point-based reconstruction may fail to make each predicted pixel/feature fully preserve its

 neighborhood context, impairing diffusion-based image synthesis. As a powerful source of automatic supervisory signal, context has been well studied for learning representations. Inspired by this, we for the first time propose ConPreDiff to improve diffusion-based image synthesis with context prediction. We explicitly reinforce each point to predict its neighborhood context (i.e., multi-stride pixels/features) with a context decoder at the end of diffusion denoising blocks in training stage, and remove the decoder for inference. In this way, each point can better reconstruct itself by preserving its semantic connections with neighborhood context. This new paradigm of ConPreDiff can generalize to arbitrary discrete and continuous diffusion backbones without introducing extra parameters in sampling procedure. Extensive experiments are conducted on unconditional image generation, text-to-image generation and image inpainting tasks. Our ConPreDiff consistently outperforms previous methods and achieves new SOTA text-to-image generation results on MS-COCO, with a zero-shot FID score of 6.21.
****

Adversarial Robustness through Random Weight Sampling

Yanxiang Ma, Minjing Dong, Chang Xu

Deep neural networks have been found to be vulnerable in a variety of tasks. Adversarial attacks can manipulate network outputs, resulting in incorrect predictions. Adversarial defense methods aim to improve the adversarial robustness of networks by countering potential attacks. In addition to traditional defense approaches, randomized defense mechanisms have recently received increasing attention from researchers. These methods introduce different types of perturbations during the inference phase to destabilize adversarial attacks.Although promising empirical results have been demonstrated by these approaches, the defense performance is quite sensitive to the randomness parameters, which are always manually tuned without further analysis. On the contrary, we propose incorporating random weights into the optimization to fully exploit the potential of randomized defense. To perform better optimization of randomness parameters, we conduct a theoretical analysis of the connections between randomness parameters and gradient similarity as well as natural performance. From these two aspects, we suggest imposing theoretically-guided constraints on random weights during optimizations, as these weights play a critical role in balancing natural performance and adversarial robustness. We derive both the upper and lower bounds of random weight parameters by considering prediction bias and gradient similarity. In this study, we introduce the Constrained Trainable Random Weight (CTRW), which adds random weight parameters to the optimization and includes a constraint guided by the upper and lower bounds to achieve better trade-offs between natural and robust accuracy. We evaluate the effectiveness of CTRW on several datasets and benchmark convolutional neural networks. Our results indicate that our model achieves a robust accuracy approximately 16% to 17% higher than the baseline model under PGD-20 and 22% to 25% higher on Auto Attack.
****

PyNeRF: Pyramidal Neural Radiance Fields

Haithem Turki, Michael Zollhöfer, Christian Richardt, Deva Ramanan

Neural Radiance Fields (NeRFs) can be dramatically accelerated by spatial grid representations. However, they do not explicitly reason about scale and so introduce aliasing artifacts when reconstructing scenes captured at different camera distances. Mip-NeRF and its extensions propose scale-aware renderers that project volumetric frustums rather than point samples. But such approaches rely on positional encodings that are not readily compatible with grid methods. We propose a simple modification to grid-based models by training model heads at different spatial grid resolutions. At render time, we simply use coarser grids to render samples that cover larger volumes. Our method can be easily applied to existing accelerated NeRF methods and significantly improves rendering quality (reducing error rates by 20–90% across synthetic and unbounded real-world scenes) while incurring minimal performance overhead (as each model head is quick to evaluate). Compared to Mip-NeRF, we reduce error rates by 20% while training over 60x faster.
****

Universal Online Learning with Gradient Variations: A Multi-layer Online Ensembl
e Approach
Yu-Hu Yan, Peng Zhao, Zhi-Hua Zhou
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Information Theoretic Lower Bounds for Information Theoretic Upper Bounds
Roi Livni
We examine the relationship between the mutual information between the output mo
del and the empirical sample and the algorithm's generalization in the context o
f stochastic convex optimization. Despite increasing interest in information-the
oretic generalization bounds, it is uncertain if these bounds can provide insigh
t into the exceptional performance of various learning algorithms. Our study of
stochastic convex optimization reveals that, for true risk minimization, dimensi
on-dependent mutual information is necessary. This indicates that existing infor
mation-theoretic generalization bounds fall short in capturing the generalizatio
n capabilities of algorithms like SGD and regularized ERM, which have dimension-
independent sample complexity.
****
CoDrug: Conformal Drug Property Prediction with Density Estimation under Covaria
te Shift
Siddhartha Laghuvarapu, Zhen Lin, Jimeng Sun
In drug discovery, it is vital to confirm the predictions of pharmaceutical prop
erties from computational models using costly wet-lab experiments. Hence, obtain
ing reliable uncertainty estimates is crucial for prioritizing drug molecules fo
r subsequent experimental validation. Conformal Prediction (CP) is a promising t
ool for creating such prediction sets for molecular properties with a coverage g
uarantee. However, the exchangeability assumption of CP is often challenged with
 covariate shift in drug discovery tasks: Most datasets contain limited labeled
data, which may not be representative of the vast chemical space from which mole
cules are drawn. To address this limitation, we propose a method called CoDrug t
hat employs an energy-based model leveraging both training data and unlabelled d
ata, and Kernel Density Estimation (KDE) to assess the densities of a molecule
set. The estimated densities are then used to weigh the molecule samples while b
uilding prediction sets and rectifying for distribution shift. In extensive expe
riments involving realistic distribution drifts in various small-molecule drug d
iscovery tasks, we demonstrate the ability of CoDrug to provide valid predictio
n sets and its utility in addressing the distribution shift arising from de novo
 drug design models. On average, using CoDrug can reduce the coverage gap by ove
r 35% when compared to conformal prediction sets not adjusted for covariate shif
t.
****
TWIGMA: A dataset of AI-Generated Images with Metadata From Twitter
Yiqun Chen, James Y. Zou
Recent progress in generative artificial intelligence (gen-AI) has enabled the g
eneration of photo-realistic and artistically-inspiring photos at a single click
, catering to millions of users online. To explore how people use gen-AI models
such as DALLE and StableDiffusion, it is critical to understand the themes, cont
ents, and variations present in the AI-generated photos. In this work, we introd
uce TWIGMA (TWItter Generative-ai images with MetadatA), a comprehensive dataset
 encompassing over 800,000 gen-AI images collected from Jan 2021 to March 2023 o
n Twitter, with associated metadata (e.g., tweet text, creation date, number of
likes). Through a comparative analysis of TWIGMA with natural images and human a
rtwork, we find that gen-AI images possess distinctive characteristics and exhib
it, on average, lower variability when compared to their non-gen-AI counterparts
. Additionally, we find that the similarity between a gen-AI image and natural i
mages is inversely correlated with the number of likes. Finally, we observe a lo
ngitudinal shift in the themes of AI-generated images on Twitter, with users inc

reasingly sharing artistically sophisticated content such as intricate human portraits, whereas their interest in simple subjects such as natural scenes and animals has decreased. Our analyses and findings underscore the significance of TWIGMA as a unique data resource for studying AI-generated images.
****
Exact Optimality of Communication-Privacy-Utility Tradeoffs in Distributed Mean Estimation
Berivan Isik, Wei-Ning Chen, Ayfer Ozgur, Tsachy Weissman, Albert No
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Estimating Generic 3D Room Structures from 2D Annotations
Denys Rozumnyi, Stefan Popov, Kevis-kokitsi Maninis, Matthias Niessner, Vittorio Ferrari
Indoor rooms are among the most common use cases in 3D scene understanding. Current state-of-the-art methods for this task are driven by large annotated datasets. Room layouts are especially important, consisting of structural elements in 3D, such as wall, floor, and ceiling. However, they are difficult to annotate, especially on pure RGB video. We propose a novel method to produce generic 3D room layouts just from 2D segmentation masks, which are easy to annotate for humans. Based on these 2D annotations, we automatically reconstruct 3D plane equations for the structural elements and their spatial extent in the scene, and connect adjacent elements at the appropriate contact edges. We annotate and publicly release 2246 3D room layouts on the RealEstate10k dataset, containing YouTube videos. We demonstrate the high quality of these 3D layouts annotations with extensive experiments.
****
Score-based Generative Modeling through Stochastic Evolution Equations in Hilbert Spaces
Sungbin Lim, EUN BI YOON, Taehyun Byun, Taewon Kang, Seungwoo Kim, Kyungjae Lee, Sungjoon Choi
Continuous-time score-based generative models consist of a pair of stochastic differential equations (SDEs)—a forward SDE that smoothly transitions data into a noise space and a reverse SDE that incrementally eliminates noise from a Gaussian prior distribution to generate data distribution samples—are intrinsically connected by the time-reversal theory on diffusion processes. In this paper, we investigate the use of stochastic evolution equations in Hilbert spaces, which expand the applicability of SDEs in two aspects: sample space and evolution operator, so they enable encompassing recent variations of diffusion models, such as generating functional data or replacing drift coefficients with image transformation. To this end, we derive a generalized time-reversal formula to build a bridge between probabilistic diffusion models and stochastic evolution equations and propose a score-based generative model called Hilbert Diffusion Model (HDM). Combining with Fourier neural operator, we verify the superiority of HDM for sampling functions from functional datasets with a power of kernel two-sample test of 4.2 on Quadratic, 0.2 on Melbourne, and 3.6 on Gridwatch, which outperforms existing diffusion models formulated in function spaces. Furthermore, the proposed method shows its strength in motion synthesis tasks by utilizing the Wiener process with values in Hilbert space. Finally, our empirical results on image datasets also validate a connection between HDM and diffusion models using heat dissipation, revealing the potential for exploring evolution operators and sample spaces.
****
Unlocking Feature Visualization for Deep Network with MAgnitude Constrained Optimization
Thomas FEL, Thibaut Boissin, Victor Boutin, Agustin PICARD, Paul Novello, Julien Colin, Drew Linsley, Tom ROUSSEAU, Remi Cadene, Lore Goetschalckx, Laurent Gardes, Thomas Serre
Feature visualization has gained significant popularity as an explainability met

hod, particularly after the influential work by Olah et al. in 2017. Despite its success, its widespread adoption has been limited due to issues in scaling to deeper neural networks and the reliance on tricks to generate interpretable images. Here, we describe MACO, a simple approach to address these shortcomings. It consists in optimizing solely an image's phase spectrum while keeping its magnitude constant to ensure that the generated explanations lie in the space of natural images. Our approach yields significantly better results -- both qualitatively and quantitatively -- unlocking efficient and interpretable feature visualizations for state-of-the-art neural networks. We also show that our approach exhibits an attribution mechanism allowing to augment feature visualizations with spatial importance. Furthermore, we enable quantitative evaluation of feature visualizations by introducing 3 metrics: transferability, plausibility, and alignment with natural images. We validate our method on various applications and we introduce a website featuring MACO visualizations for all classes of the ImageNet dataset, which will be made available upon acceptance. Overall, our study unlocks feature visualizations for the largest, state-of-the-art classification networks without resorting to any parametric prior image model, effectively advancing a field that has been stagnating since 2017 (Olah et al, 2017).
****

Exact recovery and Bregman hard clustering of node-attributed Stochastic Block Model

Maximilien Dreveton, Felipe Fernandes, Daniel Figueiredo

Classic network clustering tackles the problem of identifying sets of nodes (communities) that have similar connection patterns. However, in many scenarios nodes also have attributes that are correlated and can also be used to identify node clusters. Thus, network information (edges) and node information (attributes) can be jointly leveraged to design high-performance clustering algorithms. Under a general model for the network and node attributes, this work establishes an information-theoretic criteria for the exact recovery of community labels and characterizes a phase transition determined by the Chernoff-Hellinger divergence of the model. The criteria shows how network and attribute information can be exchanged in order to have exact recovery (e.g., more reliable network information requires less reliable attribute information). This work also presents an iterative clustering algorithm that maximizes the joint likelihood, assuming that the probability distribution of network interactions and node attributes belong to exponential families. This covers a broad range of possible interactions (e.g., edges with weights) and attributes (e.g., non-Gaussian models) while also exploring the connection between exponential families and Bregman divergences. Extensive numerical experiments using synthetic and real data indicate that the proposed algorithm outperforms algorithms that leverage only network or only attribute information as well as recently proposed algorithms that perform clustering using both sources of information. The contributions of this work provide insights into the fundamental limits and practical techniques for inferring community labels on node-attributed networks.
****

Learning to Receive Help: Intervention-Aware Concept Embedding Models

Mateo Espinosa Zarlenga, Katie Collins, Krishnamurthy Dvijotham, Adrian Weller, Zohreh Shams, Mateja Jamnik

Concept Bottleneck Models (CBMs) tackle the opacity of neural architectures by constructing and explaining their predictions using a set of high-level concepts. A special property of these models is that they permit concept interventions, wherein users can correct mispredicted concepts and thus improve the model's performance. Recent work, however, has shown that intervention efficacy can be highly dependent on the order in which concepts are intervened on and on the model's architecture and training hyperparameters. We argue that this is rooted in a CBM's lack of train-time incentives for the model to be appropriately receptive to concept interventions. To address this, we propose Intervention-aware Concept Embedding models (IntCEMs), a novel CBM-based architecture and training paradigm that improves a model's receptiveness to test-time interventions. Our model learns a concept intervention policy in an end-to-end fashion from where it can sampl

e meaningful intervention trajectories at train-time. This conditions IntCEMs to effectively select and receive concept interventions when deployed at test-time. Our experiments show that IntCEMs significantly outperform state-of-the-art concept-interpretable models when provided with test-time concept interventions, demonstrating the effectiveness of our approach.
****
Tracr: Compiled Transformers as a Laboratory for Interpretability
David Lindner, Janos Kramar, Sebastian Farquhar, Matthew Rahtz, Tom McGrath, Vladimir Mikulik
We show how to "compile" human-readable programs into standard decoder-only transformer models. Our compiler, Tracr, generates models with known structure. This structure can be used to design experiments. For example, we use it to study "superposition" in transformers that execute multi-step algorithms. Additionally, the known structure of Tracr-compiled models can serve as ground-truth for evaluating interpretability methods. Commonly, because the "programs" learned by transformers are unknown it is unclear whether an interpretation succeeded. We demonstrate our approach by implementing and examining programs including computing token frequencies, sorting, and parenthesis checking. We provide an open-source implementation of Tracr at https://github.com/google-deepmind/tracr.
****
KAKURENBO: Adaptively Hiding Samples in Deep Neural Network Training
Truong Thao Nguyen, Balazs Gerofi, Edgar Josafat Martinez-Noriega, François Trahay, Mohamed Wahib
This paper proposes a method for hiding the least-important samples during the training of deep neural networks to increase efficiency, i.e., to reduce the cost of training. Using information about the loss and prediction confidence during training, we adaptively find samples to exclude in a given epoch based on their contribution to the overall learning process, without significantly degrading accuracy. We explore the converge properties when accounting for the reduction in the number of SGD updates. Empirical results on various large-scale datasets and models used directly in image classification and segmentation show that while the with-replacement importance sampling algorithm performs poorly on large datasets, our method can reduce total training time by up to 22\% impacting accuracy only by 0.4\% compared to the baseline.
****
Mixed Samples as Probes for Unsupervised Model Selection in Domain Adaptation
Dapeng Hu, Jian Liang, Jun Hao Liew, Chuhui Xue, Song Bai, Xinchao Wang
Unsupervised domain adaptation (UDA) has been widely applied in improving model generalization on unlabeled target data. However, accurately selecting the best UDA model for the target domain is challenging due to the absence of labeled target data and domain distribution shifts. Traditional model selection approaches involve training extra models with source data to estimate the target validation risk. Recent studies propose practical methods that are based on measuring various properties of model predictions on target data. Although effective for some UDA models, these methods often lack stability and may lead to poor selections for other UDA models.In this paper, we present MixVal, an innovative model selection method that operates solely with unlabeled target data during inference. MixVal leverages mixed target samples with pseudo labels to directly probe the learned target structure by each UDA model. Specifically, MixVal employs two distinct types of probes: the intra-cluster mixed samples for evaluating neighborhood density and the inter-cluster mixed samples for investigating the classification boundary. With this comprehensive probing strategy, MixVal elegantly combines the strengths of two state-of-the-art model selection methods, Entropy and SND. We extensively evaluate MixVal on 11 UDA methods across 4 adaptation settings, including classification and segmentation tasks. Experimental results consistently demonstrate that MixVal achieves state-of-the-art performance and maintains exceptional stability in model selection. Code is available at \url{https://github.com/LHXXHB/MixVal}.
****
Payoff-based Learning with Matrix Multiplicative Weights in Quantum Games

Kyriakos Lotidis, Panayotis Mertikopoulos, Nicholas Bambos, Jose Blanchet
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Deep Stochastic Processes via Functional Markov Transition Operators
Jin Xu, Emilien Dupont, Kaspar Märtens, Thomas Rainforth, Yee Whye Teh
We introduce Markov Neural Processes (MNPs), a new class of Stochastic Processes (SPs) which are constructed by stacking sequences of neural parameterised Markov transition operators in function space. We prove that these Markov transition operators can preserve the exchangeability and consistency of SPs. Therefore, the proposed iterative construction adds substantial flexibility and expressivity to the original framework of Neural Processes (NPs) without compromising consistency or adding restrictions. Our experiments demonstrate clear advantages of MNPs over baseline models on a variety of tasks.
****
Quilt-1M: One Million Image-Text Pairs for Histopathology
Wisdom Ikezogwo, Saygin Seyfioglu, Fatemeh Ghezloo, Dylan Geva, Fatwir Sheikh Mohammed, Pavan Kumar Anand, Ranjay Krishna, Linda Shapiro
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Computation and Communication Efficient Method for Distributed Nonconvex Problems in the Partial Participation Setting
Alexander Tyurin, Peter Richtarik
We present a new method that includes three key components of distributed optimization and federated learning: variance reduction of stochastic gradients, partial participation, and compressed communication. We prove that the new method has optimal oracle complexity and state-of-the-art communication complexity in the partial participation setting. Regardless of the communication compression feature, our method successfully combines variance reduction and partial participation: we get the optimal oracle complexity, never need the participation of all nodes, and do not require the bounded gradients (dissimilarity) assumption.
****
Optimistic Active Exploration of Dynamical Systems
Bhavya , Lenart Treven, Cansu Sancaktar, Sebastian Blaes, Stelian Coros, Andreas Krause
Reinforcement learning algorithms commonly seek to optimize policies for solving one particular task. How should we explore an unknown dynamical system such that the estimated model allows us to solve multiple downstream tasks in a zero-shot manner? In this paper, we address this challenge, by developing an algorithm -- OPAX -- for active exploration. OPAX uses well-calibrated probabilistic models to quantify the epistemic uncertainty about the unknown dynamics. It optimistically---w.r.t. to plausible dynamics---maximizes the information gain between the unknown dynamics and state observations. We show how the resulting optimization problem can be reduced to an optimal control problem that can be solved at each episode using standard approaches. We analyze our algorithm for general models, and, in the case of Gaussian process dynamics, we give a sample complexity bound andshow that the epistemic uncertainty converges to zero. In our experiments, we compare OPAX with other heuristic active exploration approaches on several environments. Our experiments show that OPAX is not only theoretically sound but also performs well for zero-shot planning on novel downstream tasks.
****
HuggingGPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face
Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, Yueting Zhuang
Solving complicated AI tasks with different domains and modalities is a key step toward artificial general intelligence. While there are numerous AI models avai

lable for various domains and modalities, they cannot handle complicated AI tasks autonomously. Considering large language models (LLMs) have exhibited exceptional abilities in language understanding, generation, interaction, and reasoning, we advocate that LLMs could act as a controller to manage existing AI models to solve complicated AI tasks, with language serving as a generic interface to empower this. Based on this philosophy, we present HuggingGPT, an LLM-powered agent that leverages LLMs (e.g., ChatGPT) to connect various AI models in machine learning communities (e.g., Hugging Face) to solve AI tasks. Specifically, we use ChatGPT to conduct task planning when receiving a user request, select models according to their function descriptions available in Hugging Face, execute each subtask with the selected AI model, and summarize the response according to the execution results. By leveraging the strong language capability of ChatGPT and abundant AI models in Hugging Face, HuggingGPT can tackle a wide range of sophisticated AI tasks spanning different modalities and domains and achieve impressive results in language, vision, speech, and other challenging tasks, which paves a new way towards the realization of artificial general intelligence.
****

Multi-Step Generalized Policy Improvement by Leveraging Approximate Models
Lucas N. Alegre, Ana Bazzan, Ann Nowe, Bruno C. da Silva
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

GradOrth: A Simple yet Efficient Out-of-Distribution Detection with Orthogonal Projection of Gradients
Sima Behpour, Thang Long Doan, Xin Li, Wenbin He, Liang Gou, Liu Ren
Detecting out-of-distribution (OOD) data is crucial for ensuring the safe deployment of machine learning models in real-world applications. However, existing OOD detection approaches primarily rely on the feature maps or the full gradient space information to derive OOD scores neglecting the role of \textbf{most important parameters} of the pre-trained network over In-Distribution data. In this study, we propose a novel approach called GradOrth to facilitate OOD detection based on one intriguing observation that the important features to identify OOD data lie in the lower-rank subspace of in-distribution (ID) data.In particular, we identify OOD data by computing the norm of gradient projection on \textit{the subspaces considered \textbf{important} for the in-distribution data}. A large orthogonal projection value (i.e. a small projection value) indicates the sample as OOD as it captures a weak correlation of the in-distribution (ID) data. This simple yet effective method exhibits outstanding performance, showcasing a notable reduction in the average false positive rate at a 95\% true positive rate (FPR95) of up to 8\% when compared to the current state-of-the-art methods.
****

Learning to Modulate pre-trained Models in RL
Thomas Schmied, Markus Hofmarcher, Fabian Paischer, Razvan Pascanu, Sepp Hochreiter
Reinforcement Learning (RL) has been successful in various domains like robotics, game playing, and simulation. While RL agents have shown impressive capabilities in their specific tasks, they insufficiently adapt to new tasks. In supervised learning, this adaptation problem is addressed by large-scale pre-training followed by fine-tuning to new down-stream tasks. Recently, pre-training on multiple tasks has been gaining traction in RL. However, fine-tuning a pre-trained model often suffers from catastrophic forgetting. That is, the performance on the pre-training tasks deteriorates when fine-tuning on new tasks. To investigate the catastrophic forgetting phenomenon, we first jointly pre-train a model on datasets from two benchmark suites, namely Meta-World and DMControl. Then, we evaluate and compare a variety of fine-tuning methods prevalent in natural language processing, both in terms of performance on new tasks, and how well performance on pre-training tasks is retained. Our study shows that with most fine-tuning approaches, the performance on pre-training tasks deteriorates significantly. Therefor

e, we propose a novel method, Learning-to-Modulate (L2M), that avoids the degradation of learned skills by modulating the information flow of the frozen pre-trained model via a learnable modulation pool. Our method achieves state-of-the-art performance on the Continual-World benchmark, while retaining performance on the pre-training tasks. Finally, to aid future research in this area, we release a dataset encompassing 50 Meta-World and 16 DMControl tasks.
****
Injecting Multimodal Information into Rigid Protein Docking via Bi-level Optimization
Ruijia Wang, YiWu Sun, Yujie Luo, Shaochuan Li, Cheng Yang, Xingyi Cheng, Hui Li, Chuan Shi, Le Song
The structure of protein-protein complexes is critical for understanding binding dynamics, biological mechanisms, and intervention strategies. Rigid protein docking, a fundamental problem in this field, aims to predict the 3D structure of complexes from their unbound states without conformational changes. In this scenario, we have access to two types of valuable information: sequence-modal information, such as coevolutionary data obtained from multiple sequence alignments, and structure-modal information, including the 3D conformations of rigid structures. However, existing docking methods typically utilize single-modal information, resulting in suboptimal predictions. In this paper, we propose xTrimoBiDock (or BiDock for short), a novel rigid docking model that effectively integrates sequence- and structure-modal information through bi-level optimization. Specifically, a cross-modal transformer combines multimodal information to predict an inter-protein distance map. To achieve rigid docking, the roto-translation transformation is optimized to align the docked pose with the predicted distance map. In order to tackle this bi-level optimization problem, we unroll the gradient descent of the inner loop and further derive a better initialization for roto-translation transformation based on spectral estimation. Compared to baselines, BiDock achieves a promising result of a maximum 234% relative improvement in challenging antibody-antigen docking problem.
****
Uncertainty-Aware Alignment Network for Cross-Domain Video-Text Retrieval
Xiaoshuai Hao, Wanqian Zhang
Video-text retrieval is an important but challenging research task in the multimedia community. In this paper, we address the challenge task of Unsupervised Domain Adaptation Video-text Retrieval (UDAVR), assuming that training (source) data and testing (target) data are from different domains. Previous approaches are mostly derived from classification based domain adaptation methods, which are neither multi-modal nor suitable for retrieval task. In addition, as to the pairwise misalignment issue in target domain, i.e., no pairwise annotations between target videos and texts, the existing method assumes that a video corresponds to a text. Yet we empirically find that in the real scene, one text usually corresponds to multiple videos and vice versa. To tackle this one-to-many issue, we propose a novel method named Uncertainty-aware Alignment Network (UAN). Specifically, we first introduce the multimodal mutual information module to balance the minimization of domain shift in a smooth manner. To tackle the multimodal uncertainties pairwise misalignment in target domain, we propose the Uncertainty-aware Alignment Mechanism (UAM) to fully exploit the semantic information of both modalities in target domain. Extensive experiments in the context of domain-adaptive video-text retrieval demonstrate that our proposed method consistently outperforms multiple baselines, showing a superior generalization ability for target data.
****
Can Pre-Trained Text-to-Image Models Generate Visual Goals for Reinforcement Learning?
Jialu Gao, Kaizhe Hu, Guowei Xu, Huazhe Xu
Pre-trained text-to-image generative models can produce diverse, semantically rich, and realistic images from natural language descriptions. Compared with language, images usually convey information with more details and less ambiguity. In this study, we propose Learning from the Void (LfVoid), a method that leverages

the power of pre-trained text-to-image models and advanced image editing techniq
ues to guide robot learning. Given natural language instructions, LfVoid can edi
t the original observations to obtain goal images, such as "wiping" a stain off
a table. Subsequently, LfVoid trains an ensembled goal discriminator on the gene
rated image to provide reward signals for a reinforcement learning agent, guidin
g it to achieve the goal. The ability of LfVoid to learn with zero in-domain tra
ining on expert demonstrations or true goal observations (the void) is attribute
d to the utilization of knowledge from web-scale generative models. We evaluate
LfVoid across three simulated tasks and validate its feasibility in the correspo
nding real-world scenarios. In addition, we offer insights into the key consider
ations for the effective integration of visual generative models into robot lear
ning workflows. We posit that our work represents an initial step towards the br
oader application of pre-trained visual generative models in the robotics field.
 Our project page: https://lfvoid-rl.github.io/.
****
H3T: Efficient Integration of Memory Optimization and Parallelism for Large-scal
e Transformer Training
Yuzhong Wang, Xu Han, Weilin Zhao, Guoyang Zeng, Zhiyuan Liu, Maosong Sun
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Binarized Spectral Compressive Imaging
Yuanhao Cai, Yuxin Zheng, Jing Lin, Xin Yuan, Yulun Zhang, Haoqian Wang
Existing deep learning models for hyperspectral image (HSI) reconstruction achie
ve good performance but require powerful hardwares with enormous memory and comp
utational resources. Consequently, these methods can hardly be deployed on resou
rce-limited mobile devices. In this paper, we propose a novel method, Binarized
Spectral-Redistribution Network (BiSRNet), for efficient and practical HSI resto
ration from compressed measurement in snapshot compressive imaging (SCI) systems
. Firstly, we redesign a compact and easy-to-deploy base model to be binarized.
Then we present the basic unit, Binarized Spectral-Redistribution Convolution (B
iSR-Conv). BiSR-Conv can adaptively redistribute the HSI representations before
binarizing activation and uses a scalable hyperbolic tangent function to closer
approximate the Sign function in backpropagation. Based on our BiSR-Conv, we cus
tomize four binarized convolutional modules to address the dimension mismatch an
d propagate full-precision information throughout the whole network. Finally, ou
r BiSRNet is derived by using the proposed techniques to binarize the base model
. Comprehensive quantitative and qualitative experiments manifest that our propo
sed BiSRNet outperforms state-of-the-art binarization algorithms. Code and model
s are publicly available at https://github.com/caiyuanhao1998/BiSCI
****
When Can We Track Significant Preference Shifts in Dueling Bandits?
Joe Suk, Arpit Agarwal
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Neural Latent Geometry Search: Product Manifold Inference via Gromov-Hausdorff-I
nformed Bayesian Optimization
Haitz Sáez de Ocáriz Borde, Alvaro Arroyo, Ismael Morales, Ingmar Posner, Xiaowe
n Dong
Recent research indicates that the performance of machine learning models can be
 improved by aligning the geometry of the latent space with the underlying data
structure. Rather than relying solely on Euclidean space, researchers have propo
sed using hyperbolic and spherical spaces with constant curvature, or combinatio
ns thereof, to better model the latent space and enhance model performance. Howe
ver, little attention has been given to the problem of automatically identifying

the optimal latent geometry for the downstream task. We mathematically define this novel formulation and coin it as neural latent geometry search (NLGS). More specifically, we introduce an initial attempt to search for a latent geometry composed of a product of constant curvature model spaces with a small number of query evaluations, under some simplifying assumptions. To accomplish this, we propose a novel notion of distance between candidate latent geometries based on the Gromov-Hausdorff distance from metric geometry. In order to compute the Gromov-Hausdorff distance, we introduce a mapping function that enables the comparison of different manifolds by embedding them in a common high-dimensional ambient space. We then design a graph search space based on the notion of smoothness between latent geometries and employ the calculated distances as an additional inductive bias. Finally, we use Bayesian optimization to search for the optimal latent geometry in a query-efficient manner. This is a general method which can be applied to search for the optimal latent geometry for a variety of models and downstream tasks. We perform experiments on synthetic and real-world datasets to identify the optimal latent geometry for multiple machine learning problems.
****

Scientific Document Retrieval using Multi-level Aspect-based Queries
Jianyou (Andre) Wang, Kaicheng Wang, Xiaoyue Wang, Prudhviraj Naidu, Leon Bergen, Ramamohan Paturi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Beyond Confidence: Reliable Models Should Also Consider Atypicality
Mert Yuksekgonul, Linjun Zhang, James Y. Zou, Carlos Guestrin
While most machine learning models can provide confidence in their predictions, confidence is insufficient to understand a prediction's reliability. For instance, the model may have a low confidence prediction if the input is not well-represented in the training dataset or if the input is inherently ambiguous. In this work, we investigate the relationship between how atypical~(rare) a sample or a class is and the reliability of a model's predictions. We first demonstrate that atypicality is strongly related to miscalibration and accuracy. In particular, we empirically show that predictions for atypical inputs or atypical classes are more overconfident and have lower accuracy. Using these insights, we show incorporating atypicality improves uncertainty quantification and model performance for discriminative neural networks and large language models. In a case study, we show that using atypicality improves the performance of a skin lesion classifier across different skin tone groups without having access to the group attributes. Overall, we propose that models should use not only confidence but also atypicality to improve uncertainty quantification and performance. Our results demonstrate that simple post-hoc atypicality estimators can provide significant value.
****

Reversible and irreversible bracket-based dynamics for deep graph neural networks
Anthony Gruber, Kookjin Lee, Nathaniel Trask
Recent works have shown that physics-inspired architectures allow the training of deep graph neural networks (GNNs) without oversmoothing. The role of these physics is unclear, however, with successful examples of both reversible (e.g., Hamiltonian) and irreversible (e.g., diffusion) phenomena producing comparable results despite diametrically opposed mechanisms, and further complications arising due to empirical departures from mathematical theory. This work presents a series of novel GNN architectures based upon structure-preserving bracket-based dynamical systems, which are provably guaranteed to either conserve energy or generate positive dissipation with increasing depth. It is shown that the theoretically principled framework employed here allows for inherently explainable constructions, which contextualize departures from theory in current architectures and better elucidate the roles of reversibility and irreversibility in network performance. Code is available at the Github repository \url{https://github.com/natrask

/BracketGraphs}.
****
In Defense of Softmax Parametrization for Calibrated and Consistent Learning to Defer
Yuzhou Cao, Hussein Mozannar, Lei Feng, Hongxin Wei, Bo An
Enabling machine learning classifiers to defer their decision to a downstream expert when the expert is more accurate will ensure improved safety and performance. This objective can be achieved with the learning-to-defer framework which aims to jointly learn how to classify and how to defer to the expert. In recent studies, it has been theoretically shown that popular estimators for learning to defer parameterized with softmax provide unbounded estimates for the likelihood of deferring which makes them uncalibrated. However, it remains unknown whether this is due to the widely used softmax parameterization and if we can find a softmax-based estimator that is both statistically consistent and possesses a valid probability estimator. In this work, we first show that the cause of the miscalibrated and unbounded estimator in prior literature is due to the symmetric nature of the surrogate losses used and not due to softmax. We then propose a novel statistically consistent asymmetric softmax-based surrogate loss that can produce valid estimates without the issue of unboundedness. We further analyze the non-asymptotic properties of our proposed method and empirically validate its performance and calibration on benchmark datasets.
****
Leveraging Vision-Centric Multi-Modal Expertise for 3D Object Detection
Linyan Huang, Zhiqi Li, Chonghao Sima, Wenhai Wang, Jingdong Wang, Yu Qiao, Hongyang Li
Current research is primarily dedicated to advancing the accuracy of camera-only 3D object detectors (apprentice) through the knowledge transferred from LiDAR- or multi-modal-based counterparts (expert). However, the presence of the domain gap between LiDAR and camera features, coupled with the inherent incompatibility in temporal fusion, significantly hinders the effectiveness of distillation-based enhancements for apprentices. Motivated by the success of uni-modal distillation, an apprentice-friendly expert model would predominantly rely on camera features, while still achieving comparable performance to multi-modal models. To this end, we introduce VCD, a framework to improve the camera-only apprentice model, including an apprentice-friendly multi-modal expert and temporal-fusion-friendly distillation supervision. The multi-modal expert VCD-E adopts an identical structure as that of the camera-only apprentice in order to alleviate the feature disparity, and leverages LiDAR input as a depth prior to reconstruct the 3D scene, achieving the performance on par with other heterogeneous multi-modal experts. Additionally, a fine-grained trajectory-based distillation module is introduced with the purpose of individually rectifying the motion misalignment for each object in the scene. With those improvements, our camera-only apprentice VCD-A sets new state-of-the-art on nuScenes with a score of 63.1% NDS. The code will be released at https://github.com/OpenDriveLab/Birds-eye-view-Perception.
****
No-Regret Online Reinforcement Learning with Adversarial Losses and Transitions
Tiancheng Jin, Junyan Liu, Chloé Rouyer, William Chang, Chen-Yu Wei, Haipeng Luo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Massively Multilingual Corpus of Sentiment Datasets and Multi-faceted Sentiment Classification Benchmark
Lukasz Augustyniak, Szymon Wo■niak, Marcin Gruza, Piotr Gramacki, Krzysztof Rajda, Miko■aj Morzy, Tomasz Kajdanowicz
Despite impressive advancements in multilingual corpora collection and model training, developing large-scale deployments of multilingual models still presents a significant challenge. This is particularly true for language tasks that are culture-dependent. One such example is the area of multilingual sentiment analysi

s, where affective markers can be subtle and deeply ensconced in culture.This work presents the most extensive open massively multilingual corpus of datasets for training sentiment models. The corpus consists of 79 manually selected datasets from over 350 datasets reported in the scientific literature based on strict quality criteria. The corpus covers 27 languages representing 6 language families. Datasets can be queried using several linguistic and functional features. In addition, we present a multi-faceted sentiment classification benchmark summarizing hundreds of experiments conducted on different base models, training objectives, dataset collections, and fine-tuning strategies.
****

Generalizable Lightweight Proxy for Robust NAS against Diverse Perturbations
Hyeonjeong Ha, Minseon Kim, Sung Ju Hwang
Recent neural architecture search (NAS) frameworks have been successful in finding optimal architectures for given conditions (e.g., performance or latency). However, they search for optimal architectures in terms of their performance on clean images only, while robustness against various types of perturbations or corruptions is crucial in practice. Although there exist several robust NAS frameworks that tackle this issue by integrating adversarial training into one-shot NAS, however, they are limited in that they only consider robustness against adversarial attacks and require significant computational resources to discover optimal architectures for a single task, which makes them impractical in real-world scenarios. To address these challenges, we propose a novel lightweight robust zero-cost proxy that considers the consistency across features, parameters, and gradients of both clean and perturbed images at the initialization state. Our approach facilitates an efficient and rapid search for neural architectures capable of learning generalizable features that exhibit robustness across diverse perturbations. The experimental results demonstrate that our proxy can rapidly and efficiently search for neural architectures that are consistently robust against various perturbations on multiple benchmark datasets and diverse search spaces, largely outperforming existing clean zero-shot NAS and robust NAS with reduced search cost.
****

Ignorance is Bliss: Robust Control via Information Gating
Manan Tomar, Riashat Islam, Matthew Taylor , Sergey Levine, Philip Bachman
Informational parsimony provides a useful inductive bias for learning representations that achieve better generalization by being robust to noise and spurious correlations. We propose information gating as a way to learn parsimonious representations that identify the minimal information required for a task. When gating information, we can learn to reveal as little information as possible so that a task remains solvable, or hide as little information as possible so that a task becomes unsolvable. We gate information using a differentiable parameterization of the signal-to-noise ratio, which can be applied to arbitrary values in a network, e.g., erasing pixels at the input layer or activations in some intermediate layer. When gating at the input layer, our models learn which visual cues matter for a given task. When gating intermediate layers, our models learn which activations are needed for subsequent stages of computation. We call our approach InfoGating. We apply InfoGating to various objectives such as multi-step forward and inverse dynamics models, Q-learning, and behavior cloning, highlighting how InfoGating can naturally help in discarding information not relevant for control. Results show that learning to identify and use minimal information can improve generalization in downstream tasks. Policies based on InfoGating are considerably more robust to irrelevant visual features, leading to improved pretraining and finetuning of RL models.
****

Reduced Policy Optimization for Continuous Control with Hard Constraints
Shutong Ding, Jingya Wang, Yali Du, Ye Shi
Recent advances in constrained reinforcement learning (RL) have endowed reinforcement learning with certain safety guarantees. However, deploying existing constrained RL algorithms in continuous control tasks with general hard constraints remains challenging, particularly in those situations with non-convex hard constr

aints. Inspired by the generalized reduced gradient (GRG) algorithm, a classical constrained optimization technique, we propose a reduced policy optimization (RPO) algorithm that combines RL with GRG to address general hard constraints. RPO partitions actions into basic actions and nonbasic actions following the GRG method and outputs the basic actions via a policy network. Subsequently, RPO calculates the nonbasic actions by solving equations based on equality constraints using the obtained basic actions. The policy network is then updated by implicitly differentiating nonbasic actions with respect to basic actions. Additionally, we introduce an action projection procedure based on the reduced gradient and apply a modified Lagrangian relaxation technique to ensure inequality constraints are satisfied. To the best of our knowledge, RPO is the first attempt that introduces GRG to RL as a way of efficiently handling both equality and inequality hard constraints. It is worth noting that there is currently a lack of RL environments with complex hard constraints, which motivates us to develop three new benchmarks: two robotics manipulation tasks and a smart grid operation control task. With these benchmarks, RPO achieves better performance than previous constrained RL algorithms in terms of both cumulative reward and constraint violation. We believe RPO, along with the new benchmarks, will open up new opportunities for applying RL to real-world problems with complex constraints.
****

ALIM: Adjusting Label Importance Mechanism for Noisy Partial Label Learning
Mingyu Xu, Zheng Lian, Lei Feng, Bin Liu, Jianhua Tao
Noisy partial label learning (noisy PLL) is an important branch of weakly supervised learning. Unlike PLL where the ground-truth label must conceal in the candidate label set, noisy PLL relaxes this constraint and allows the ground-truth label may not be in the candidate label set. To address this challenging problem, most of the existing works attempt to detect noisy samples and estimate the ground-truth label for each noisy sample. However, detection errors are unavoidable. These errors can accumulate during training and continuously affect model optimization. To this end, we propose a novel framework for noisy PLL with theoretical interpretations, called ``Adjusting Label Importance Mechanism (ALIM)''. It aims to reduce the negative impact of detection errors by trading off the initial candidate set and model outputs. ALIM is a plug-in strategy that can be integrated with existing PLL approaches. Experimental results on multiple benchmark datasets demonstrate that our method can achieve state-of-the-art performance on noisy PLL. Our code is available at: https://github.com/zeroQiaoba/ALIM.
****

Conditional Score Guidance for Text-Driven Image-to-Image Translation
Hyunsoo Lee, Minsoo Kang, Bohyung Han
We present a novel algorithm for text-driven image-to-image translation based on a pretrained text-to-image diffusion model. Our method aims to generate a target image by selectively editing regions of interest in a source image, defined by a modifying text, while preserving the remaining parts.In contrast to existing techniques that solely rely on a target prompt, we introduce a new score function that additionally considers both the source image and the source text prompt, tailored to address specific translation tasks. To this end, we derive the conditional score function in a principled way, decomposing it into the standard score and a guiding term for target image generation.For the gradient computation about the guiding term, we assume a Gaussian distribution for the posterior distribution and estimate its mean and variance to adjust the gradient without additional training.In addition, to improve the quality of the conditional score guidance, we incorporate a simple yet effective mixup technique, which combines two cross-attention maps derived from the source and target latents.This strategy is effective for promoting a desirable fusion of the invariant parts in the source image and the edited regions aligned with the target prompt, leading to high-fidelity target image generation.Through comprehensive experiments, we demonstrate that our approach achieves outstanding image-to-image translation performance on various tasks.Code is available at https://github.com/Hleephilip/CSG.
****

A Unified Approach to Count-Based Weakly Supervised Learning

Vinay Shukla, Zhe Zeng, Kareem Ahmed, Guy Van den Broeck
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Transformers are uninterpretable with myopic methods: a case study with bounded Dyck grammars
Kaiyue Wen, Yuchen Li, Bingbin Liu, Andrej Risteski
Transformer interpretability aims to understand the algorithm implemented by a learned Transformer by examining various aspects of the model, such as the weight matrices or the attention patterns.In this work, through a combination of theoretical results and carefully controlled experiments on synthetic data, we take a critical viewof methods that exclusively focus on individual parts of the model, rather than consider the network as a whole.We consider a simple synthetic setup of learning a (bounded) Dyck language. Theoretically, we show that the set of models that (exactly or approximately) solve this task satisfy a structural characterization derived from ideas in formal languages (the pumping lemma).We use this characterization to show that the set of optima is qualitatively rich; in particular, the attention pattern of a single layer can be "nearly randomized", while preserving the functionality of the network.We also show via extensive experiments that these constructions are not merely a theoretical artifact: even with severe constraints to the architecture of the model, vastly different solutions can be reached via standard training. Thus, interpretability claims based on inspecting individual heads or weight matrices in the Transformer can be misleading.
****
GEQ: Gaussian Kernel Inspired Equilibrium Models
Mingjie Li, Yisen Wang, Zhouchen Lin
Despite the connection established by optimization-induced deep equilibrium models (OptEqs) between their output and the underlying hidden optimization problems, the performance of it along with its related works is still not good enough especially when compared to deep networks. One key factor responsible for this performance limitation is the use of linear kernels to extract features in these models. To address this issue, we propose a novel approach by replacing its linear kernel with a new function that can readily capture nonlinear feature dependencies in the input data. Drawing inspiration from classical machine learning algorithms, we introduce Gaussian kernels as the alternative function and then propose our new equilibrium model, which we refer to as GEQ. By leveraging Gaussian kernels, GEQ can effectively extract the nonlinear information embedded within the input features, surpassing the performance of the original OptEqs. Moreover, GEQ can be perceived as a weight-tied neural network with infinite width and depth. GEQ also enjoys better theoretical properties and improved overall performance. Additionally, our GEQ exhibits enhanced stability when confronted with various samples. We further substantiate the effectiveness and stability of GEQ through a series of comprehensive experiments.
****
Efficient Potential-based Exploration in Reinforcement Learning using Inverse Dynamic Bisimulation Metric
Yiming Wang, Ming Yang, Renzhi Dong, Binbin Sun, Furui Liu, Leong Hou U
Reward shaping is an effective technique for integrating domain knowledge into reinforcement learning (RL). However, traditional approaches like potential-based reward shaping totally rely on manually designing shaping reward functions, which significantly restricts exploration efficiency and introduces human cognitive biases.While a number of RL methods have been proposed to boost exploration by designing an intrinsic reward signal as exploration bonus. Nevertheless, these methods heavily rely on the count-based episodic term in their exploration bonus which falls short in scalability. To address these limitations, we propose a general end-to-end potential-based exploration bonus for deep RL via potentials of state discrepancy, which motivates the agent to discover novel states and provid

es them with denser rewards without manual intervention. Specifically, we measure the novelty of adjacent states by calculating their distance using the bisimulation metric-based potential function, which enhances agent's exploration and ensures policy invariance. In addition, we offer a theoretical guarantee on our inverse dynamic bisimulation metric, bounding the value difference and ensuring that the agent explores states with higher TD error, thus significantly improving training efficiency. The proposed approach is named \textbf{LIBERTY} (exp\textbf{L}oration v\textbf{I}a \textbf{B}isimulation m\textbf{E}t\textbf{R}ic-based s\textbf{T}ate discrepanc\textbf{Y}) which is comprehensively evaluated on the MuJoCo and the Arcade Learning Environments. Extensive experiments have verified the superiority and scalability of our algorithm compared with other competitive methods.
****

What’s Left? Concept Grounding with Logic-Enhanced Foundation Models

Joy Hsu, Jiayuan Mao, Josh Tenenbaum, Jiajun Wu

Recent works such as VisProg and ViperGPT have smartly composed foundation models for visual reasoning—using large language models (LLMs) to produce programs that can be executed by pre-trained vision-language models. However, they operate in limited domains, such as 2D images, not fully exploiting the generalization of language: abstract concepts like “left” can also be grounded in 3D, temporal, and action data, as in moving to your left. This limited generalization stems from these inference-only methods’ inability to learn or adapt pre-trained models to a new domain. We propose the Logic-Enhanced FoundaTion Model (LEFT), a unified framework that learns to ground and reason with concepts across domains with a differentiable, domain-independent, first-order logic-based program executor. LEFT has an LLM interpreter that outputs a program represented in a general, logic-based reasoning language, which is shared across all domains and tasks. LEFT’s executor then executes the program with trainable domain-specific grounding modules. We show that LEFT flexibly learns concepts in four domains: 2D images, 3D scenes, human motions, and robotic manipulation. It exhibits strong reasoning ability in a wide variety of tasks, including those that are complex and not seen during training, and can be easily applied to new domains.
****

Recovering from Out-of-sample States via Inverse Dynamics in Offline Reinforcement Learning

Ke Jiang, Jia-Yu Yao, Xiaoyang Tan

In this paper we deal with the state distributional shift problem commonly encountered in offline reinforcement learning during test, where the agent tends to take unreliable actions at out-of-sample (unseen) states. Our idea is to encourage the agent to follow the so called state recovery principle when taking actions, i.e., besides long-term return, the immediate consequences of the current action should also be taken into account and those capable of recovering the state distribution of the behavior policy are preferred. For this purpose, an inverse dynamics model is learned and employed to guide the state recovery behavior of the new policy. Theoretically, we show that the proposed method helps aligning the transited state distribution of the new policy with the offline dataset at out-of-sample states, without the need of explicitly predicting the transited state distribution, which is usually difficult in high-dimensional and complicated environments. The effectiveness and feasibility of the proposed method is demonstrated with the state-of-the-art performance on the general offline RL benchmarks.
****

VTaC: A Benchmark Dataset of Ventricular Tachycardia Alarms from ICU Monitors

Li-wei Lehman, Benjamin Moody, Harsh Deep, Feng Wu, Hasan Saeed, Lucas McCullum, Diane Perry, Tristan Struja, Qiao Li, Gari Clifford, Roger Mark

False arrhythmia alarms in intensive care units (ICUs) are a continuing problem despite considerable effort from industrial and academic algorithm developers. Of all life-threatening arrhythmias, ventricular tachycardia (VT) stands out as the most challenging arrhythmia to detect reliably. We introduce a new annotated VT alarm database, VTaC (Ventricular Tachycardia annotated alarms from ICUs) consisting of over 5,000 waveform recordings with VT alarms triggered by bedside

monitors in the ICU. Each VT alarm waveform in the dataset has been labeled by a
t least two independent human expert annotators. The dataset encompasses data co
llected from ICUs in two major US hospitals and includes data from three leading
 bedside monitor manufacturers, providing a diverse and representative collectio
n of alarm waveform data. Each waveform recording comprises at least two electro
cardiogram (ECG) leads and one or more pulsatile waveforms, such as photoplethys
mogram (PPG or PLETH) and arterial blood pressure (ABP) waveforms. We demonstrat
e the utility of this new benchmark dataset for the task of false arrhythmia ala
rm reduction, and present performance of multiple machine learning approaches, i
ncluding conventional supervised machine learning, deep learning, semi-supervise
d learning, and generative approaches for the task of VT false alarm reduction.
****
TMT-VIS: Taxonomy-aware Multi-dataset Joint Training for Video Instance Segmenta
tion
Rongkun Zheng, Lu Qi, Xi Chen, Yi Wang, Kun Wang, Yu Qiao, Hengshuang Zhao
Training on large-scale datasets can boost the performance of video instance seg
mentation while the annotated datasets for VIS are hard to scale up due to the h
igh labor cost. What we possess are numerous isolated filed-specific datasets, t
hus, it is appealing to jointly train models across the aggregation of datasets
to enhance data volume and diversity. However, due to the heterogeneity in categ
ory space, as mask precision increase with the data volume, simply utilizing mul
tiple datasets will dilute the attention of models on different taxonomy. Thus,
increasing the data scale and enriching taxonomy space while improving classific
ation precision is important. In this work, we analyze that providing extra taxo
nomy information can help models concentrate on specific taxonomy, and propose o
ur model named Taxonomy-aware Multi-dataset Joint Training for Video Instance Se
gmentation (TMT-VIS) to address this vital challenge. Specifically, we design a
two-stage taxonomy aggregation module that first compiles taxonomy information f
rom input videos and then aggregates these taxonomy priors into instance queries
 before the transformer decoder. We conduct extensive experimental evaluations o
n four popular and challenging benchmarks, including YouTube-VIS 2019, YouTube-V
IS 2021, OVIS, and UVO. Our model shows significant improvement over the baselin
e solutions, and sets new state-of-the-art records on all these benchmarks. Thes
e appealing and encouraging results demonstrate the effectiveness and generality
 of our proposed approach. The code and trained models will be publicly availabl
e.
****
Ego4D Goal-Step: Toward Hierarchical Understanding of Procedural Activities
Yale Song, Eugene Byrne, Tushar Nagarajan, Huiyu Wang, Miguel Martin, Lorenzo To
rresani
Human activities are goal-oriented and hierarchical, comprising primary goals at
 the top level, sequences of steps and substeps in the middle, and atomic action
s at the lowest level. Recognizing human activities thus requires relating atomi
c actions and steps to their functional objectives (what the actions contribute
to) and modeling their sequential and hierarchical dependencies towards achievin
g the goals. Current activity recognition research has primarily focused on only
 the lowest levels of this hierarchy, i.e., atomic or low-level actions, often i
n trimmed videos with annotations spanning only a few seconds. In this work, we
introduce Ego4D Goal-Step, a new set of annotations on the recently released Ego
4D with a novel hierarchical taxonomy of goal-oriented activity labels. It provi
des dense annotations for 48K procedural step segments (430 hours) and high-leve
l goal annotations for 2,807 hours of Ego4D videos. Compared to existing procedu
ral video datasets, it is substantially larger in size, contains hierarchical ac
tion labels (goals - steps - substeps), and provides goal-oriented auxiliary inf
ormation including natural language summary description, step completion status,
 and step-to-goal relevance information. We take a data-driven approach to build
 our taxonomy, resulting in dense step annotations that do not suffer from poor
label-data alignment issues resulting from a taxonomy defined a priori. Through
comprehensive evaluations and analyses, we demonstrate how Ego4D Goal-Step suppo
rts exploring various questions in procedural activity understanding, including

goal inference, step prediction, hierarchical relation learning, and long-term temporal modeling.
****
The Emergence of Essential Sparsity in Large Pre-trained Models: The Weights that Matter
AJAY JAISWAL, Shiwei Liu, Tianlong Chen, Zhangyang "Atlas" Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Hypervolume Maximization: A Geometric View of Pareto Set Learning
Xiaoyuan Zhang, Xi Lin, Bo Xue, Yifan Chen, Qingfu Zhang
This paper presents a novel approach to multiobjective algorithms aimed at modeling the Pareto set using neural networks. Whereas previous methods mainly focused on identifying a finite number of solutions, our approach allows for the direct modeling of the entire Pareto set. Furthermore, we establish an equivalence between learning the complete Pareto set and maximizing the associated hypervolume, which enables the convergence analysis of hypervolume (as a new metric) for Pareto set learning. Specifically, our new analysis framework reveals the connection between the learned Pareto solution and its representation in a polar coordinate system. We evaluate our proposed approach on various benchmark problems and real-world problems, and the encouraging results make it a potentially viable alternative to existing multiobjective algorithms. Code is available at \url{https://github.com/xzhang2523/hvpsl/tree/master}.
****
Equivariant Neural Simulators for Stochastic Spatiotemporal Dynamics
Koen Minartz, Yoeri Poels, Simon Koop, Vlado Menkovski
Neural networks are emerging as a tool for scalable data-driven simulation of high-dimensional dynamical systems, especially in settings where numerical methods are infeasible or computationally expensive. Notably, it has been shown that incorporating domain symmetries in deterministic neural simulators can substantially improve their accuracy, sample efficiency, and parameter efficiency. However, to incorporate symmetries in probabilistic neural simulators that can simulate stochastic phenomena, we need a model that produces equivariant distributions over trajectories, rather than equivariant function approximations. In this paper, we propose Equivariant Probabilistic Neural Simulation (EPNS), a framework for autoregressive probabilistic modeling of equivariant distributions over system evolutions. We use EPNS to design models for a stochastic n-body system and stochastic cellular dynamics. Our results show that EPNS considerably outperforms existing neural network-based methods for probabilistic simulation. More specifically, we demonstrate that incorporating equivariance in EPNS improves simulation quality, data efficiency, rollout stability, and uncertainty quantification. We conclude that EPNS is a promising method for efficient and effective data-driven probabilistic simulation in a diverse range of domains.
****
Collaborative Alignment of NLP Models
Fereshte Khani, Marco Tulio Ribeiro
Despite substantial advancements, Natural Language Processing (NLP) models often require post-training adjustments to enforce business rules, rectify undesired behavior, and align with user values. These adjustments involve operationalizing "concepts"—dictating desired model responses to certain inputs. However, it's difficult for a single entity to enumerate and define all possible concepts, indicating a need for a multi-user, collaborative model alignment framework. Moreover, the exhaustive delineation of a concept is challenging, and an improper approach can create shortcuts or interfere with original data or other concepts.To address these challenges, we introduce CoAlign, a framework that enables multi-user interaction with the model, thereby mitigating individual limitations. CoAlign aids users in operationalizing their concepts using Large Language Models, and relying on the principle that NLP models exhibit simpler behaviors in local reg

ions. Our main insight is learning a \emph{local} model for each concept, and a \emph{global} model to integrate the original data with all concepts.We then steer a large language model to generate instances within concept boundaries where local and global disagree.Our experiments show CoAlign is effective at helping multiple users operationalize concepts and avoid interference for a variety of scenarios, tasks, and models.
****
PlanBench: An Extensible Benchmark for Evaluating Large Language Models on Planning and Reasoning about Change
Karthik Valmeekam, Matthew Marquez, Alberto Olmo, Sarath Sreedharan, Subbarao Kambhampati
Generating plans of action, and reasoning about change have long been considered a core competence of intelligent agents. It is thus no surprise that evaluating the planning and reasoning capabilities of large language models (LLMs) has become a hot topic of research. Most claims about LLM planning capabilities are however based on common sense tasks–where it becomes hard to tell whether LLMs are planning or merely retrieving from their vast world knowledge. There is a strong need for systematic and extensible planning benchmarks with sufficient diversity to evaluate whether LLMs have innate planning capabilities. Motivated by this, we propose PlanBench, an extensible benchmark suite based on the kinds of domains used in the automated planning community, especially in the International Planning Competition, to test the capabilities of LLMs in planning or reasoning about actions and change. PlanBench provides sufficient diversity in both the task domains and the specific planning capabilities. Our studies also show that on many critical capabilities–including plan generation–LLM performance falls quite short, even with the SOTA models. PlanBench can thus function as a useful marker of progress of LLMs in planning and reasoning.
****
Extraction and Recovery of Spatio-Temporal Structure in Latent Dynamics Alignment with Diffusion Models
Yule Wang, Zijing Wu, Chengrui Li, Anqi Wu
In the field of behavior-related brain computation, it is necessary to align raw neural signals against the drastic domain shift among them. A foundational framework within neuroscience research posits that trial-based neural population activities rely on low-dimensional latent dynamics, thus focusing on the latter greatly facilitates the alignment procedure. Despite this field's progress, existing methods ignore the intrinsic spatio-temporal structure during the alignment phase. Hence, their solutions usually lead to poor quality in latent dynamics structures and overall performance. To tackle this problem, we propose an alignment method ERDiff, which leverages the expressivity of the diffusion model to preserve the spatio-temporal structure of latent dynamics. Specifically, the latent dynamics structures of the source domain are first extracted by a diffusion model. Then, under the guidance of this diffusion model, such structures are well-recovered through a maximum likelihood alignment procedure in the target domain. We first demonstrate the effectiveness of our proposed method on a synthetic dataset. Then, when applied to neural recordings from the non-human primate motor cortex, under both cross-day and inter-subject settings, our method consistently manifests its capability of preserving the spatio-temporal structure of latent dynamics and outperforms existing approaches in alignment goodness-of-fit and neural decoding performance.
****
Closing the gap between the upper bound and lower bound of Adam's iteration complexity
Bohan Wang, Jingwen Fu, Huishuai Zhang, Nanning Zheng, Wei Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Deep Patch Visual Odometry

Zachary Teed, Lahav Lipson, Jia Deng
We propose Deep Patch Visual Odometry (DPVO), a new deep learning system for monocular Visual Odometry (VO). DPVO uses a novel recurrent network architecture designed for tracking image patches across time. Recent approaches to VO have significantly improved the state-of-the-art accuracy by using deep networks to predict dense flow between video frames. However, using dense flow incurs a large computational cost, making these previous methods impractical for many use cases. Despite this, it has been assumed that dense flow is important as it provides additional redundancy against incorrect matches. DPVO disproves this assumption, showing that it is possible to get the best accuracy and efficiency by exploiting the advantages of sparse patch-based matching over dense flow. DPVO introduces a novel recurrent update operator for patch based correspondence coupled with differentiable bundle adjustment. On Standard benchmarks, DPVO outperforms all prior work, including the learning-based state-of-the-art VO-system (DROID) using a third of the memory while running 3x faster on average. Code is available at https://github.com/princeton-vl/DPVO
****
BoardgameQA: A Dataset for Natural Language Reasoning with Contradictory Information
Mehran Kazemi, Quan Yuan, Deepti Bhatia, Najoung Kim, Xin Xu, Vaiva Imbrasaite, Deepak Ramachandran
Automated reasoning with unstructured natural text is a key requirement for many potential applications of NLP and for developing robust AI systems. Recently, Language Models (LMs) have demonstrated complex reasoning capacities even without any finetuning. However, existing evaluation for automated reasoning assumes access to a consistent and coherent set of information over which models reason. When reasoning in the real-world, the available information is frequently inconsistent or contradictory, and therefore models need to be equipped with a strategy to resolve such conflicts when they arise. One widely-applicable way of resolving conflicts is to impose preferences over information sources (e.g., based on source credibility or information recency) and adopt the source with higher preference. In this paper, we formulate the problem of reasoning with contradictory information guided by preferences over sources as the classical problem of defeasible reasoning, and develop a dataset called BoardgameQA for measuring the reasoning capacity of LMs in this setting. BoardgameQA also incorporates reasoning with implicit background knowledge, to better reflect reasoning problems in downstream applications. We benchmark various LMs on BoardgameQA and the results reveal a significant gap in the reasoning capacity of state-of-the-art LMs on this problem, showing that reasoning with conflicting information does not surface out-of-the-box in LMs. While performance can be improved with finetuning, it nevertheless remains poor.
****
Isometric Quotient Variational Auto-Encoders for Structure-Preserving Representation Learning
In Huh, changwook jeong, Jae Myung Choe, YOUNGGU KIM, Daesin Kim
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SpokenWOZ: A Large-Scale Speech-Text Benchmark for Spoken Task-Oriented Dialogue Agents
Shuzheng Si, Wentao Ma, Haoyu Gao, Yuchuan Wu, Ting-En Lin, Yinpei Dai, Hangyu Li, Rui Yan, Fei Huang, Yongbin Li
Task-oriented dialogue (TOD) models have made significant progress in recent years. However, previous studies primarily focus on datasets written by annotators, which has resulted in a gap between academic research and real-world spoken con - versation scenarios. While several small-scale spoken TOD datasets are proposed to address robustness issues such as ASR errors, they ignore the unique challenges in spoken conversation. To tackle the limitations, we introduce SpokenWOZ,

a large-scale speech-text dataset for spoken TOD, containing 8 domains, 203k turns, 5.7k dialogues and 249 hours of audios from human-to-human spoken conversations. SpokenWOZ further incorporates common spoken characteristics such as word-by-word processing and reasoning in spoken language. Based on these characteristics, we present cross-turn slot and reasoning slot detection as new challenges. We conduct experiments on various baselines, including text-modal models, newly proposed dual-modal models, and LLMs, e.g., ChatGPT. The results show that the current models still have substantial room for improvement in spoken conversation, where the most advanced dialogue state tracker only achieves 25.65% in joint goal accuracy and the SOTA end-to-end model only correctly completes the user request in 52.1% of dialogues. Our dataset, code, and leaderboard are available at https://spokenwoz.github.io/SpokenWOZ-github.io/.
****

Fast Partitioned Learned Bloom Filter

Atsuki Sato, Yusuke Matsui

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Instructing Goal-Conditioned Reinforcement Learning Agents with Temporal Logic Objectives

Wenjie Qiu, Wensen Mao, He Zhu

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Neural Multi-Objective Combinatorial Optimization with Diversity Enhancement

Jinbiao Chen, Zizhen Zhang, Zhiguang Cao, Yaoxin Wu, Yining Ma, Te Ye, Jiahai Wang

Most of existing neural methods for multi-objective combinatorial optimization (MOCO) problems solely rely on decomposition, which often leads to repetitive solutions for the respective subproblems, thus a limited Pareto set. Beyond decomposition, we propose a novel neural heuristic with diversity enhancement (NHDE) to produce more Pareto solutions from two perspectives. On the one hand, to hinder duplicated solutions for different subproblems, we propose an indicator-enhanced deep reinforcement learning method to guide the model, and design a heterogeneous graph attention mechanism to capture the relations between the instance graph and the Pareto front graph. On the other hand, to excavate more solutions in the neighborhood of each subproblem, we present a multiple Pareto optima strategy to sample and preserve desirable solutions. Experimental results on classic MOCO problems show that our NHDE is able to generate a Pareto front with higher diversity, thereby achieving superior overall performance. Moreover, our NHDE is generic and can be applied to different neural methods for MOCO.
****

Multi-Agent First Order Constrained Optimization in Policy Space

Youpeng Zhao, Yaodong Yang, Zhenbo Lu, Wengang Zhou, Houqiang Li

In the realm of multi-agent reinforcement learning (MARL), achieving high performance is crucial for a successful multi-agent system.Meanwhile, the ability to avoid unsafe actions is becoming an urgent and imperative problem to solve for real-life applications. Whereas, it is still challenging to develop a safety-aware method for multi-agent systems in MARL. In this work, we introduce a novel approach called Multi-Agent First Order Constrained Optimization in Policy Space (MAFOCOPS), which effectively addresses the dual objectives of attaining satisfactory performance and enforcing safety constraints. Using data generated from the current policy, MAFOCOPS first finds the optimal update policy by solving a constrained optimization problem in the nonparameterized policy space. Then, the update policy is projected back into the parametric policy space to achieve a feasible policy. Notably, our method is first-order in nature, ensuring the ease of im

plementation, and exhibits an approximate upper bound on the worst-case constraint violation. Empirical results show that our approach achieves remarkable performance while satisfying safe constraints on several safe MARL benchmarks.
****

This Looks Like Those: Illuminating Prototypical Concepts Using Multiple Visualizations

Chiyu Ma, Brandon Zhao, Chaofan Chen, Cynthia Rudin

We present ProtoConcepts, a method for interpretable image classification combining deep learning and case-based reasoning using prototypical parts. Existing work in prototype-based image classification uses a "this looks like that'' reasoning process, which dissects a test image by finding prototypical parts and combining evidence from these prototypes to make a final classification. However, all of the existing prototypical part-based image classifiers provide only one-to-one comparisons, where a single training image patch serves as a prototype to compare with a part of our test image. With these single-image comparisons, it can often be difficult to identify the underlying concept being compared (e.g., "is it comparing the color or the shape?''). Our proposed method modifies the architecture of prototype-based networks to instead learn prototypical concepts which are visualized using multiple image patches. Having multiple visualizations of the same prototype allows us to more easily identify the concept captured by that prototype (e.g., "the test image and the related training patches are all the same shade of blue''), and allows our model to create richer, more interpretable visual explanations. Our experiments show that our ``this looks like those'' reasoning process can be applied as a modification to a wide range of existing prototypical image classification networks while achieving comparable accuracy on benchmark datasets.
****

Speculative Decoding with Big Little Decoder

Sehoon Kim, Karttikeya Mangalam, Suhong Moon, Jitendra Malik, Michael W. Mahoney, Amir Gholami, Kurt Keutzer

The recent emergence of Large Language Models based on the Transformer architecture has enabled dramatic advancements in the field of Natural Language Processing. However, these models have long inference latency, which limits their deployment and makes them prohibitively expensive for various real-time applications. The inference latency is further exacerbated by autoregressive generative tasks, as models need to run iteratively to generate tokens sequentially without leveraging token-level parallelization. To address this, we propose Big Little Decoder (BiLD), a framework that can improve inference efficiency and latency for a wide range of text generation applications. The BiLD framework contains two models with different sizes that collaboratively generate text. The small model runs autoregressively to generate text with a low inference cost, and the large model is only invoked occasionally to refine the small model’s inaccurate predictions in a non-autoregressive manner. To coordinate the small and large models, BiLD introduces two simple yet effective policies: (1) the fallback policy that determines when to hand control over to the large model; and (2) the rollback policy that determines when the large model needs to correct the small model's inaccurate predictions. To evaluate our framework across different tasks and models, we apply BiLD to various text generation scenarios encompassing machine translation on IWSLT 2017 De-En and WMT 2014 De-En, and summarization on XSUM and CNN/DailyMail. On an NVIDIA T4 GPU, our framework achieves a speedup of up to 2.12x speedup with minimal generation quality degradation. Furthermore, our framework is fully plug-and-play and can be applied without any modifications in the training process or model architecture. Our code is open-sourced.
****

Intrinsic Dimension Estimation for Robust Detection of AI-Generated Texts

Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Sergey Nikolenko, Evgeny Burnaev, Serguei Barannikov, Irina Piontkovskaya

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Replicable Clustering
Hossein Esfandiari, Amin Karbasi, Vahab Mirrokni, Grigoris Velegkas, Felix Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Counterfactual Memorization in Neural Language Models
Chiyuan Zhang, Daphne Ippolito, Katherine Lee, Matthew Jagielski, Florian Tramer, Nicholas Carlini
Modern neural language models that are widely used in various NLP tasks risk memorizing sensitive information from their training data.Understanding this memorization is important in real world applications and also from a learning-theoretical perspective. An open question in previous studies of language model memorization is how to filter out ``common'' memorization. In fact, most memorization criteria strongly correlate with the number of occurrences in the training set, capturing memorized familiar phrases, public knowledge, templated texts, or other repeated data.We formulate a notion of counterfactual memorization which characterizes how a model's predictions change if a particular document is omitted during training.We identify and study counterfactually-memorized training examples in standard text datasets.We estimate the influence of each memorized training example on the validation set and on generated texts, showing how this can provide direct evidence of the source of memorization at test time.
****
Learning Generalizable Agents via Saliency-guided Features Decorrelation
Sili Huang, Yanchao Sun, Jifeng Hu, Siyuan Guo, Hechang Chen, Yi Chang, Lichao Sun, Bo Yang
In visual-based Reinforcement Learning (RL), agents often struggle to generalize well to environmental variations in the state space that were not observed during training. The variations can arise in both task-irrelevant features, such as background noise, and task-relevant features, such as robot configurations, that are related to the optimal decisions. To achieve generalization in both situations, agents are required to accurately understand the impact of changed features on the decisions, i.e., establishing the true associations between changed features and decisions in the policy model. However, due to the inherent correlations among features in the state space, the associations between features and decisions become entangled, making it difficult for the policy to distinguish them. To this end, we propose Saliency-Guided Features Decorrelation (SGFD) to eliminate these correlations through sample reweighting. Concretely, SGFD consists of two core techniques: Random Fourier Functions (RFF) and the saliency map. RFF is utilized to estimate the complex non-linear correlations in high-dimensional images, while the saliency map is designed to identify the changed features. Under the guidance of the saliency map, SGFD employs sample reweighting to minimize the estimated correlations related to changed features, thereby achieving decorrelation in visual RL tasks. Our experimental results demonstrate that SGFD can generalize well on a wide range of test environments and significantly outperforms state-of-the-art methods in handling both task-irrelevant variations and task-relevant variations.
****
You Only Condense Once: Two Rules for Pruning Condensed Datasets
Yang He, Lingao Xiao, Joey Tianyi Zhou
Dataset condensation is a crucial tool for enhancing training efficiency by reducing the size of the training dataset, particularly in on-device scenarios. However, these scenarios have two significant challenges: 1) the varying computational resources available on the devices require a dataset size different from the pre-defined condensed dataset, and 2) the limited computational resources often preclude the possibility of conducting additional condensation processes. We introduce You Only Condense Once (YOCO) to overcome these limitations. On top of on

e condensed dataset, YOCO produces smaller condensed datasets with two embarrassingly simple dataset pruning rules: Low LBPE Score and Balanced Construction. YOCO offers two key advantages: 1) it can flexibly resize the dataset to fit varying computational constraints, and 2) it eliminates the need for extra condensation processes, which can be computationally prohibitive. Experiments validate our findings on networks including ConvNet, ResNet and DenseNet, and datasets including CIFAR-10, CIFAR-100 and ImageNet. For example, our YOCO surpassed various dataset condensation and dataset pruning methods on CIFAR-10 with ten Images Per Class (IPC), achieving 6.98-8.89% and 6.31-23.92% accuracy gains, respectively. The code is available at: https://github.com/he-y/you-only-condense-once.
****
Provably Efficient Offline Reinforcement Learning in Regular Decision Processes
Roberto Cipollone, Anders Jonsson, Alessandro Ronca, Mohammad Sadegh Talebi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CP-SLAM: Collaborative Neural Point-based SLAM System
Jiarui Hu, Mao Mao, Hujun Bao, Guofeng Zhang, Zhaopeng Cui
This paper presents a collaborative implicit neural simultaneous localization and mapping (SLAM) system with RGB-D image sequences, which consists of complete front-end and back-end modules including odometry, loop detection, sub-map fusion, and global refinement. In order to enable all these modules in a unified framework, we propose a novel neural point based 3D scene representation in which each point maintains a learnable neural feature for scene encoding and is associated with a certain keyframe. Moreover, a distributed-to-centralized learning strategy is proposed for the collaborative implicit SLAM to improve consistency and cooperation. A novel global optimization framework is also proposed to improve the system accuracy like traditional bundle adjustment. Experiments on various datasets demonstrate the superiority of the proposed method in both camera tracking and mapping.
****
The Surprising Effectiveness of Diffusion Models for Optical Flow and Monocular Depth Estimation
Saurabh Saxena, Charles Herrmann, Junhwa Hur, Abhishek Kar, Mohammad Norouzi, Deqing Sun, David J. Fleet
Denoising diffusion probabilistic models have transformed image generation with their impressive fidelity and diversity.We show that they also excel in estimating optical flow and monocular depth, surprisingly without task-specific architectures and loss functions that are predominant for these tasks. Compared to the point estimates of conventional regression-based methods, diffusion models also enable Monte Carlo inference, e.g., capturing uncertainty and ambiguity in flow and depth.With self-supervised pre-training, the combined use of synthetic and real data for supervised training, and technical innovations (infilling and step-unrolled denoising diffusion training) to handle noisy-incomplete training data, one can train state-of-the-art diffusion models for depth and optical flow estimation, with additional zero-shot coarse-to-fine refinement for high resolution estimates. Extensive experiments focus on quantitative performance against benchmarks, ablations, and the model's ability to capture uncertainty and multimodality, and impute missing values. Our model obtains a state-of-the-art relative depth error of 0.074 on the indoor NYU benchmark and an Fl-all score of 3.26\% on the KITTI optical flow benchmark, about 25\% better than the best published method.
****
Efficient Testable Learning of Halfspaces with Adversarial Label Noise
Ilias Diakonikolas, Daniel Kane, Vasilis Kontonis, Sihan Liu, Nikos Zarifis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Achieving $\mathcal{O}(\epsilon^{-1.5})$ Complexity in Hessian/Jacobian-free Stochastic Bilevel Optimization
Yifan Yang, Peiyao Xiao, Kaiyi Ji
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Robust and Actively Secure Serverless Collaborative Learning
Nicholas Franzese, Adam Dziedzic, Christopher A. Choquette-Choo, Mark R Thomas, Muhammad Ahmad Kaleem, Stephan Rabanser, Congyu Fang, Somesh Jha, Nicolas Papernot, Xiao Wang
Collaborative machine learning (ML) is widely used to enable institutions to learn better models from distributed data. While collaborative approaches to learning intuitively protect user data, they remain vulnerable to either the server, the clients, or both, deviating from the protocol. Indeed, because the protocol is asymmetric, a malicious server can abuse its power to reconstruct client data points. Conversely, malicious clients can corrupt learning with malicious updates. Thus, both clients and servers require a guarantee when the other cannot be trusted to fully cooperate. In this work, we propose a peer-to-peer (P2P) learning scheme that is secure against malicious servers and robust to malicious clients. Our core contribution is a generic framework that transforms any (compatible) algorithm for robust aggregation of model updates to the setting where servers and clients can act maliciously. Finally, we demonstrate the computational efficiency of our approach even with 1-million parameter models trained by 100s of peers on standard datasets.
****
Birder: Communication-Efficient 1-bit Adaptive Optimizer for Practical Distributed DNN Training
Hanyang Peng, Shuang Qin, Yue Yu, Jin Wang, Hui Wang, Ge Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
MIMONets: Multiple-Input-Multiple-Output Neural Networks Exploiting Computation in Superposition
Nicolas Menet, Michael Hersche, Geethan Karunaratne, Luca Benini, Abu Sebastian, Abbas Rahimi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
C-Disentanglement: Discovering Causally-Independent Generative Factors under an Inductive Bias of Confounder
Xiaoyu Liu, Jiaxin Yuan, Bang An, Yuancheng Xu, Yifan Yang, Furong Huang
Representation learning assumes that real-world data is generated by a few semantically meaningful generative factors (i.e., sources of variation) and aims to discover them in the latent space. These factors are expected to be causally disentangled, meaning that distinct factors are encoded into separate latent variables, and changes in one factor will not affect the values of the others. Compared to statistical independence, causal disentanglement allows more controllable data generation, improved robustness, and better generalization. However, most existing work assumes unconfoundedness in the discovery process, that there are no common causes to the generative factors and thus obtain only statistical independence. In this paper, we recognize the importance of modeling confounders in discovering causal generative factors. Unfortunately, such factors are not identifi

able without proper inductive bias. We fill the gap by introducing a framework entitled Confounded-Disentanglement (C-Disentanglement), the first framework that explicitly introduces the inductive bias of confounder via labels from domain expertise. In addition, we accordingly propose an approach to sufficiently identify the causally-disentangled factors under any inductive bias of the confounder. We conduct extensive experiments on both synthetic and real-world datasets. Our method demonstrates competitive results compared to various SOTA baselines in obtaining causally disentangled features and downstream tasks under domain shifts.
****
Representation Learning via Consistent Assignment of Views over Random Partitions
Thalles Santos Silva, Adín Ramírez Rivera
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Federated Multi-Objective Learning
Haibo Yang, Zhuqing Liu, Jia Liu, Chaosheng Dong, Michinari Momma
In recent years, multi-objective optimization (MOO) emerges as a foundational problem underpinning many multi-agent multi-task learning applications. However, existing algorithms in MOO literature remain limited to centralized learning settings, which do not satisfy the distributed nature and data privacy needs of such multi-agent multi-task learning applications. This motivates us to propose a new federated multi-objective learning (FMOL) framework with multiple clients distributively and collaboratively solving an MOO problem while keeping their training data private. Notably, our FMOL framework allows a different set of objective functions across different clients to support a wide range of applications, which advances and generalizes the MOO formulation to the federated learning paradigm for the first time. For this FMOL framework, we propose two new federated multi-objective optimization (FMOO) algorithms called federated multi-gradient descent averaging (FMGDA) and federated stochastic multi-gradient descent averaging (FSMGDA). Both algorithms allow local updates to significantly reduce communication costs, while achieving the {\em same} convergence rates as those of their algorithmic counterparts in the single-objective federated learning. Our extensive experiments also corroborate the efficacy of our proposed FMOO algorithms.
****
MARBLE: Music Audio Representation Benchmark for Universal Evaluation
Ruibin Yuan, Yinghao Ma, Yizhi Li, Ge Zhang, Xingran Chen, Hanzhi Yin, zhuo le, Yiqi Liu, Jiawen Huang, Zeyue Tian, Binyue Deng, Ningzhi Wang, Chenghua Lin, Emmanouil Benetos, Anton Ragni, Norbert Gyenge, Roger Dannenberg, Wenhu Chen, Gus Xia, Wei Xue, Si Liu, Shi Wang, Ruibo Liu, Yike Guo, Jie Fu
In the era of extensive intersection between art and Artificial Intelligence (AI), such as image generation and fiction co-creation, AI for music remains relatively nascent, particularly in music understanding. This is evident in the limited work on deep music representations, the scarcity of large-scale datasets, and the absence of a universal and community-driven benchmark. To address this issue, we introduce the Music Audio Representation Benchmark for universaL Evaluation, termed MARBLE. It aims to provide a benchmark for various Music Information Retrieval (MIR) tasks by defining a comprehensive taxonomy with four hierarchy levels, including acoustic, performance, score, and high-level description. We then establish a unified protocol based on 18 tasks on 12 public-available datasets, providing a fair and standard assessment of representations of all open-sourced pre-trained models developed on music recordings as baselines. Besides, MARBLE offers an easy-to-use, extendable, and reproducible suite for the community, with a clear statement on copyright issues on datasets. Results suggest recently proposed large-scale pre-trained musical language models perform the best in most tasks, with room for further improvement. The leaderboard and toolkit repository are published to promote future music AI research.

****
Language Models can Solve Computer Tasks
Geunwoo Kim, Pierre Baldi, Stephen McAleer
Agents capable of carrying out general tasks on a computer can improve efficiency and productivity by automating repetitive tasks and assisting in complex problem-solving. Ideally, such agents should be able to solve new computer tasks presented to them through natural language commands. However, previous approaches to this problem require large amounts of expert demonstrations and task-specific reward functions, both of which are impractical for new tasks. In this work, we show that a pre-trained large language model (LLM) agent can execute computer tasks guided by natural language using a simple prompting scheme where the agent \textbf{R}ecursively \textbf{C}riticizes and \textbf{I}mproves its output (RCI). The RCI approach significantly outperforms existing LLM methods for automating computer tasks and surpasses supervised learning (SL) and reinforcement learning (RL) approaches on the MiniWoB++ benchmark. We compare multiple LLMs and find that RCI with the InstructGPT-3+RLHF LLM is state-of-the-art on MiniWoB++, using only a handful of demonstrations per task rather than tens of thousands, and without a task-specific reward function. Furthermore, we demonstrate RCI prompting's effectiveness in enhancing LLMs' reasoning abilities on a suite of natural language reasoning tasks, outperforming chain of thought (CoT) prompting with external feedback. We find that RCI combined with CoT performs better than either separately. Our code can be found here: https://github.com/posgnu/rci-agent.
****
An NLP Benchmark Dataset for Assessing Corporate Climate Policy Engagement
Gaku Morio, Christopher D Manning
As societal awareness of climate change grows, corporate climate policy engagements are attracting attention.We propose a dataset to estimate corporate climate policy engagement from various PDF-formatted documents.Our dataset comes from LobbyMap (a platform operated by global think tank InfluenceMap) that provides engagement categories and stances on the documents.To convert the LobbyMap data into the structured dataset, we developed a pipeline using text extraction and OCR. Our contributions are: (i) Building an NLP dataset including 10K documents on corporate climate policy engagement. (ii) Analyzing the properties and challenges of the dataset. (iii) Providing experiments for the dataset using pre-trained language models.The results show that while Longformer outperforms baselines and other pre-trained models, there is still room for significant improvement.We hope our work begins to bridge research on NLP and climate change.
****
Robustness Guarantees for Adversarially Trained Neural Networks
Poorya Mianjy, Raman Arora
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Pre-training Contextualized World Models with In-the-wild Videos for Reinforcement Learning
Jialong Wu, Haoyu Ma, Chaoyi Deng, Mingsheng Long
Unsupervised pre-training methods utilizing large and diverse datasets have achieved tremendous success across a range of domains. Recent work has investigated such unsupervised pre-training methods for model-based reinforcement learning (MBRL) but is limited to domain-specific or simulated data. In this paper, we study the problem of pre-training world models with abundant in-the-wild videos for efficient learning of downstream visual control tasks. However, in-the-wild videos are complicated with various contextual factors, such as intricate backgrounds and textured appearance, which precludes a world model from extracting shared world knowledge to generalize better. To tackle this issue, we introduce Contextualized World Models (ContextWM) that explicitly separate context and dynamics modeling to overcome the complexity and diversity of in-the-wild videos and facilitate knowledge transfer between distinct scenes. Specifically, a contextualized

extension of the latent dynamics model is elaborately realized by incorporating a context encoder to retain contextual information and empower the image decoder, which encourages the latent dynamics model to concentrate on essential temporal variations. Our experiments show that in-the-wild video pre-training equipped with ContextWM can significantly improve the sample efficiency of MBRL in various domains, including robotic manipulation, locomotion, and autonomous driving. Code is available at this repository: https://github.com/thuml/ContextWM.
****
Strategyproof Voting under Correlated Beliefs
Daniel Halpern, Rachel Li, Ariel D. Procaccia
In voting theory, when voters have ranked preferences over candidates, the celebrated Gibbard-Satterthwaite Theorem essentially rules out the existence of reasonable strategyproof methods for picking a winner. What if we weaken strategyproofness to only hold for Bayesian voters with beliefs over others' preferences? When voters believe other participants' rankings are drawn independently from a fixed distribution, the impossibility persists. However, it is quite reasonable for a voter to believe that other votes are correlated, either to each other or to their own ranking. We consider such beliefs induced by classic probabilistic models in social choice such as the Mallows, Placket-Luce, and Thurstone-Mosteller models. We single out the plurality rule (choosing the candidate ranked first most often) as a particularly promising choice as it is strategyproof for a large class of beliefs containing the specific ones we introduce. Further, we show that plurality is unique among positional scoring rules in having this property: no other scoring rule is strategyproof for beliefs induced by the Mallows model when there are a sufficient number of voters. Finally, we give examples of prominent non-scoring voting rules failing to be strategyproof on beliefs in this class, further bolstering the case for plurality.
****
PCF-GAN: generating sequential data via the characteristic function of measures on the path space
Hang Lou, Siran Li, Hao Ni
Generating high-fidelity time series data using generative adversarial networks (GANs) remains a challenging task, as it is difficult to capture the temporal dependence of joint probability distributions induced by time-series data. Towards this goal, a key step is the development of an effective discriminator to distinguish between time series distributions. We propose the so-called PCF-GAN, a novel GAN that incorporates the path characteristic function (PCF) as the principled representation of time series distribution into the discriminator to enhance its generative performance. On the one hand, we establish theoretical foundations of the PCF distance by proving its characteristicity, boundedness, differentiability with respect to generator parameters, and weak continuity, which ensure the stability and feasibility of training the PCF-GAN. On the other hand, we design efficient initialisation and optimisation schemes for PCFs to strengthen the discriminative power and accelerate training efficiency. To further boost the capabilities of complex time series generation, we integrate the auto-encoder structure via sequential embedding into the PCF-GAN, which provides additional reconstruction functionality. Extensive numerical experiments on various datasets demonstrate the consistently superior performance of PCF-GAN over state-of-the-art baselines, in both generation and reconstruction quality.
****
A Rigorous Link between Deep Ensembles and (Variational) Bayesian Methods
Veit David Wild, Sahra Ghalebikesabi, Dino Sejdinovic, Jeremias Knoblauch
We establish the first mathematically rigorous link between Bayesian, variational Bayesian, and ensemble methods. A key step towards this it to reformulate the non-convex optimisation problem typically encountered in deep learning as a convex optimisation in the space of probability measures. On a technical level, our contribution amounts to studying generalised variational inference through the lense of Wasserstein gradient flows. The result is a unified theory of various seemingly disconnected approaches that are commonly used for uncertainty quantification in deep learning---including deep ensembles and (variational) Bayesian met

hods. This offers a fresh perspective on the reasons behind the success of deep ensembles over procedures based on parameterised variational inference, and allows the derivation of new ensembling schemes with convergence guarantees. We showcase this by proposing a family of interacting deep ensembles with direct parallels to the interactions of particle systems in thermodynamics, and use our theory to prove the convergence of these algorithms to a well-defined global minimiser on the space of probability measures.
****
Markovian Sliced Wasserstein Distances: Beyond Independent Projections
Khai Nguyen, Tongzheng Ren, Nhat Ho
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Generative Modelling of Stochastic Actions with Arbitrary Constraints in Reinforcement Learning
Changyu CHEN, Ramesha Karunasena, Thanh Nguyen, Arunesh Sinha, Pradeep Varakantham
Many problems in Reinforcement Learning (RL) seek an optimal policy with large discrete multidimensional yet unordered action spaces; these include problems in randomized allocation of resources such as placements of multiple security resources and emergency response units, etc. A challenge in this setting is that the underlying action space is categorical (discrete and unordered) and large, for which existing RL methods do not perform well. Moreover, these problems require validity of the realized action (allocation); this validity constraint is often difficult to express compactly in a closed mathematical form. The allocation nature of the problem also prefers stochastic optimal policies, if one exists. In this work, we address these challenges by (1) applying a (state) conditional normalizing flow to compactly represent the stochastic policy — the compactness arises due to the network only producing one sampled action and the corresponding log probability of the action, which is then used by an actor-critic method; and (2) employing an invalid action rejection method (via a valid action oracle) to update the base policy. The action rejection is enabled by a modified policy gradient that we derive. Finally, we conduct extensive experiments to show the scalability of our approach compared to prior methods and the ability to enforce arbitrary state-conditional constraints on the support of the distribution of actions in any state.
****
On the impact of activation and normalization in obtaining isometric embeddings at initialization
Amir Joudaki, Hadi Daneshmand, Francis Bach
In this paper, we explore the structure of the penultimate Gram matrix in deep neural networks, which contains the pairwise inner products of outputs corresponding to a batch of inputs. In several architectures it has been observed that this Gram matrix becomes degenerate with depth at initialization, which dramatically slows training. Normalization layers, such as batch or layer normalization, play a pivotal role in preventing the rank collapse issue. Despite promising advances, the existing theoretical results do not extend to layer normalization, which is widely used in transformers, and can not quantitatively characterize the role of non-linear activations. To bridge this gap, we prove that layer normalization, in conjunction with activation layers, biases the Gram matrix of a multilayer perceptron towards the identity matrix at an exponential rate with depth at initialization. We quantify this rate using the Hermite expansion of the activation function.
****
Uni3DETR: Unified 3D Detection Transformer
Zhenyu Wang, Ya-Li Li, Xi Chen, Hengshuang Zhao, Shengjin Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SceneScape: Text-Driven Consistent Scene Generation
Rafail Fridman, Amit Abecasis, Yoni Kasten, Tali Dekel
We present a method for text-driven perpetual view generation -- synthesizing long-term videos of various scenes solely, given an input text prompt describing the scene and camera poses. We introduce a novel framework that generates such videos in an online fashion by combining the generative power of a pre-trained text-to-image model with the geometric priors learned by a pre-trained monocular depth prediction model. To tackle the pivotal challenge of achieving 3D consistency, i.e., synthesizing videos that depict geometrically-plausible scenes, we deploy an online test-time training to encourage the predicted depth map of the current frame to be geometrically consistent with the synthesized scene. The depth maps are used to construct a \emph{unified} mesh representation of the scene, which is progressively constructed along the video generation process. In contrast to previous works, which are applicable only to limited domains, our method generates diverse scenes, such as walkthroughs in spaceships, caves, or ice castles.
****
RDumb: A simple approach that questions our progress in continual test-time adaptation
Ori Press, Steffen Schneider, Matthias Kümmerer, Matthias Bethge
Test-Time Adaptation (TTA) allows to update pre-trained models to changing data distributions at deployment time. While early work tested these algorithms for individual fixed distribution shifts, recent work proposed and applied methods for continual adaptation over long timescales. To examine the reported progress in the field, we propose the Continually Changing Corruptions (CCC) benchmark to measure asymptotic performance of TTA techniques. We find that eventually all but one state-of-the-art methods collapse and perform worse than a non-adapting model, including models specifically proposed to be robust to performance collapse. In addition, we introduce a simple baseline, "RDumb", that periodically resets the model to its pretrained state. RDumb performs better or on par with the previously proposed state-of-the-art in all considered benchmarks.Our results show that previous TTA approaches are neither effective at regularizing adaptation to avoid collapse nor able to outperform a simplistic resetting strategy.
****
Swap Agnostic Learning, or Characterizing Omniprediction via Multicalibration
Parikshit Gopalan, Michael Kim, Omer Reingold
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
AR-Diffusion: Auto-Regressive Diffusion Model for Text Generation
Tong Wu, Zhihao Fan, Xiao Liu, Hai-Tao Zheng, Yeyun Gong, yelong shen, Jian Jiao, Juntao Li, zhongyu wei, Jian Guo, Nan Duan, Weizhu Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Adaptive Uncertainty Estimation via High-Dimensional Testing on Latent Representations
Tsai Hor Chan, Kin Wai Lau, Jiajun Shen, Guosheng Yin, Lequan Yu
Uncertainty estimation aims to evaluate the confidence of a trained deep neural network. However, existing uncertainty estimation approaches rely on low-dimensional distributional assumptions and thus suffer from the high dimensionality of latent features. Existing approaches tend to focus on uncertainty on discrete classification probabilities, which leads to poor generalizability to uncertainty estimation for other tasks. Moreover, most of the literature requires seeing the

 out-of-distribution (OOD) data in the training for better estimation of uncerta
inty, which limits the uncertainty estimation performance in practice because th
e OOD data are typically unseen. To overcome these limitations, we propose a ne
w framework using data-adaptive high-dimensional hypothesis testing for uncertai
nty estimation, which leverages the statistical properties of the feature repres
entations. Our method directly operates on latent representations and thus does
not require retraining the feature encoder under a modified objective. The test
statistic relaxes the feature distribution assumptions to high dimensionality, a
nd it is more discriminative to uncertainties in the latent representations. We
demonstrate that encoding features with Bayesian neural networks can enhance tes
ting performance and lead to more accurate uncertainty estimation. We further in
troduce a family-wise testing procedure to determine the optimal threshold of OO
D detection, which minimizes the false discovery rate (FDR). Extensive experimen
ts validate the satisfactory performance of our framework on uncertainty estimat
ion and task-specific prediction over a variety of competitors. The experiments
on the OOD detection task also show satisfactory performance of our method when
the OOD data are unseen in the training. Codes are available at https://github.c
om/HKU-MedAI/bnn_uncertainty.
****
Algorithmic Regularization in Tensor Optimization: Towards a Lifted Approach in
Matrix Sensing
Ziye Ma, Javad Lavaei, Somayeh Sojoudi
Gradient descent (GD) is crucial for generalization in machine learning models,
as it induces implicit regularization, promoting compact representations. In thi
s work, we examine the role of GD in inducing implicit regularization for tensor
 optimization, particularly within the context of the lifted matrix sensing fram
ework. This framework has been recently proposed to address the non-convex matri
x sensing problem by transforming spurious solutions into strict saddles when op
timizing over symmetric, rank-1 tensors. We show that, with sufficiently small i
nitialization scale, GD applied to this lifted problem results in approximate ra
nk-1 tensors and critical points with escape directions. Our findings underscore
 the significance of the tensor parametrization of matrix sensing, in combinatio
n with first-order methods, in achieving global optimality in such problems.
****
A General Theory of Correct, Incorrect, and Extrinsic Equivariance
Dian Wang, Xupeng Zhu, Jung Yeon Park, Mingxi Jia, Guanang Su, Robert Platt, Rob
in Walters
Although equivariant machine learning has proven effective at many tasks, succes
s depends heavily on the assumption that the ground truth function is symmetric
over the entire domain matching the symmetry in an equivariant neural network. A
 missing piece in the equivariant learning literature is the analysis of equivar
iant networks when symmetry exists only partially in the domain. In this work, w
e present a general theory for such a situation. We propose pointwise definition
s of correct, incorrect, and extrinsic equivariance, which allow us to quantify
continuously the degree of each type of equivariance a function displays. We the
n study the impact of various degrees of incorrect or extrinsic symmetry on mode
l error. We prove error lower bounds for invariant or equivariant networks in cl
assification or regression settings with partially incorrect symmetry. We also a
nalyze the potentially harmful effects of extrinsic equivariance. Experiments va
lidate these results in three different environments.
****
Analyzing Vision Transformers for Image Classification in Class Embedding Space
Martina G. Vilas, Timothy Schaumlöffel, Gemma Roig
Despite the growing use of transformer models in computer vision, a mechanistic
understanding of these networks is still needed. This work introduces a method t
o reverse-engineer Vision Transformers trained to solve image classification tas
ks. Inspired by previous research in NLP, we demonstrate how the inner represent
ations at any level of the hierarchy can be projected onto the learned class emb
edding space to uncover how these networks build categorical representations for
 their predictions. We use our framework to show how image tokens develop class-

specific representations that depend on attention mechanisms and contextual information, and give insights on how self-attention and MLP layers differentially contribute to this categorical composition. We additionally demonstrate that this method (1) can be used to determine the parts of an image that would be important for detecting the class of interest, and (2) exhibits significant advantages over traditional linear probing approaches. Taken together, our results position our proposed framework as a powerful tool for mechanistic interpretability and explainability research.
****
Toward Re-Identifying Any Animal
Bingliang Jiao, Lingqiao Liu, Liying Gao, Ruiqi Wu, Guosheng Lin, PENG WANG, Yanning Zhang
The current state of re-identification (ReID) models poses limitations to their applicability in the open world, as they are primarily designed and trained for specific categories like person or vehicle. In light of the importance of ReID technology for tracking wildlife populations and migration patterns, we propose a new task called ``Re-identify Any Animal in the Wild'' (ReID-AW). This task aims to develop a ReID model capable of handling any unseen wildlife category it encounters. To address this challenge, we have created a comprehensive dataset called Wildlife-71, which includes ReID data from 71 different wildlife categories. This dataset is the first of its kind to encompass multiple object categories in the realm of ReID. Furthermore, we have developed a universal re-identification model named UniReID specifically for the ReID-AW task. To enhance the model's adaptability to the target category, we employ a dynamic prompting mechanism using category-specific visual prompts. These prompts are generated based on knowledge gained from a set of pre-selected images within the target category. Additionally, we leverage explicit semantic knowledge derived from the large-scale pre-trained language model, GPT-4. This allows UniReID to focus on regions that are particularly useful for distinguishing individuals within the target category. Extensive experiments have demonstrated the remarkable generalization capability of our UniReID model. It showcases promising performance in handling arbitrary wildlife categories, offering significant advancements in the field of ReID for wildlife conservation and research purposes.
****
Critical Initialization of Wide and Deep Neural Networks using Partial Jacobians: General Theory and Applications
Darshil Doshi, Tianyu He, Andrey Gromov
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Trading-off price for data quality to achieve fair online allocation
Mathieu Molina, Nicolas Gast, Patrick Loiseau, Vianney Perchet
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Asymptotics of Bayesian Uncertainty Estimation in Random Features Regression
Youngsoo Baek, Samuel Berchuck, Sayan Mukherjee
In this paper we compare and contrast the behavior of the posterior predictive distribution to the risk of the the maximum a posteriori estimator for the random features regression model in the overparameterized regime. We will focus on the variance of the posterior predictive distribution (Bayesian model average) and compare its asymptotics to that of the risk of the MAP estimator. In the regime where the model dimensions grow faster than any constant multiple of the number of samples, asymptotic agreement between these two quantities is governed by the phase transition in the signal-to-noise ratio. They also asymptotically agree with each other when the number of samples grow faster than any constant multip

le of model dimensions. Numerical simulations illustrate finer distributional properties of the two quantities for finite dimensions. We conjecture they have Gaussian fluctuations and exhibit similar properties as found by previous authors in a Gaussian sequence model, this is of independent theoretical interest.
****

Decentralized Matrix Sensing: Statistical Guarantees and Fast Convergence
Marie Maros, Gesualdo Scutari
We explore the matrix sensing problem from near-isotropic linear measurements, distributed across a network of agents modeled as an undirected graph, with no centralized node. We provide the first study of statistical, computational/communication guarantees for a decentralized gradient algorithm that solves the (nonconvex) Burer-Monteiro type decomposition associated to the low-rank matrix estimation. With small random initialization, the algorithm displays an approximate two-phase convergence: (i) a spectral phase that aligns the iterates' column space with the underlying low-rank matrix, mimicking centralized spectral initialization (not directly implementable over networks); and (ii) a local refinement phase that diverts the iterates from certain degenerate saddle points, while ensuring swift convergence to the underlying low-rank matrix. Central to our analysis is a novel "in-network" Restricted Isometry Property which accommodates for the decentralized nature of the optimization, revealing an intriguing interplay between sample complexity and network connectivity, topology, and communication complexity.
****

Dynamics Generalisation in Reinforcement Learning via Adaptive Context-Aware Policies
Michael Beukman, Devon Jarvis, Richard Klein, Steven James, Benjamin Rosman
While reinforcement learning has achieved remarkable successes in several domains, its real-world application is limited due to many methods failing to generalise to unfamiliar conditions. In this work, we consider the problem of generalising to new transition dynamics, corresponding to cases in which the environment's response to the agent's actions differs. For example, the gravitational force exerted on a robot depends on its mass and changes the robot's mobility. Consequently, in such cases, it is necessary to condition an agent's actions on extrinsic state information and pertinent contextual information reflecting how the environment responds. While the need for context-sensitive policies has been established, the manner in which context is incorporated architecturally has received less attention. Thus, in this work, we present an investigation into how context information should be incorporated into behaviour learning to improve generalisation. To this end, we introduce a neural network architecture, the Decision Adapter, which generates the weights of an adapter module and conditions the behaviour of an agent on the context information. We show that the Decision Adapter is a useful generalisation of a previously proposed architecture and empirically demonstrate that it results in superior generalisation performance compared to previous approaches in several environments. Beyond this, the Decision Adapter is more robust to irrelevant distractor variables than several alternative methods.
****

Goal Driven Discovery of Distributional Differences via Language Descriptions
Ruiqi Zhong, Peter Zhang, Steve Li, Jinwoo Ahn, Dan Klein, Jacob Steinhardt
Exploring large corpora can generate useful discoveries but is time-consuming for humans.   We formulate a new task, D5, that automatically discovers differences between two large corpora in a goal-driven way.   The task input is a problem comprising a user-specified research goal (“comparing the side effects of drug A and drug”) and a corpus pair (two large collections of patients' self-reported reactions after taking each drug).   The output is a goal-related description (discovery) of how these corpora differ (patients taking drug A “mention feelings of paranoia” more often).   We build a D5 system, and to quantitatively evaluate its performance, we 1) build a diagnostic benchmark, SynD5, to test whether it can recover known differences between two synthetic corpora, and 2) contribute a meta-dataset, OpenD5, aggregating 675 open-ended problems ranging across business, social sciences, humanities, machine learning, and health.   With bot

h synthetic and real datasets, we confirm that language models can leverage the user-specified goals to propose more relevant candidate discoveries, and they sometimes produce discoveries previously unknown to the authors, including demographic differences in discussion topics, political stances in speech, insights in commercial reviews, and error patterns in NLP models.   Finally, we discuss the limitations of the current D5 system, which discovers correlation rather than causation and has the potential to reinforce societal biases when misused; therefore, practitioners should treat the outputs of our system with caution.
****

Convex and Non-convex Optimization Under Generalized Smoothness
Haochuan Li, Jian Qian, Yi Tian, Alexander Rakhlin, Ali Jadbabaie
Classical analysis of convex and non-convex optimization methods often requires the Lipschitz continuity of the gradient, which limits the analysis to functions bounded by quadratics. Recent work relaxed this requirement to a non-uniform smoothness condition with the Hessian norm bounded by an affine function of the gradient norm, and proved convergence in the non-convex setting via gradient clipping, assuming bounded noise. In this paper, we further generalize this non-uniform smoothness condition and develop a simple, yet powerful analysis technique that bounds the gradients along the trajectory, thereby leading to stronger results for both convex and non-convex optimization problems. In particular, we obtain the classical convergence rates for (stochastic) gradient descent and Nesterov's accelerated gradient method in the convex and/or non-convex setting under this general smoothness condition. The new analysis approach does not require gradient clipping and allows heavy-tailed noise with bounded variance in the stochastic setting.
****

DOSE: Diffusion Dropout with Adaptive Prior for Speech Enhancement
Wenxin Tai, Yue Lei, Fan Zhou, Goce Trajcevski, Ting Zhong
Speech enhancement (SE) aims to improve the intelligibility and quality of speech in the presence of non-stationary additive noise. Deterministic deep learning models have traditionally been used for SE, but recent studies have shown that generative approaches, such as denoising diffusion probabilistic models (DDPMs), can also be effective. However, incorporating condition information into DDPMs for SE remains a challenge. We propose a model-agnostic method called DOSE that employs two efficient condition-augmentation techniques to address this challenge, based on two key insights: (1) We force the model to prioritize the condition factor when generating samples by training it with dropout operation; (2) We inject the condition information into the sampling process by providing an informative adaptive prior. Experiments demonstrate that our approach yields substantial improvements in high-quality and stable speech generation, consistency with the condition factor, and inference efficiency. Codes are publicly available at https://github.com/ICDM-UESTC/DOSE.
****

ISP: Multi-Layered Garment Draping with Implicit Sewing Patterns
Ren Li, Benoît Guillard, Pascal Fua
Many approaches to draping individual garments on human body models are realistic, fast, and yield outputs that are differentiable with respect to the body shape on which they are draped. However, they are either unable to handle multi-layered clothing, which is prevalent in everyday dress, or restricted to bodies in T-pose. In this paper, we introduce a parametric garment representation model that addresses these limitations. As in models used by clothing designers, each garment consists of individual 2D panels. Their 2D shape is defined by a Signed Distance Function and 3D shape by a 2D to 3D mapping. The 2D parameterization enables easy detection of potential collisions and the 3D parameterization handles complex shapes effectively. We show that this combination is faster and yields higher quality reconstructions than purely implicit surface representations, and makes the recovery of layered garments from images possible thanks to its differentiability. Furthermore, it supports rapid editing of garment shapes and texture by modifying individual 2D panels.
****

Optimality of Message-Passing Architectures for Sparse Graphs
Aseem Baranwal, Kimon Fountoulakis, Aukosh Jagannath
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Distribution-Free Statistical Dispersion Control for Societal Applications
Zhun Deng, Thomas Zollo, Jake Snell, Toniann Pitassi, Richard Zemel
Explicit finite-sample statistical guarantees on model performance are an important ingredient in responsible machine learning. Previous work has focused mainly on bounding either the expected loss of a predictor or the probability that an individual prediction will incur a loss value in a specified range. However, for many high-stakes applications it is crucial to understand and control the \textit{dispersion} of a loss distribution, or the extent to which different members of a population experience unequal effects of algorithmic decisions. We initiate the study of distribution-free control of statistical dispersion measures with societal implications and propose a simple yet flexible framework that allows us to handle a much richer class of statistical functionals beyond previous work. Our methods are verified through experiments in toxic comment detection, medical imaging, and film recommendation.
****
Switching Temporary Teachers for Semi-Supervised Semantic Segmentation
Jaemin Na, Jung-Woo Ha, Hyung Jin Chang, Dongyoon Han, Wonjun Hwang
The teacher-student framework, prevalent in semi-supervised semantic segmentation, mainly employs the exponential moving average (EMA) to update a single teacher's weights based on the student's. However, EMA updates raise a problem in that the weights of the teacher and student are getting coupled, causing a potential performance bottleneck. Furthermore, this problem may become more severe when training with more complicated labels such as segmentation masks but with few annotated data. This paper introduces Dual Teacher, a simple yet effective approach that employs dual temporary teachers aiming to alleviate the coupling problem for the student. The temporary teachers work in shifts and are progressively improved, so consistently prevent the teacher and student from becoming excessively close. Specifically, the temporary teachers periodically take turns generating pseudo-labels to train a student model and maintain the distinct characteristics of the student model for each epoch. Consequently, Dual Teacher achieves competitive performance on the PASCAL VOC, Cityscapes, and ADE20K benchmarks with remarkably shorter training times than state-of-the-art methods. Moreover, we demonstrate that our approach is model-agnostic and compatible with both CNN- and Transformer-based models. Code is available at https://github.com/naver-ai/dual-teacher.
****
Extremal Domain Translation with Neural Optimal Transport
Milena Gazdieva, Alexander Korotin, Daniil Selikhanovych, Evgeny Burnaev
In many unpaired image domain translation problems, e.g., style transfer or super-resolution, it is important to keep the translated image similar to its respective input image. We propose the extremal transport (ET) which is a mathematical formalization of the theoretically best possible unpaired translation between a pair of domains w.r.t. the given similarity function. Inspired by the recent advances in neural optimal transport (OT), we propose a scalable algorithm to approximate ET maps as a limit of partial OT maps. We test our algorithm on toy examples and on the unpaired image-to-image translation task. The code is publicly available at https://github.com/milenagazdieva/ExtremalNeuralOptimalTransport
****
Recaptured Raw Screen Image and Video Demoiréing via Channel and Spatial Modulations
Yijia Cheng, Xin Liu, Jingyu Yang
Capturing screen contents by smartphone cameras has become a common way for information sharing. However, these images and videos are often degraded by moiré pa

tterns, which are caused by frequency aliasing between the camera filter array and digital display grids. We observe that the moiré patterns in raw domain is simpler than those in sRGB domain, and the moiré patterns in raw color channels have different properties. Therefore, we propose an image and video demoiréing network tailored for raw inputs. We introduce a color-separated feature branch, and it is fused with the traditional feature-mixed branch via channel and spatial modulations. Specifically, the channel modulation utilizes modulated color-separated features to enhance the color-mixed features. The spatial modulation utilizes the feature with large receptive field to modulate the feature with small receptive field. In addition, we build the first well-aligned raw video demoiréing (RawVDemoiré) dataset and propose an efficient temporal alignment method by inserting alternating patterns. Experiments demonstrate that our method achieves state-of-the-art performance for both image and video demoiréing. Our dataset and code will be released after the acceptance of this work.
****
On Imitation in Mean-field Games
Giorgia Ramponi, Pavel Kolev, Olivier Pietquin, Niao He, Mathieu Lauriere, Matthieu Geist
We explore the problem of imitation learning (IL) in the context of mean-field games (MFGs), where the goal is to imitate the behavior of a population of agents following a Nash equilibrium policy according to some unknown payoff function. IL in MFGs presents new challenges compared to single-agent IL, particularly when both the reward function and the transition kernel depend on the population distribution. In this paper, departing from the existing literature on IL for MFGs, we introduce a new solution concept called the Nash imitation gap. Then we show that when only the reward depends on the population distribution, IL in MFGs can be reduced to single-agent IL with similar guarantees. However, when the dynamics is population-dependent, we provide a novel upper-bound that suggests IL is harder in this setting. To address this issue, we propose a new adversarial formulation where the reinforcement learning problem is replaced by a mean-field control (MFC) problem, suggesting progress in IL within MFGs may have to build upon MFC.
****
CluB: Cluster Meets BEV for LiDAR-Based 3D Object Detection
Yingjie Wang, Jiajun Deng, Yuenan Hou, Yao Li, Yu Zhang, Jianmin Ji, Wanli Ouyang, Yanyong Zhang
Currently, LiDAR-based 3D detectors are broadly categorized into two groups, namely, BEV-based detectors and cluster-based detectors.BEV-based detectors capture the contextual information from the Bird's Eye View (BEV) and fill their center voxels via feature diffusion with a stack of convolution layers, which, however, weakens the capability of presenting an object with the center point.On the other hand, cluster-based detectors exploit the voting mechanism and aggregate the foreground points into object-centric clusters for further prediction.In this paper, we explore how to effectively combine these two complementary representations into a unified framework.Specifically, we propose a new 3D object detection framework, referred to as CluB, which incorporates an auxiliary cluster-based branch into the BEV-based detector by enriching the object representation at both feature and query levels.Technically, CluB is comprised of two steps.First, we construct a cluster feature diffusion module to establish the association between cluster features and BEV features in a subtle and adaptive fashion. Based on that, an imitation loss is introduced to distill object-centric knowledge from the cluster features to the BEV features.Second, we design a cluster query generation module to leverage the voting centers directly from the cluster branch, thus enriching the diversity of object queries.Meanwhile, a direction loss is employed to encourage a more accurate voting center for each cluster.Extensive experiments are conducted on Waymo and nuScenes datasets, and our CluB achieves state-of-the-art performance on both benchmarks.
****
Probabilistic Exponential Integrators
Nathanael Bosch, Philipp Hennig, Filip Tronarp

Probabilistic solvers provide a flexible and efficient framework for simulation, uncertainty quantification, and inference in dynamical systems. However, like standard solvers, they suffer performance penalties for certain stiff systems, where small steps are required not for reasons of numerical accuracy but for the sake of stability. This issue is greatly alleviated in semi-linear problems by the probabilistic exponential integrators developed in this paper. By including the fast, linear dynamics in the prior, we arrive at a class of probabilistic integrators with favorable properties. Namely, they are proven to be L-stable, and in a certain case reduce to a classic exponential integrator---with the added benefit of providing a probabilistic account of the numerical error. The method is also generalized to arbitrary non-linear systems by imposing piece-wise semi-linearity on the prior via Jacobians of the vector field at the previous estimates, resulting in probabilistic exponential Rosenbrock methods. We evaluate the proposed methods on multiple stiff differential equations and demonstrate their improved stability and efficiency over established probabilistic solvers. The present contribution thus expands the range of problems that can be effectively tackled within probabilistic numerics.
****
Understanding Neural Network Binarization with Forward and Backward Proximal Quantizers
Yiwei Lu, Yaoliang Yu, Xinlin Li, Vahid Partovi Nia
In neural network binarization, BinaryConnect (BC) and its variants are considered the standard. These methods apply the sign function in their forward pass and their respective gradients are backpropagated to update the weights. However, the derivative of the sign function is zero whenever defined, which consequently freezes training. Therefore, implementations of BC (e.g., BNN) usually replace the derivative of sign in the backward computation with identity or other approximate gradient alternatives. Although such practice works well empirically, it is largely a heuristic or ``training trick.'' We aim at shedding some light on these training tricks from the optimization perspective. Building from existing theory on ProxConnect (PC, a generalization of BC), we (1) equip PC with different forward-backward quantizers and obtain ProxConnect++ (PC++) that includes existing binarization techniques as special cases; (2) derive a principled way to synthesize forward-backward quantizers with automatic theoretical guarantees; (3) illustrate our theory by proposing an enhanced binarization algorithm BNN++; (4) conduct image classification experiments on CNNs and vision transformers, and empirically verify that BNN++ generally achieves competitive results on binarizing these models.
****
QH9: A Quantum Hamiltonian Prediction Benchmark for QM9 Molecules
Haiyang Yu, Meng Liu, Youzhi Luo, Alex Strasser, Xiaofeng Qian, Xiaoning Qian, Shuiwang Ji
Supervised machine learning approaches have been increasingly used in accelerating electronic structure prediction as surrogates of first-principle computational methods, such as density functional theory (DFT). While numerous quantum chemistry datasets focus on chemical properties and atomic forces, the ability to achieve accurate and efficient prediction of the Hamiltonian matrix is highly desired, as it is the most important and fundamental physical quantity that determines the quantum states of physical systems and chemical properties. In this work, we generate a new Quantum Hamiltonian dataset, named as QH9, to provide precise Hamiltonian matrices for 2,399 molecular dynamics trajectories and 130,831 stable molecular geometries, based on the QM9 dataset. By designing benchmark tasks with various molecules, we show that current machine learning models have the capacity to predict Hamiltonian matrices for arbitrary molecules. Both the QH9 dataset and the baseline models are provided to the community through an open-source benchmark, which can be highly valuable for developing machine learning methods and accelerating molecular and materials design for scientific and technological applications. Our benchmark is publicly available at \url{https://github.com/divelab/AIRS/tree/main/OpenDFT/QHBench}.
****

CorresNeRF: Image Correspondence Priors for Neural Radiance Fields
Yixing Lao, Xiaogang Xu, zhipeng cai, Xihui Liu, Hengshuang Zhao
Neural Radiance Fields (NeRFs) have achieved impressive results in novel view synthesis and surface reconstruction tasks. However, their performance suffers under challenging scenarios with sparse input views. We present CorresNeRF, a novel method that leverages image correspondence priors computed by off-the-shelf methods to supervise NeRF training. We design adaptive processes for augmentation and filtering to generate dense and high-quality correspondences. The correspondences are then used to regularize NeRF training via the correspondence pixel reprojection and depth loss terms. We evaluate our methods on novel view synthesis and surface reconstruction tasks with density-based and SDF-based NeRF models on different datasets. Our method outperforms previous methods in both photometric and geometric metrics. We show that this simple yet effective technique of using correspondence priors can be applied as a plug-and-play module across different NeRF variants. The project page is at https://yxlao.github.io/corres-nerf/.
****
Score-based Data Assimilation
François Rozet, Gilles Louppe
Data assimilation, in its most comprehensive form, addresses the Bayesian inverse problem of identifying plausible state trajectories that explain noisy or incomplete observations of stochastic dynamical systems. Various approaches have been proposed to solve this problem, including particle-based and variational methods. However, most algorithms depend on the transition dynamics for inference, which becomes intractable for long time horizons or for high-dimensional systems with complex dynamics, such as oceans or atmospheres. In this work, we introduce score-based data assimilation for trajectory inference. We learn a score-based generative model of state trajectories based on the key insight that the score of an arbitrarily long trajectory can be decomposed into a series of scores over short segments. After training, inference is carried out using the score model, in a non-autoregressive manner by generating all states simultaneously. Quite distinctively, we decouple the observation model from the training procedure and use it only at inference to guide the generative process, which enables a wide range of zero-shot observation scenarios. We present theoretical and empirical evidence supporting the effectiveness of our method.
****
Mr. HiSum: A Large-scale Dataset for Video Highlight Detection and Summarization
Jinhwan Sul, Jihoon Han, Joonseok Lee
Video highlight detection is a task to automatically select the most engaging moments from a long video. This problem is highly challenging since it aims to learn a general way of finding highlights from a variety of videos in the real world.The task has an innate subjectivity because the definition of a highlight differs across individuals. Therefore, to detect consistent and meaningful highlights, prior benchmark datasets have been labeled by multiple (5-20) raters. Due to the high cost of manual labeling, most existing public benchmarks are in extremely small scale, containing only a few tens or hundreds of videos. This insufficient benchmark scale causes multiple issues such as unstable evaluation or high sensitivity in traintest splits. We present Mr. HiSum, a large-scale dataset for video highlight detection and summarization, containing 31,892 videos and reliable labels aggregated over 50,000+ users per video. We empirically prove reliability of the labels as frame importance by cross-dataset transfer and user study.
****
Sharp Bounds for Generalized Causal Sensitivity Analysis
Dennis Frauen, Valentyn Melnychuk, Stefan Feuerriegel
Causal inference from observational data is crucial for many disciplines such as medicine and economics. However, sharp bounds for causal effects under relaxations of the unconfoundedness assumption (causal sensitivity analysis) are subject to ongoing research. So far, works with sharp bounds are restricted to fairly simple settings (e.g., a single binary treatment). In this paper, we propose a unified framework for causal sensitivity analysis under unobserved confounding in various settings. For this, we propose a flexible generalization of the marginal

sensitivity model (MSM) and then derive sharp bounds for a large class of causal effects. This includes (conditional) average treatment effects, effects for mediation analysis and path analysis, and distributional effects. Furthermore, our sensitivity model is applicable to discrete, continuous, and time-varying treatments. It allows us to interpret the partial identification problem under unobserved confounding as a distribution shift in the latent confounders while evaluating the causal effect of interest. In the special case of a single binary treatment, our bounds for (conditional) average treatment effects coincide with recent optimality results for causal sensitivity analysis. Finally, we propose a scalable algorithm to estimate our sharp bounds from observational data.
****
Supported Value Regularization for Offline Reinforcement Learning
Yixiu Mao, Hongchang Zhang, Chen Chen, Yi Xu, Xiangyang Ji
Offline reinforcement learning suffers from the extrapolation error and value overestimation caused by out-of-distribution (OOD) actions. To mitigate this issue, value regularization approaches aim to penalize the learned value functions to assign lower values to OOD actions. However, existing value regularization methods lack a proper distinction between the regularization effects on in-distribution (ID) and OOD actions, and fail to guarantee optimal convergence results of the policy. To this end, we propose Supported Value Regularization (SVR), which penalizes the Q-values for all OOD actions while maintaining standard Bellman updates for ID ones. Specifically, we utilize the bias of importance sampling to compute the summation of Q-values over the entire OOD region, which serves as the penalty for policy evaluation. This design automatically separates the regularization for ID and OOD actions without manually distinguishing between them. In tabular MDP, we show that the policy evaluation operator of SVR is a contraction, whose fixed point outputs unbiased Q-values for ID actions and underestimated Q-values for OOD actions. Furthermore, the policy iteration with SVR guarantees strict policy improvement until convergence to the optimal support-constrained policy in the dataset. Empirically, we validate the theoretical properties of SVR in a tabular maze environment and demonstrate its state-of-the-art performance on a range of continuous control tasks in the D4RL benchmark.
****
Revisit Weakly-Supervised Audio-Visual Video Parsing from the Language Perspective
Yingying Fan, Yu Wu, Bo Du, Yutian Lin
We focus on the weakly-supervised audio-visual video parsing task (AVVP), which aims to identify and locate all the events in audio/visual modalities. Previous works only concentrate on video-level overall label denoising across modalities, but overlook the segment-level label noise, where adjacent video segments (i.e., 1-second video clips) may contain different events. However, recognizing events on the segment is challenging because its label could be any combination of events that occur in the video. To address this issue, we consider tackling AVVP from the language perspective, since language could freely describe how various events appear in each segment beyond fixed labels. Specifically, we design language prompts to describe all cases of event appearance for each video. Then, the similarity between language prompts and segments is calculated, where the event of the most similar prompt is regarded as the segment-level label. In addition, to deal with the mislabeled segments, we propose to perform dynamic re-weighting on the unreliable segments to adjust their labels. Experiments show that our simple yet effective approach outperforms state-of-the-art methods by a large margin.
****
Digital Typhoon: Long-term Satellite Image Dataset for the Spatio-Temporal Modeling of Tropical Cyclones
Asanobu Kitamoto, Jared Hwang, Bastien Vuillod, Lucas Gautier, Yingtao Tian, Tarin Clanuwat
This paper presents the official release of the Digital Typhoon dataset, the longest typhoon satellite image dataset for 40+ years aimed at benchmarking machine learning models for long-term spatio-temporal data. To build the dataset, we de

veloped a workflow to create an infrared typhoon-centered image for cropping usi
ng Lambert azimuthal equal-area projection referring to the best track data. We
also address data quality issues such as inter-satellite calibration to create a
 homogeneous dataset. To take advantage of the dataset, we organized machine lea
rning tasks by the types and targets of inference, with other tasks for meteorol
ogical analysis, societal impact, and climate change. The benchmarking results o
n the analysis, forecasting, and reanalysis for the intensity suggest that the d
ataset is challenging for recent deep learning models, due to many choices that
affect the performance of various models. This dataset reduces the barrier for m
achine learning researchers to meet large-scale real-world events called tropica
l cyclones and develop machine learning models that may contribute to advancing
scientific knowledge on tropical cyclones as well as solving societal and sustai
nability issues such as disaster reduction and climate change. The dataset is pu
blicly available at http://agora.ex.nii.ac.jp/digital-typhoon/dataset/ and https
://github.com/kitamoto-lab/digital-typhoon/.
****
Maximum Independent Set: Self-Training through Dynamic Programming
Lorenzo Brusca, Lars C.P.M. Quaedvlieg, Stratis Skoulakis, Grigorios Chrysos, Vo
lkan Cevher
This work presents a graph neural network (GNN) framework for solving the maximu
m independent set (MIS) problem, inspired by dynamic programming (DP). Specifica
lly, given a graph, we propose a DP-like recursive algorithm based on GNNs that
firstly constructs two smaller sub-graphs, predicts the one with the larger MIS,
 and then uses it in the next recursive call. To train our algorithm, we require
 annotated comparisons of different graphs concerning their MIS size. Annotating
 the comparisons with the output of our algorithm leads to a self-training proce
ss that results in more accurate self-annotation of the comparisons and vice ver
sa. We provide numerical evidence showing the superiority of our method vs prior
 methods in multiple synthetic and real-world datasets.
****
Reference-Based POMDPs
Edward Kim, Yohan Karunanayake, Hanna Kurniawati
Making good decisions in partially observable and non-deterministic scenarios is
 a crucial capability for robots. A Partially Observable Markov Decision Process
 (POMDP) is a general framework for the above problem. Despite advances in POMDP
 solving, problems with long planning horizons and evolving environments remain
difficult to solve even by the best approximate solvers today. To alleviate this
 difficulty, we propose a slightly modified POMDP problem, called a Reference-Ba
sed POMDP, where the objective is to balance between maximizing the expected tot
al reward and being close to a given reference (stochastic) policy. The optimal
policy of a Reference-Based POMDP can be computed via iterative expectations usi
ng the given reference policy, thereby avoiding exhaustive enumeration of action
s at each belief node of the search tree. We demonstrate theoretically that the
standard POMDP under stochastic policies is related to the Reference-Based POMDP
. To demonstrate the feasibility of exploiting the formulation, we present a bas
ic algorithm RefSolver. Results from experiments on long-horizon navigation prob
lems indicate that this basic algorithm substantially outperforms POMCP.
****
Siamese Masked Autoencoders
Agrim Gupta, Jiajun Wu, Jia Deng, Fei-Fei Li
Establishing correspondence between images or scenes is a significant challenge
in computer vision, especially given occlusions, viewpoint changes, and varying
object appearances. In this paper, we present Siamese Masked Autoencoders (SiamM
AE), a simple extension of Masked Autoencoders (MAE) for learning visual corresp
ondence from videos. SiamMAE operates on pairs of randomly sampled video frames
and asymmetrically masks them. These frames are processed independently by an en
coder network, and a decoder composed of a sequence of cross-attention layers is
 tasked with predicting the missing patches in the future frame. By masking a la
rge fraction (95%) of patches in the future frame while leaving the past frame u
nchanged, SiamMAE encourages the network to focus on object motion and learn obj

ect-centric representations. Despite its conceptual simplicity, features learned via SiamMAE outperform state-of-the-art self-supervised methods on video object segmentation, pose keypoint propagation, and semantic part propagation tasks. SiamMAE achieves competitive results without relying on data augmentation, handcrafted tracking-based pretext tasks, or other techniques to prevent representational collapse.
****
Score-based Generative Models with Lévy Processes
EUN BI YOON, Keehun Park, Sungwoong Kim, Sungbin Lim
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
3D Indoor Instance Segmentation in an Open-World
Mohamed El Amine Boudjoghra, Salwa Al Khatib, Jean Lahoud, Hisham Cholakkal, Rao Anwer, Salman H. Khan, Fahad Shahbaz Khan
Existing 3D instance segmentation methods typically assume that all semantic classes to be segmented would be available during training and only seen categories are segmented at inference. We argue that such a closed-world assumption is restrictive and explore for the first time 3D indoor instance segmentation in an open-world setting, where the model is allowed to distinguish a set of known classes as well as identify an unknown object as unknown and then later incrementally learning the semantic category of the unknown when the corresponding category labels are available. To this end, we introduce an open-world 3D indoor instance segmentation method, where an auto-labeling scheme is employed to produce pseudo-labels during training and induce separation to separate known and unknown category labels. We further improve the pseudo-labels quality at inference by adjusting the unknown class probability based on the objectness score distribution. We also introduce carefully curated open-world splits leveraging realistic scenarios based on inherent object distribution, region-based indoor scene exploration and randomness aspect of open-world classes. Extensive experiments reveal the efficacy of the proposed contributions leading to promising open-world 3D instance segmentation performance. Code and splits are available at: https://github.com/aminebdj/3D-OWIS.
****
AbDiffuser: full-atom generation of in-vitro functioning antibodies
Karolis Martinkus, Jan Ludwiczak, WEI-CHING LIANG, Julien Lafrance-Vanasse, Isidro Hotzel, Arvind Rajpal, Yan Wu, Kyunghyun Cho, Richard Bonneau, Vladimir Gligorijevic, Andreas Loukas
We introduce AbDiffuser, an equivariant and physics-informed diffusion model for the joint generation of antibody 3D structures and sequences. AbDiffuser is built on top of a new representation of protein structure, relies on a novel architecture for aligned proteins, and utilizes strong diffusion priors to improve the denoising process. Our approach improves protein diffusion by taking advantage of domain knowledge and physics-based constraints; handles sequence-length changes; and reduces memory complexity by an order of magnitude, enabling backbone and side chain generation. We validate AbDiffuser in silico and in vitro. Numerical experiments showcase the ability of AbDiffuser to generate antibodies that closely track the sequence and structural properties of a reference set. Laboratory experiments confirm that all 16 HER2 antibodies discovered were expressed at high levels and that 57.1% of the selected designs were tight binders.
****
Structure Learning with Adaptive Random Neighborhood Informed MCMC
Xitong Liang, Alberto Caron, Samuel Livingstone, Jim Griffin
In this paper, we introduce a novel MCMC sampler, PARNI-DAG, for a fully-Bayesian approach to the problem of structure learning under observational data. Under the assumption of causal sufficiency, the algorithm allows for approximate sampling directly from the posterior distribution on Directed Acyclic Graphs (DAGs). PARNI-DAG performs efficient sampling of DAGs via locally informed, adaptive ran

dom neighborhood proposal that results in better mixing properties. In addition, to ensure better scalability with the number of nodes, we couple PARNI-DAG with a pre-tuning procedure of the sampler's parameters that exploits a skeleton graph derived through some constraint-based or scoring-based algorithms. Thanks to these novel features, PARNI-DAG quickly converges to high-probability regions and is less likely to get stuck in local modes in the presence of high correlation between nodes in high-dimensional settings. After introducing the technical novelties in PARNI-DAG, we empirically demonstrate its mixing efficiency and accuracy in learning DAG structures on a variety of experiments.
****
Reining Generalization in Offline Reinforcement Learning via Representation Distinction
Yi Ma, Hongyao Tang, Dong Li, Zhaopeng Meng
Offline Reinforcement Learning (RL) aims to address the challenge of distribution shift between the dataset and the learned policy, where the value of out-of-distribution (OOD) data may be erroneously estimated due to overgeneralization. It has been observed that a considerable portion of the benefits derived from the conservative terms designed by existing offline RL approaches originates from their impact on the learned representation. This observation prompts us to scrutinize the learning dynamics of offline RL, formalize the process of generalization, and delve into the prevalent overgeneralization issue in offline RL. We then investigate the potential to rein the generalization from the representation perspective to enhance offline RL. Finally, we present Representation Distinction (RD), an innovative plug-in method for improving offline RL algorithm performance by explicitly differentiating between the representations of in-sample and OOD state-action pairs generated by the learning policy. Considering scenarios in which the learning policy mirrors the behavioral policy and similar samples may be erroneously distinguished, we suggest a dynamic adjustment mechanism for RD based on an OOD data generator to prevent data representation collapse and further enhance policy performance. We demonstrate the efficacy of our approach by applying RD to specially-designed backbone algorithms and widely-used offline RL algorithms. The proposed RD method significantly improves their performance across various continuous control tasks on D4RL datasets, surpassing several state-of-the-art offline RL algorithms.
****
BIRD: Generalizable Backdoor Detection and Removal for Deep Reinforcement Learning
Xuan Chen, Wenbo Guo, Guanhong Tao, Xiangyu Zhang, Dawn Song
Backdoor attacks pose a severe threat to the supply chain management of deep reinforcement learning (DRL) policies. Despite initial defenses proposed in recent studies, these methods have very limited generalizability and scalability. To address this issue, we propose BIRD, a technique to detect and remove backdoors from a pretrained DRL policy in a clean environment without requiring any knowledge about the attack specifications and accessing its training process. By analyzing the unique properties and behaviors of backdoor attacks, we formulate trigger restoration as an optimization problem and design a novel metric to detect backdoored policies. We also design a finetuning method to remove the backdoor, while maintaining the agent's performance in the clean environment. We evaluate BIRD against three backdoor attacks in ten different single-agent or multi-agent environments. Our results verify the effectiveness, efficiency, and generalizability of BIRD, as well as its robustness to different attack variations and adaptions.
****
Cluster-aware Semi-supervised Learning: Relational Knowledge Distillation Provably Learns Clustering
Yijun Dong, Kevin Miller, Qi Lei, Rachel Ward
Despite the empirical success and practical significance of (relational) knowledge distillation that matches (the relations of) features between teacher and student models, the corresponding theoretical interpretations remain limited for various knowledge distillation paradigms. In this work, we take an initial step to

ward a theoretical understanding of relational knowledge distillation (RKD), wit
h a focus on semi-supervised classification problems. We start by casting RKD as
 spectral clustering on a population-induced graph unveiled by a teacher model.
Via a notion of clustering error that quantifies the discrepancy between the pre
dicted and ground truth clusterings, we illustrate that RKD over the population
provably leads to low clustering error. Moreover, we provide a sample complexity
 bound for RKD with limited unlabeled samples. For semi-supervised learning, we
further demonstrate the label efficiency of RKD through a general framework of c
luster-aware semi-supervised learning that assumes low clustering errors. Finall
y, by unifying data augmentation consistency regularization into this cluster-aw
are framework, we show that despite the common effect of learning accurate clust
erings, RKD facilitates a "global" perspective through spectral clustering, wher
eas consistency regularization focuses on a "local" perspective via expansion.
****
Bicriteria Multidimensional Mechanism Design with Side Information
Siddharth Prasad, Maria-Florina F. Balcan, Tuomas Sandholm
We develop a versatile new methodology for multidimensional mechanism design tha
t incorporates side information about agent types to generate high social welfar
e and high revenue simultaneously. Prominent sources of side information in prac
tice include predictions from a machine-learning model trained on historical age
nt data, advice from domain experts, and even the mechanism designer's own gut i
nstinct. In this paper we adopt a prior-free perspective that makes no assumptio
ns on the correctness, accuracy, or source of the side information. First, we de
sign a meta-mechanism that integrates input side information with an improvement
 of the classical VCG mechanism. The welfare, revenue, and incentive properties
of our meta-mechanism are characterized by novel constructions we introduce base
d on the notion of a weakest competitor, which is an agent that has the smallest
 impact on welfare. We show that our meta-mechanism, when carefully instantiated
, simultaneously achieves strong welfare and revenue guarantees parameterized by
 errors in the side information. When the side information is highly informative
 and accurate, our mechanism achieves welfare and revenue competitive with the t
otal social surplus, and its performance decays continuously and gradually as th
e quality of the side information decreases. Finally, we apply our meta-mechanis
m to a setting where each agent's type is determined by a constant number of par
ameters. Specifically, agent types lie on constant-dimensional subspaces (of the
 potentially high-dimensional ambient type space) that are known to the mechanis
m designer. We use our meta-mechanism to obtain the first known welfare and reve
nue guarantees in this setting.
****
Are These the Same Apple? Comparing Images Based on Object Intrinsics
Klemen Kotar, Stephen Tian, Hong-Xing Yu, Dan Yamins, Jiajun Wu
The human visual system can effortlessly recognize an object under different ext
rinsic factors such as lighting, object poses, and background, yet current compu
ter vision systems often struggle with these variations. An important step to un
derstanding and improving artificial vision systems is to measure image similari
ty purely based on intrinsic object properties that define object identity. This
 problem has been studied in the computer vision literature as re-identification
, though mostly restricted to specific object categories such as people and cars
. We propose to extend it to general object categories, exploring an image simil
arity metric based on object intrinsics. To benchmark such measurements, we coll
ect the Common paired objects Under differenT Extrinsics (CUTE) dataset of 18, 0
00 images of 180 objects under different extrinsic factors such as lighting, pos
es, and imaging conditions. While existing methods such as LPIPS and CLIP scores
 do not measure object intrinsics well, we find that combining deep features lea
rned from contrastive self-supervised learning with foreground filtering is a si
mple yet effective approach to approximating the similarity. We conduct an exten
sive survey of pre-trained features and foreground extraction methods to arrive
at a strong baseline that best measures intrinsic object-centric image similarit
y among current methods. Finally, we demonstrate that our approach can aid in do
wnstream applications such as acting as an analog for human subjects and improvi

ng generalizable re-identification. Please see our project website at https://s-tian.github.io/projects/cute/ for visualizations of the data and demos of our me tric.
****

The ToMCAT Dataset

Adarsh Pyarelal, Eric Duong, Caleb Shibu, Paulo Soares, Savannah Boyd, Payal Kho sla, Valeria A. Pfeifer, Diheng Zhang, Eric Andrews, Rick Champlin, Vincent Raym ond, Meghavarshini Krishnaswamy, Clayton Morrison, Emily Butler, Kobus Barnard

We present a rich, multimodal dataset consisting of data from 40 teams of three humans conducting simulated urban search-and-rescue (SAR) missions in a Minecraf t-based testbed, collected for the Theory of Mind-based Cognitive Architecture f or Teams (ToMCAT) project. Modalities include two kinds of brain scan data---fun ctional near-infrared spectroscopy (fNIRS) and electroencephalography (EEG), as well as skin conductance, heart rate, eye tracking, face images, spoken dialog a udio data with automatic speech recognition (ASR) transcriptions, game screensho ts, gameplay data, game performance data, demographic data, and self-report ques tionnaires. Each team undergoes up to six consecutive phases: three behavioral t asks, one mission training session, and two collaborative SAR missions. As time-synchronized multimodal data collected under a variety of circumstances, this da taset will support studying a large variety of research questions on topics incl uding teamwork, coordination, plan recognition, affective computing, physiologic al linkage, entrainment, and dialog understanding. We provide an initial public release of the de-identified data, along with analyses illustrating the utility of this dataset to both computer scientists and social scientists.
****

Exploring Diverse In-Context Configurations for Image Captioning

Xu Yang, Yongliang Wu, Mingzhuo Yang, Haokun Chen, Xin Geng

After discovering that Language Models (LMs) can be good in-context few-shot lea rners, numerous strategies have been proposed to optimize in-context sequence co nfigurations. Recently, researchers in Vision-Language (VL) domains also develop their few-shot learners, while they only use the simplest way, \ie, randomly sa mpling, to configure in-context image-text pairs. In order to explore the effect s of varying configurations on VL in-context learning, we devised four strategie s for image selection and four for caption assignment to configure in-context im age-text pairs for image captioning. Here Image Captioning is used as the case s tudy since it can be seen as the visually-conditioned LM. Our comprehensive expe riments yield two counter-intuitive but valuable insights, highlighting the dist inct characteristics of VL in-context learning due to multi-modal synergy, as co mpared to the NLP case. Furthermore, in our exploration of optimal combination s trategies, we observed an average performance enhancement of 20.9 in CIDEr score s compared to the baseline. The code is given in https://github.com/yongliang-wu /ExploreCfg.
****

DELIFFAS: Deformable Light Fields for Fast Avatar Synthesis

Youngjoong Kwon, Lingjie Liu, Henry Fuchs, Marc Habermann, Christian Theobalt

Generating controllable and photorealistic digital human avatars is a long-stand ing and important problem in Vision and Graphics. Recent methods have shown grea t progress in terms of either photorealism or inference speed while the combinat ion of the two desired properties still remains unsolved. To this end, we propos e a novel method, called DELIFFAS, which parameterizes the appearance of the hum an as a surface light field that is attached to a controllable and deforming hum an mesh model. At the core, we represent the light field around the human with a deformable two-surface parameterization, which enables fast and accurate infere nce of the human appearance. This allows perceptual supervision on the full imag e compared to previous approaches that could only supervise individual pixels or small patches due to their slow runtime. Our carefully designed human represent ation and supervision strategy leads to state-of-the-art synthesis results and i nference time. The video results and code are available at https://vcai.mpi-inf. mpg.de/projects/DELIFFAS.
****

Zero-Shot Anomaly Detection via Batch Normalization

Aodong Li, Chen Qiu, Marius Kloft, Padhraic Smyth, Maja Rudolph, Stephan Mandt

Anomaly detection (AD) plays a crucial role in many safety-critical application domains. The challenge of adapting an anomaly detector to drift in the normal data distribution, especially when no training data is available for the "new normal," has led to the development of zero-shot AD techniques. In this paper, we propose a simple yet effective method called Adaptive Centered Representations (ACR) for zero-shot batch-level AD. Our approach trains off-the-shelf deep anomaly detectors (such as deep SVDD) to adapt to a set of inter-related training data distributions in combination with batch normalization, enabling automatic zero-shot generalization for unseen AD tasks. This simple recipe, batch normalization plus meta-training, is a highly effective and versatile tool. Our results demonstrate the first zero-shot AD results for tabular data and outperform existing methods in zero-shot anomaly detection and segmentation on image data from specialized domains.

****

What Makes Data Suitable for a Locally Connected Neural Network? A Necessary and Sufficient Condition Based on Quantum Entanglement.

■Yotam Alexander■■, Nimrod De La Vega, Noam Razin, Nadav Cohen

The question of what makes a data distribution suitable for deep learning is a fundamental open problem. Focusing on locally connected neural networks (a prevalent family of architectures that includes convolutional and recurrent neural networks as well as local self-attention models), we address this problem by adopting theoretical tools from quantum physics. Our main theoretical result states that a certain locally connected neural network is capable of accurate prediction over a data distribution if and only if the data distribution admits low quantum entanglement under certain canonical partitions of features. As a practical application of this result, we derive a preprocessing method for enhancing the suitability of a data distribution to locally connected neural networks. Experiments with widespread models over various datasets demonstrate our findings. We hope that our use of quantum entanglement will encourage further adoption of tools from physics for formally reasoning about the relation between deep learning and real-world data.

****

Parameter and Computation Efficient Transfer Learning for Vision-Language Pre-trained Models

Qiong Wu, Wei Yu, Yiyi Zhou, Shubin Huang, Xiaoshuai Sun, Rongrong Ji

With ever increasing parameters and computation, vision-language pre-trained (VLP) models exhibit prohibitive expenditure in downstream task adaption. Recent endeavors mainly focus on parameter efficient transfer learning (PETL) for VLP models by only updating a small number of parameters. However, excessive computational overhead still plagues the application of VLPs. In this paper, we aim at parameter and computation efficient transfer learning (PCETL) for VLP models. In particular, PCETL not only needs to limit the number of trainable parameters in VLP models, but also to reduce the computational redundancy during inference, thus enabling a more efficient transfer. To approach this target, we propose a novel dynamic architecture skipping (DAS) approach towards effective PCETL. Instead of directly optimizing the intrinsic architectures of VLP models, DAS first observes the significances of their modules to downstream tasks via a reinforcement learning (RL) based process, and then skips the redundant ones with lightweight networks, i.e. adapters, according to the obtained rewards. In this case, the VLP model can well maintain the scale of trainable parameters while speeding up its inference on downstream tasks. To validate DAS, we apply it to two representative VLP models, namely ViLT and METER, and conduct extensive experiments on a bunch of VL tasks. The experimental results not only show the great advantages of DAS in reducing computational complexity, e.g. -11.97% FLOPs of METER on VQA2.0, but also confirm its competitiveness against existing PETL methods in terms of parameter scale and performance. Our source code is given in our appendix.

****

A Dynamical System View of Langevin-Based Non-Convex Sampling

Mohammad Reza Karimi Jaghargh, Ya-Ping Hsieh, Andreas Krause

Non-convex sampling is a key challenge in machine learning, central to non-convex optimization in deep learning as well as to approximate probabilistic inference. Despite its significance, theoretically there remain some important challenges: Existing guarantees suffer from the drawback of lacking guarantees for the last-iterates, and little is known beyond the elementary schemes of stochastic gradient Langevin dynamics. To address these issues, we develop a novel framework that lifts the above issues by harnessing several tools from the theory of dynamical systems. Our key result is that, for a large class of state-of-the-art sampling schemes, their last-iterate convergence in Wasserstein distances can be reduced to the study of their continuous-time counterparts, which is much better understood. Coupled with standard assumptions of MCMC sampling, our theory immediately yields the last-iterate Wasserstein convergence of many advanced sampling schemes such as mirror Langevin, proximal, randomized mid-point, and Runge-Kutta methods.

****

OKRidge: Scalable Optimal k-Sparse Ridge Regression

Jiachang Liu, Sam Rosen, Chudi Zhong, Cynthia Rudin

We consider an important problem in scientific discovery, namely identifying sparse governing equations for nonlinear dynamical systems. This involves solving sparse ridge regression problems to provable optimality in order to determine which terms drive the underlying dynamics. We propose a fast algorithm, OKRidge, for sparse ridge regression, using a novel lower bound calculation involving, first, a saddle point formulation, and from there, either solving (i) a linear system or (ii) using an ADMM-based approach, where the proximal operators can be efficiently evaluated by solving another linear system and an isotonic regression problem. We also propose a method to warm-start our solver, which leverages a beam search. Experimentally, our methods attain provable optimality with run times that are orders of magnitude faster than those of the existing MIP formulations solved by the commercial solver Gurobi.

****

Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise

Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie Li, Hamid Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, Tom Goldstein

Standard diffusion models involve an image transform -- adding Gaussian noise -- and an image restoration operator that inverts this degradation. We observe that the generative behavior of diffusion models is not strongly dependent on the choice of image degradation, and in fact, an entire family of generative models can be constructed by varying this choice. Even when using completely deterministic degradations (e.g., blur, masking, and more), the training and test-time update rules that underlie diffusion models can be easily generalized to create generative models. The success of these fully deterministic models calls into question the community's understanding of diffusion models, which relies on noise in either gradient Langevin dynamics or variational inference and paves the way for generalized diffusion models that invert arbitrary processes.

****

Towards the Difficulty for a Deep Neural Network to Learn Concepts of Different Complexities

Dongrui Liu, Huiqi Deng, Xu Cheng, Qihan Ren, Kangrui Wang, Quanshi Zhang

This paper theoretically explains the intuition that simple concepts are more likely to be learned by deep neural networks (DNNs) than complex concepts. In fact, recent studies have observed [24, 15] and proved [26] the emergence of interactive concepts in a DNN, i.e., it is proven that a DNN usually only encodes a small number of interactive concepts, and can be considered to use their interaction effects to compute inference scores. Each interactive concept is encoded by the DNN to represent the collaboration between a set of input variables. Therefore, in this study, we aim to theoretically explain that interactive concepts involving more input variables (i.e., more complex concepts) are more difficult to learn. Our finding clarifies the exact conceptual complexity that boosts the learning difficulty.

****
Limits, approximation and size transferability for GNNs on sparse graphs via gra
phops
Thien Le, Stefanie Jegelka
Can graph neural networks generalize to graphs that are different from the graph
s they were trained on, e.g., in size? In this work, we study this question from
 a theoretical perspective. While recent work established such transferability a
nd approximation results via graph limits, e.g., via graphons, these only apply
nontrivially to dense graphs. To include frequently encountered sparse graphs su
ch as bounded-degree or power law graphs, we take a perspective of taking limits
 of operators derived from graphs, such as the aggregation operation that makes
up GNNs. This leads to the recently introduced limit notion of graphops (Backhau
sz and Szegedy, 2022). We demonstrate how the operator perspective allows us to
develop quantitative bounds on the distance between a finite GNN and its limit o
n an infinite graph, as well as the distance between the GNN on graphs of differ
ent sizes that share structural properties, under a regularity assumption verifi
ed for various graph sequences. Our results hold for dense and sparse graphs, an
d various notions of graph limits.
****
The Adversarial Consistency of Surrogate Risks for Binary Classification
Natalie Frank, Jonathan Niles-Weed
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
The Cambridge Law Corpus: A Corpus for Legal AI Research
Andreas Östling, Holli Sargeant, Huiyuan Xie, Ludwig Bull, Alexander Terenin, Le
if Jonsson, Måns Magnusson, Felix Steffek
We introduce the Cambridge Law Corpus (CLC), a corpus for legal AI research. It
consists of over 250 000 court cases from the UK. Most cases are from the 21st c
entury, but the corpus includes cases as old as the 16th century. This paper pre
sents the first release of the corpus, containing the raw text and meta-data. To
gether with the corpus, we provide annotations on case outcomes for 638 cases, d
one by legal experts. Using our annotated data, we have trained and evaluated ca
se outcome extraction with GPT-3, GPT-4 and RoBERTa models to provide benchmarks
. We include an extensive legal and ethical discussion to address the potentiall
y sensitive nature of this material. As a consequence, the corpus will only be r
eleased for research purposes under certain restrictions.
****
Large Language Models of Code Fail at Completing Code with Potential Bugs
Tuan Dinh, Jinman Zhao, Samson Tan, Renato Negrinho, Leonard Lausen, Sheng Zha,
George Karypis
Large language models of code (Code-LLMs) have recently brought tremendous advan
ces to code completion, a fundamental feature of programming assistance and code
 intelligence. However, most existing works ignore the possible presence of bugs
 in the code context for generation, which are inevitable in software developmen
t. Therefore, we introduce and study the buggy-code completion problem, inspired
 by the realistic scenario of real-time code suggestion where the code context c
ontains potential bugs – anti-patterns that can become bugs in the completed pro
gram. To systematically study the task, we introduce two datasets: one with synt
hetic bugs derived from semantics-altering operator changes (buggy-HumanEval) an
d one with realistic bugs derived from user submissions to coding problems (bugg
y-FixEval). We find that the presence of potential bugs significantly degrades t
he generation performance of the high-performing Code-LLMs. For instance, the pa
ssing rates of CODEGEN-2B-MONO on test cases of buggy-HumanEval drop more than 5
0% given a single potential bug in the context. Finally, we investigate several
post-hoc methods for mitigating the adverse effect of potential bugs and find th
at there remains a large gap in post-mitigation performance.
****

Doubly-Robust Self-Training
Banghua Zhu, Mingyu Ding, Philip Jacobson, Ming Wu, Wei Zhan, Michael Jordan, Jiantao Jiao
Self-training is a well-established technique in semi-supervised learning, which leverages unlabeled data by generating pseudo-labels and incorporating them with a limited labeled dataset for training. The effectiveness of self-training heavily relies on the accuracy of these pseudo-labels. In this paper, we introduce doubly-robust self-training, an innovative semi-supervised algorithm that provably balances between two extremes. When pseudo-labels are entirely incorrect, our method reduces to a training process solely using labeled data. Conversely, when pseudo-labels are completely accurate, our method transforms into a training process utilizing all pseudo-labeled data and labeled data, thus increasing the effective sample size. Through empirical evaluations on both the ImageNet dataset for image classification and the nuScenes autonomous driving dataset for 3D object detection, we demonstrate the superiority of the doubly-robust loss over the self-training baseline.
****
FairLISA: Fair User Modeling with Limited Sensitive Attributes Information
zheng zhang, Qi Liu, Hao Jiang, Fei Wang, Yan Zhuang, Le Wu, Weibo Gao, Enhong Chen
User modeling techniques profile users' latent characteristics (e.g., preference) from their observed behaviors, and play a crucial role in decision-making. Unfortunately, traditional user models may unconsciously capture biases related to sensitive attributes (e.g., gender) from behavior data, even when this sensitive information is not explicitly provided. This can lead to unfair issues and discrimination against certain groups based on these sensitive attributes. Recent studies have been proposed to improve fairness by explicitly decorrelating user modeling results and sensitive attributes. However, most existing approaches assume that fully sensitive attribute labels are available in the training set, which is unrealistic due to collection limitations like privacy concerns, and hence bear the limitation of performance. In this paper, we focus on a practical situation with limited sensitive data and propose a novel FairLISA framework, which can efficiently utilize data with known and unknown sensitive attributes to facilitate fair model training. We first propose a novel theoretical perspective to build the relationship between data with both known and unknown sensitive attributes with the fairness objective. Then, based on this, we provide a general adversarial framework to effectively leverage the whole user data for fair user modeling. We conduct experiments on representative user modeling tasks including recommender system and cognitive diagnosis. The results demonstrate that our FairLISA can effectively improve fairness while retaining high accuracy in scenarios with different ratios of missing sensitive attributes.
****
Inference-Time Intervention: Eliciting Truthful Answers from a Language Model
Kenneth Li, Oam Patel, Fernanda Viégas, Hanspeter Pfister, Martin Wattenberg
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Composable Coresets for Determinant Maximization: Greedy is Almost Optimal
Siddharth Gollapudi, Sepideh Mahabadi, Varun Sivashankar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ProBio: A Protocol-guided Multimodal Dataset for Molecular Biology Lab
Jieming Cui, Ziren Gong, Baoxiong Jia, Siyuan Huang, Zilong Zheng, Jianzhu Ma, Yixin Zhu
The challenge of replicating research results has posed a significant impediment

to the field of molecular biology. The advent of modern intelligent systems has led to notable progress in various domains. Consequently, we embarked on an investigation of intelligent monitoring systems as a means of tackling the issue of the reproducibility crisis. Specifically, we first curate a comprehensive multimodal dataset, named ProBio, as an initial step towards this objective. This dataset comprises fine-grained hierarchical annotations intended for the purpose of studying activity understanding in BioLab. Next, we devise two challenging benchmarks, transparent solution tracking and multimodal action recognition, to emphasize the unique characteristics and difficulties associated with activity understanding in BioLab settings. Finally, we provide a thorough experimental evaluation of contemporary video understanding models and highlight their limitations in this specialized domain to identify potential avenues for future research. We hope \dataset with associated benchmarks may garner increased focus on modern AI techniques in the realm of molecular biology.
****
Spuriosity Rankings: Sorting Data to Measure and Mitigate Biases
Mazda Moayeri, Wenxiao Wang, Sahil Singla, Soheil Feizi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
NEO-KD: Knowledge-Distillation-Based Adversarial Training for Robust Multi-Exit Neural Networks
Seokil Ham, Jungwuk Park, Dong-Jun Han, Jaekyun Moon
While multi-exit neural networks are regarded as a promising solution for making efficient inference via early exits, combating adversarial attacks remains a challenging problem. In multi-exit networks, due to the high dependency among different submodels, an adversarial example targeting a specific exit not only degrades the performance of the target exit but also reduces the performance of all other exits concurrently. This makes multi-exit networks highly vulnerable to simple adversarial attacks. In this paper, we propose NEO-KD, a knowledge-distillation-based adversarial training strategy that tackles this fundamental challenge based on two key contributions. NEO-KD first resorts to neighbor knowledge distillation to guide the output of the adversarial examples to tend to the ensemble outputs of neighbor exits of clean data. NEO-KD also employs exit-wise orthogonal knowledge distillation for reducing adversarial transferability across different submodels. The result is a significantly improved robustness against adversarial attacks. Experimental results on various datasets/models show that our method achieves the best adversarial accuracy with reduced computation budgets, compared to the baselines relying on existing adversarial training or knowledge distillation techniques for multi-exit networks.
****
Self-Evaluation Guided Beam Search for Reasoning
Yuxi Xie, Kenji Kawaguchi, Yiran Zhao, James Xu Zhao, Min-Yen Kan, Junxian He, Michael Xie
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ViSt3D: Video Stylization with 3D CNN
Ayush Pande, Gaurav Sharma
Visual stylization has been a very popular research area in recent times. While image stylization has seen a rapid advancement in the recent past, video stylization, while being more challenging, is relatively less explored. The immediate method of stylizing videos by stylizing each frame independently has been tried with some success. To the best of our knowledge, we present the first approach to video stylization using 3D CNN directly, building upon insights from 2D image stylization. Stylizing video is highly challenging, as the appearance and video m

otion, which includes both camera and subject motions, are inherently entangled in the representations learnt by a 3D CNN. Hence, a naive extension of 2D CNN stylization methods to 3D CNN does not work. To perform stylization with 3D CNN, we propose to explicitly disentangle motion and appearance, stylize the appearance part, and then add back the motion component and decode the final stylized video. In addition, we propose a dataset, curated from existing datasets, to train video stylization networks. We also provide an independently collected test set to study the generalization of video stylization methods. We provide results on this test dataset comparing the proposed method with 2D stylization methods applied frame by frame. We show successful stylization with 3D CNN for the first time, and obtain better stylization in terms of texture cf.\ the existing 2D methods.
****
Smoothed Online Learning for Prediction in Piecewise Affine Systems
Adam Block, Max Simchowitz, Russ Tedrake
The problem of piecewise affine (PWA) regression and planning is of foundational importance to the study of online learning, control, and robotics, where it provides a theoretically and empirically tractable setting to study systems undergoing sharp changes in the dynamics. Unfortunately, due to the discontinuities that arise when crossing into different ``pieces,'' learning in general sequential settings is impossible and practical algorithms are forced to resort to heuristic approaches. This paper builds on the recently developed smoothed online learning framework and provides the first algorithms for prediction and simulation in PWA systems whose regret is polynomial in all relevant problem parameters under a weak smoothness assumption; moreover, our algorithms are efficient in the number of calls to an optimization oracle. We further apply our results to the problems of one-step prediction and multi-step simulation regret in piecewise affine dynamical systems, where the learner is tasked with simulating trajectories and regret is measured in terms of the Wasserstein distance between simulated and true data. Along the way, we develop several technical tools of more general interest.
****
Adversarial Attacks on Online Learning to Rank with Click Feedback
Jinhang Zuo, Zhiyao Zhang, Zhiyong Wang, Shuai Li, Mohammad Hajiesmaili, Adam Wierman
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
RAPHAEL: Text-to-Image Generation via Large Mixture of Diffusion Paths
Zeyue Xue, Guanglu Song, Qiushan Guo, Boxiao Liu, Zhuofan Zong, Yu Liu, Ping Luo
Text-to-image generation has recently witnessed remarkable achievements. We introduce a text-conditional image diffusion model, termed RAPHAEL, to generate highly artistic images, which accurately portray the text prompts, encompassing multiple nouns, adjectives, and verbs. This is achieved by stacking tens of mixture-of-experts (MoEs) layers, i.e., space-MoE and time-MoE layers, enabling billions of diffusion paths (routes) from the network input to the output. Each path intuitively functions as a "painter" for depicting a particular textual concept onto a specified image region at a diffusion timestep. Comprehensive experiments reveal that RAPHAEL outperforms recent cutting-edge models, such as Stable Diffusion, ERNIE-ViLG 2.0, DeepFloyd, and DALL-E 2, in terms of both image quality and aesthetic appeal. Firstly, RAPHAEL exhibits superior performance in switching images across diverse styles, such as Japanese comics, realism, cyberpunk, and ink illustration. Secondly, a single model with three billion parameters, trained on 1,000 A100 GPUs for two months, achieves a state-of-the-art zero-shot FID score of 6.61 on the COCO dataset. Furthermore, RAPHAEL significantly surpasses its counterparts in human evaluation on the ViLG-300 benchmark. We believe that RAPHAEL holds the potential to propel the frontiers of image generation research in both academia and industry, paving the way for future breakthroughs in this

rapidly evolving field. More details can be found on a webpage: https://raphael-painter.github.io/.
****
Towards Evaluating Transfer-based Attacks Systematically, Practically, and Fairly
Qizhang Li, Yiwen Guo, Wangmeng Zuo, Hao Chen
The adversarial vulnerability of deep neural networks (DNNs) has drawn great attention due to the security risk of applying these models in real-world applications. Based on transferability of adversarial examples, an increasing number of transfer-based methods have been developed to fool black-box DNN models whose architecture and parameters are inaccessible. Although tremendous effort has been exerted, there still lacks a standardized benchmark that could be taken advantage of to compare these methods systematically, fairly, and practically. Our investigation shows that the evaluation of some methods needs to be more reasonable and more thorough to verify their effectiveness, to avoid, for example, unfair comparison and insufficient consideration of possible substitute/victim models. Therefore, we establish a transfer-based attack benchmark (TA-Bench) which implements 30+ methods. In this paper, we evaluate and compare them comprehensively on 10 popular substitute/victim models on ImageNet. New insights about the effectiveness of these methods are gained and guidelines for future evaluations are provided.
****
Automatic Clipping: Differentially Private Deep Learning Made Easier and Stronger
Zhiqi Bu, Yu-Xiang Wang, Sheng Zha, George Karypis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Error Discovery By Clustering Influence Embeddings
Fulton Wang, Julius Adebayo, Sarah Tan, Diego Garcia-Olano, Narine Kokhlikyan
We present a method for identifying groups of test examples---slices---on which a model under-performs, a task now known as slice discovery. We formalize coherence---a requirement that erroneous predictions, within a slice, should be wrong for the same reason---as a key property that any slice discovery method should satisfy. We then use influence functions to derive a new slice discovery method, InfEmbed, which satisfies coherence by returning slices whose examples are influenced similarly by the training data. InfEmbed is simple, and consists of applying K-Means clustering to a novel representation we deem influence embeddings. We show InfEmbed outperforms current state-of-the-art methods on 2 benchmarks, and is effective for model debugging across several case studies.
****
BadTrack: A Poison-Only Backdoor Attack on Visual Object Tracking
Bin Huang, Jiaqian Yu, Yiwei Chen, Siyang Pan, Qiang Wang, Zhi Wang
Visual object tracking (VOT) is one of the most fundamental tasks in computer vision community. State-of-the-art VOT trackers extract positive and negative examples that are used to guide the tracker to distinguish the object from the background. In this paper, we show that this characteristic can be exploited to introduce new threats and hence propose a simple yet effective poison-only backdoor attack. To be specific, we poison a small part of the training data by attaching a predefined trigger pattern to the background region of each video frame, so that the trigger appears almost exclusively in the extracted negative examples. To the best of our knowledge, this is the first work that reveals the threat of poison-only backdoor attack on VOT trackers. We experimentally show that our backdoor attack can significantly degrade the performance of both two-stream Siamese and one-stream Transformer trackers on the poisoned data while gaining comparable performance with the benign trackers on the clean data.
****
Spiking PointNet: Spiking Neural Networks for Point Clouds

Dayong Ren, Zhe Ma, Yuanpei Chen, Weihang Peng, Xiaode Liu, Yuhan Zhang, Yufei Guo
Recently, Spiking Neural Networks (SNNs), enjoying extreme energy efficiency, have drawn much research attention on 2D visual recognition and shown gradually increasing application potential. However, it still remains underexplored whether SNNs can be generalized to 3D recognition. To this end, we present Spiking PointNet in the paper, the first spiking neural model for efficient deep learning on point clouds. We discover that the two huge obstacles limiting the application of SNNs in point clouds are: the intrinsic optimization obstacle of SNNs that impedes the training of a big spiking model with large time steps, and the expensive memory and computation cost of PointNet that makes training a big spiking point model unrealistic. To solve the problems simultaneously, we present a trained-less but learning-more paradigm for Spiking PointNet with theoretical justifications and in-depth experimental analysis. In specific, our Spiking PointNet is trained with only a single time step but can obtain better performance with multiple time steps inference, compared to the one trained directly with multiple time steps. We conduct various experiments on ModelNet10, ModelNet40 to demonstrate the effectiveness of Sipiking PointNet. Notably, our Spiking PointNet even can outperform its ANN counterpart, which is rare in the SNN field thus providing a potential research direction for the following work. Moreover, Spiking PointNet shows impressive speedup and storage saving in the training phase. Our code is open-sourced at https://github.com/DayongRen/Spiking-PointNet.
****
A Sublinear-Time Spectral Clustering Oracle with Improved Preprocessing Time
Ranran Shen, Pan Peng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Boosting with Tempered Exponential Measures
Richard Nock, Ehsan Amid, Manfred Warmuth
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DP-HyPO: An Adaptive Private Framework for Hyperparameter Optimization
Hua Wang, Sheng Gao, Huanyu Zhang, Weijie Su, Milan Shen
Hyperparameter optimization, also known as hyperparameter tuning, is a widely recognized technique for improving model performance. Regrettably, when training private ML models, many practitioners often overlook the privacy risks associated with hyperparameter optimization, which could potentially expose sensitive information about the underlying dataset.Currently, the sole existing approach to allow privacy-preserving hyperparameter optimization is to uniformly and randomly select hyperparameters for a number of runs, subsequently reporting the best-performing hyperparameter.In contrast, in non-private settings, practitioners commonly utilize "adaptive" hyperparameter optimization methods such as Gaussian Process-based optimization, which select the next candidate based on information gathered from previous outputs.This substantial contrast between private and non-private hyperparameter optimization underscores a critical concern. In our paper, we introduce DP-HyPO, a pioneering framework for "adaptive" private hyperparameter optimization, aiming to bridge the gap between private and non-private hyperparameter optimization. To accomplish this, we provide a comprehensive differential privacy analysis of our framework. Furthermore, we empirically demonstrate the effectiveness of DP-HyPO on a diverse set of real-world datasets.
****
Active Learning-Based Species Range Estimation
Christian Lange, Elijah Cole, Grant Van Horn, Oisin Mac Aodha
We propose a new active learning approach for efficiently estimating the geograp

hic range of a species from a limited number of on the ground observations. We m odel the range of an unmapped species of interest as the weighted combination of estimated ranges obtained from a set of different species. We show that it is p ossible to generate this candidate set of ranges by using models that have been trained on large weakly supervised community collected observation data. From th is, we develop a new active querying approach that sequentially selects geograph ic locations to visit that best reduce our uncertainty over an unmapped species' range. We conduct a detailed evaluation of our approach and compare it to exist ing active learning methods using an evaluation dataset containing expert-derive d ranges for one thousand species. Our results demonstrate that our method outpe rforms alternative active learning methods and approaches the performance of end -to-end trained models, even when only using a fraction of the data. This highli ghts the utility of active learning via transfer learned spatial representations for species range estimation. It also emphasizes the value of leveraging emergi ng large-scale crowdsourced datasets, not only for modeling a species' range, bu t also for actively discovering them.
****
One-Step Diffusion Distillation via Deep Equilibrium Models
Zhengyang Geng, Ashwini Pokle, J. Zico Kolter
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Discrete-Smoothness in Online Algorithms with Predictions
Yossi Azar, Debmalya Panigrahi, Noam Touitou
In recent years, there has been an increasing focus on designing online algorith ms with (machine-learned) predictions. The ideal learning-augmented algorithm is comparable to the optimum when given perfect predictions (consistency), to the best online approximation for arbitrary predictions (robustness), and should int erpolate between these extremes as a smooth function of the prediction error. In this paper, we quantify these guarantees in terms of a general property that we call discrete-smoothness, and achieve discrete-smooth algorithms for online cov ering, specifically the facility location and set cover problems. For set cover, our work improves the results of Bamas, Maggiori, and Svensson (2020) by augmen ting consistency and robustness with smoothness guarantees. For facility locatio n, our work improves on prior work by Almanza et al. (2021) by generalizing to n onuniform costs and also providing smoothness guarantees by augmenting consisten cy and robustness.
****
A Performance-Driven Benchmark for Feature Selection in Tabular Deep Learning
Valeriia Cherepanova, Roman Levin, Gowthami Somepalli, Jonas Geiping, C. Bayan B russ, Andrew G. Wilson, Tom Goldstein, Micah Goldblum
Academic tabular benchmarks often contain small sets of curated features. In con trast, data scientists typically collect as many features as possible into their datasets, and even engineer new features from existing ones. To prevent over-fi tting in subsequent downstream modeling, practitioners commonly use automated fe ature selection methods that identify a reduced subset of informative features. Existing benchmarks for tabular feature selection consider classical downstream models, toy synthetic datasets, or do not evaluate feature selectors on the basi s of downstream performance. We construct a challenging feature selection benchm ark evaluated on downstream neural networks including transformers, using real d atasets and multiple methods for generating extraneous features. We also propose an input-gradient-based analogue of LASSO for neural networks that outperforms classical feature selection methods on challenging problems such as selecting fr om corrupted or second-order features.
****
Riemannian Projection-free Online Learning
Zihao Hu, Guanghui Wang, Jacob D. Abernethy
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SwiFT: Swin 4D fMRI Transformer
Peter Kim, Junbeom Kwon, Sunghwan Joo, Sangyoon Bae, Donggyu Lee, Yoonho Jung, Shinjae Yoo, Jiook Cha, Taesup Moon
Modeling spatiotemporal brain dynamics from high-dimensional data, such as functional Magnetic Resonance Imaging (fMRI), is a formidable task in neuroscience. Existing approaches for fMRI analysis utilize hand-crafted features, but the process of feature extraction risks losing essential information in fMRI scans. To address this challenge, we present SwiFT (Swin 4D fMRI Transformer), a Swin Transformer architecture that can learn brain dynamics directly from fMRI volumes in a memory and computation-efficient manner. SwiFT achieves this by implementing a 4D window multi-head self-attention mechanism and absolute positional embeddings. We evaluate SwiFT using multiple large-scale resting-state fMRI datasets, including the Human Connectome Project (HCP), Adolescent Brain Cognitive Development (ABCD), and UK Biobank (UKB) datasets, to predict sex, age, and cognitive intelligence. Our experimental outcomes reveal that SwiFT consistently outperforms recent state-of-the-art models. Furthermore, by leveraging its end-to-end learning capability, we show that contrastive loss-based self-supervised pre-training of SwiFT can enhance performance on downstream tasks. Additionally, we employ an explainable AI method to identify the brain regions associated with sex classification. To our knowledge, SwiFT is the first Swin Transformer architecture to process dimensional spatiotemporal brain functional data in an end-to-end fashion. Our work holds substantial potential in facilitating scalable learning of functional brain imaging in neuroscience research by reducing the hurdles associated with applying Transformer models to high-dimensional fMRI.
****
Consistent Diffusion Models: Mitigating Sampling Drift by Learning to be Consistent
Giannis Daras, Yuval Dagan, Alex Dimakis, Constantinos Daskalakis
Imperfect score-matching leads to a shift between the training and the sampling distribution of diffusion models. Due to the recursive nature of the generation process, errors in previous steps yield sampling iterates that drift away from the training distribution. However, the standard training objective via Denoising Score Matching (DSM) is only designed to optimize over non-drifted data. To train on drifted data, we propose to enforce a \emph{Consistency} property (CP) which states that predictions of the model on its owngenerated data are consistent across time. Theoretically, we show that the differential equation that describes CP together with the one that describes a conservative vector field, have a unique solution given some initial condition. Consequently, if the score is learned well on non-drifted points via DSM (enforcing the true initial condition) then enforcing CP on drifted points propagates true score values. Empirically, we show that enforcing CP improves the generation quality for conditional and unconditional generation on CIFAR-10, and in AFHQ and FFHQ. We open-source our code and models: https://github.com/giannisdaras/cdm.
****
A High-Resolution Dataset for Instance Detection with Multi-View Object Capture
QIANQIAN SHEN, Yunhan Zhao, Nahyun Kwon, Jeeeun Kim, Yanan Li, Shu Kong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
AVIDa-hIL6: A Large-Scale VHH Dataset Produced from an Immunized Alpaca for Predicting Antigen-Antibody Interactions
Hirofumi Tsuruta, Hiroyuki Yamazaki, Ryota Maeda, Ryotaro Tamura, Jennifer Wei, Zelda E. Mariet, Poomarin Phloyphisut, Hidetoshi Shimokawa, Joseph R. Ledsam, Lucy Colwell, Akihiro Imura

Antibodies have become an important class of therapeutic agents to treat human diseases.To accelerate therapeutic antibody discovery, computational methods, especially machine learning, have attracted considerable interest for predicting specific interactions between antibody candidates and target antigens such as viruses and bacteria.However, the publicly available datasets in existing works have notable limitations, such as small sizes and the lack of non-binding samples and exact amino acid sequences.To overcome these limitations, we have developed AVIDa-hIL6, a large-scale dataset for predicting antigen-antibody interactions in the variable domain of heavy chain of heavy chain antibodies (VHHs), produced from an alpaca immunized with the human interleukin-6 (IL-6) protein, as antigens. By leveraging the simple structure of VHHs, which facilitates identification of full-length amino acid sequences by DNA sequencing technology, AVIDa-hIL6 contains 573,891 antigen-VHH pairs with amino acid sequences.All the antigen-VHH pairs have reliable labels for binding or non-binding, as generated by a novel labeling method.Furthermore, via introduction of artificial mutations, AVIDa-hIL6 contains 30 different mutants in addition to wild-type IL-6 protein.This characteristic provides opportunities to develop machine learning models for predicting changes in antibody binding by antigen mutations.We report experimental benchmark results on AVIDa-hIL6 by using machine learning models.The results indicate that the existing models have potential, but further research is needed to generalize them to predict effective antibodies against unknown mutants.The dataset is available at https://avida-hil6.cognanous.com.
****
Token-Scaled Logit Distillation for Ternary Weight Generative Language Models
Minsoo Kim, Sihwa Lee, Janghwan Lee, Sukjin Hong, Du-Seong Chang, Wonyong Sung, Jungwook Choi
Generative Language Models (GLMs) have shown impressive performance in tasks such as text generation, understanding, and reasoning. However, the large model size poses challenges for practical deployment. To solve this problem, Quantization-Aware Training (QAT) has become increasingly popular. However, current QAT methods for generative models have resulted in a noticeable loss of accuracy. To counteract this issue, we propose a novel knowledge distillation method specifically designed for GLMs. Our method, called token-scaled logit distillation, prevents overfitting and provides superior learning from the teacher model and ground truth. This research marks the first evaluation of ternary weight quantization-aware training of large-scale GLMs with less than 1.0 degradation in perplexity and achieves enhanced accuracy in tasks like common-sense QA and arithmetic reasoning as well as natural language understanding. Our code is available at https://github.com/aiha-lab/TSLD.
****
Efficient Exploration in Continuous-time Model-based Reinforcement Learning
Lenart Treven, Jonas Hübotter, Bhavya , Florian Dorfler, Andreas Krause
Reinforcement learning algorithms typically consider discrete-time dynamics, even though the underlying systems are often continuous in time. In this paper, we introduce a model-based reinforcement learning algorithm that represents continuous-time dynamics using nonlinear ordinary differential equations (ODEs). We capture epistemic uncertainty using well-calibrated probabilistic models, and use the optimistic principle for exploration. Our regret bounds surface the importance of the measurement selection strategy (MSS), since in continuous time we not only must decide how to explore, but also when to observe the underlying system. Our analysis demonstrates that the regret is sublinear when modeling ODEs with Gaussian Processes (GP) for common choices of MSS, such as equidistant sampling. Additionally, we propose an adaptive, data-dependent, practical MSS that, when combined with GP dynamics, also achieves sublinear regret with significantly fewer samples. We showcase the benefits of continuous-time modeling over its discrete-time counterpart, as well as our proposed adaptive MSS over standard baselines, on several applications.
****
On Learning Necessary and Sufficient Causal Graphs
Hengrui Cai, Yixin Wang, Michael Jordan, Rui Song

The causal revolution has stimulated interest in understanding complex relationships in various fields. Most of the existing methods aim to discover causal relationships among all variables within a complex large-scale graph. However, in practice, only a small subset of variables in the graph are relevant to the outcomes of interest. Consequently, causal estimation with the full causal graph---particularly given limited data---could lead to numerous falsely discovered, spurious variables that exhibit high correlation with, but exert no causal impact on, the target outcome. In this paper, we propose learning a class of necessary and sufficient causal graphs (NSCG) that exclusively comprises causally relevant variables for an outcome of interest, which we term causal features. The key idea is to employ probabilities of causation to systematically evaluate the importance of features in the causal graph, allowing us to identify a subgraph relevant to the outcome of interest. To learn NSCG from data, we develop a necessary and sufficient causal structural learning (NSCSL) algorithm, by establishing theoretical properties and relationships between probabilities of causation and natural causal effects of features. Across empirical studies of simulated and real data, we demonstrate that NSCSL outperforms existing algorithms and can reveal crucial yeast genes for target heritable traits of interest.
****
Renku: a platform for sustainable data science
Rok Roškar, Chandrasekhar Ramakrishnan, Michele Volpi, Fernando Perez-Cruz, Lilian Gasser, Firat Ozdemir, Patrick Paitz, Mohammad Alisafaee, Philipp Fischer, Ralf Grubenmann, Eliza Harris, Tasko Olevski, Carl Remlinger, Luis Salamanca, Elisabet Capon Garcia, Lorenzo Cavazzi, Jakub Chrobasik, Darlin Cordoba Osnas, Alessandro Degano, Jimena Dupre, Wesley Johnson, Eike Kettner, Laura Kinkead, Sean D. Murphy, Flora Thiebaut, Olivier Verscheure
Data and code working together is fundamental to machine learning (ML), but the context around datasets and interactions between datasets and code are in general captured only rudimentarily. Context such as how the dataset was prepared and created, what source data were used, what code was used in processing, how the dataset evolved, and where it has been used and reused can provide much insight, but this information is often poorly documented. That is unfortunate since it makes datasets into black-boxes with potentially hidden characteristics that have downstream consequences. We argue that making dataset preparation more accessible and dataset usage easier to record and document would have significant benefits for the ML community: it would allow for greater diversity in datasets by inviting modification to published sources, simplify use of alternative datasets and, in doing so, make results more transparent and robust, while allowing for all contributions to be adequately credited. We present a platform, Renku, designed to support and encourage such sustainable development and use of data, datasets, and code, and we demonstrate its benefits through a few illustrative projects which span the spectrum from dataset creation to dataset consumption and showcasing.
****
Slimmed Asymmetrical Contrastive Learning and Cross Distillation for Lightweight Model Training
Jian Meng, Li Yang, Kyungmin Lee, Jinwoo Shin, Deliang Fan, Jae-sun Seo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Beyond Unimodal: Generalising Neural Processes for Multimodal Uncertainty Estimation
Myong Chol Jung, He Zhao, Joanna Dipnall, Lan Du
Uncertainty estimation is an important research area to make deep neural networks (DNNs) more trustworthy. While extensive research on uncertainty estimation has been conducted with unimodal data, uncertainty estimation for multimodal data remains a challenge. Neural processes (NPs) have been demonstrated to be an effective uncertainty estimation method for unimodal data by providing the reliabili

ty of Gaussian processes with efficient and powerful DNNs. While NPs hold significant potential for multimodal uncertainty estimation, the adaptation of NPs for multimodal data has not been carefully studied. To bridge this gap, we propose Multimodal Neural Processes (MNPs) by generalising NPs for multimodal uncertainty estimation. Based on the framework of NPs, MNPs consist of several novel and principled mechanisms tailored to the characteristics of multimodal data. In extensive empirical evaluation, our method achieves state-of-the-art multimodal uncertainty estimation performance, showing its appealing robustness against noisy samples and reliability in out-of-distribution detection with faster computation time compared to the current state-of-the-art multimodal uncertainty estimation method.
****

Trans-Dimensional Generative Modeling via Jump Diffusion Models

Andrew Campbell, William Harvey, Christian Weilbach, Valentin De Bortoli, Thomas Rainforth, Arnaud Doucet

We propose a new class of generative model that naturally handles data of varying dimensionality by jointly modeling the state and dimension of each datapoint. The generative process is formulated as a jump diffusion process that makes jumps between different dimensional spaces. We first define a dimension destroying forward noising process, before deriving the dimension creating time-reversed generative process along with a novel evidence lower bound training objective for learning to approximate it.Simulating our learned approximation to the time-reversed generative process then provides an effective way of sampling data of varying dimensionality by jointly generating state values and dimensions. We demonstrate our approach on molecular and video datasets of varying dimensionality, reporting better compatibility with test-time diffusion guidance imputation tasks and improved interpolation capabilities versus fixed dimensional models that generate state values and dimensions separately.
****

Plug-and-Play Stability for Intracortical Brain-Computer Interfaces: A One-Year Demonstration of Seamless Brain-to-Text Communication

Chaofei Fan, Nick Hahn, Foram Kamdar, Donald Avansino, Guy Wilson, Leigh Hochberg, Krishna V Shenoy, Jaimie Henderson, Francis Willett

Intracortical brain-computer interfaces (iBCIs) have shown promise for restoring rapid communication to people with neurological disorders such as amyotrophic lateral sclerosis (ALS). However, to maintain high performance over time, iBCIs typically need frequent recalibration to combat changes in the neural recordings that accrue over days. This requires iBCI users to stop using the iBCI and engage in supervised data collection, making the iBCI system hard to use. In this paper, we propose a method that enables self-recalibration of communication iBCIs without interrupting the user. Our method leverages large language models (LMs) to automatically correct errors in iBCI outputs. The self-recalibration process uses these corrected outputs ("pseudo-labels") to continually update the iBCI decoder online. Over a period of more than one year (403 days), we evaluated our Continual Online Recalibration with Pseudo-labels (CORP) framework with one clinical trial participant. CORP achieved a stable decoding accuracy of 93.84% in an online handwriting iBCI task, significantly outperforming other baseline methods. Notably, this is the longest-running iBCI stability demonstration involving a human participant. Our results provide the first evidence for long-term stabilization of a plug-and-play, high-performance communication iBCI, addressing a major barrier for the clinical translation of iBCIs.
****

Towards robust and generalizable representations of extracellular data using contrastive learning

Ankit Vishnubhotla, Charlotte Loh, Akash Srivastava, Liam Paninski, Cole Hurwitz

Contrastive learning is quickly becoming an essential tool in neuroscience for extracting robust and meaningful representations of neural activity. Despite numerous applications to neuronal population data, there has been little exploration of how these methods can be adapted to key primary data analysis tasks such as spike sorting or cell-type classification. In this work, we propose a novel cont

rastive learning framework, CEED (Contrastive Embeddings for Extracellular Data), for high-density extracellular recordings. We demonstrate that through careful design of the network architecture and data augmentations, it is possible to generically extract representations that far outperform current specialized approaches. We validate our method across multiple high-density extracellular recordings. All code used to run CEED can be found at https://github.com/ankitvishnu23/CEED.
****
Rethinking Conditional Diffusion Sampling with Progressive Guidance
Anh-Dung Dinh, Daochang Liu, Chang Xu
This paper tackles two critical challenges encountered in classifier guidance for diffusion generative models, i.e., the lack of diversity and the presence of adversarial effects. These issues often result in a scarcity of diverse samples or the generation of non-robust features. The underlying cause lies in the mechanism of classifier guidance, where discriminative gradients push samples to be recognized as conditions aggressively. This inadvertently suppresses information with common features among relevant classes, resulting in a limited pool of features with less diversity or the absence of robust features for image construction.We propose a generalized classifier guidance method called Progressive Guidance, which mitigates the problems by allowing relevant classes' gradients to contribute to shared information construction when the image is noisy in early sampling steps. In the later sampling stage, we progressively enhance gradients to refine the details in the image toward the primary condition. This helps to attain a high level of diversity and robustness compared to the vanilla classifier guidance. Experimental results demonstrate that our proposed method further improves the image quality while offering a significant level of diversity as well as robust features.
****
State-Action Similarity-Based Representations for Off-Policy Evaluation
Brahma Pavse, Josiah Hanna
In reinforcement learning, off-policy evaluation (OPE) is the problem of estimating the expected return of an evaluation policy given a fixed dataset that was collected by running one or more different policies. One of the more empirically successful algorithms for OPE has been the fitted q-evaluation (FQE) algorithm that uses temporal difference updates to learn an action-value function, which is then used to estimate the expected return of the evaluation policy. Typically, the original fixed dataset is fed directly into FQE to learn the action-value function of the evaluation policy. Instead, in this paper, we seek to enhance the data-efficiency of FQE by first transforming the fixed dataset using a learned encoder, and then feeding the transformed dataset into FQE. To learn such an encoder, we introduce an OPE-tailored state-action behavioral similarity metric, and use this metric and the fixed dataset to learn an encoder that models this metric. Theoretically, we show that this metric allows us to bound the error in the resulting OPE estimate. Empirically, we show that other state-action similarity metrics lead to representations that cannot represent the action-value function of the evaluation policy, and that our state-action representation method boosts the data-efficiency of FQE and lowers OPE error relative to other OPE-based representation learning methods on challenging OPE tasks. We also empirically show that the learned representations significantly mitigate divergence of FQE under varying distribution shifts. Our code is available here: https://github.com/Badger-RL/ROPE.
****
Can LLM Already Serve as A Database Interface? A BIg Bench for Large-Scale Database Grounded Text-to-SQLs
Jinyang Li, Binyuan Hui, Ge Qu, Jiaxi Yang, Binhua Li, Bowen Li, Bailin Wang, Bowen Qin, Ruiying Geng, Nan Huo, Xuanhe Zhou, Ma Chenhao, Guoliang Li, Kevin Chang, Fei Huang, Reynold Cheng, Yongbin Li
Text-to-SQL parsing, which aims at converting natural language instructions into executable SQLs, has gained increasing attention in recent years. In particular, GPT-4 and Claude-2 have shown impressive results in this task. However, most

of the prevalent benchmarks, i.e., Spider, and WikiSQL, focus on database schema with few rows of database contents leaving the gap between academic study and real-world applications. To mitigate this gap, we present BIRD, a BIg benchmark for laRge-scale Database grounded in text-to-SQL tasks, containing 12,751 pairs of text-to-SQL data and 95 databases with a total size of 33.4 GB, spanning 37 professional domains. Our emphasis on database values highlights the new challenges of dirty database contents, external knowledge between NL questions and database contents, and SQL efficiency, particularly in the context of massive databases. To solve these problems, text-to-SQL models must feature database value comprehension in addition to semantic parsing. The experimental results demonstrate the significance of database values in generating accurate text-to-SQLs for big databases. Furthermore, even the most popular and effective text-to-SQL models, i.e. GPT-4, only achieve 54.89% in execution accuracy, which is still far from the human result of 92.96%, proving that challenges still stand. We also provide an efficiency analysis to offer insights into generating text-to-efficient-SQLs that are beneficial to industries. We believe that BIRD will contribute to advancing real-world applications of text-to-SQL research.The leaderboard and source code are available: https://bird-bench.github.io/.
****
CARE-MI: Chinese Benchmark for Misinformation Evaluation in Maternity and Infant Care
Tong Xiang, Liangzhi Li, Wangyue Li, Mingbai Bai, Lu Wei, Bowen Wang, Noa Garcia
The recent advances in natural language processing (NLP), have led to a new trend of applying large language models (LLMs) to real-world scenarios. While the latest LLMs are astonishingly fluent when interacting with humans, they suffer from the misinformation problem by unintentionally generating factually false statements. This can lead to harmful consequences, especially when produced within sensitive contexts, such as healthcare. Yet few previous works have focused on evaluating misinformation in the long-form (LF) generation of LLMs, especially for knowledge-intensive topics. Moreover, although LLMs have been shown to perform well in different languages, misinformation evaluation has been mostly conducted in English. To this end, we present a benchmark, CARE-MI, for evaluating LLM misinformation in: 1) a sensitive topic, specifically the maternity and infant care domain; and 2) a language other than English, namely Chinese. Most importantly, we provide an innovative paradigm for building LF generation evaluation benchmarks that can be transferred to other knowledge-intensive domains and low-resourced languages. Our proposed benchmark fills the gap between the extensive usage of LLMs and the lack of datasets for assessing the misinformation generated by these models. It contains 1,612 expert-checked questions, accompanied with human-selected references. Using our benchmark, we conduct extensive experiments and found that current Chinese LLMs are far from perfect in the topic of maternity and infant care. In an effort to minimize the reliance on human resources for performance evaluation, we offer off-the-shelf judgment models for automatically assessing the LF output of LLMs given benchmark questions. Moreover, we compare potential solutions for LF generation evaluation and provide insights for building better automated metrics.
****
Explore In-Context Learning for 3D Point Cloud Understanding
Zhongbin Fang, Xiangtai Li, Xia Li, Joachim M Buhmann, Chen Change Loy, Mengyuan Liu
With the rise of large-scale models trained on broad data, in-context learning has become a new learning paradigm that has demonstrated significant potential in natural language processing and computer vision tasks. Meanwhile, in-context learning is still largely unexplored in the 3D point cloud domain. Although masked modeling has been successfully applied for in-context learning in 2D vision, directly extending it to 3D point clouds remains a formidable challenge. In the case of point clouds, the tokens themselves are the point cloud positions (coordinates) that are masked during inference. Moreover, position embedding in previous works may inadvertently introduce information leakage. To address these challenges, we introduce a novel framework, named Point-In-Context, designed especially

for in-context learning in 3D point clouds, where both inputs and outputs are modeled as coordinates for each task. Additionally, we propose the Joint Sampling module, carefully designed to work in tandem with the general point sampling operator, effectively resolving the aforementioned technical issues. We conduct extensive experiments to validate the versatility and adaptability of our proposed methods in handling a wide range of tasks. Furthermore, with a more effective prompt selection strategy, our framework surpasses the results of individually trained models.
****
Learning Re-sampling Methods with Parameter Attribution for Image Super-resolution
Xiaotong Luo, Yuan Xie, Yanyun Qu
Single image super-resolution (SISR) has made a significant breakthrough benefiting from the prevalent rise of deep neural networks and large-scale training samples. The mainstream deep SR models primarily focus on network architecture design as well as optimization schemes, while few pay attention to the training data. In fact, most of the existing SR methods train the model on uniformly sampled patch pairs from the whole image. However, the uneven image content makes the training data present an unbalanced distribution, i.e., the easily reconstructed region (smooth) occupies the majority of the data, while the hard reconstructed region (edge or texture) has rarely few samples. Based on this phenomenon, we consider rethinking the current paradigm of merely using uniform data sampling way for training SR models. In this paper, we propose a simple yet effective Bi-Sampling Parameter Attribution (BSPA) method for accurate image SR. Specifically, the bi-sampling consists of uniform sampling and inverse sampling, which is introduced to reconcile the unbalanced inherent data bias. The former aims to keep the intrinsic data distribution, and the latter is designed to enhance the feature extraction ability of the model on the hard samples. Moreover, integrated gradient is introduced to attribute the contribution of each parameter in the alternate models trained by both sampling data so as to filter the trivial parameters for further dynamic refinement. By progressively decoupling the allocation of parameters, the SR model can learn a more compact representation. Extensive experiments on publicly available datasets demonstrate that our proposal can effectively boost the performance of baseline methods from the data re-sampling view.
****
Interactive Visual Reasoning under Uncertainty
Manjie Xu, Guangyuan Jiang, Wei Liang, Chi Zhang, Yixin Zhu
One of the fundamental cognitive abilities of humans is to quickly resolve uncertainty by generating hypotheses and testing them via active trials. Encountering a novel phenomenon accompanied by ambiguous cause-effect relationships, humans make hypotheses against data, conduct inferences from observation, test their theory via experimentation, and correct the proposition if inconsistency arises. These iterative processes persist until the underlying mechanism becomes clear. In this work, we devise the IVRE (pronounced as "ivory") environment for evaluating artificial agents' reasoning ability under uncertainty. IVRE is an interactive environment featuring rich scenarios centered around Blicket detection. Agents in IVRE are placed into environments with various ambiguous action-effect pairs and asked to determine each object's role. They are encouraged to propose effective and efficient experiments to validate their hypotheses based on observations and actively gather new information. The game ends when all uncertainties are resolved or the maximum number of trials is consumed. By evaluating modern artificial agents in IVRE, we notice a clear failure of today's learning methods compared to humans. Such inefficacy in interactive reasoning ability under uncertainty calls for future research in building human-like intelligence.
****
Generative Modeling through the Semi-dual Formulation of Unbalanced Optimal Transport
Jaemoo Choi, Jaewoong Choi, Myungjoo Kang
Optimal Transport (OT) problem investigates a transport map that bridges two distributions while minimizing a given cost function. In this regard, OT between tr

actable prior distribution and data has been utilized for generative modeling tasks. However, OT-based methods are susceptible to outliers and face optimization challenges during training. In this paper, we propose a novel generative model based on the semi-dual formulation of Unbalanced Optimal Transport (UOT). Unlike OT, UOT relaxes the hard constraint on distribution matching. This approach provides better robustness against outliers, stability during training, and faster convergence. We validate these properties empirically through experiments. Moreover, we study the theoretical upper-bound of divergence between distributions in UOT. Our model outperforms existing OT-based generative models, achieving FID scores of 2.97 on CIFAR-10 and 6.36 on CelebA-HQ-256. The code is available at \url{https://github.com/Jae-Moo/UOTM}.
****
Generalized Belief Transport
Junqi Wang, PEI WANG, Patrick Shafto
Human learners have ability to adopt appropriate learning approaches depending on constraints such as prior on the hypothesis, urgency of decision, and drift of the environment. However, existing learning models are typically considered individually rather than in relation to one and other. To build agents that have the ability to move between different modes of learning over time, it is important to understand how learning models are related as points in a broader space of possibilities. We introduce a mathematical framework, Generalized Belief Transport (GBT), that unifies and generalizes prior models, including Bayesian inference, cooperative communication and classification, as parameterizations of three learning constraints within Unbalanced Optimal Transport (UOT). We visualize the space of learning models encoded by GBT as a cube which includes classic learning models as special points. We derive critical properties of this parameterized space including proving continuity and differentiability which is the basis for model interpolation, and study limiting behavior of the parameters, which allows attaching learning models on the boundaries. Moreover, we investigate the long-run behavior of GBT, explore convergence properties of models in GBT mathematical and computationally, document the ability to learn in the presence of distribution drift, and formulate conjectures about general behavior. We conclude with open questions and implications for more unified models of learning.
****
Lie Point Symmetry and Physics-Informed Networks
Tara Akhound-Sadegh, Laurence Perreault-Levasseur, Johannes Brandstetter, Max Welling, Siamak Ravanbakhsh
Symmetries have been leveraged to improve the generalization of neural networks through different mechanisms from data augmentation to equivariant architectures. However, despite their potential, their integration into neural solvers for partial differential equations (PDEs) remains largely unexplored. We explore the integration of PDE symmetries, known as Lie point symmetries, in a major family of neural solvers known as physics-informed neural networks (PINNs). We propose a loss function that informs the network about Lie point symmetries in the same way that PINN models try to enforce the underlying PDE through a loss function. Intuitively, our symmetry loss ensures that the infinitesimal generators of the Lie group conserve the PDE solutions.. Effectively, this means that once the network learns a solution, it also learns the neighbouring solutions generated by Lie point symmetries.Empirical evaluations indicate that the inductive bias introduced by the Lie point symmetries of the PDEs greatly boosts the sample efficiency of PINNs.
****
Norm-based Generalization Bounds for Sparse Neural Networks
Tomer Galanti, Mengjia Xu, Liane Galanti, Tomaso Poggio
In this paper, we derive norm-based generalization bounds for sparse ReLU neural networks, including convolutional neural networks. These bounds differ from previous ones because they consider the sparse structure of the neural network architecture and the norms of the convolutional filters, rather than the norms of the (Toeplitz) matrices associated with the convolutional layers. Theoretically, we demonstrate that these bounds are significantly tighter than standard norm-bas

ed generalization bounds. Empirically, they offer relatively tight estimations o
f generalization for various simple classification problems. Collectively, these
 findings suggest that the sparsity of the underlying target function and the mo
del's architecture plays a crucial role in the success of deep learning.
****
Intelligent Knee Sleeves: A Real-time Multimodal Dataset for 3D Lower Body Motio
n Estimation Using Smart Textile
Wenwen Zhang, Arvin Tashakori, Zenan Jiang, Amir Servati, Harishkumar Narayana,
Saeid Soltanian, Rou Yi Yeap, Menghan Ma, Lauren Toy, Peyman Servati
The kinematics of human movements and locomotion are closely linked to the activ
ation and contractions of muscles. To investigate this, we present a multimodal
dataset with benchmarks collected using a novel pair of Intelligent Knee Sleeves
 (Texavie MarsWear Knee Sleeves) for human pose estimation. Our system utilizes
synchronized datasets that comprise time-series data from the Knee Sleeves and t
he corresponding ground truth labels from visualized motion capture camera syste
m. We employ these to generate 3D human models solely based on the wearable data
 of individuals performing different activities. We demonstrate the effectivenes
s of this camera-free system and machine learning algorithms in the assessment o
f various movements and exercises, including extension to unseen exercises and i
ndividuals. The results show an average error of 7.21 degrees across all eight l
ower body joints when compared to the ground truth, indicating the effectiveness
 and reliability of the Knee Sleeve system for the prediction of different lower
 body joints beyond knees. The results enable human pose estimation in a seamles
s manner without being limited by visual occlusion or the field of view of camer
as. Our results show the potential of multimodal wearable sensing in a variety o
f applications from home fitness to sports, healthcare, and physical rehabilitat
ion focusing on pose and movement estimation.
****
Neural Injective Functions for Multisets, Measures and Graphs via a Finite Witne
ss Theorem
Tal Amir, Steven Gortler, Ilai Avni, Ravina Ravina, Nadav Dym
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Online Ad Procurement in Non-stationary Autobidding Worlds
Jason Cheuk Nam Liang, Haihao Lu, Baoyu Zhou
Today's online advertisers procure digital ad impressions through interacting wi
th autobidding platforms: advertisers convey high level procurement goals via se
tting levers such as budget, target return-on-investment, max cost per click, et
c.. Then ads platforms subsequently procure impressions on advertisers' behalf,
and report final procurement conversions (e.g. click) to advertisers. In practic
e, advertisers may receive minimal information on platforms' procurement details
, and procurement outcomes are subject to non-stationary factors like seasonal
patterns, occasional system corruptions, and market trends which make it difficu
lt for advertisers to optimize lever decisions effectively. Motivated by this,
we present an online learning framework that helps advertisers dynamically optim
ize ad platform lever decisions while subject to general long-term constraints i
n a realistic bandit feedback environment with non-stationary procurement outcom
es. In particular, we introduce a primal-dual algorithm for online decision maki
ng with multi-dimension decision variables, bandit feedback and long-term uncert
ain constraints. We show that our algorithm achieves low regret in many worlds w
hen procurement outcomes are generated through procedures that are stochastic, a
dversarial, adversarially corrupted, periodic, and ergodic, respectively, withou
t having to know which procedure is the ground truth. Finally, we emphasize that
 our proposed algorithm and theoretical results extend beyond the applications o
f online advertising.
****
Precise asymptotic generalization for multiclass classification with overparamet

erized linear models
David Wu, Anant Sahai
We study the asymptotic generalization of an overparameterized linear model for multiclass classification under the Gaussian covariates bi-level model introduced in Subramanian et al. (NeurIPS'22), where the number of data points, features, and classes all grow together. We fully resolve the conjecture posed in Subramanian et al. '22, matching the predicted regimes for which the model does and does not generalize. Furthermore, our new lower bounds are akin to an information-theoretic strong converse: they establish that the misclassification rate goes to 0 or 1 asymptotically. One surprising consequence of our tight results is that the min-norm interpolating classifier can be asymptotically suboptimal relative to noninterpolating classifiers in the regime where the min-norm interpolating regressor is known to be optimal. The key to our tight analysis is a new variant of the Hanson-Wright inequality which is broadly useful for multiclass problems with sparse labels. As an application, we show that the same type of analysis can be used to analyze the related multi-label classification problem under the same bi-level ensemble.
****
Break It Down: Evidence for Structural Compositionality in Neural Networks
Michael Lepori, Thomas Serre, Ellie Pavlick
Though modern neural networks have achieved impressive performance in both vision and language tasks, we know little about the functions that they implement. One possibility is that neural networks implicitly break down complex tasks into subroutines, implement modular solutions to these subroutines, and compose them into an overall solution to a task --- a property we term structural compositionality. Another possibility is that they may simply learn to match new inputs to learned templates, eliding task decomposition entirely. Here, we leverage model pruning techniques to investigate this question in both vision and language across a variety of architectures, tasks, and pretraining regimens. Our results demonstrate that models oftentimes implement solutions to subroutines via modular subnetworks, which can be ablated while maintaining the functionality of other subnetworks. This suggests that neural networks may be able to learn compositionality, obviating the need for specialized symbolic mechanisms.
****
Focused Transformer: Contrastive Training for Context Scaling
Szymon Tworkowski, Konrad Staniszewski, Miko■aj Pacek, Yuhuai Wu, Henryk Michalewski, Piotr Mi■o■
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Res-Tuning: A Flexible and Efficient Tuning Paradigm via Unbinding Tuner from Backbone
Zeyinzi Jiang, Chaojie Mao, Ziyuan Huang, Ao Ma, Yiliang Lv, Yujun Shen, Deli Zhao, Jingren Zhou
Parameter-efficient tuning has become a trend in transferring large-scale foundation models to downstream applications. Existing methods typically embed some light-weight tuners into the backbone, where both the design and the learning of the tuners are highly dependent on the base model. This work offers a new tuning paradigm, dubbed Res-Tuning, which intentionally unbinds tuners from the backbone. With both theoretical and empirical evidence, we show that popular tuning approaches have their equivalent counterparts under our unbinding formulation, and hence can be integrated into our framework effortlessly. Thanks to the structural disentanglement, we manage to free the design of tuners from the network architecture, facilitating flexible combination of various tuning strategies. We further propose a memory-efficient variant of Res-Tuning, where the bypass i.e., formed by a sequence of tuners) is effectively detached from the main branch, such that the gradients are back-propagated only to the tuners but not to the backbone. Such a detachment also allows one-time backbone forward for multi-task infere

nce. Extensive experiments on both discriminative and generative tasks demonstrate the superiority of our method over existing alternatives from the perspectives of efficacy and efficiency. Project page: https://res-tuning.github.io/.
****
Learning Energy-Based Prior Model with Diffusion-Amortized MCMC
Peiyu Yu, Yaxuan Zhu, Sirui Xie, Xiaojian (Shawn) Ma, Ruiqi Gao, Song-Chun Zhu, Ying Nian Wu
Latent space EBMs, also known as energy-based priors, have drawn growing interests in the field of generative modeling due to its flexibility in the formulation and strong modeling power of the latent space. However, the common practice of learning latent space EBMs with non-convergent short-run MCMC for prior and posterior sampling is hindering the model from further progress; the degenerate MCMC sampling quality in practice often leads to degraded generation quality and instability in training, especially with highly multi-modal and/or high-dimensional target distributions. To remedy this sampling issue, in this paper we introduce a simple but effective diffusion-based amortization method for long-run MCMC sampling and develop a novel learning algorithm for the latent space EBM based on it. We provide theoretical evidence that the learned amortization of MCMC is a valid long-run MCMC sampler. Experiments on several image modeling benchmark datasets demonstrate the superior performance of our method compared with strong counterparts.
****
Perception Test: A Diagnostic Benchmark for Multimodal Video Models
Viorica Patraucean, Lucas Smaira, Ankush Gupta, Adria Recasens, Larisa Markeeva, Dylan Banarse, Skanda Koppula, joseph heyward, Mateusz Malinowski, Yi Yang, Carl Doersch, Tatiana Matejovicova, Yury Sulsky, Antoine Miech, Alexandre Fréchette, Hanna Klimczak, Raphael Koster, Junlin Zhang, Stephanie Winkler, Yusuf Aytar, Simon Osindero, Dima Damen, Andrew Zisserman, Joao Carreira
We propose a novel multimodal video benchmark - the Perception Test - to evaluate the perception and reasoning skills of pre-trained multimodal models (e.g. Flamingo, BEiT-3, or GPT-4). Compared to existing benchmarks that focus on computational tasks (e.g. classification, detection or tracking), the Perception Test focuses on skills (Memory, Abstraction, Physics, Semantics) and types of reasoning (descriptive, explanatory, predictive, counterfactual) across video, audio, and text modalities, to provide a comprehensive and efficient evaluation tool. The benchmark probes pre-trained models for their transfer capabilities, in a zero-shot / few-shot or limited finetuning regime. For these purposes, the Perception Test introduces 11.6k real-world videos, 23s average length, designed to show perceptually interesting situations, filmed by around 100 participants worldwide. The videos are densely annotated with six types of labels (multiple-choice and grounded video question-answers, object and point tracks, temporal action and sound segments), enabling both language and non-language evaluations. The fine-tuning and validation splits of the benchmark are publicly available (CC-BY license), in addition to a challenge server with a held-out test split. Human baseline results compared to state-of-the-art video QA models show a significant gap in performance (91.4% vs 45.8%), suggesting that there is significant room for improvement in multimodal video understanding.Dataset, baselines code, and challenge server are available at https://github.com/deepmind/perception_test
****
Directed Cyclic Graph for Causal Discovery from Multivariate Functional Data
Saptarshi Roy, Raymond K. W. Wong, Yang Ni
Discovering causal relationship using multivariate functional data has received a significant amount of attention very recently. In this article, we introduce a functional linear structural equation model for causal structure learning when the underlying graph involving the multivariate functions may have cycles. To enhance interpretability, our model involves a low-dimensional causal embedded space such that all the relevant causal information in the multivariate functional data is preserved in this lower-dimensional subspace. We prove that the proposed model is causally identifiable under standard assumptions that are often made in the causal discovery literature. To carry out inference of our model, we devel

op a fully Bayesian framework with suitable prior specifications and uncertainty quantification through posterior summaries. We illustrate the superior performance of our method over existing methods in terms of causal graph estimation through extensive simulation studies. We also demonstrate the proposed method using a brain EEG dataset.
****
Do SSL Models Have Déjà Vu? A Case of Unintended Memorization in Self-supervised Learning
Casey Meehan, Florian Bordes, Pascal Vincent, Kamalika Chaudhuri, Chuan Guo
Self-supervised learning (SSL) algorithms can produce useful image representations by learning to associate different parts of natural images with one another. However, when taken to the extreme, SSL models can unintendedly memorize specific parts in individual training samples rather than learning semantically meaningful associations. In this work, we perform a systematic study of the unintended memorization of image-specific information in SSL models -- which we refer to as déjà vu memorization. Concretely, we show that given the trained model and a crop of a training image containing only the background (e.g., water, sky, grass), it is possible to infer the foreground object with high accuracy or even visually reconstruct it. Furthermore, we show that déjà vu memorization is common to different SSL algorithms, is exacerbated by certain design choices, and cannot be detected by conventional techniques for evaluating representation quality. Our study of déjà vu memorization reveals previously unknown privacy risks in SSL models, as well as suggests potential practical mitigation strategies.
****
Differentiable and Stable Long-Range Tracking of Multiple Posterior Modes
Ali Younis, Erik Sudderth
Particle filters flexibly represent multiple posterior modes nonparametrically, via a collection of weighted samples, but have classically been applied to tracking problems with known dynamics and observation likelihoods. Such generative models may be inaccurate or unavailable for high-dimensional observations like images. We instead leverage training data to discriminatively learn particle-based representations of uncertainty in latent object states, conditioned on arbitrary observations via deep neural network encoders.   While prior discriminative particle filters have used heuristic relaxations of discrete particle resampling, or biased learning by truncating gradients at resampling steps, we achieve unbiased and low-variance gradient estimates by representing posteriors as continuous mixture densities. Our theory and experiments expose dramatic failures of existing reparameterization-based estimators for mixture gradients, an issue we address via an importance-sampling gradient estimator. Unlike standard recurrent neural networks, our mixture density particle filter represents multimodal uncertainty in continuous latent states, improving accuracy and robustness. On a range of challenging tracking and robot localization problems, our approach achieves dramatic improvements in accuracy, will also showing much greater stability across multiple training runs.
****
Benchmarking Robustness to Adversarial Image Obfuscations
Florian Stimberg, Ayan Chakrabarti, Chun-Ta Lu, Hussein Hazimeh, Otilia Stretcu, Wei Qiao, Yintao Liu, Merve Kaya, Cyrus Rashtchian, Ariel Fuxman, Mehmet Tek, Sven Gowal
Automated content filtering and moderation is an important tool that allows online platforms to build striving user communities that facilitate cooperation and prevent abuse. Unfortunately, resourceful actors try to bypass automated filters in a bid to post content that violate platform policies and codes of conduct. To reach this goal, these malicious actors may obfuscate policy violating images (e.g., overlay harmful images by carefully selected benign images or visual patterns) to prevent machine learning models from reaching the correct decision. In this paper, we invite researchers to tackle this specific issue and present a new image benchmark. This benchmark, based on ImageNet, simulates the type of obfuscations created by malicious actors. It goes beyond Image-Net-C and ImageNet-C-bar by proposing general, drastic, adversarial modifications that preserve the o

riginal content intent. It aims to tackle a more common adversarial threat than the one considered by lp-norm bounded adversaries. We evaluate 33 pretrained models on the benchmark and train models with different augmentations, architectures and training methods on subsets of the obfuscations to measure generalization. Our hope is that this benchmark will encourage researchers to test their models and methods and try to find new approaches that are more robust to these obfuscations.
****
Learning Curves for Deep Structured Gaussian Feature Models
Jacob Zavatone-Veth, Cengiz Pehlevan
In recent years, significant attention in deep learning theory has been devoted to analyzing when models that interpolate their training data can still generalize well to unseen examples. Many insights have been gained from studying models with multiple layers of Gaussian random features, for which one can compute precise generalization asymptotics. However, few works have considered the effect of weight anisotropy; most assume that the random features are generated using independent and identically distributed Gaussian weights, and allow only for structure in the input data. Here, we use the replica trick from statistical physics to derive learning curves for models with many layers of structured Gaussian features. We show that allowing correlations between the rows of the first layer of features can aid generalization, while structure in later layers is generally detrimental. Our results shed light on how weight structure affects generalization in a simple class of solvable models.
****
Mirror Diffusion Models for Constrained and Watermarked Generation
Guan-Horng Liu, Tianrong Chen, Evangelos Theodorou, Molei Tao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Training Transitive and Commutative Multimodal Transformers with LoReTTa
Manuel Tran, Yashin Dicente Cid, Amal Lahiani, Fabian Theis, Tingying Peng, Eldad Klaiman
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SaVeNet: A Scalable Vector Network for Enhanced Molecular Representation Learning
Sarp Aykent, Tian Xia
Geometric representation learning of molecules is challenging yet essential for applications in multiple domains. Despite the impressive breakthroughs made by geometric deep learning in various molecular representation learning tasks, effectively capturing complicated geometric features across spatial dimensions is still underexplored due to the significant difficulties in modeling efficient geometric representations and learning the inherent correlation in 3D structural modeling. These include computational inefficiency, underutilization of vectorial embeddings, and limited generalizability to integrate various geometric properties. To address the raised concerns, we introduce an efficient and effective framework, Scalable Vector Network (SaVeNet), designed to accommodate a range of geometric requirements without depending on costly embeddings. In addition, the proposed framework scales effectively with introduced direction noise. Theoretically, we analyze the desired properties (i.e., invariance and equivariant) and framework efficiency of the SaVeNet. Empirically, we conduct a comprehensive series of experiments to evaluate the efficiency and expressiveness of the proposed model. Our efficiency-focused experiments underscore the model's empirical superiority over existing methods. Experimental results on synthetic and real-world datasets demonstrate the expressiveness of our model, which achieves state-of-the-art

performance across various tasks within molecular representation learning.
****
Beyond Pretrained Features: Noisy Image Modeling Provides Adversarial Defense
Zunzhi You, Daochang Liu, Bohyung Han, Chang Xu
Recent advancements in masked image modeling (MIM) have made it a prevailing framework for self-supervised visual representation learning. The MIM pretrained models, like most deep neural network methods, remain vulnerable to adversarial attacks, limiting their practical application, and this issue has received little research attention. In this paper, we investigate how this powerful self-supervised learning paradigm can provide adversarial robustness to downstream classifiers. During the exploration, we find that noisy image modeling (NIM), a simple variant of MIM that adopts denoising as the pre-text task, reconstructs noisy images surprisingly well despite severe corruption. Motivated by this observation, we propose an adversarial defense method, referred to as De^3, by exploiting the pretrained decoder for denoising. Through De^3, NIM is able to enhance adversarial robustness beyond providing pretrained features. Furthermore, we incorporate a simple modification, sampling the noise scale hyperparameter from random distributions, and enable the defense to achieve a better and tunable trade-off between accuracy and robustness. Experimental results demonstrate that, in terms of adversarial robustness, NIM is superior to MIM thanks to its effective denoising capability. Moreover, the defense provided by NIM achieves performance on par with adversarial training while offering the extra tunability advantage. Source code and models are available at https://github.com/youzunzhi/NIM-AdvDef.
****
UniControl: A Unified Diffusion Model for Controllable Visual Generation In the Wild
Can Qin, Shu Zhang, Ning Yu, Yihao Feng, Xinyi Yang, Yingbo Zhou, Huan Wang, Juan Carlos Niebles, Caiming Xiong, Silvio Savarese, Stefano Ermon, Yun Fu, Ran Xu
Achieving machine autonomy and human control often represent divergent objectives in the design of interactive AI systems. Visual generative foundation models such as Stable Diffusion show promise in navigating these goals, especially when prompted with arbitrary languages. However, they often fall short in generating images with spatial, structural, or geometric controls. The integration of such controls, which can accommodate various visual conditions in a single unified model, remains an unaddressed challenge. In response, we introduce UniControl, a new generative foundation model that consolidates a wide array of controllable condition-to-image (C2I) tasks within a singular framework, while still allowing for arbitrary language prompts. UniControl enables pixel-level-precise image generation, where visual conditions primarily influence the generated structures and language prompts guide the style and context. To equip UniControl with the capacity to handle diverse visual conditions, we augment pretrained text-to-image diffusion models and introduce a task-aware HyperNet to modulate the diffusion models, enabling the adaptation to different C2I tasks simultaneously. Trained on nine unique C2I tasks, UniControl demonstrates impressive zero-shot generation abilities with unseen visual conditions. Experimental results show that UniControl often surpasses the performance of single-task-controlled methods of comparable model sizes. This control versatility positions UniControl as a significant advancement in the realm of controllable visual generation.
****
Unlocking Deterministic Robustness Certification on ImageNet
Kai Hu, Andy Zou, Zifan Wang, Klas Leino, Matt Fredrikson
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Bayesian Optimisation of Functions on Graphs
Xingchen Wan, Pierre Osselin, Henry Kenlay, Binxin Ru, Michael A Osborne, Xiaowen Dong
The increasing availability of graph-structured data motivates the task of optim

ising over functions defined on the node set of graphs. Traditional graph search algorithms can be applied in this case, but they may be sample-inefficient and do not make use of information about the function values; on the other hand, Bayesian optimisation is a class of promising black-box solvers with superior sample efficiency, but it has scarcely been applied to such novel setups. To fill this gap, we propose a novel Bayesian optimisation framework that optimises over functions defined on generic, large-scale and potentially unknown graphs. Through the learning of suitable kernels on graphs, our framework has the advantage of adapting to the behaviour of the target function. The local modelling approach further guarantees the efficiency of our method. Extensive experiments on both synthetic and real-world graphs demonstrate the effectiveness of the proposed optimisation framework.
****
RIO: A Benchmark for Reasoning Intention-Oriented Objects in Open Environments
Mengxue Qu, Yu Wu, Wu Liu, Xiaodan Liang, Jingkuan Song, Yao Zhao, Yunchao Wei
Intention-oriented object detection aims to detect desired objects based on specific intentions or requirements. For instance, when we desire to "lie down and rest", we instinctively seek out a suitable option such as a "bed" or a "sofa" that can fulfill our needs. Previous work in this area is limited either by the number of intention descriptions or by the affordance vocabulary available for intention objects. These limitations make it challenging to handle intentions in open environments effectively. To facilitate this research, we construct a comprehensive dataset called Reasoning Intention-Oriented Objects (RIO). In particular, RIO is specifically designed to incorporate diverse real-world scenarios and a wide range of object categories. It offers the following key features: 1) intention descriptions in RIO are represented as natural sentences rather than a mere word or verb phrase, making them more practical and meaningful; 2) the intention descriptions are contextually relevant to the scene, enabling a broader range of potential functionalities associated with the objects; 3) the dataset comprises a total of 40,214 images and 130,585 intention-object pairs. With the proposed RIO, we evaluate the ability of some existing models to reason intention-oriented objects in open environments.
****
Supervised Pretraining Can Learn In-Context Reinforcement Learning
Jonathan Lee, Annie Xie, Aldo Pacchiano, Yash Chandak, Chelsea Finn, Ofir Nachum, Emma Brunskill
Large transformer models trained on diverse datasets have shown a remarkable ability to learn in-context, achieving high few-shot performance on tasks they were not explicitly trained to solve. In this paper, we study the in-context learning capabilities of transformers in decision-making problems, i.e., reinforcement learning (RL) for bandits and Markov decision processes. To do so, we introduce and study the Decision-Pretrained Transformer (DPT), a supervised pretraining method where a transformer predicts an optimal action given a query state and an in-context dataset of interactions from a diverse set of tasks. While simple, this procedure produces a model with several surprising capabilities. We find that the trained transformer can solve a range of RL problems in-context, exhibiting both exploration online and conservatism offline, despite not being explicitly trained to do so. The model also generalizes beyond the pretraining distribution to new tasks and automatically adapts its decision-making strategies to unknown structure. Theoretically, we show DPT can be viewed as an efficient implementation of Bayesian posterior sampling, a provably sample-efficient RL algorithm. We further leverage this connection to provide guarantees on the regret of the in-context algorithm yielded by DPT, and prove that it can learn faster than algorithms used to generate the pretraining data. These results suggest a promising yet simple path towards instilling strong in-context decision-making abilities in transformers.
****
L2T-DLN: Learning to Teach with Dynamic Loss Network
Zhaoyang Hai, Liyuan Pan, Xiabi Liu, Zhengzheng Liu, Mirna Yunita
With the concept of teaching being introduced to the machine learning community,

a teacher model start using dynamic loss functions to teach the training of a student model. The dynamic intends to set adaptive loss functions to different phases of student model learning. In existing works, the teacher model 1) merely determines the loss function based on the present states of the student model, e.g., disregards the experience of the teacher; 2) only utilizes the states of the student model, e.g., training iteration number and loss/accuracy from training/validation sets, while ignoring the states of the loss function. In this paper, we first formulate the loss adjustment as a temporal task by designing a teacher model with memory units, and, therefore, enables the student learning to be guided by the experience of the teacher model. Then, with a Dynamic Loss Network, we can additionally use the states of the loss to assist the teacher learning in enhancing the interactions between the teacher and the student model. Extensive experiments demonstrate our approach can enhance student learning and improve the performance of various deep models on real-world tasks, including classification, objective detection, and semantic segmentation scenario.
****
Structured Federated Learning through Clustered Additive Modeling
Jie Ma, Tianyi Zhou, Guodong Long, Jing Jiang, Chengqi Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
An Inductive Bias for Tabular Deep Learning
Ege Beyazit, Jonathan Kozaczuk, Bo Li, Vanessa Wallace, Bilal Fadlallah
Deep learning methods have achieved state-of-the-art performance in most modeling tasks involving images, text and audio, however, they typically underperform tree-based methods on tabular data. In this paper, we hypothesize that a significant contributor to this performance gap is the interaction between irregular target functions resulting from the heterogeneous nature of tabular feature spaces, and the well-known tendency of neural networks to learn smooth functions. Utilizing tools from spectral analysis, we show that functions described by tabular datasets often have high irregularity, and that they can be smoothed by transformations such as scaling and ranking in order to improve performance. However, because these transformations tend to lose information or negatively impact the loss landscape during optimization, they need to be rigorously fine-tuned for each feature to achieve performance gains. To address these problems, we propose introducing frequency reduction as an inductive bias. We realize this bias as a neural network layer that promotes learning low-frequency representations of the input features, allowing the network to operate in a space where the target function is more regular. Our proposed method introduces less computational complexity than a fully connected layer, while significantly improving neural network performance, and speeding up its convergence on 14 tabular datasets.
****
Fairness-guided Few-shot Prompting for Large Language Models
Huan Ma, Changqing Zhang, Yatao Bian, Lemao Liu, Zhirui Zhang, Peilin Zhao, Shu Zhang, Huazhu Fu, Qinghua Hu, Bingzhe Wu
Large language models have demonstrated surprising ability to perform in-context learning, i.e., these models can be directly applied to solve numerous downstream tasks by conditioning on a prompt constructed by a few input-output examples. However, prior research has shown that in-context learning can suffer from high instability due to variations in training examples, example order, and prompt formats. Therefore, the construction of an appropriate prompt is essential for improving the performance of in-context learning. In this paper, we revisit this problem from the view of predictive bias. Specifically, we introduce a metric to evaluate the predictive bias of a fixed prompt against labels or a given attributes. Then we empirically show that prompts with higher bias always lead to unsatisfactory predictive quality. Based on this observation, we propose a novel search strategy based on the greedy search to identify the near-optimal prompt for improving the performance of in-context learning. We perform comprehensive exper

iments with state-of-the-art mainstream models such as GPT-3 on various downstre
am tasks. Our results indicate that our method can enhance the model's in-contex
t learning performance in an effective and interpretable manner.
****
Efficient Robust Bayesian Optimization for Arbitrary Uncertain inputs
Lin Yang, Junlong Lyu, Wenlong Lyu, Zhitang Chen
Bayesian Optimization (BO) is a sample-efficient optimization algorithm widely e
mployed across various applications. In some challenging BO tasks, input uncerta
inty arises due to the inevitable randomness in the optimization process, such a
s machining errors, execution noise, or contextual variability. This uncertainty
 deviates the input from the intended value before evaluation, resulting in sign
ificant performance fluctuations in the final result. In this paper, we introduc
e a novel robust Bayesian Optimization algorithm, AIRBO, which can effectively i
dentify a robust optimum that performs consistently well under arbitrary input u
ncertainty. Our method directly models the uncertain inputs of arbitrary distrib
utions by empowering the Gaussian Process with the Maximum Mean Discrepancy (MMD
) and further accelerates the posterior inference via Nystrom approximation. Rig
orous theoretical regret bound is established under MMD estimation error and ext
ensive experiments on synthetic functions and real problems demonstrate that our
 approach can handle various input uncertainties and achieve a state-of-the-art
performance.
****
HyenaDNA: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution
Eric Nguyen, Michael Poli, Marjan Faizi, Armin Thomas, Michael Wornow, Callum Bi
rch-Sykes, Stefano Massaroli, Aman Patel, Clayton Rabideau, Yoshua Bengio, Stefa
no Ermon, Christopher Ré, Stephen Baccus
Genomic (DNA) sequences encode an enormous amount of information for gene regula
tion and protein synthesis. Similar to natural language models, researchers have
 proposed foundation models in genomics to learn generalizable features from unl
abeled genome data that can then be fine-tuned for downstream tasks such as iden
tifying regulatory elements. Due to the quadratic scaling of attention, previous
 Transformer-based genomic models have used 512 to 4k tokens as context (<0.001%
 of the human genome), significantly limiting the modeling of long-range interac
tions in DNA. In addition, these methods rely on tokenizers or fixed k-mers to a
ggregate meaningful DNA units, losing single nucleotide resolution (i.e. DNA "ch
aracters") where subtle genetic variations can completely alter protein function
 via single nucleotide polymorphisms (SNPs). Recently, Hyena, a large language m
odel based on implicit convolutions was shown to match attention in quality whil
e allowing longer context lengths and lower time complexity. Leveraging Hyena’s
new long-range capabilities, we present HyenaDNA, a genomic foundation model pre
trained on the human reference genome with context lengths of up to 1 million to
kens at the single nucleotide-level – an up to 500x increase over previous dense
 attention-based models. HyenaDNA scales sub-quadratically in sequence length (t
raining up to 160x faster than Transformer), uses single nucleotide tokens, and
has full global context at each layer. We explore what longer context enables -
including the first use of in-context learning in genomics for simple adaptation
 to novel tasks without updating pretrained model weights. On fine-tuned benchma
rks from the Nucleotide Transformer, HyenaDNA reaches state-of-the-art (SotA) on
 12 of 18 datasets using a model with orders of magnitude less parameters and pr
etraining data.1 On the GenomicBenchmarks, HyenaDNA surpasses SotA on 7 of 8 dat
asets on average by +10 accuracy points. Code at https://github.com/HazyResearch
/hyena-dna.
****
Learning a 1-layer conditional generative model in total variation
Ajil Jalal, Justin Kang, Ananya Uppal, Kannan Ramchandran, Eric Price
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****

Model Shapley: Equitable Model Valuation with Black-box Access
Xinyi Xu, Thanh Lam, Chuan Sheng Foo, Bryan Kian Hsiang Low
Valuation methods of data and machine learning (ML) models are essential to the establishment of AI marketplaces. Importantly, certain practical considerations (e.g., operational constraints, legal restrictions) favor the use of model valuation over data valuation. Also, existing marketplaces that involve trading of pre-trained ML models call for an equitable model valuation method to price them. In particular, we investigate the black-box access setting which allows querying a model (to observe predictions) without disclosing model-specific information (e.g., architecture and parameters). By exploiting a Dirichlet abstraction of a model's predictions, we propose a novel and equitable model valuation method called model Shapley. We also leverage a Lipschitz continuity of model Shapley to design a learning approach for predicting the model Shapley values (MSVs) of many vendors' models (e.g., 150) in a large-scale marketplace. We perform extensive empirical validation on the effectiveness of model Shapley using various real-world datasets and heterogeneous model types.
****
Robust Concept Erasure via Kernelized Rate-Distortion Maximization
Somnath Basu Roy Chowdhury, Nicholas Monath, Kumar Avinava Dubey, Amr Ahmed, Snigdha Chaturvedi
Distributed representations provide a vector space that captures meaningful relationships between data instances. The distributed nature of these representations, however, entangles together multiple attributes or concepts of data instances (e.g., the topic or sentiment of a text, characteristics of the author (age, gender, etc), etc). Recent work has proposed the task of concept erasure, in which rather than making a concept predictable, the goal is to remove an attribute from distributed representations while retaining other information from the original representation space as much as possible. In this paper, we propose a new distance metric learning-based objective, the Kernelized Rate-Distortion Maximizer (KRaM), for performing concept erasure. KRaM fits a transformation of representations to match a specified distance measure (defined by a labeled concept to erase) using a modified rate-distortion function. Specifically, KRaM's objective function aims to make instances with similar concept labels dissimilar in the learned representation space while retaining other information. We find that optimizing KRaM effectively erases various types of concepts—categorical, continuous, and vector-valued variables—from data representations across diverse domains. We also provide a theoretical analysis of several properties of KRaM's objective. To assess the quality of the learned representations, we propose an alignment score to evaluate their similarity with the original representation space. Additionally, we conduct experiments to showcase KRaM's efficacy in various settings, from erasing binary gender variables in word embeddings to vector-valued variables in GPT-3 representations.
****
BiMatting: Efficient Video Matting via Binarization
Haotong Qin, Lei Ke, Xudong Ma, Martin Danelljan, Yu-Wing Tai, Chi-Keung Tang, Xianglong Liu, Fisher Yu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
One Fits All: Power General Time Series Analysis by Pretrained LM
Tian Zhou, Peisong Niu, xue wang, Liang Sun, Rong Jin
Although we have witnessed great success of pre-trained models in natural language processing (NLP) and computer vision (CV), limited progress has been made for general time series analysis. Unlike NLP and CV where a unified model can be used to perform different tasks, specially designed approach still dominates in each time series analysis task such as classification, anomaly detection, forecasting, and few-shot learning. The main challenge that blocks the development of pre-trained model for time series analysis is the lack of a large amount of data f

or training. In this work, we address this challenge by leveraging language or C
V models, pre-trained from billions of tokens, for time series analysis. Specifi
cally, we refrain from altering the self-attention and feedforward layers of the
 residual blocks in the pre-trained language or image model. This model, known a
s the Frozen Pretrained Transformer (FPT), is evaluated through fine-tuning on a
ll major types of tasks involving time series. Our results demonstrate that pre-
trained models on natural language or images can lead to a comparable or state-o
f-the-art performance in all main time series analysis tasks, as illustrated in
Figure1. We also found both theoretically and empirically that the self-attentio
n module behaviors similarly to principle component analysis (PCA), an observati
on that helps explains how transformer bridges the domain gap and a crucial step
 towards understanding the universality of a pre-trained transformer. The code i
s publicly available at https://anonymous.4open.science/r/Pretrained-LM-for-TSFo
rcasting-C561.
****
Parameterizing Non-Parametric Meta-Reinforcement Learning Tasks via Subtask Deco
mposition
Suyoung Lee, Myungsik Cho, Youngchul Sung
Meta-reinforcement learning (meta-RL) techniques have demonstrated remarkable su
ccess in generalizing deep reinforcement learning across a range of tasks. Never
theless, these methods often struggle to generalize beyond tasks with parametric
 variations. To overcome this challenge, we propose Subtask Decomposition and Vi
rtual Training (SDVT), a novel meta-RL approach that decomposes each non-paramet
ric task into a collection of elementary subtasks and parameterizes the task bas
ed on its decomposition. We employ a Gaussian mixture VAE to meta-learn the dec
omposition process, enabling the agent to reuse policies acquired from common su
btasks. Additionally, we propose a virtual training procedure, specifically desi
gned for non-parametric task variability, which generates hypothetical subtask c
ompositions, thereby enhancing generalization to previously unseen subtask compo
sitions. Our method significantly improves performance on the Meta-World ML-10 a
nd ML-45 benchmarks, surpassing current state-of-the-art techniques.
****
Near-Optimal Algorithms for Gaussians with Huber Contamination: Mean Estimation
and Linear Regression
Ilias Diakonikolas, Daniel Kane, Ankit Pensia, Thanasis Pittas
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Causal Discovery from Subsampled Time Series with Proxy Variables
Mingzhou Liu, Xinwei Sun, Lingjing Hu, Yizhou Wang
Inferring causal structures from time series data is the central interest of man
y scientific inquiries. A major barrier to such inference is the problem of subs
ampling, i.e., the frequency of measurement is much lower than that of causal in
fluence. To overcome this problem, numerous methods have been proposed, yet eith
er was limited to the linear case or failed to achieve identifiability. In this
paper, we propose a constraint-based algorithm that can identify the entire caus
al structure from subsampled time series, without any parametric constraint. Our
 observation is that the challenge of subsampling arises mainly from hidden vari
ables at the unobserved time steps. Meanwhile, every hidden variable has an obse
rved proxy, which is essentially itself at some observable time in the future, b
enefiting from the temporal structure. Based on these, we can leverage the proxi
es to remove the bias induced by the hidden variables and hence achieve identifi
ability. Following this intuition, we propose a proxy-based causal discovery alg
orithm. Our algorithm is nonparametric and can achieve full causal identificatio
n. Theoretical advantages are reflected in synthetic and real-world experiments.
****
“Why Not Looking backward?” A Robust Two-Step Method to Automatically Terminate
Bayesian Optimization

Shuang Li, Ke Li, Wei Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Chameleon: Plug-and-Play Compositional Reasoning with Large Language Models
Pan Lu, Baolin Peng, Hao Cheng, Michel Galley, Kai-Wei Chang, Ying Nian Wu, Song-Chun Zhu, Jianfeng Gao
Large language models (LLMs) have achieved remarkable progress in solving various natural language processing tasks due to emergent reasoning abilities. However, LLMs have inherent limitations as they are incapable of accessing up-to-date information (stored on the Web or in task-specific knowledge bases), using external tools, and performing precise mathematical and logical reasoning. In this paper, we present Chameleon, an AI system that mitigates these limitations by augmenting LLMs with plug-and-play modules for compositional reasoning. Chameleon synthesizes programs by composing various tools (e.g., LLMs, off-the-shelf vision models, web search engines, Python functions, and heuristic-based modules) for accomplishing complex reasoning tasks. At the heart of Chameleon is an LLM-based planner that assembles a sequence of tools to execute to generate the final response. We showcase the effectiveness of Chameleon on two multi-modal knowledge-intensive reasoning tasks: ScienceQA and TabMWP. Chameleon, powered by GPT-4, achieves an 86.54% overall accuracy on ScienceQA, improving the best published few-shot result by 11.37%. On TabMWP, GPT-4-powered Chameleon improves the accuracy by 17.0%, lifting the state of the art to 98.78%. Our analysis also shows that the GPT-4-powered planner exhibits more consistent and rational tool selection via inferring potential constraints from instructions, compared to a ChatGPT-powered planner.
****
Diffusion Models and Semi-Supervised Learners Benefit Mutually with Few Labels
Zebin You, Yong Zhong, Fan Bao, Jiacheng Sun, Chongxuan LI, Jun Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Pre-Training Protein Encoder via Siamese Sequence-Structure Diffusion Trajectory Prediction
Zuobai Zhang, Minghao Xu, Aurelie C. Lozano, Vijil Chenthamarakshan, Payel Das, Jian Tang
Self-supervised pre-training methods on proteins have recently gained attention, with most approaches focusing on either protein sequences or structures, neglecting the exploration of their joint distribution, which is crucial for a comprehensive understanding of protein functions by integrating co-evolutionary information and structural characteristics. In this work, inspired by the success of denoising diffusion models in generative tasks, we propose the DiffPreT approach to pre-train a protein encoder by sequence-structure joint diffusion modeling. DiffPreT guides the encoder to recover the native protein sequences and structures from the perturbed ones along the joint diffusion trajectory, which acquires the joint distribution of sequences and structures. Considering the essential protein conformational variations, we enhance DiffPreT by a method called Siamese Diffusion Trajectory Prediction (SiamDiff) to capture the correlation between different conformers of a protein. SiamDiff attains this goal by maximizing the mutual information between representations of diffusion trajectories of structurally-correlated conformers. We study the effectiveness of DiffPreT and SiamDiff on both atom- and residue-level structure-based protein understanding tasks. Experimental results show that the performance of DiffPreT is consistently competitive on all tasks, and SiamDiff achieves new state-of-the-art performance, considering the mean ranks on all tasks. Code will be released upon acceptance.
****

Progressive Ensemble Distillation: Building Ensembles for Efficient Inference
Don Dennis, Abhishek Shetty, Anish Prasad Sevekari, Kazuhito Koishida, Virginia
Smith
Knowledge distillation is commonly used to compress an ensemble of models into a
 single model. In this work we study the problem of progressive ensemble distill
ation: Given a large, pretrained teacher model , we seek to decompose the model
into an ensemble of smaller, low-inference cost student models . The resulting e
nsemble allows for flexibly tuning accuracy vs. inference cost, which can be use
ful for a multitude of applications in efficient inference. Our method, B-DISTIL
, uses a boosting procedure that allows function composition based aggregation r
ules to construct expressive ensembles with similar performance as using much sm
aller student models. We demonstrate the effectiveness of B-DISTIL by decomposin
g pretrained models across a variety of image, speech, and sensor datasets. Our
method comes with strong theoretical guarantees in terms of convergence as well
as generalization.
****
Differentially Private Approximate Near Neighbor Counting in High Dimensions
Alexandr Andoni, Piotr Indyk, Sepideh Mahabadi, Shyam Narayanan
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Reinforcement-Enhanced Autoregressive Feature Transformation: Gradient-steered S
earch in Continuous Space for Postfix Expressions
Dongjie Wang, Meng Xiao, Min Wu, pengfei wang, Yuanchun Zhou, Yanjie Fu
Feature transformation aims to generate new pattern-discriminative feature space
 from original features to improve downstream machine learning (ML) task perfor
mances. However, the discrete search space for the optimal feature explosively g
rows on the basis of combinations of features and operations from low-order form
s to high-order forms. Existing methods, such as exhaustive search, expansion re
duction, evolutionary algorithms, reinforcement learning, and iterative greedy,
suffer from large search space. Overly emphasizing efficiency in algorithm desig
n usually sacrifice stability or robustness. To fundamentally fill this gap, we
reformulate discrete feature transformation as a continuous space optimization t
ask and develop an embedding-optimization-reconstruction framework. This framewo
rk includes four steps: 1) reinforcement-enhanced data preparation, aiming to pr
epare high-quality transformation-accuracy training data; 2) feature transformat
ion operation sequence embedding, intending to encapsulate the knowledge of prep
ared training data within a continuous space; 3) gradient-steered optimal embedd
ing search, dedicating to uncover potentially superior embeddings within the lea
rned space; 4) transformation operation sequence reconstruction, striving to rep
roduce the feature transformation solution to pinpoint the optimal feature space
. Finally, extensive experiments and case studies are performed to demonstrate t
he effectiveness and robustness of the proposed method. The code and data are pu
blicly accessible https://www.dropbox.com/sh/imh8ckui7va3k5u/AACulQegVx0MuywYyoC
qSdVPa?dl=0.
****
FD-Align: Feature Discrimination Alignment for Fine-tuning Pre-Trained Models in
 Few-Shot Learning
Kun Song, Huimin Ma, Bochao Zou, Huishuai Zhang, Weiran Huang
Due to the limited availability of data, existing few-shot learning methods trai
ned from scratch fail to achieve satisfactory performance. In contrast, large-sc
ale pre-trained models such as CLIP demonstrate remarkable few-shot and zero-sho
t capabilities. To enhance the performance of pre-trained models for downstream
tasks, fine-tuning the model on downstream data is frequently necessary. However
, fine-tuning the pre-trained model leads to a decrease in its generalizability
in the presence of distribution shift, while the limited number of samples in fe
w-shot learning makes the model highly susceptible to overfitting. Consequently,
 existing methods for fine-tuning few-shot learning primarily focus on fine-tuni

ng the model's classification head or introducing additional structure. In this paper, we introduce a fine-tuning approach termed Feature Discrimination Alignment (FD-Align). Our method aims to bolster the model's generalizability by preserving the consistency of spurious features across the fine-tuning process. Extensive experimental results validate the efficacy of our approach for both ID and OOD tasks. Once fine-tuned, the model can seamlessly integrate with existing methods, leading to performance improvements. Our code can be found in https://github.com/skingorz/FD-Align.
****
A Step Towards Worldwide Biodiversity Assessment: The BIOSCAN-1M Insect Dataset
Zahra Gharaee, ZeMing Gong, Nicholas Pellegrino, Iuliia Zarubiieva, Joakim Bruslund Haurum, Scott Lowe, Jaclyn McKeown, Chris Ho, Joschka McLeod, Yi-Yun Wei, Jireh Agda, Sujeevan Ratnasingham, Dirk Steinke, Angel Chang, Graham W. Taylor, Paul Fieguth
In an effort to catalog insect biodiversity, we propose a new large dataset of hand-labelled insect images, the BIOSCAN-1M Insect Dataset. Each record is taxonomically classified by an expert, and also has associated genetic information including raw nucleotide barcode sequences and assigned barcode index numbers, which are genetic-based proxies for species classification. This paper presents a curated million-image dataset, primarily to train computer-vision models capable of providing image-based taxonomic assessment, however, the dataset also presents compelling characteristics, the study of which would be of interest to the broader machine learning community. Driven by the biological nature inherent to the dataset, a characteristic long-tailed class-imbalance distribution is exhibited. Furthermore, taxonomic labelling is a hierarchical classification scheme, presenting a highly fine-grained classification problem at lower levels. Beyond spurring interest in biodiversity research within the machine learning community, progress on creating an image-based taxonomic classifier will also further the ultimate goal of all BIOSCAN research: to lay the foundation for a comprehensive survey of global biodiversity. This paper introduces the dataset and explores the classification task through the implementation and analysis of a baseline classifier. The code repository of the BIOSCAN-1M-Insect dataset is available at https://github.com/zahrag/BIOSCAN-1M
****
D-Separation for Causal Self-Explanation
Wei Liu, Jun Wang, Haozhao Wang, Ruixuan Li, Zhiying Deng, YuanKai Zhang, Yang Qiu
Rationalization aims to strengthen the interpretability of NLP models by extracting a subset of human-intelligible pieces of their inputting texts. Conventional works generally employ the maximum mutual information (MMI) criterion to find the rationale that is most indicative of the target label. However, this criterion can be influenced by spurious features that correlate with the causal rationale or the target label. Instead of attempting to rectify the issues of the MMI criterion, we propose a novel criterion to uncover the causal rationale, termed the Minimum Conditional Dependence (MCD) criterion, which is grounded on our finding that the non-causal features and the target label are \emph{d-separated} by the causal rationale. By minimizing the dependence between the non-selected parts of the input and the target label conditioned on the selected rationale candidate, all the causes of the label are compelled to be selected. In this study, we employ a simple and practical measure for dependence, specifically the KL-divergence, to validate our proposed MCD criterion. Empirically, we demonstrate that MCD improves the F1 score by up to 13.7% compared to previous state-of-the-art MMI-based methods.Our code is in an anonymous repository: https://anonymous.4open.science/r/MCD-CE88.
****
History Filtering in Imperfect Information Games: Algorithms and Complexity
Christopher Solinas, Doug Rebstock, Nathan Sturtevant, Michael Buro
Historically applied exclusively to perfect information games, depth-limited search with value functions has been key to recent advances in AI for imperfect information games. Most prominent approaches with strong theoretical guarantees req

uire subgame decomposition - a process in which a subgame is computed from public information and player beliefs. However, subgame decomposition can itself require non-trivial computations, and its tractability depends on the existence of efficient algorithms for either full enumeration or generation of the histories that form the root of the subgame. Despite this, no formal analysis of the tractability of such computations has been established in prior work, and application domains have often consisted of games, such as poker, for which enumeration is trivial on modern hardware.Applying these ideas to more complex domains requires understanding their cost. In this work, we introduce and analyze the computational aspects and tractability of filtering histories for subgame decomposition. We show that constructing a single history from the root of the subgame is generally intractable, and then provide a necessary and sufficient condition for efficient enumeration. We also introduce a novel Markov Chain Monte Carlo-based generation algorithm for trick-taking card games - a domain where enumeration is often prohibitively expensive. Our experiments demonstrate its improved scalability in the trick-taking card game Oh Hell.These contributions clarify when and how depth-limited search via subgame decomposition can be an effective tool for sequential decision-making in imperfect information settings.
****

Constant Approximation for Individual Preference Stable Clustering
Anders Aamand, Justin Chen, Allen Liu, Sandeep Silwal, Pattara Sukprasert, Ali Vakilian, Fred Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Intervention Generalization: A View from Factor Graph Models
Gecia Bravo-Hermsdorff, David Watson, Jialin Yu, Jakob Zeitler, Ricardo Silva
One of the goals of causal inference is to generalize from past experiments and observational data to novel conditions. While it is in principle possible to eventually learn a mapping from a novel experimental condition to an outcome of interest, provided a sufficient variety of experiments is available in the training data, coping with a large combinatorial space of possible interventions is hard. Under a typical sparse experimental design, this mapping is ill-posed without relying on heavy regularization or prior distributions. Such assumptions may or may not be reliable, and can be hard to defend or test. In this paper, we take a close look at how to warrant a leap from past experiments to novel conditions based on minimal assumptions about the factorization of the distribution of the manipulated system, communicated in the well-understood language of factor graph models. A postulated interventional factor model (IFM) may not always be informative, but it conveniently abstracts away a need for explicitly modeling unmeasured confounding and feedback mechanisms, leading to directly testable claims. Given an IFM and datasets from a collection of experimental regimes, we derive conditions for identifiability of the expected outcomes of new regimes never observed in these training data. We implement our framework using several efficient algorithms, and apply them on a range of semi-synthetic experiments.
****

Decision Tree for Locally Private Estimation with Public Data
Yuheng Ma, Han Zhang, Yuchao Cai, Hanfang Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

DeepACO: Neural-enhanced Ant Systems for Combinatorial Optimization
Haoran Ye, Jiarui Wang, Zhiguang Cao, Helan Liang, Yong Li
Ant Colony Optimization (ACO) is a meta-heuristic algorithm that has been successfully applied to various Combinatorial Optimization Problems (COPs). Traditionally, customizing ACO for a specific problem requires the expert design of knowle

dge-driven heuristics. In this paper, we propose DeepACO, a generic framework th
at leverages deep reinforcement learning to automate heuristic designs. DeepACO
serves to strengthen the heuristic measures of existing ACO algorithms and dispe
nse with laborious manual design in future ACO applications. As a neural-enhance
d meta-heuristic, DeepACO consistently outperforms its ACO counterparts on eight
 COPs using a single neural model and a single set of hyperparameters. As a Neur
al Combinatorial Optimization method, DeepACO performs better than or on par wit
h problem-specific methods on canonical routing problems. Our code is publicly a
vailable at https://github.com/henry-yeh/DeepACO.
****
Offline Imitation Learning with Variational Counterfactual Reasoning
Zexu Sun, Bowei He, Jinxin Liu, Xu Chen, Chen Ma, Shuai Zhang
In offline imitation learning (IL), an agent aims to learn an optimal expert beh
avior policy without additional online environment interactions. However, in man
y real-world scenarios, such as robotics manipulation, the offline dataset is co
llected from suboptimal behaviors without rewards. Due to the scarce expert data
, the agents usually suffer from simply memorizing poor trajectories and are vul
nerable to the variations in the environments, lacking the capability of general
izing to new environments.To automatically generate high-quality expert data and
 improve the generalization ability of the agent, we propose a framework named \
underline{O}ffline \underline{I}mitation \underline{L}earning with \underline{C}
ounterfactual data \underline{A}ugmentation (OILCA) by doing counterfactual infe
rence. In particular, we leverage identifiable variational autoencoder to genera
te \textit{counterfactual} samples for expert data augmentation. We theoreticall
y analyze the influence of the generated expert data and the improvement of gene
ralization. Moreover, we conduct extensive experiments to demonstrate that our a
pproach significantly outperforms various baselines on both \textsc{DeepMind Con
trol Suite} benchmark for in-distribution performance and \textsc{CausalWorld} b
enchmark for out-of-distribution generalization.
****
LART: Neural Correspondence Learning with Latent Regularization Transformer for
3D Motion Transfer
Haoyu Chen, Hao Tang, Radu Timofte, Luc V Gool, Guoying Zhao
3D motion transfer aims at transferring the motion from a dynamic input sequence
 to a static 3D object and outputs an identical motion of the target with high-f
idelity and realistic visual effects. In this work, we propose a novel 3D Transf
ormer framework called LART for 3D motion transfer. With carefully-designed arch
itectures, LART is able to implicitly learn the correspondence via a flexible ge
ometry perception. Thus, unlike other existing methods, LART does not require an
y key point annotations or pre-defined correspondence between the motion source
and target meshes and can also handle large-size full-detailed unseen 3D targets
. Besides, we introduce a novel latent metric regularization on the Transformer
for better motion generation. Our rationale lies in the observation that the dec
oded motions can be approximately expressed as linearly geometric distortion at
the frame level. The metric preservation of motions could be translated to the f
ormation of linear paths in the underlying latent space as a rigorous constraint
 to control the synthetic motions occurring in the construction of the latent sp
ace. The proposed LART shows a high learning efficiency with the need for a few
samples from the AMASS dataset to generate motions with plausible visual effects
. The experimental results verify the potential of our generative model in appli
cations of motion transfer, content generation, temporal interpolation, and moti
on denoising. The code is made available: https://github.com/mikecheninoulu/LART
.
****
Neural Relation Graph: A Unified Framework for Identifying Label Noise and Outli
er Data
Jang-Hyun Kim, Sangdoo Yun, Hyun Oh Song
Diagnosing and cleaning data is a crucial step for building robust machine learn
ing systems. However, identifying problems within large-scale datasets with real
-world distributions is challenging due to the presence of complex issues such a

s label errors, under-representation, and outliers. In this paper, we propose a unified approach for identifying the problematic data by utilizing a largely ignored source of information: a relational structure of data in the feature-embedded space. To this end, we present scalable and effective algorithms for detecting label errors and outlier data based on the relational graph structure of data. We further introduce a visualization tool that provides contextual information of a data point in the feature-embedded space, serving as an effective tool for interactively diagnosing data. We evaluate the label error and outlier/out-of-distribution (OOD) detection performances of our approach on the large-scale image, speech, and language domain tasks, including ImageNet, ESC-50, and SST2. Our approach achieves state-of-the-art detection performance on all tasks considered and demonstrates its effectiveness in debugging large-scale real-world datasets across various domains. We release codes at https://github.com/snu-mllab/Neural-Relation-Graph.
****
Lift Yourself Up: Retrieval-augmented Text Generation with Self-Memory
Xin Cheng, Di Luo, Xiuying Chen, Lemao Liu, Dongyan Zhao, Rui Yan
With direct access to human-written reference as memory, retrieval-augmented generation has achieved much progress in a wide range of text generation tasks. Since better memory would typically prompt better generation (we define this as primal problem). The traditional approach for memory retrieval involves selecting memory that exhibits the highest similarity to the input. However, this method is constrained by the quality of the fixed corpus from which memory is retrieved. In this paper, by exploring the duality of the primal problem: better generation also prompts better memory, we propose a novel framework, selfmem, which addresses this limitation by iteratively employing a retrieval-augmented generator to create an unbounded memory pool and using a memory selector to choose one output as memory for the subsequent generation round. This enables the model to leverage its own output, referred to as self-memory, for improved generation. We evaluate the effectiveness of selfmem on three distinct text generation tasks: neural machine translation, abstractive text summarization, and dialogue generation, under two generation paradigms: fine-tuned small model and few-shot LLM. Our approach achieves state-of-the-art results in four directions in JRC-Acquis translation dataset, 50.3 ROUGE-1 in XSum, and 62.9 ROUGE-1 in BigPatent, demonstrating the potential of self-memory in enhancing retrieval-augmented generation models. Furthermore, we conduct thorough analyses of each component in the selfmem framework to identify current system bottlenecks and provide insights for future research.
****
Front-door Adjustment Beyond Markov Equivalence with Limited Graph Knowledge
Abhin Shah, Karthikeyan Shanmugam, Murat Kocaoglu
Causal effect estimation from data typically requires assumptions about the cause-effect relations either explicitly in the form of a causal graph structure within the Pearlian framework, or implicitly in terms of (conditional) independence statements between counterfactual variables within the potential outcomes framework. When the treatment variable and the outcome variable are confounded, front-door adjustment is an important special case where, given the graph, causal effect of the treatment on the target can be estimated using post-treatment variables. However, the exact formula for front-door adjustment depends on the structure of the graph, which is difficult to learn in practice. In this work, we provide testable conditional independence statements to compute the causal effect using front-door-like adjustment without knowing the graph under limited structural side information. We show that our method is applicable in scenarios where knowing the Markov equivalence class is not sufficient for causal effect estimation. We demonstrate the effectiveness of our method on a class of random graphs as well as real causal fairness benchmarks.
****
Training Energy-Based Normalizing Flow with Score-Matching Objectives
Chen-Hao Chao, Wei-Fang Sun, Yen-Chang Hsu, Zsolt Kira, Chun-Yi Lee
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

LayoutPrompter: Awaken the Design Ability of Large Language Models
Jiawei Lin, Jiaqi Guo, Shizhao Sun, Zijiang Yang, Jian-Guang Lou, Dongmei Zhang
Conditional graphic layout generation, which automatically maps user constraints to high-quality layouts, has attracted widespread attention today. Although recent works have achieved promising performance, the lack of versatility and data efficiency hinders their practical applications. In this work, we propose LayoutPrompter, which leverages large language models (LLMs) to address the above problems through in-context learning. LayoutPrompter is made up of three key components, namely input-output serialization, dynamic exemplar selection and layout ranking. Specifically, the input-output serialization component meticulously designs the input and output formats for each layout generation task. Dynamic exemplar selection is responsible for selecting the most helpful prompting exemplars for a given input. And a layout ranker is used to pick the highest quality layout from multiple outputs of LLMs. We conduct experiments on all existing layout generation tasks using four public datasets. Despite the simplicity of our approach, experimental results show that LayoutPrompter can compete with or even outperform state-of-the-art approaches on these tasks without any model training or fine-tuning. This demonstrates the effectiveness of this versatile and training-free approach. In addition, the ablation studies show that LayoutPrompter is significantly superior to the training-based baseline in a low-data regime, further indicating the data efficiency of LayoutPrompter. Our project is available at https://github.com/microsoft/LayoutGeneration/tree/main/LayoutPrompter.
****

Classification of Heavy-tailed Features in High Dimensions: a Superstatistical Approach
Urte Adomaityte, Gabriele Sicuro, Pierpaolo Vivo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

CoLA: Exploiting Compositional Structure for Automatic and Efficient Numerical Linear Algebra
Andres Potapczynski, Marc Finzi, Geoff Pleiss, Andrew G. Wilson
Many areas of machine learning and science involve large linear algebra problems, such as eigendecompositions, solving linear systems, computing matrix exponentials, and trace estimation. The matrices involved often have Kronecker, convolutional, block diagonal, sum, or product structure. In this paper, we propose a simple but general framework for large-scale linear algebra problems in machine learning, named CoLA (Compositional Linear Algebra). By combining a linear operator abstraction with compositional dispatch rules, CoLA automatically constructs memory and runtime efficient numerical algorithms. Moreover, CoLA provides memory efficient automatic differentiation, low precision computation, and GPU acceleration in both JAX and PyTorch, while also accommodating new objects, operations, and rules in downstream packages via multiple dispatch. CoLA can accelerate many algebraic operations, while making it easy to prototype matrix structures and algorithms, providing an appealing drop-in tool for virtually any computational effort that requires linear algebra. We showcase its efficacy across a broad range of applications, including partial differential equations, Gaussian processes, equivariant model construction, and unsupervised learning.
****

Large language models implicitly learn to straighten neural sentence trajectories to construct a predictive representation of natural language.
Eghbal Hosseini, Evelina Fedorenko
Predicting upcoming events is critical to our ability to effectively interact with ourenvironment and conspecifics. In natural language processing, transformer

models,which are trained on next-word prediction, appear to construct a general-purposerepresentation of language that can support diverse downstream tasks. However, westill lack an understanding of how a predictive objective shapes such representations.Inspired by recent work in vision neuroscience Hénaff et al. (2019), here we test ahypothesis about predictive representations of autoregressive transformer models.In particular, we test whether the neural trajectory of a sequence of words in asentence becomes progressively more straight as it passes through the layers of thenetwork. The key insight behind this hypothesis is that straighter trajectories shouldfacilitate prediction via linear extrapolation. We quantify straightness using a 1-dimensional curvature metric, and present four findings in support of the trajectorystraightening hypothesis: i) In trained models, the curvature progressively decreasesfrom the first to the middle layers of the network. ii) Models that perform better onthe next-word prediction objective, including larger models and models trained onlarger datasets, exhibit greater decreases in curvature, suggesting that this improvedability to straighten sentence neural trajectories may be the underlying driver ofbetter language modeling performance. iii) Given the same linguistic context, thesequences that are generated by the model have lower curvature than the groundtruth (the actual continuations observed in a language corpus), suggesting thatthe model favors straighter trajectories for making predictions. iv) A consistentrelationship holds between the average curvature and the average surprisal ofsentences in the middle layers of models, such that sentences with straighter neuraltrajectories also have lower surprisal. Importantly, untrained models don’t exhibitthese behaviors. In tandem, these results support the trajectory straighteninghypothesis and provide a possible mechanism for how the geometry of the internalrepresentations of autoregressive models supports next word prediction.
****

Weakly Coupled Deep Q-Networks

Ibrahim El Shar, Daniel Jiang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Provably Fast Convergence of Independent Natural Policy Gradient for Markov Potential Games

Youbang Sun, Tao Liu, Ruida Zhou, P. R. Kumar, Shahin Shahrampour

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

MMGP: a Mesh Morphing Gaussian Process-based machine learning method for regression of physical problems under nonparametrized geometrical variability

Fabien Casenave, Brian Staber, Xavier Roynard

When learning simulations for modeling physical phenomena in industrial designs, geometrical variabilities are of prime interest. While classical regression techniques prove effective for parameterized geometries, practical scenarios often involve the absence of shape parametrization during the inference stage, leaving us with only mesh discretizations as available data. Learning simulations from such mesh-based representations poses significant challenges, with recent advances relying heavily on deep graph neural networks to overcome the limitations of conventional machine learning approaches. Despite their promising results, graph neural networks exhibit certain drawbacks, including their dependency on extensive datasets and limitations in providing built-in predictive uncertainties or handling large meshes. In this work, we propose a machine learning method that do not rely on graph neural networks. Complex geometrical shapes and variations with fixed topology are dealt with using well-known mesh morphing onto a common support, combined with classical dimensionality reduction techniques and Gaussian processes. The proposed methodology can easily deal with large meshes without th

e need for explicit shape parameterization and provides crucial predictive uncertainties, which are essential for informed decision-making. In the considered numerical experiments, the proposed method is competitive with respect to existing graph neural networks, regarding training efficiency and accuracy of the predictions.
****

Adaptive Online Replanning with Diffusion Models

Siyuan Zhou, Yilun Du, Shun Zhang, Mengdi Xu, Yikang Shen, Wei Xiao, Dit-Yan Yeung, Chuang Gan

Diffusion models have risen a promising approach to data-driven planning, and have demonstrated impressive robotic control, reinforcement learning, and video planning performance. Given an effective planner, an important question to consider is replanning -- when given plans should be regenerated due to both action execution error and external environment changes. Direct plan execution, without replanning, is problematic as errors from individual actions rapidly accumulate and environments are partially observable and stochastic. Simultaneously, replanning at each timestep incurs a substantial computational cost, and may prevent successful task execution, as different generated plans prevent consistent progress to any particular goal. In this paper, we explore how we may effectively replan with diffusion models. We propose a principled approach to determine when to replan, based on the diffusion model's estimated likelihood of existing generated plans. We further present an approach to replan existing trajectories to ensure that new plans follow the same goal state as the original trajectory, which may efficiently bootstrap off previously generated plans. We illustrate how a combination of our proposed additions significantly improves the performance of diffusion planners leading to 38\% gains over past diffusion planning approaches on Maze2D and further enables handling of stochastic and long-horizon robotic control tasks.
****

SODA: Robust Training of Test-Time Data Adaptors

Zige Wang, Yonggang Zhang, Zhen Fang, Long Lan, Wenjing Yang, Bo Han

Adapting models deployed to test distributions can mitigate the performance degradation caused by distribution shifts. However, privacy concerns may render model parameters inaccessible. One promising approach involves utilizing zeroth-order optimization (ZOO) to train a data adaptor to adapt the test data to fit the deployed models. Nevertheless, the data adaptor trained with ZOO typically brings restricted improvements due to the potential corruption of data features caused by the data adaptor. To address this issue, we revisit ZOO in the context of test-time data adaptation. We find that the issue directly stems from the unreliable estimation of the gradients used to optimize the data adaptor, which is inherently due to the unreliable nature of the pseudo-labels assigned to the test data. Based on this observation, we propose pseudo-label-robust data adaptation (SODA) to improve the performance of data adaptation. Specifically, SODA leverages high-confidence predicted labels as reliable labels to optimize the data adaptor with ZOO for label prediction. For data with low-confidence predictions, SODA encourages the adaptor to preserve data information to mitigate data corruption. Empirical results indicate that SODA can significantly enhance the performance of deployed models in the presence of distribution shifts without requiring access to model parameters.
****

Training Neural Networks is NP-Hard in Fixed Dimension

Vincent Froese, Christoph Hertrich

We study the parameterized complexity of training two-layer neural networks with respect to the dimension of the input data and the number of hidden neurons, considering ReLU and linear threshold activation functions. Albeit the computational complexity of these problems has been studied numerous times in recent years, several questions are still open. We answer questions by Arora et al. (ICLR 2018) and Khalife and Basu (IPCO 2022) showing that both problems are NP-hard for two dimensions, which excludes any polynomial-time algorithm for constant dimension. We also answer a question by Froese et al. (JAIR 2022) proving W[1]-hardness

for four ReLUs (or two linear threshold neurons) with zero training error. Finally, in the ReLU case, we show fixed-parameter tractability for the combined parameter number of dimensions and number of ReLUs if the network is assumed to compute a convex map. Our results settle the complexity status regarding these parameters almost completely.
****
GlyphControl: Glyph Conditional Control for Visual Text Generation
Yukang Yang, Dongnan Gui, YUHUI YUAN, Weicong Liang, Haisong Ding, Han Hu, Kai Chen
Recently, there has been an increasing interest in developing diffusion-based text-to-image generative models capable of generating coherent and well-formed visual text. In this paper, we propose a novel and efficient approach called GlyphControl to address this task. Unlike existing methods that rely on character-aware text encoders like ByT5 and require retraining of text-to-image models, our approach leverages additional glyph conditional information to enhance the performance of the off-the-shelf Stable-Diffusion model in generating accurate visual text. By incorporating glyph instructions, users can customize the content, location, and size of the generated text according to their specific requirements. To facilitate further research in visual text generation, we construct a training benchmark dataset called LAION-Glyph. We evaluate the effectiveness of our approach by measuring OCR-based metrics, CLIP score, and FID of the generated visual text. Our empirical evaluations demonstrate that GlyphControl outperforms the recent DeepFloyd IF approach in terms of OCR accuracy, CLIP score, and FID, highlighting the efficacy of our method.
****
Domain Adaptive Imitation Learning with Visual Observation
Sungho Choi, Seungyul Han, Woojun Kim, Jongseong Chae, Whiyoung Jung, Youngchul Sung
In this paper, we consider domain-adaptive imitation learning with visual observation, where an agent in a target domain learns to perform a task by observing expert demonstrations in a source domain. Domain adaptive imitation learning arises in practical scenarios where a robot, receiving visual sensory data, needs to mimic movements by visually observing other robots from different angles or observing robots of different shapes. To overcome the domain shift in cross-domain imitation learning with visual observation, we propose a novel framework for extracting domain-independent behavioral features from input observations that can be used to train the learner, based on dual feature extraction and image reconstruction. Empirical results demonstrate that our approach outperforms previous algorithms for imitation learning from visual observation with domain shift.
****
Discriminative Feature Attributions: Bridging Post Hoc Explainability and Inherent Interpretability
Usha Bhalla, Suraj Srinivas, Himabindu Lakkaraju
With the increased deployment of machine learning models in various real-world applications, researchers and practitioners alike have emphasized the need for explanations of model behaviour. To this end, two broad strategies have been outlined in prior literature to explain models. Post hoc explanation methods explain the behaviour of complex black-box models by identifying features critical to model predictions; however, prior work has shown that these explanations may not be faithful, in that they incorrectly attribute high importance to features that are unimportant or non-discriminative for the underlying task. Inherently interpretable models, on the other hand, circumvent these issues by explicitly encoding explanations into model architecture, meaning their explanations are naturally faithful, but they often exhibit poor predictive performance due to their limited expressive power. In this work, we identify a key reason for the lack of faithfulness of feature attributions: the lack of robustness of the underlying black-box models, especially the erasure of unimportant distractor features in the input. To address this issue, we propose Distractor Erasure Tuning (DiET), a method that adapts black-box models to be robust to distractor erasure, thus providing discriminative and faithful attributions. This strategy naturally combines the

ease-of-use of post hoc explanations with the faithfulness of inherently interp
retable models. We perform extensive experiments on semi-synthetic and real-worl
d datasets, and show that DiET produces models that (1) closely approximate the
original black-box models they are intended to explain, and (2) yield explanatio
ns that match approximate ground truths available by construction.
****
LegalBench: A Collaboratively Built Benchmark for Measuring Legal Reasoning in L
arge Language Models
Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Aditya K, Ale
x Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano,
Dmitry Talisman, Enam Hoque, Faiz Surani, Frank Fagan, Galit Sarfaty, Gregory Di
ckinson, Haggai Porat, Jason Hegland, Jessica Wu, Joe Nudell, Joel Niklaus, John
 Nay, Jonathan Choi, Kevin Tobia, Margaret Hagan, Megan Ma, Michael Livermore, N
ikon Rasumov-Rahe, Nils Holzenberger, Noam Kolt, Peter Henderson, Sean Rehaag, S
harad Goel, Shang Gao, Spencer Williams, Sunny Gandhi, Tom Zur, Varun Iyer, Zehu
a Li
The advent of large language models (LLMs) and their adoption by the legal commu
nity has given rise to the question: what types of legal reasoning can LLMs perf
orm? To enable greater study of this question, we present LegalBench: a collabor
atively constructed legal reasoning benchmark consisting of 162 tasks covering s
ix different types of legal reasoning. LegalBench was built through an interdisc
iplinary process, in which we collected tasks designed and hand-crafted by legal
 professionals. Because these subject matter experts took a leading role in cons
truction, tasks either measure legal reasoning capabilities that are practically
 useful, or measure reasoning skills that lawyers find interesting. To enable cr
oss-disciplinary conversations about LLMs in the law, we additionally show how p
opular legal frameworks for describing legal reasoning—which distinguish between
 its many forms—correspond to LegalBench tasks, thus giving lawyers and LLM deve
lopers a common vocabulary. This paper describes LegalBench, presents an empiric
al evaluation of 20 open-source and commercial LLMs, and illustrates the types o
f research explorations LegalBench enables.
****
Detecting hidden confounding in observational data using multiple environments
Rickard Karlsson, Jesse Krijthe
A common assumption in causal inference from observational data is that there is
 no hidden confounding. Yet it is, in general, impossible to verify the presence
 of hidden confounding factors from a single dataset. Under the assumption of in
dependent causal mechanisms underlying the data-generating process, we demonstra
te a way to detect unobserved confounders when having multiple observational dat
asets coming from different environments. We present a theory for testable condi
tional independencies that are only absent when there is hidden confounding and
examine cases where we violate its assumptions: degenerate & dependent mechanism
s, and faithfulness violations. Additionally, we propose a procedure to test the
se independencies and study its empirical finite-sample behavior using simulatio
n studies and semi-synthetic data based on a real-world dataset. In most cases,
the proposed procedure correctly predicts the presence of hidden confounding, pa
rticularly when the confounding bias is large.
****
Information Geometry of the Retinal Representation Manifold
Xuehao Ding, Dongsoo Lee, Joshua Melander, George Sivulka, Surya Ganguli, Stephe
n Baccus
The ability for the brain to discriminate among visual stimuli is constrained by
 their retinal representations. Previous studies of visual discriminability have
 been limited to either low-dimensional artificial stimuli or pure theoretical c
onsiderations without a realistic encoding model. Here we propose a novel framew
ork for understanding stimulus discriminability achieved by retinal representati
ons of naturalistic stimuli with the method of information geometry. To model th
e joint probability distribution of neural responses conditioned on the stimulus
, we created a stochastic encoding model of a population of salamander retinal g
anglion cells based on a three-layer convolutional neural network model. This mo

del not only accurately captured the mean response to natural scenes but also a variety of second-order statistics. With the model and the proposed theory, we c omputed the Fisher information metric over stimuli to study the most discriminab le stimulus directions. We found that the most discriminable stimulus varied sub stantially across stimuli, allowing an examination of the relationship between t he most discriminable stimulus and the current stimulus. By examining responses generated by the most discriminable stimuli we further found that the most discr iminative response mode is often aligned with the most stochastic mode. This fin ding carries the important implication that under natural scenes, retinal noise correlations are information-limiting rather than increasing information transmi ssion as has been previously speculated. We additionally observed that sensitivi ty saturates less in the population than for single cells and that as a function of firing rate, Fisher information varies less than sensitivity. We conclude th at under natural scenes, population coding benefits from complementary coding an d helps to equalize the information carried by different firing rates, which may facilitate decoding of the stimulus under principles of information maximizatio n.
****
RoboHive: A Unified Framework for Robot Learning
Vikash Kumar, Rutav Shah, Gaoyue Zhou, Vincent Moens, Vittorio Caggiano, Abhishe k Gupta, Aravind Rajeswaran
We present RoboHive, a comprehensive software platform and ecosystem for researc h in the field of Robot Learning and Embodied Artificial Intelligence. Our platf orm encompasses a diverse range of pre-existing and novel environments, includin g dexterous manipulation with the Shadow Hand, whole-arm manipulation tasks with Franka and Fetch robots, quadruped locomotion, among others. Included environme nts are organized within and cover multiple domains such as hand manipulation, l ocomotion, multi-task, multi-agent, muscles, etc. In comparison to prior works, RoboHive offers a streamlined and unified task interface taking dependency on on ly a minimal set of well-maintained packages, features tasks with high physics f idelity and rich visual diversity, and supports common hardware drivers for real -world deployment. The unified interface of RoboHive offers a convenient and acc essible abstraction for algorithmic research in imitation, reinforcement, multi- task, and hierarchical learning. Furthermore, RoboHive includes expert demonstra tions and baseline results for most environments, providing a standard for bench marking and comparisons. Details: https://sites.google.com/view/robohive
****
Sequential Memory with Temporal Predictive Coding
Mufeng Tang, Helen Barron, Rafal Bogacz
Forming accurate memory of sequential stimuli is a fundamental function of biolo gical agents. However, the computational mechanism underlying sequential memory in the brain remains unclear. Inspired by neuroscience theories and recent succe sses in applying predictive coding (PC) to \emph{static} memory tasks, in this w ork we propose a novel PC-based model for \emph{sequential} memory, called \emph {temporal predictive coding} (tPC). We show that our tPC models can memorize and retrieve sequential inputs accurately with a biologically plausible neural impl ementation. Importantly, our analytical study reveals that tPC can be viewed as a classical Asymmetric Hopfield Network (AHN) with an implicit statistical white ning process, which leads to more stable performance in sequential memory tasks of structured inputs. Moreover, we find that tPC exhibits properties consistent with behavioral observations and theories in neuroscience, thereby strengthening its biological relevance. Our work establishes a possible computational mechani sm underlying sequential memory in the brain that can also be theoretically inte rpreted using existing memory model frameworks.
****
Transportability for Bandits with Data from Different Environments
Alexis Bellot, Alan Malek, Silvia Chiappa
A unifying theme in the design of intelligent agents is to efficiently optimize a policy based on what prior knowledge of the problem is available and what acti ons can be taken to learn more about it. Bandits are a canonical instance of thi

s task that has been intensely studied in the literature. Most methods, however, typically rely solely on an agent's experimentation in a single environment (or multiple closely related environments). In this paper, we relax this assumption and consider the design of bandit algorithms from a combination of batch data and qualitative assumptions about the relatedness across different environments, represented in the form of causal models. In particular, we show that it is possible to exploit invariances across environments, wherever they may occur in the underlying causal model, to consistently improve learning. The resulting bandit algorithm has a sub-linear regret bound with an explicit dependency on a term that captures how informative related environments are for the task at hand; and may have substantially lower regret than experimentation-only bandit instances.
****

Students Parrot Their Teachers: Membership Inference on Model Distillation
Matthew Jagielski, Milad Nasr, Katherine Lee, Christopher A. Choquette-Choo, Nicholas Carlini, Florian Tramer
Model distillation is frequently proposed as a technique to reduce the privacy leakage of machine learning. These empirical privacy defenses rely on the intuition that distilled student'' models protect the privacy of training data, as they only interact with this data indirectly through ateacher'' model. In this work, we design membership inference attacks to systematically study the privacy provided by knowledge distillation to both the teacher and student training sets. Our new attacks show that distillation alone provides only limited privacy across a number of domains. We explain the success of our attacks on distillation by showing that membership inference attacks on a private dataset can succeed even if the target model is never queried on any actual training points, but only on inputs whose predictions are highly influenced by training data. Finally, we show that our attacks are strongest when student and teacher sets are similar, or when the attacker can poison the teacher set.
****

DiffuseBot: Breeding Soft Robots With Physics-Augmented Generative Diffusion Models
Tsun-Hsuan Johnson Wang, Juntian Zheng, Pingchuan Ma, Yilun Du, Byungchul Kim, Andrew Spielberg, Josh Tenenbaum, Chuang Gan, Daniela Rus
Nature evolves creatures with a high complexity of morphological and behavioral intelligence, meanwhile computational methods lag in approaching that diversity and efficacy. Co-optimization of artificial creatures' morphology and control in silico shows promise for applications in physical soft robotics and virtual character creation; such approaches, however, require developing new learning algorithms that can reason about function atop pure structure. In this paper, we present DiffuseBot, a physics-augmented diffusion model that generates soft robot morphologies capable of excelling in a wide spectrum of tasks. \name bridges the gap between virtually generated content and physical utility by (i) augmenting the diffusion process with a physical dynamical simulation which provides a certificate of performance, and (ii) introducing a co-design procedure that jointly optimizes physical design and control by leveraging information about physical sensitivities from differentiable simulation. We showcase a range of simulated and fabricated robots along with their capabilities. Check our website: https://diffusebot.github.io/
****

Hidden Poison: Machine Unlearning Enables Camouflaged Poisoning Attacks
Jimmy Di, Jack Douglas, Jayadev Acharya, Gautam Kamath, Ayush Sekhari
We introduce camouflaged data poisoning attacks, a new attack vector that arises in the context of machine unlearning and other settings when model retraining may be induced. An adversary first adds a few carefully crafted points to the training dataset such that the impact on the model's predictions is minimal. The adversary subsequently triggers a request to remove a subset of the introduced points at which point the attack is unleashed and the model's predictions are negatively affected. In particular, we consider clean-label targeted attacks (in which the goal is to cause the model to misclassify a specific test point) on datasets including CIFAR-10, Imagenette, and Imagewoof. This attack is realized by con

structing camouflage datapoints that mask the effect of a poisoned dataset. We d
emonstrate efficacy of our attack when unlearning is performed via retraining fr
om scratch, the idealized setting of machine unlearning which other efficient me
thods attempt to emulate, as well as against the approximate unlearning approach
 of Graves et al. (2021).
****
Thought Cloning: Learning to Think while Acting by Imitating Human Thinking
Shengran Hu, Jeff Clune
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
SNEkhorn: Dimension Reduction with Symmetric Entropic Affinities
Hugues Van Assel, Titouan Vayer, Rémi Flamary, Nicolas Courty
Many approaches in machine learning rely on a weighted graph to encode thesimila
rities between samples in a dataset. Entropic affinities (EAs), which are notabl
y used in the popular Dimensionality Reduction (DR) algorithm t-SNE, are particu
lar instances of such graphs. To ensure robustness to heterogeneous sampling den
sities, EAs assign a kernel bandwidth parameter to every sample in such a way th
at the entropy of each row in the affinity matrix is kept constant at a specific
 value, whose exponential is known as perplexity. EAs are inherently asymmetric
and row-wise stochastic, but they are used in DR approaches after undergoing heu
ristic symmetrization methods that violate both the row-wise constant entropy an
d stochasticity properties. In this work, we uncover a novel characterization of
 EA as an optimal transport problem, allowing a natural symmetrization that can
be computed efficiently using dual ascent. The corresponding novel affinity matr
ix derives advantages from symmetric doubly stochastic normalization in terms of
 clustering performance, while also effectively controlling the entropy of each
row thus making it particularly robust to varying noise levels. Following, we pr
esent a new DR algorithm, SNEkhorn, that leverages this new affinity matrix. We
show its clear superiority to state-of-the-art approaches with several indicator
s on both synthetic and real-world datasets.
****
Hyperbolic Graph Neural Networks at Scale: A Meta Learning Approach
Nurendra Choudhary, Nikhil Rao, Chandan Reddy
The progress in hyperbolic neural networks (HNNs) research is hindered by their
absence of inductive bias mechanisms, which are essential for generalizing to ne
w tasks and facilitating scalable learning over large datasets. In this paper, w
e aim to alleviate these issues by learning generalizable inductive biases from
the nodes’ local subgraph and transfer them for faster learning over new subgrap
hs with a disjoint set of nodes, edges, and labels in a few-shot setting. We int
roduce a novel method, Hyperbolic GRAph Meta Learner (H-GRAM), that, for the tas
ks of node classification and link prediction, learns transferable information f
rom a set of support local subgraphs in the form of hyperbolic meta gradients an
d label hyperbolic protonets to enable faster learning over a query set of new t
asks dealing with disjoint subgraphs. Furthermore, we show that an extension of
our meta-learning framework also mitigates the scalability challenges seen in HN
Ns faced by existing approaches. Our comparative analysis shows that H-GRAM effe
ctively learns and transfers information in multiple challenging few-shot settin
gs compared to other state-of-the-art baselines. Additionally, we demonstrate th
at, unlike standard HNNs, our approach is able to scale over large graph dataset
s and improve performance over its Euclidean counterparts.
****
FELM: Benchmarking Factuality Evaluation of Large Language Models
shiqi chen, Yiran Zhao, Jinghan Zhang, I-Chun Chern, Siyang Gao, Pengfei Liu, Ju
nxian He
Assessing factuality of text generated by large language models (LLMs) is an eme
rging yet crucial research area, aimed at alerting users to potential errors and
 guiding the development of more reliable LLMs. Nonetheless, the evaluators asse

ssing factuality necessitate suitable evaluation themselves to gauge progress and foster advancements. This direction remains under-explored, resulting in substantial impediments to the progress of factuality evaluators. To mitigate this issue, we introduce a benchmark for Factuality Evaluation of large Language Models, referred to as FELM. In this benchmark, we collect responses generated from LLMs and annotate factuality labels in a fine-grained manner. Contrary to previous studies that primarily concentrate on the factuality of world knowledge (e.g. information from Wikipedia), FELM focuses on factuality across diverse domains, spanning from world knowledge to math and reasoning. Our annotation is based on text segments, which can help pinpoint specific factual errors. The factuality annotations are further supplemented by predefined error types and reference links that either support or contradict the statement. In our experiments, we investigate the performance of several LLM-based factuality evaluators on FELM, including both vanilla LLMs and those augmented with retrieval mechanisms and chain-of-thought processes. Our findings reveal that while retrieval aids factuality evaluation, current LLMs are far from satisfactory to faithfully detect factual errors.
****
AircraftVerse: A Large-Scale Multimodal Dataset of Aerial Vehicle Designs
Adam Cobb, Anirban Roy, Daniel Elenius, Frederick Heim, Brian Swenson, Sydney Whittington, James Walker, Theodore Bapty, Joseph Hite, Karthik Ramani, Christopher McComb, Susmit Jha
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Separate Normalization in Self-supervised Transformers
Xiaohui Chen, Yinkai Wang, Yuanqi Du, Soha Hassoun, Liping Liu
Self-supervised training methods for transformers have demonstrated remarkable performance across various domains. Previous transformer-based models, such as masked autoencoders (MAE), typically utilize a single normalization layer for both the [CLS] symbol and the tokens. We propose in this paper a simple modification that employs separate normalization layers for the tokens and the [CLS] symbol to better capture their distinct characteristics and enhance downstream task performance. Our method aims to alleviate the potential negative effects of using the same normalization statistics for both token types, which may not be optimally aligned with their individual roles. We empirically show that by utilizing a separate normalization layer, the [CLS] embeddings can better encode the global contextual information and are distributed more uniformly in its anisotropic space. When replacing the conventional normalization layer with the two separate layers, we observe an average 2.7% performance improvement over the image, natural language, and graph domains.
****
PAD: A Dataset and Benchmark for Pose-agnostic Anomaly Detection
Qiang Zhou, Weize Li, Lihan Jiang, Guoliang Wang, Guyue Zhou, Shanghang Zhang, Hao Zhao
Object anomaly detection is an important problem in the field of machine vision and has seen remarkable progress recently. However, two significant challenges hinder its research and application. First, existing datasets lack comprehensive visual information from various pose angles. They usually have an unrealistic assumption that the anomaly-free training dataset is pose-aligned, and the testing samples have the same pose as the training data. However, in practice, anomaly may exist in any regions on a object, the training and query samples may have different poses, calling for the study on pose-agnostic anomaly detection. Second, the absence of a consensus on experimental protocols for pose-agnostic anomaly detection leads to unfair comparisons of different methods, hindering the research on pose-agnostic anomaly detection. To address these issues, we develop Multi-pose Anomaly Detection (MAD) dataset and Pose-agnostic Anomaly Detection (PAD) benchmark, which takes the first step to address the pose-agnostic anomaly detec

tion problem. Specifically, we build MAD using 20 complex-shaped LEGO toys inclu
ding 4K views with various poses, and high-quality and diverse 3D anomalies in b
oth simulated and real environments. Additionally, we propose a novel method Omn
iposeAD, trained using MAD, specifically designed for pose-agnostic anomaly dete
ction. Through comprehensive evaluations, we demonstrate the relevance of our da
taset and method. Furthermore, we provide an open-source benchmark library, incl
uding dataset and baseline methods that cover 8 anomaly detection paradigms, to
facilitate future research and application in this domain. Code, data, and model
s are publicly available at https://github.com/EricLee0224/PAD.
****
Modulated Neural ODEs
Ilze Amanda Auzina, Ça■atay Y■ld■z, Sara Magliacane, Matthias Bethge, Efstratios
 Gavves
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
DrugCLIP: Contrastive Protein-Molecule Representation Learning for Virtual Scree
ning
Bowen Gao, Bo Qiang, Haichuan Tan, Yinjun Jia, Minsi Ren, Minsi Lu, Jingjing Liu
, Wei-Ying Ma, Yanyan Lan
Virtual screening, which identifies potential drugs from vast compound databases
 to bind with a particular protein pocket, is a critical step in AI-assisted dru
g discovery. Traditional docking methods are highly time-consuming, and can only
 work with a restricted search library in real-life applications. Recent supervi
sed learning approaches using scoring functions for binding-affinity prediction,
 although promising, have not yet surpassed docking methods due to their strong
dependency on limited data with reliable binding-affinity labels. In this paper,
 we propose a novel contrastive learning framework, DrugCLIP, by reformulating v
irtual screening as a dense retrieval task and employing contrastive learning to
 align representations of binding protein pockets and molecules from a large qua
ntity of pairwise data without explicit binding-affinity scores. We also introdu
ce a biological-knowledge inspired data augmentation strategy to learn better pr
otein-molecule representations. Extensive experiments show that DrugCLIP signifi
cantly outperforms traditional docking and supervised learning methods on divers
e virtual screening benchmarks with highly reduced computation time, especially
in zero-shot setting.
****
On the Convergence of Black-Box Variational Inference
Kyurae Kim, Jisu Oh, Kaiwen Wu, Yian Ma, Jacob Gardner
We provide the first convergence guarantee for black-box variational inference (
BBVI) with the reparameterization gradient. While preliminary investigations wo
rked on simplified versions of BBVI (e.g., bounded domain, bounded support, only
 optimizing for the scale, and such), our setup does not need any such algorithm
ic modifications. Our results hold for log-smooth posterior densities with and
without strong log-concavity and the location-scale variational family. Notably
, our analysis reveals that certain algorithm design choices commonly employed i
n practice, such as nonlinear parameterizations of the scale matrix, can result
in suboptimal convergence rates. Fortunately, running BBVI with proximal stocha
stic gradient descent fixes these limitations and thus achieves the strongest kn
own convergence guarantees. We evaluate this theoretical insight by comparing p
roximal SGD against other standard implementations of BBVI on large-scale Bayesi
an inference problems.
****
DRAUC: An Instance-wise Distributionally Robust AUC Optimization Framework
Siran Dai, Qianqian Xu, Zhiyong Yang, Xiaochun Cao, Qingming Huang
The Area Under the ROC Curve (AUC) is a widely employed metric in long-tailed cl
assification scenarios. Nevertheless, most existing methods primarily assume tha
t training and testing examples are drawn i.i.d. from the same distribution, whi

ch is often unachievable in practice. Distributionally Robust Optimization (DRO) enhances model performance by optimizing it for the local worst-case scenario, but directly integrating AUC optimization with DRO results in an intractable optimization problem. To tackle this challenge, methodically we propose an instance-wise surrogate loss of Distributionally Robust AUC (DRAUC) and build our optimization framework on top of it. Moreover, we highlight that conventional DRAUC may induce label bias, hence introducing distribution-aware DRAUC as a more suitable metric for robust AUC learning. Theoretically, we affirm that the generalization gap between the training loss and testing error diminishes if the training set is sufficiently large. Empirically, experiments on corrupted benchmark datasets demonstrate the effectiveness of our proposed method. Code is available at: https://github.com/EldercatSAM/DRAUC.
****
iSCAN: Identifying Causal Mechanism Shifts among Nonlinear Additive Noise Models
Tianyu Chen, Kevin Bello, Bryon Aragam, Pradeep Ravikumar
Structural causal models (SCMs) are widely used in various disciplines to represent causal relationships among variables in complex systems.Unfortunately, the underlying causal structure is often unknown, and estimating it from data remains a challenging task. In many situations, however, the end goal is to localize the changes (shifts) in the causal mechanisms between related datasets instead of learning the full causal structure of the individual datasets. Some applications include root cause analysis, analyzing gene regulatory network structure changes between healthy and cancerous individuals, or explaining distribution shifts. This paper focuses on identifying the causal mechanism shifts in two or more related datasets over the same set of variables---without estimating the entire DAG structure of each SCM.Prior work under this setting assumed linear models with Gaussian noises; instead, in this work we assume that each SCM belongs to the more general class of nonlinear additive noise models (ANMs).A key technical contribution of this work is to show that the Jacobian of the score function for the mixture distribution allows for the identification of shifts under general non-parametric functional mechanisms.Once the shifted variables are identified, we leverage recent work to estimate the structural differences, if any, for the shifted variables.Experiments on synthetic and real-world data are provided to showcase the applicability of this approach.Code implementing the proposed method is open-source and publicly available at https://github.com/kevinsbello/iSCAN.
****
Optimal Learners for Realizable Regression: PAC Learning and Online Learning
Idan Attias, Steve Hanneke, Alkis Kalavasis, Amin Karbasi, Grigoris Velegkas
In this work, we aim to characterize the statistical complexity of realizable regression both in the PAC learning setting and the online learning setting. Previous work had established the sufficiency of finiteness of the fat shattering dimension for PAC learnability and the necessity of finiteness of the scaled Natarajan dimension, but little progress had been made towards a more complete characterization since the work of Simon 1997 (SICOMP '97). To this end, we first introduce a minimax instance optimal learner for realizable regression and propose a novel dimension that both qualitatively and quantitatively characterizes which classes of real-valued predictors are learnable. We then identify a combinatorial dimension related to the graph dimension that characterizes ERM learnability in the realizable setting. Finally, we establish a necessary condition for learnability based on a combinatorial dimension related to the DS dimension, and conjecture that it may also be sufficient in this context. Additionally, in the context of online learning we provide a dimension that characterizes the minimax instance optimal cumulative loss up to a constant factor and design an optimal online learner for realizable regression, thus resolving an open question raised by Daskalakis and Golowich in STOC '22.
****
Sounding Bodies: Modeling 3D Spatial Sound of Humans Using Body Pose and Audio
Xudong XU, Dejan Markovic, Jacob Sandakly, Todd Keebler, Steven Krenn, Alexander Richard
While 3D human body modeling has received much attention in computer vision, mod

eling the acoustic equivalent, i.e. modeling 3D spatial audio produced by body m
otion and speech, has fallen short in the community. To close this gap, we prese
nt a model that can generate accurate 3D spatial audio for full human bodies. Th
e system consumes, as input, audio signals from headset microphones and body pos
e, and produces, as output, a 3D sound field surrounding the transmitter's body,
 from which spatial audio can be rendered at any arbitrary position in the 3D sp
ace. We collect a first-of-its-kind multimodal dataset of human bodies, recorded
 with multiple cameras and a spherical array of 345 microphones. In an empirical
 evaluation, we demonstrate that our model can produce accurate body-induced sou
nd fields when trained with a suitable loss. Dataset and code are available onli
ne.
****
Large Language Models for Automated Data Science: Introducing CAAFE for Context-
Aware Automated Feature Engineering
Noah Hollmann, Samuel Müller, Frank Hutter
As the field of automated machine learning (AutoML) advances, it becomes increas
ingly important to incorporate domain knowledge into these systems.We present an
 approach for doing so by harnessing the power of large language models (LLMs).
Specifically, we introduce Context-Aware Automated Feature Engineering (CAAFE),
a feature engineering method for tabular datasets that utilizes an LLM to iterat
ively generate additional semantically meaningful features for tabular datasets
based on the description of the dataset. The method produces both Python code fo
r creating new features and explanations for the utility of the generated featur
es.Despite being methodologically simple, CAAFE improves performance on 11 out o
f 14 datasets -- boosting mean ROC AUC performance from 0.798 to 0.822 across al
l dataset - similar to the improvement achieved by using a random forest instead
 of logistic regression on our datasets. Furthermore, CAAFE is interpretable by
providing a textual explanation for each generated feature.CAAFE paves the way f
or more extensive semi-automation in data science tasks and emphasizes the signi
ficance of context-aware solutions that can extend the scope of AutoML systems t
o semantic AutoML. We release our code, a simple demo and a python package.
****
LIBERO: Benchmarking Knowledge Transfer for Lifelong Robot Learning
Bo Liu, Yifeng Zhu, Chongkai Gao, Yihao Feng, Qiang Liu, Yuke Zhu, Peter Stone
Lifelong learning offers a promising paradigm of building a generalist agent tha
t learns and adapts over its lifespan. Unlike traditional lifelong learning prob
lems in image and text domains, which primarily involve the transfer of declarat
ive knowledge of entities and concepts, lifelong learning in decision-making (LL
DM) also necessitates the transfer of procedural knowledge, such as actions and
behaviors. To advance research in LLDM, we introduce LIBERO, a novel benchmark o
f lifelong learning for robot manipulation. Specifically, LIBERO highlights five
 key research topics in LLDM: 1) how to efficiently transfer declarative knowled
ge, procedural knowledge, or the mixture of both; 2) how to design effective pol
icy architectures and 3) effective algorithms for LLDM; 4) the robustness of a l
ifelong learner with respect to task ordering; and 5) the effect of model pretra
ining for LLDM. We develop an extendible procedural generation pipeline that can
 in principle generate infinitely many tasks. For benchmarking purpose, we creat
e four task suites (130 tasks in total) that we use to investigate the above-men
tioned research topics. To support sample-efficient learning, we provide high-qu
ality human-teleoperated demonstration data for all tasks. Our extensive experim
ents present several insightful or even unexpected discoveries: sequential finet
uning outperforms existing lifelong learning methods in forward transfer, no sin
gle visual encoder architecture excels at all types of knowledge transfer, and n
aive supervised pretraining can hinder agents' performance in the subsequent LLD
M.
****
On quantum backpropagation, information reuse, and cheating measurement collapse
Amira Abbas, Robbie King, Hsin-Yuan Huang, William J. Huggins, Ramis Movassagh,
Dar Gilboa, Jarrod McClean
The success of modern deep learning hinges on the ability to train neural networ

ks at scale. Through clever reuse of intermediate information, backpropagation facilitates training through gradient computation at a total cost roughly proportional to running the function, rather than incurring an additional factor proportional to the number of parameters -- which can now be in the trillions. Naively, one expects that quantum measurement collapse entirely rules out the reuse of quantum information as in backpropagation. But recent developments in shadow tomography, which assumes access to multiple copies of a quantum state, have challenged that notion. Here, we investigate whether parameterized quantum models can train as efficiently as classical neural networks. We show that achieving backpropagation scaling is impossible without access to multiple copies of a state. With this added ability, we introduce an algorithm with foundations in shadow tomography that matches backpropagation scaling in quantum resources while reducing classical auxiliary computational costs to open problems in shadow tomography. These results highlight the nuance of reusing quantum information for practical purposes and clarify the unique difficulties in training large quantum models, which could alter the course of quantum machine learning.
****
First Order Methods with Markovian Noise: from Acceleration to Variational Inequalities
Aleksandr Beznosikov, Sergey Samsonov, Marina Sheshukova, Alexander Gasnikov, Alexey Naumov, Eric Moulines
This paper delves into stochastic optimization problems that involve Markovian noise. We present a unified approach for the theoretical analysis of first-order gradient methods for stochastic optimization and variational inequalities. Our approach covers scenarios for both non-convex and strongly convex minimization problems. To achieve an optimal (linear) dependence on the mixing time of the underlying noise sequence, we use the randomized batching scheme, which is based on the multilevel Monte Carlo method. Moreover, our technique allows us to eliminate the limiting assumptions of previous research on Markov noise, such as the need for a bounded domain and uniformly bounded stochastic gradients. Our extension to variational inequalities under Markovian noise is original. Additionally, we provide lower bounds that match the oracle complexity of our method in the case of strongly convex optimization problems.
****
Fair, Polylog-Approximate Low-Cost Hierarchical Clustering
Marina Knittel, Max Springer, John Dickerson, MohammadTaghi Hajiaghayi
Research in fair machine learning, and particularly clustering, has been crucial in recent years given the many ethical controversies that modern intelligent systems have posed. Ahmadian et al. [2020] established the study of fairness in hierarchical clustering, a stronger, more structured variant of its well-known flat counterpart, though their proposed algorithm that optimizes for Dasgupta's [2016] famous cost function was highly theoretical. Knittel et al. [2023] then proposed the first practical fair approximation for cost, however they were unable to break the polynomial-approximate barrier they posed as a hurdle of interest. We break this barrier, proposing the first truly polylogarithmic-approximate low-cost fair hierarchical clustering, thus greatly bridging the gap between the best fair and vanilla hierarchical clustering approximations.
****
A Novel Approach for Effective Multi-View Clustering with Information-Theoretic Perspective
Chenhang Cui, Yazhou Ren, Jingyu Pu, Jiawei Li, Xiaorong Pu, Tianyi Wu, Yutao Shi, Lifang He
Multi-view clustering (MVC) is a popular technique for improving clustering performance using various data sources. However, existing methods primarily focus on acquiring consistent information while often neglecting the issue of redundancy across multiple views.This study presents a new approach called Sufficient Multi-View Clustering (SUMVC) that examines the multi-view clustering framework from an information-theoretic standpoint. Our proposed method consists of two parts. Firstly, we develop a simple and reliable multi-view clustering method SCMVC (simple consistent multi-view clustering) that employs variational analysis to gen

erate consistent information. Secondly, we propose a sufficient representation lower bound to enhance consistent information and minimise unnecessary information among views. The proposed SUMVC method offers a promising solution to the problem of multi-view clustering and provides a new perspective for analyzing multi-view data. To verify the effectiveness of our model, we conducted a theoretical analysis based on the Bayes Error Rate, and experiments on multiple multi-view datasets demonstrate the superior performance of SUMVC.
****
OpenShape: Scaling Up 3D Shape Representation Towards Open-World Understanding
Minghua Liu, Ruoxi Shi, Kaiming Kuang, Yinhao Zhu, Xuanlin Li, Shizhong Han, Hong Cai, Fatih Porikli, Hao Su
We introduce OpenShape, a method for learning multi-modal joint representations of text, image, and point clouds. We adopt the commonly used multi-modal contrastive learning framework for representation alignment, but with a specific focus on scaling up 3D representations to enable open-world 3D shape understanding. To achieve this, we scale up training data by ensembling multiple 3D datasets and propose several strategies to automatically filter and enrich noisy text descriptions. We also explore and compare strategies for scaling 3D backbone networks and introduce a novel hard negative mining module for more efficient training. We evaluate OpenShape on zero-shot 3D classification benchmarks and demonstrate its superior capabilities for open-world recognition. Specifically, OpenShape achieves a zero-shot accuracy of 46.8% on the 1,156-category Objaverse-LVIS benchmark, compared to less than 10% for existing methods. OpenShape also achieves an accuracy of 85.3% on ModelNet40, outperforming previous zero-shot baseline methods by 20% and performing on par with some fully-supervised methods. Furthermore, we show that our learned embeddings encode a wide range of visual and semantic concepts (e.g., subcategories, color, shape, style) and facilitate fine-grained text-3D and image-3D interactions. Due to their alignment with CLIP embeddings, our learned shape representations can also be integrated with off-the-shelf CLIP-based models for various applications, such as point cloud captioning and point cloud-conditioned image generation.
****
KuaiSim: A Comprehensive Simulator for Recommender Systems
Kesen Zhao, Shuchang Liu, Qingpeng Cai, Xiangyu Zhao, Ziru Liu, Dong Zheng, Peng Jiang, Kun Gai
Reinforcement Learning (RL)-based recommender systems (RSs) have garnered considerable attention due to their ability to learn optimal recommendation policies and maximize long-term user rewards. However, deploying RL models directly in online environments and generating authentic data through A/B tests can pose challenges and require substantial resources. Simulators offer an alternative approach by providing training and evaluation environments for RS models, reducing reliance on real-world data. Existing simulators have shown promising results but also have limitations such as simplified user feedback, lacking consistency with real-world data, the challenge of simulator evaluation, and difficulties in migration and expansion across RSs.To address these challenges, we propose KuaiSim, a comprehensive user environment that provides user feedback with multi-behavior and cross-session responses.The resulting simulator can support three levels of recommendation problems: the request level list-wise recommendation task, the whole-session level sequential recommendation task, and the cross-session level retention optimization task. For each task, KuaiSim also provides evaluation protocols and baseline recommendation algorithms that further serve as benchmarks for future research. We also restructure existing competitive simulators on the Kuairand Dataset and compare them against KuaiSim to future assess their performance and behavioral differences. Furthermore, to showcase KuaiSim's flexibility in accommodating different datasets, we demonstrate its versatility and robustness when deploying it on the ML-1m dataset. The implementation code is available online to ease reproducibility \footnote{https://github.com/Applied-Machine-Learning-Lab/KuaiSim}.
****
Optimizing over trained GNNs via symmetry breaking

Shiqiang Zhang, Juan Campos, Christian Feldmann, David Walz, Frederik Sandfort, Miriam Mathea, Calvin Tsay, Ruth Misener

Optimization over trained machine learning models has applications including: verification, minimizing neural acquisition functions, and integrating a trained surrogate into a larger decision-making problem. This paper formulates and solves optimization problems constrained by trained graph neural networks (GNNs). To circumvent the symmetry issue caused by graph isomorphism, we propose two types of symmetry-breaking constraints: one indexing a node 0 and one indexing the remaining nodes by lexicographically ordering their neighbor sets. To guarantee that adding these constraints will not remove all symmetric solutions, we construct a graph indexing algorithm and prove that the resulting graph indexing satisfies the proposed symmetry-breaking constraints. For the classical GNN architectures considered in this paper, optimizing over a GNN with a fixed graph is equivalent to optimizing over a dense neural network. Thus, we study the case where the input graph is not fixed, implying that each edge is a decision variable, and develop two mixed-integer optimization formulations. To test our symmetry-breaking strategies and optimization formulations, we consider an application in molecular design.

****

REx: Data-Free Residual Quantization Error Expansion

Edouard YVINEC, Arnaud Dapogny, Matthieu Cord, Kevin Bailly

Deep neural networks (DNNs) are ubiquitous in computer vision and natural language processing, but suffer from high inference cost. This problem can be addressed by quantization, which consists in converting floating point operations into a lower bit-width format. With the growing concerns on privacy rights, we focus our efforts on data-free methods. However, such techniques suffer from their lack of adaptability to the target devices, as a hardware typically only supports specific bit widths. Thus, to adapt to a variety of devices, a quantization method shall be flexible enough to find good accuracy v.s. speed trade-offs for every bit width and target device. To achieve this, we propose REx, a quantization method that leverages residual error expansion, along with group sparsity. We show experimentally that REx enables better trade-offs (in terms of accuracy given any target bit-width) on both convnets and transformers for computer vision, as well as NLP models. In particular, when applied to large language models, we show that REx elegantly solves the outlier problem that hinders state-of-the-art quantization methods.In addition, REx is backed off by strong theoretical guarantees on the preservation of the predictive function of the original model. Lastly, we show that REx is agnostic to the quantization operator and can be used in combination with previous quantization work.

****

A Unified, Scalable Framework for Neural Population Decoding

Mehdi Azabou, Vinam Arora, Venkataramana Ganesh, Ximeng Mao, Santosh Nachimuthu, Michael Mendelson, Blake Richards, Matthew Perich, Guillaume Lajoie, Eva Dyer

Our ability to use deep learning approaches to decipher neural activity would likely benefit from greater scale, in terms of both the model size and the datasets. However, the integration of many neural recordings into one unified model is challenging, as each recording contains the activity of different neurons from different individual animals. In this paper, we introduce a training framework and architecture designed to model the population dynamics of neural activity across diverse, large-scale neural recordings. Our method first tokenizes individual spikes within the dataset to build an efficient representation of neural events that captures the fine temporal structure of neural activity. We then employ cross-attention and a PerceiverIO backbone to further construct a latent tokenization of neural population activities. Utilizing this architecture and training framework, we construct a large-scale multi-session model trained on large datasets from seven nonhuman primates, spanning over 158 different sessions of recording from over 27,373 neural units and over 100 hours of recordings. In a number of different tasks, we demonstrate that our pretrained model can be rapidly adapted to new, unseen sessions with unspecified neuron correspondence, enabling few-shot performance with minimal labels. This work presents a powerful new approach

for building deep learning tools to analyze neural data and stakes out a clear p
ath to training at scale for neural decoding models.
****
Species196: A One-Million Semi-supervised Dataset for Fine-grained Species Recog
nition
Wei He, Kai Han, Ying Nie, Chengcheng Wang, Yunhe Wang
The development of foundation vision models has pushed the general visual recogn
ition to a high level, but cannot well address the fine-grained recognition in s
pecialized domain such as invasive species classification. Identifying and manag
ing invasive species has strong social and ecological value. Currently, most inv
asive species datasets are limited in scale and cover a narrow range of species,
 which restricts the development of deep-learning based invasion biometrics syst
ems. To fill the gap of this area, we introduced Species196, a large-scale semi-
supervised dataset of 196-category invasive species. It collects over 19K images
 with expert-level accurate annotations (Species196-L), and 1.2M unlabeled image
s of invasive species (Species196-U). The dataset provides four experimental set
tings for benchmarking the existing models and algorithms, namely, supervised le
arning, semi-supervised learning and self-supervised pretraining. To facilitate
future research on these four learning paradigms, we conduct an empirical study
of the representative methods on the introduced dataset. The dataset will be mad
e publicly available at https://species-dataset.github.io/.
****
PTADisc: A Cross-Course Dataset Supporting Personalized Learning in Cold-Start S
cenarios
Liya Hu, Zhiang Dong, Jingyuan Chen, Guifeng Wang, Zhihua Wang, Zhou Zhao, Fei W
u
The focus of our work is on diagnostic tasks in personalized learning, such as c
ognitive diagnosis and knowledge tracing. The goal of these tasks is to assess s
tudents' latent proficiency on knowledge concepts through analyzing their histor
ical learning records. However, existing research has been limited to single-cou
rse scenarios; cross-course studies have not been explored due to a lack of data
set. We address this issue by constructing PTADisc, a Diverse, Immense, Student-
centered dataset that emphasizes its sufficient Cross-course information for per
sonalized learning. PTADisc includes 74 courses, 1,530,100 students, 4,054 conce
pts, 225,615 problems, and over 680 million student response logs. Based on PTAD
isc, we developed a model-agnostic Cross-Course Learner Modeling Framework (CCLM
F) which utilizes relationships between students' proficiency across courses to
alleviate the difficulty of diagnosing student knowledge state in cold-start sce
narios. CCLMF uses a meta network to generate personalized mapping functions bet
ween courses. The experimental results on PTADisc verify the effectiveness of CC
LMF with an average improvement of 4.2% on AUC. We also report the performance o
f baseline models for cognitive diagnosis and knowledge tracing over PTADisc, de
monstrating that our dataset supports a wide scope of research in personalized l
earning. Additionally, PTADisc contains valuable programming logs and student-gr
oup information that are worth exploring in the future.
****
Adaptive Contextual Perception: How To Generalize To New Backgrounds and Ambiguo
us Objects
Zhuofan Ying, Peter Hase, Mohit Bansal
Biological vision systems make adaptive use of context to recognize objects in n
ew settings with novel contexts as well as occluded or blurry objects in familia
r settings. In this paper, we investigate how vision models adaptively use conte
xt for out-of-distribution (OOD) generalization and leverage our analysis result
s to improve model OOD generalization. First, we formulate two distinct OOD sett
ings where the contexts are either beneficial Object-Disambiguation or irrelevan
t Background-Invariance, reflecting the diverse contextual challenges faced in b
iological vision. We then analyze model performance in these two different OOD s
ettings and demonstrate that models that excel in one setting tend to struggle i
n the other. Notably, prior works on learning causal features improve on one set
ting but hurt on the other. This underscores the importance of generalizing acro

ss both OOD settings, as this ability is crucial for both human cognition and robust AI systems. Next, to better understand the model properties contributing to OOD generalization, we use representational geometry analysis and our own probing methods to examine a population of models, and we discover that those with more factorized representations and appropriate feature weighting are more successful in handling Object-Disambiguation and Background-Invariance tests. We further validate these findings through causal intervention, manipulating representation factorization and feature weighting to demonstrate their causal effect on performance. Motivated by our analysis results, we propose new augmentation methods aimed at enhancing model generalization. The proposed methods outperform strong baselines, yielding improvements in both in-distribution and OOD tests. We conclude that, in order to replicate the generalization abilities of biological vision, computer vision models must have factorized object vs. background representations and appropriately weigh both kinds of features.
****
PUG: Photorealistic and Semantically Controllable Synthetic Data for Representation Learning
Florian Bordes, Shashank Shekhar, Mark Ibrahim, Diane Bouchacourt, Pascal Vincent, Ari Morcos
Synthetic image datasets offer unmatched advantages for designing and evaluating deep neural networks: they make it possible to (i) render as many data samples as needed, (ii) precisely control each scene and yield granular ground truth labels (and captions), (iii) precisely control distribution shifts between training and testing to isolate variables of interest for sound experimentation.Despite such promise, the use of synthetic image data is still limited -- and often played down -- mainly due to their lack of realism. Most works therefore rely on datasets of real images, which have often been scraped from public images on the internet, and may have issues with regards to privacy, bias, and copyright, while offering little control over how objects precisely appear.In this work, we present a path to democratize the use of photorealistic synthetic data: we develop a new generation of interactive environments for representation learning research, that offer both controllability and realism. We use the Unreal Engine, a powerful game engine well known in the entertainment industry, to produce PUG (Photorealistic Unreal Graphics) environments and datasets for representation learning. Using PUG for evaluation and fine-tuning, we demonstrate the potential of PUG to both enable more rigorous evaluations and to improve model training.
****
On the Gini-impurity Preservation For Privacy Random Forests
XinRan Xie, Man-Jie Yuan, Xuetong Bai, Wei Gao, Zhi-Hua Zhou
Random forests have been one successful ensemble algorithms in machine learning. Various techniques have been utilized to preserve the privacy of random forests from anonymization, differential privacy, homomorphic encryption, etc., whereas it rarely takes into account some crucial ingredients of learning algorithm. This work presents a new encryption to preserve data's Gini impurity, which plays a crucial role during the construction of random forests. Our basic idea is to modify the structure of binary search tree to store several examples in each node, and encrypt data features by incorporating label and order information. Theoretically, we prove that our scheme preserves the minimum Gini impurity in ciphertexts without decrypting, and present the security guarantee for encryption. For random forests, we encrypt data features based on our Gini-impurity-preserving scheme, and take the homomorphic encryption scheme CKKS to encrypt data labels due to their importance and privacy. We conduct extensive experiments to show the effectiveness, efficiency and security of our proposed method.
****
Debiasing Pretrained Generative Models by Uniformly Sampling Semantic Attributes
Walter Gerych, Kevin Hickey, Luke Buquicchio, Kavin Chandrasekaran, Abdulaziz Alajaji, Elke A. Rundensteiner, Emmanuel Agu
Generative models are being increasingly used in science and industry applications. Unfortunately, they often perpetuate the biases present in their training sets, such as societal biases causing certain groups to be underrepresented in th

e data. For instance, image generators may overwhelmingly produce images of white people due to few non-white samples in their training data. It is imperative to debias generative models so they synthesize an equal number of instances for each group, while not requiring retraining of the model to avoid prohibitive expense. We thus propose a distribution mapping module that produces samples from a fair noise distribution, such that the pretrained generative model produces semantically uniform outputs - an equal number of instances for each group - when conditioned on these samples. This does not involve retraining the generator, nor does it require any real training data. Experiments on debiasing generators trained on popular real-world datasets show that our method outperforms existing approaches.
****
Improved Algorithms for Stochastic Linear Bandits Using Tail Bounds for Martingale Mixtures
Hamish Flynn, David Reeb, Melih Kandemir, Jan R. Peters
We present improved algorithms with worst-case regret guarantees for the stochastic linear bandit problem. The widely used "optimism in the face of uncertainty" principle reduces a stochastic bandit problem to the construction of a confidence sequence for the unknown reward function. The performance of the resulting bandit algorithm depends on the size of the confidence sequence, with smaller confidence sets yielding better empirical performance and stronger regret guarantees. In this work, we use a novel tail bound for adaptive martingale mixtures to construct confidence sequences which are suitable for stochastic bandits. These confidence sequences allow for efficient action selection via convex programming. We prove that a linear bandit algorithm based on our confidence sequences is guaranteed to achieve competitive worst-case regret. We show that our confidence sequences are tighter than competitors, both empirically and theoretically. Finally, we demonstrate that our tighter confidence sequences give improved performance in several hyperparameter tuning tasks.
****
Tame a Wild Camera: In-the-Wild Monocular Camera Calibration
Shengjie Zhu, Abhinav Kumar, Masa Hu, Xiaoming Liu
3D sensing for monocular in-the-wild images, e.g., depth estimation and 3D object detection, has become increasingly important.However, the unknown intrinsic parameter hinders their development and deployment.Previous methods for the monocular camera calibration rely on specific 3D objects or strong geometry prior, such as using a checkerboard or imposing a Manhattan World assumption.This work instead calibrates intrinsic via exploiting the monocular 3D prior.Given an undistorted image as input, our method calibrates the complete 4 Degree-of-Freedom (DoF) intrinsic parameters.First, we show intrinsic is determined by the two well-studied monocular priors: monocular depthmap and surface normal map.However, this solution necessitates a low-bias and low-variance depth estimation.Alternatively, we introduce the incidence field, defined as the incidence rays between points in 3D space and pixels in the 2D imaging plane.We show that: 1) The incidence field is a pixel-wise parametrization of the intrinsic invariant to image cropping and resizing.2) The incidence field is a learnable monocular 3D prior, determined pixel-wisely by up-to-sacle monocular depthmap and surface normal.With the estimated incidence field, a robust RANSAC algorithm recovers intrinsic.We show the effectiveness of our method through superior performance on synthetic and zero-shot testing datasets.Beyond calibration, we demonstrate downstream applications in image manipulation detection \& restoration, uncalibrated two-view pose estimation, and 3D sensing.
****
ATTA: Anomaly-aware Test-Time Adaptation for Out-of-Distribution Detection in Segmentation
Zhitong Gao, Shipeng Yan, Xuming He
Recent advancements in dense out-of-distribution (OOD) detection have primarily focused on scenarios where the training and testing datasets share a similar domain, with the assumption that no domain shift exists between them. However, in real-world situations, domain shift often exits and significantly affects the acc

uracy of existing out-of-distribution (OOD) detection models. In this work, we propose a dual-level OOD detection framework to handle domain shift and semantic shift jointly. The first level distinguishes whether domain shift exists in the image by leveraging global low-level features, while the second level identifies pixels with semantic shift by utilizing dense high-level feature maps. In this way, we can selectively adapt the model to unseen domains as well as enhance model's capacity in detecting novel classes. We validate the efficacy of our proposed method on several OOD segmentation benchmarks, including those with significant domain shifts and those without, observing consistent performance improvements across various baseline models. Code is available at https://github.com/gaozhitong/ATTA.
****
Regression with Cost-based Rejection
Xin Cheng, Yuzhou Cao, Haobo Wang, Hongxin Wei, Bo An, Lei Feng
Learning with rejection is an important framework that can refrain from making predictions to avoid critical mispredictions by balancing between prediction and rejection. Previous studies on cost-based rejection only focused on the classification setting, which cannot handle the continuous and infinite target space in the regression setting. In this paper, we investigate a novel regression problem called regression with cost-based rejection, where the model can reject to make predictions on some examples given certain rejection costs. To solve this problem, we first formulate the expected risk for this problem and then derive the Bayes optimal solution, which shows that the optimal model should reject to make predictions on the examples whose variance is larger than the rejection cost when the mean squared error is used as the evaluation metric. Furthermore, we propose to train the model by a surrogate loss function that considers rejection as binary classification and we provide conditions for the model consistency, which implies that the Bayes optimal solution can be recovered by our proposed surrogate loss. Extensive experiments demonstrate the effectiveness of our proposed method.
****
A State Representation for Diminishing Rewards
Ted Moskovitz, Samo Hromadka, Ahmed Touati, Diana Borsa, Maneesh Sahani
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Unified Segment-to-Segment Framework for Simultaneous Sequence Generation
Shaolei Zhang, Yang Feng
Simultaneous sequence generation is a pivotal task for real-time scenarios, such as streaming speech recognition, simultaneous machine translation and simultaneous speech translation, where the target sequence is generated while receiving the source sequence. The crux of achieving high-quality generation with low latency lies in identifying the optimal moments for generating, accomplished by learning a mapping between the source and target sequences. However, existing methods often rely on task-specific heuristics for different sequence types, limiting the model’s capacity to adaptively learn the source-target mapping and hindering the exploration of multi-task learning for various simultaneous tasks. In this paper, we propose a unified segment-to-segment framework (Seg2Seg) for simultaneous sequence generation, which learns the mapping in an adaptive and unified manner. During the process of simultaneous generation, the model alternates between waiting for a source segment and generating a target segment, making the segment serve as the natural bridge between the source and target. To accomplish this, Seg2Seg introduces a latent segment as the pivot between source to target and explores all potential source-target mappings via the proposed expectation training, thereby learning the optimal moments for generating. Experiments on multiple simultaneous generation tasks demonstrate that Seg2Seg achieves state-of-the-art performance and exhibits better generality across various tasks.
****

DYffusion: A Dynamics-informed Diffusion Model for Spatiotemporal Forecasting
Salva Rühling Cachay, Bo Zhao, Hailey Joren, Rose Yu
While diffusion models can successfully generate data and make predictions, they are predominantly designed for static images. We propose an approach for training diffusion models for dynamics forecasting that leverages the temporal dynamics encoded in the data, directly coupling it with the diffusion steps in the network. We train a stochastic, time-conditioned interpolator and a backbone forecaster networkthat mimic the forward and reverse processes of conventional diffusion models, respectively. This design choice naturally encodes multi-step and long-range forecasting capabilities, allowing for highly flexible, continuous-time sampling trajectories and the ability to trade-off performance with accelerated sampling at inference time. In addition, the dynamics-informed diffusion process imposes a strong inductive bias, allowing for improved computational efficiency compared to traditional Gaussian noise-based diffusion models. Our approach performs competitively on probabilistic skill score metrics in complex dynamics forecasting of sea surface temperatures, Navier-Stokes flows, and spring mesh systems.
****
Towards a Comprehensive Benchmark for High-Level Synthesis Targeted to FPGAs
Yunsheng Bai, Atefeh Sohrabizadeh, Zongyue Qin, Ziniu Hu, Yizhou Sun, Jason Cong
High-level synthesis (HLS) aims to raise the abstraction layer in hardware design, enabling the design of domain-specific accelerators (DSAs) like field-programmable gate arrays (FPGAs) using C/C++ instead of hardware description languages (HDLs). Compiler directives in the form of pragmas play a crucial role in modifying the microarchitecture within the HLS framework. However, the space of possible microarchitectures grows exponentially with the number of pragmas. Moreover, the evaluation of each candidate design using the HLS tool consumes significant time, ranging from minutes to hours, leading to a time-consuming optimization process. To accelerate this process, machine learning models have been used to predict design quality in milliseconds. However, existing open-source datasets for training such models are limited in terms of design complexity and available optimizations. In this paper, we present HLSyn, the first benchmark that addresses these limitations. It contains more complex programs with a wider range of optimization pragmas, making it a comprehensive dataset for training and evaluating design quality prediction models. The HLSyn benchmark consists of 42 unique programs/kernels, resulting in over 42,000 labeled designs. We conduct an extensive comparison of state-of-the-art baselines to assess their effectiveness in predicting design quality. As an ongoing project, we anticipate expanding the HLSyn benchmark in terms of both quantity and variety of programs to further support the development of this field.
****
Energy Discrepancies: A Score-Independent Loss for Energy-Based Models
Tobias Schröder, Zijing Ou, Jen Lim, Yingzhen Li, Sebastian Vollmer, Andrew Duncan
Energy-based models are a simple yet powerful class of probabilistic models, but their widespread adoption has been limited by the computational burden of training them. We propose a novel loss function called Energy Discrepancy (ED) which does not rely on the computation of scores or expensive Markov chain Monte Carlo. We show that energy discrepancy approaches the explicit score matching and negative log-likelihood loss under different limits, effectively interpolating between both. Consequently, minimum energy discrepancy estimation overcomes the problem of nearsightedness encountered in score-based estimation methods, while also enjoying theoretical guarantees. Through numerical experiments, we demonstrate that ED learns low-dimensional data distributions faster and more accurately than explicit score matching or contrastive divergence. For high-dimensional image data, we describe how the manifold hypothesis puts limitations on our approach and demonstrate the effectiveness of energy discrepancy by training the energy-based model as a prior of a variational decoder model.
****
Learning to Group Auxiliary Datasets for Molecule

Tinglin Huang, Ziniu Hu, Rex Ying
The limited availability of annotations in small molecule datasets presents a ch
allenge to machine learning models. To address this, one common strategy is to c
ollaborate with additional auxiliary datasets. However, having more data does no
t always guarantee improvements. Negative transfer can occur when the knowledge
in the target dataset differs or contradicts that of the auxiliary molecule data
sets. In light of this, identifying the auxiliary molecule datasets that can ben
efit the target dataset when jointly trained remains a critical and unresolved p
roblem. Through an empirical analysis, we observe that combining graph structure
 similarity and task similarity can serve as a more reliable indicator for ident
ifying high-affinity auxiliary datasets. Motivated by this insight, we propose M
olGroup, which separates the dataset affinity into task and structure affinity t
o predict the potential benefits of each auxiliary molecule dataset. MolGroup ac
hieves this by utilizing a routing mechanism optimized through a bi-level optimi
zation framework. Empowered by the meta gradient, the routing mechanism is optim
ized toward maximizing the target dataset's performance and quantifies the affin
ity as the gating score. As a result, MolGroup is capable of predicting the opti
mal combination of auxiliary datasets for each target dataset. Our extensive exp
eriments demonstrate the efficiency and effectiveness of MolGroup, showing an av
erage improvement of 4.41%/3.47% for GIN/Graphormer trained with the group of mo
lecule datasets selected by MolGroup on 11 target molecule datasets.
****
Equivariant Spatio-Temporal Attentive Graph Networks to Simulate Physical Dynami
cs
Liming Wu, Zhichao Hou, Jirui Yuan, Yu Rong, Wenbing Huang
Learning to represent and simulate the dynamics of physical systems is a crucial
 yet challenging task. Existing equivariant Graph Neural Network (GNN) based met
hods have encapsulated the symmetry of physics, \emph{e.g.}, translations, rotat
ions, etc, leading to better generalization ability. Nevertheless, their frame-t
o-frame formulation of the task overlooks the non-Markov property mainly incurre
d by unobserved dynamics in the environment. In this paper, we reformulate dynam
ics simulation as a spatio-temporal prediction task, by employing the trajectory
 in the past period to recover the Non-Markovian interactions. We propose Equiva
riant Spatio-Temporal Attentive Graph Networks (ESTAG), an equivariant version o
f spatio-temporal GNNs, to fulfil our purpose. At its core, we design a novel Eq
uivariant Discrete Fourier Transform (EDFT) to extract periodic patterns from th
e history frames, and then construct an Equivariant Spatial Module (ESM) to acco
mplish spatial message passing, and an Equivariant Temporal Module (ETM) with th
e forward attention and equivariant pooling mechanisms to aggregate temporal mes
sage. We evaluate our model on three real datasets corresponding to the molecula
r-, protein- and macro-level. Experimental results verify the effectiveness of E
STAG compared to typical spatio-temporal GNNs and equivariant GNNs.
****
Differentially Private Decoupled Graph Convolutions for Multigranular Topology P
rotection
Eli Chien, Wei-Ning Chen, Chao Pan, Pan Li, Ayfer Ozgur, Olgica Milenkovic
Graph Neural Networks (GNNs) have proven to be highly effective in solving real-
world learning problems that involve graph-structured data. However, GNNs can al
so inadvertently expose sensitive user information and interactions through thei
r model predictions. To address these privacy concerns, Differential Privacy (DP
) protocols are employed to control the trade-off between provable privacy prote
ction and model utility. Applying standard DP approaches to GNNs directly is not
 advisable due to two main reasons. First, the prediction of node labels, which
relies on neighboring node attributes through graph convolutions, can lead to pr
ivacy leakage. Second, in practical applications, the privacy requirements for n
ode attributes and graph topology may differ. In the latter setting, existing DP
-GNN models fail to provide multigranular trade-offs between graph topology priv
acy, node attribute privacy, and GNN utility. To address both limitations, we pr
opose a new framework termed Graph Differential Privacy (GDP), specifically tail
ored to graph learning. GDP ensures both provably private model parameters as we

ll as private predictions. Additionally, we describe a novel unified notion of g raph dataset adjacency to analyze the properties of GDP for different levels of graph topology privacy. Our findings reveal that DP-GNNs, which rely on graph co nvolutions, not only fail to meet the requirements for multigranular graph topol ogy privacy but also necessitate the injection of DP noise that scales at least linearly with the maximum node degree. In contrast, our proposed Differentially Private Decoupled Graph Convolutions (DPDGCs) represent a more flexible and effi cient alternative to graph convolutions that still provides the necessary guaran tees of GDP. To validate our approach, we conducted extensive experiments on sev en node classification benchmarking and illustrative synthetic datasets. The res ults demonstrate that DPDGCs significantly outperform existing DP-GNNs in terms of privacy-utility trade-offs.
****
Team-PSRO for Learning Approximate TMECor in Large Team Games via Cooperative Re inforcement Learning
Stephen McAleer, Gabriele Farina, Gaoyue Zhou, Mingzhi Wang, Yaodong Yang, Tuoma s Sandholm
Recent algorithms have achieved superhuman performance at a number of two-player zero-sum games such as poker and go. However, many real-world situations are mu lti-player games. Zero-sum two-team games, such as bridge and football, involve two teams where each member of the team shares the same reward with every other member of that team, and each team has the negative of the reward of the other t eam. A popular solution concept in this setting, called TMECor, assumes that tea ms can jointly correlate their strategies before play, but are not able to commu nicate during play. This setting is harder than two-player zero-sum games becaus e each player on a team has different information and must use their public acti ons to signal to other members of the team. Prior works either have game-theoret ic guarantees but only work in very small games, or are able to scale to large g ames but do not have game-theoretic guarantees. In this paper we introduce two a lgorithms: Team-PSRO, an extension of PSRO from two-player games to team games, and Team-PSRO Mix-and-Match which improves upon Team PSRO by better using popula tion policies. In Team-PSRO, in every iteration both teams learn a joint best re sponse to the opponent's meta-strategy via reinforcement learning. As the reinfo rcement learning joint best response approaches the optimal best response, Team-PSRO is guaranteed to converge to a TMECor. In experiments on Kuhn poker and Lia r's Dice, we show that a tabular version of Team-PSRO converges to TMECor, and a version of Team PSRO using deep cooperative reinforcement learning beats self-p lay reinforcement learning in the large game of Google Research Football.
****
Learning Linear Causal Representations from Interventions under General Nonlinea r Mixing
Simon Buchholz, Goutham Rajendran, Elan Rosenfeld, Bryon Aragam, Bernhard Schölk opf, Pradeep Ravikumar
We study the problem of learning causal representations from unknown, latent int erventions in a general setting, where the latent distribution is Gaussian but t he mixing function is completely general. We prove strong identifiability result s given unknown single-node interventions, i.e., without having access to the in tervention targets. This generalizes prior works which have focused on weaker cl asses, such as linear maps or paired counterfactual data. This is also the first instance of identifiability from non-paired interventions for deep neural netwo rk embeddings and general causal structures. Our proof relies on carefully uncov ering the high-dimensional geometric structure present in the data distribution after a non-linear density transformation, which we capture by analyzing quadrat ic forms of precision matrices of the latent distributions. Finally, we propose a contrastive algorithm to identify the latent variables in practice and evaluat e its performance on various tasks.
****
Disentanglement via Latent Quantization
Kyle Hsu, William Dorrell, James Whittington, Jiajun Wu, Chelsea Finn
In disentangled representation learning, a model is asked to tease apart a datas

et's underlying sources of variation and represent them independently of one another. Since the model is provided with no ground truth information about these sources, inductive biases take a paramount role in enabling disentanglement. In this work, we construct an inductive bias towards encoding to and decoding from an organized latent space. Concretely, we do this by (i) quantizing the latent space into discrete code vectors with a separate learnable scalar codebook per dimension and (ii) applying strong model regularization via an unusually high weight decay. Intuitively, the latent space design forces the encoder to combinatorially construct codes from a small number of distinct scalar values, which in turn enables the decoder to assign a consistent meaning to each value. Regularization then serves to drive the model towards this parsimonious strategy. We demonstrate the broad applicability of this approach by adding it to both basic data-reconstructing (vanilla autoencoder) and latent-reconstructing (InfoGAN) generative models. For reliable evaluation, we also propose InfoMEC, a new set of metrics for disentanglement that is cohesively grounded in information theory and fixes well-established shortcomings in previous metrics. Together with regularization, latent quantization dramatically improves the modularity and explicitness of learned representations on a representative suite of benchmark datasets. In particular, our quantized-latent autoencoder (QLAE) consistently outperforms strong methods from prior work in these key disentanglement properties without compromising data reconstruction.
****
Variance-Reduced Gradient Estimation via Noise-Reuse in Online Evolution Strategies
Oscar Li, James Harrison, Jascha Sohl-Dickstein, Virginia Smith, Luke Metz
Unrolled computation graphs are prevalent throughout machine learning but present challenges to automatic differentiation (AD) gradient estimation methods when their loss functions exhibit extreme local sensitivtiy, discontinuity, or blackbox characteristics. In such scenarios, online evolution strategies methods are a more capable alternative, while being more parallelizable than vanilla evolution strategies (ES) by interleaving partial unrolls and gradient updates. In this work, we propose a general class of unbiased online evolution strategies methods. We analytically and empirically characterize the variance of this class of gradient estimators and identify the one with the least variance, which we term Noise-Reuse Evolution Strategies (NRES). Experimentally, we show NRES results in faster convergence than existing AD and ES methods in terms of wall-clock time and number of unroll steps across a variety of applications, including learning dynamical systems, meta-training learned optimizers, and reinforcement learning.
****
Beyond Black-Box Advice: Learning-Augmented Algorithms for MDPs with Q-Value Predictions
Tongxin Li, Yiheng Lin, Shaolei Ren, Adam Wierman
We study the tradeoff between consistency and robustness in the context of a single-trajectory time-varying Markov Decision Process (MDP) with untrusted machine-learned advice. Our work departs from the typical approach of treating advice as coming from black-box sources by instead considering a setting where additional information about how the advice is generated is available. We prove a first-of-its-kind consistency and robustness tradeoff given Q-value advice under a general MDP model that includes both continuous and discrete state/action spaces. Our results highlight that utilizing Q-value advice enables dynamic pursuit of the better of machine-learned advice and a robust baseline, thus result in near-optimal performance guarantees, which provably improves what can be obtained solely with black-box advice.
****
Graph Contrastive Learning with Stable and Scalable Spectral Encoding
Deyu Bo, Yuan Fang, Yang Liu, Chuan Shi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****
A Tale of Two Features: Stable Diffusion Complements DINO for Zero-Shot Semantic Correspondence
Junyi Zhang, Charles Herrmann, Junhwa Hur, Luisa Polania Cabrera, Varun Jampani, Deqing Sun, Ming-Hsuan Yang
Text-to-image diffusion models have made significant advances in generating and editing high-quality images. As a result, numerous approaches have explored the ability of diffusion model features to understand and process single images for downstream tasks, e.g., classification, semantic segmentation, and stylization. However, significantly less is known about what these features reveal across multiple, different images and objects. In this work, we exploit Stable Diffusion (SD) features for semantic and dense correspondence and discover that with simple post-processing, SD features can perform quantitatively similar to SOTA representations. Interestingly, the qualitative analysis reveals that SD features have very different properties compared to existing representation learning features, such as the recently released DINOv2: while DINOv2 provides sparse but accurate matches, SD features provide high-quality spatial information but sometimes inaccurate semantic matches. We demonstrate that a simple fusion of these two features works surprisingly well, and a zero-shot evaluation using nearest neighbors on these fused features provides a significant performance gain over state-of-the-art methods on benchmark datasets, e.g., SPair-71k, PF-Pascal, and TSS. We also show that these correspondences can enable interesting applications such as instance swapping in two images. Project page: https://sd-complements-dino.github.io/.
****
SatLM: Satisfiability-Aided Language Models Using Declarative Prompting
Xi Ye, Qiaochu Chen, Isil Dillig, Greg Durrett
Prior work has combined chain-of-thought prompting in large language models (LLMs) with programmatic representations to perform effective and transparent reasoning. While such an approach works well for tasks that only require forward reasoning (e.g., straightforward arithmetic), it is less effective for constraint solving problems that require more sophisticated planning and search. In this paper, we propose a new satisfiability-aided language modeling (SatLM) approach for improving the reasoning capabilities of LLMs. We use an LLM to generate a declarative task specification rather than an imperative program and leverage an off-the-shelf automated theorem prover to derive the final answer. This approach has two key advantages. The declarative specification is closer to the problem description than the reasoning steps are, so the LLM can parse it out of the description more accurately. Furthermore, by offloading the actual reasoning task to an automated theorem prover, our approach can guarantee the correctness of the answer with respect to the parsed specification and avoid planning errors in the solving process. We evaluate SATLM on 8 different datasets and show that it consistently outperforms program-aided LMs in the imperative paradigm. In particular, SATLM outperforms program-aided LMs by 23% on a challenging subset of the GSM arithmetic reasoning dataset; SATLM also achieves a new SoTA on LSAT and BoardgameQA, surpassing previous models that are trained on the respective training sets.
****
A normative theory of social conflict
Sergey Shuvaev, Evgeny Amelchenko, Dmitry Smagin, Natalia Kudryavtseva, Grigori Enikolopov, Alex Koulakov
Social conflict is a survival mechanism yielding both normal and pathological behaviors. To understand its underlying principles, we collected behavioral and whole-brain neural data from mice advancing through stages of social conflict. We modeled the animals’ interactions as a normal-form game using Bayesian inference to account for the partial observability of animals’ strengths. We find that our behavioral and neural data are consistent with the first-level Theory of Mind (1-ToM) model where mice form “primary” beliefs about the strengths of all mice involved and “secondary” beliefs that estimate the beliefs of their opponents. Our model identifies the brain regions that carry the information about these beliefs and offers a framework for studies of social behaviors in partially observa

ble settings.
****
Learning Invariant Representations of Graph Neural Networks via Cluster Generalization
Donglin Xia, Xiao Wang, Nian Liu, Chuan Shi
Graph neural networks (GNNs) have become increasingly popular in modeling graph-structured data due to their ability to learn node representations by aggregating local structure information. However, it is widely acknowledged that the test graph structure may differ from the training graph structure, resulting in a structure shift. In this paper, we experimentally find that the performance of GNNs drops significantly when the structure shift happens, suggesting that the learned models may be biased towards specific structure patterns. To address this challenge, we propose the Cluster Information Transfer (\textbf{CIT}) mechanism, which can learn invariant representations for GNNs, thereby improving their generalization ability to various and unknown test graphs with structure shift. The CIT mechanism achieves this by combining different cluster information with the nodes while preserving their cluster-independent information. By generating nodes across different clusters, the mechanism significantly enhances the diversity of the nodes and helps GNNs learn the invariant representations. We provide a theoretical analysis of the CIT mechanism, showing that the impact of changing clusters during structure shift can be mitigated after transfer. Additionally, the proposed mechanism is a plug-in that can be easily used to improve existing GNNs. We comprehensively evaluate our proposed method on three typical structure shift scenarios, demonstrating its effectiveness in enhancing GNNs' performance.
****
Transformers learn to implement preconditioned gradient descent for in-context learning
Kwangjun Ahn, Xiang Cheng, Hadi Daneshmand, Suvrit Sra
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Linear Time Algorithms for k-means with Multi-Swap Local Search
Junyu Huang, Qilong Feng, Ziyun Huang, Jinhui Xu, Jianxin Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
VaRT: Variational Regression Trees
Sebastian Salazar
Decision trees are a well-established tool in machine learning for classification and regression tasks. In this paper, we introduce a novel non-parametric Bayesian model that uses variational inference to approximate a posterior distribution over the space of stochastic decision trees. We evaluate the model's performance on 18 datasets and demonstrate its competitiveness with other state-of-the-art methods in regression tasks. We also explore its application to causal inference problems. We provide a fully vectorized implementation of our algorithm in PyTorch.
****
STREAMER: Streaming Representation Learning and Event Segmentation in a Hierarchical Manner
Ramy Mounir, Sujal Vijayaraghavan, Sudeep Sarkar
We present a novel self-supervised approach for hierarchical representation learning and segmentation of perceptual inputs in a streaming fashion. Our research addresses how to semantically group streaming inputs into chunks at various levels of a hierarchy while simultaneously learning, for each chunk, robust global representations throughout the domain. To achieve this, we propose STREAMER, an architecture that is trained layer-by-layer, adapting to the complexity of the in

put domain. In our approach, each layer is trained with two primary objectives: making accurate predictions into the future and providing necessary information to other levels for achieving the same objective. The event hierarchy is constructed by detecting prediction error peaks at different levels, where a detected boundary triggers a bottom-up information flow. At an event boundary, the encoded representation of inputs at one layer becomes the input to a higher-level layer. Additionally, we design a communication module that facilitates top-down and bottom-up exchange of information during the prediction process. Notably, our model is fully self-supervised and trained in a streaming manner, enabling a single pass on the training data. This means that the model encounters each input only once and does not store the data. We evaluate the performance of our model on the egocentric EPIC-KITCHENS dataset, specifically focusing on temporal event segmentation. Furthermore, we conduct event retrieval experiments using the learned representations to demonstrate the high quality of our video event representations. Illustration videos and code are available on our project page: https://rammounir.com/publications/streamer
****
Pgx: Hardware-Accelerated Parallel Game Simulators for Reinforcement Learning
Sotetsu Koyamada, Shinri Okano, Soichiro Nishimori, Yu Murata, Keigo Habara, Haruka Kita, Shin Ishii
We propose Pgx, a suite of board game reinforcement learning (RL) environments written in JAX and optimized for GPU/TPU accelerators. By leveraging JAX's auto-vectorization and parallelization over accelerators, Pgx can efficiently scale to thousands of simultaneous simulations over accelerators. In our experiments on a DGX-A100 workstation, we discovered that Pgx can simulate RL environments 10-100x faster than existing implementations available in Python. Pgx includes RL environments commonly used as benchmarks in RL research, such as backgammon, chess, shogi, and Go. Additionally, Pgx offers miniature game sets and baseline models to facilitate rapid research cycles. We demonstrate the efficient training of the Gumbel AlphaZero algorithm with Pgx environments. Overall, Pgx provides high-performance environment simulators for researchers to accelerate their RL experiments. Pgx is available at https://github.com/sotetsuk/pgx.
****
Robust Distributed Learning: Tight Error Bounds and Breakdown Point under Data Heterogeneity
Youssef Allouah, Rachid Guerraoui, Nirupam Gupta, Rafael Pinot, Geovani Rizk
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Pre-RMSNorm and Pre-CRMSNorm Transformers: Equivalent and Efficient Pre-LN Transformers
Zixuan Jiang, Jiaqi Gu, Hanqing Zhu, David Pan
Transformers have achieved great success in machine learning applications.Normalization techniques, such as Layer Normalization (LayerNorm, LN) and Root Mean Square Normalization (RMSNorm), play a critical role in accelerating and stabilizing the training of Transformers.While LayerNorm recenters and rescales input vectors, RMSNorm only rescales the vectors by their RMS value.Despite being more computationally efficient, RMSNorm may compromise the representation ability of Transformers.There is currently no consensus regarding the preferred normalization technique, as some models employ LayerNorm while others utilize RMSNorm, especially in recent large language models.It is challenging to convert Transformers with one normalization to the other type.While there is an ongoing disagreement between the two normalization types,we propose a solution to unify two mainstream Transformer architectures, Pre-LN and Pre-RMSNorm Transformers.By removing the inherent redundant mean information in the main branch of Pre-LN Transformers, we can reduce LayerNorm to RMSNorm, achieving higher efficiency.We further propose the Compressed RMSNorm (CRMSNorm) and Pre-CRMSNorm Transformer based on a lossless compression of the zero-mean vectors.We formally establish the equivalence

of Pre-LN, Pre-RMSNorm, and Pre-CRMSNorm Transformer variants in both training a nd inference.It implies that Pre-LN Transformers can be substituted with Pre-(C) RMSNorm counterparts at almost no cost, offering the same arithmetic functionali ty along with free efficiency improvement.Experiments demonstrate that we can re duce the training and inference time of Pre-LN Transformers by 1% - 10%.
****
Multimodal Clinical Benchmark for Emergency Care (MC-BEC): A Comprehensive Bench mark for Evaluating Foundation Models in Emergency Medicine
Emma Chen, Aman Kansal, Julie Chen, Boyang Tom Jin, Julia Reisler, David E. Kim, Pranav Rajpurkar
We propose the Multimodal Clinical Benchmark for Emergency Care (MC-BEC), a comp rehensive benchmark for evaluating foundation models in Emergency Medicine using a dataset of 100K+ continuously monitored Emergency Department visits from 2020 -2022. MC-BEC focuses on clinically relevant prediction tasks at timescales from minutes to days, including predicting patient decompensation, disposition, and emergency department (ED) revisit, and includes a standardized evaluation framew ork with train-test splits and evaluation metrics. The multimodal dataset includ es a wide range of detailed clinical data, including triage information, prior d iagnoses and medications, continuously measured vital signs, electrocardiogram a nd photoplethysmograph waveforms, orders placed and medications administered thr oughout the visit, free-text reports of imaging studies, and information on ED d iagnosis, disposition, and subsequent revisits. We provide performance baselines for each prediction task to enable the evaluation of multimodal, multitask mode ls. We believe that MC-BEC will encourage researchers to develop more effective, generalizable, and accessible foundation models for multimodal clinical data.
****
AGD: an Auto-switchable Optimizer using Stepwise Gradient Difference for Precond itioning Matrix
Yun Yue, Zhiling Ye, Jiadi Jiang, Yongchao Liu, Ke Zhang
Adaptive optimizers, such as Adam, have achieved remarkable success in deep lear ning. A key component of these optimizers is the so-called preconditioning matri x, providing enhanced gradient information and regulating the step size of each gradient direction. In this paper, we propose a novel approach to designing the preconditioning matrix by utilizing the gradient difference between two successi ve steps as the diagonal elements. These diagonal elements are closely related t o the Hessian and can be perceived as an approximation of the inner product betw een the Hessian row vectors and difference of the adjacent parameter vectors. Ad ditionally, we introduce an auto-switching function that enables the preconditio ning matrix to switch dynamically between Stochastic Gradient Descent (SGD) and the adaptive optimizer. Based on these two techniques, we develop a new optimize r named AGD that enhances the generalization performance. We evaluate AGD on pub lic datasets of Natural Language Processing (NLP), Computer Vision (CV), and Rec ommendation Systems (RecSys). Our experimental results demonstrate that AGD outp erforms the state-of-the-art (SOTA) optimizers, achieving highly competitive or significantly better predictive performance. Furthermore, we analyze how AGD is able to switch automatically between SGD and the adaptive optimizer and its actu al effects on various scenarios. The code is available at https://github.com/int elligent-machine-learning/dlrover/tree/master/atorch/atorch/optimizers.
****
PDP: Parameter-free Differentiable Pruning is All You Need
Minsik Cho, Saurabh Adya, Devang Naik
DNN pruning is a popular way to reduce the size of a model, improve the inferenc elatency, and minimize the power consumption on DNN accelerators. However,existi ng approaches might be too complex, expensive or ineffective to apply toa variet y of vision/language tasks, DNN architectures and to honor structuredpruning con straints. In this paper, we propose an efficient yet effective train-timepruning scheme, Parameter-free Differentiable Pruning (PDP), which offers state-of-the- art qualities in model size, accuracy, and training cost. PDP uses a dynamicfunc tion of weights during training to generate soft pruning masks for the weightsin a parameter-free manner for a given pruning target. While differentiable, thesi

mplicity and efficiency of PDP make it universal enough to deliver state-of-the-artrandom/structured/channel pruning results on various vision and natural languagetasks. For example, for MobileNet-v1, PDP can achieve 68.2% top-1 ImageNet1kaccuracy at 86.6% sparsity, which is 1.7% higher accuracy than those from thestate-of-the-art algorithms. Also, PDP yields over 83.1% accuracy on Multi-GenreNatural Language Inference with 90% sparsity for BERT, while the next best fromthe existing techniques shows 81.5% accuracy. In addition, PDP can be applied tostructured pruning, such as N:M pruning and channel pruning. For 1:4 structuredpruning of ResNet18, PDP improved the top-1 ImageNet1k accuracy by over 3.6%over the state-of-the-art. For channel pruning of ResNet50, PDP reduced the top-1ImageNet1k accuracy by 0.6% from the state-of-the-art.
****
ExPT: Synthetic Pretraining for Few-Shot Experimental Design
Tung Nguyen, Sudhanshu Agrawal, Aditya Grover
Experimental design is a fundamental problem in many science and engineering fields. In this problem, sample efficiency is crucial due to the time, money, and safety costs of real-world design evaluations. Existing approaches either rely on active data collection or access to large, labeled datasets of past experiments, making them impractical in many real-world scenarios. In this work, we address the more challenging yet realistic setting of few-shot experimental design, where only a few labeled data points of input designs and their corresponding values are available. We approach this problem as a conditional generation task, where a model conditions on a few labeled examples and the desired output to generate an optimal input design. To this end, we introduce Experiment Pretrained Transformers (ExPT), a foundation model for few-shot experimental design that employs a novel combination of synthetic pretraining with in-context learning. In ExPT, we only assume knowledge of a finite collection of unlabelled data points from the input domain and pretrain a transformer neural network to optimize diverse synthetic functions defined over this domain. Unsupervised pretraining allows ExPT to adapt to any design task at test time in an in-context fashion by conditioning on a few labeled data points from the target task and generating the candidate optima. We evaluate ExPT on few-shot experimental design in challenging domains and demonstrate its superior generality and performance compared to existing methods. The source code is available at https://github.com/tung-nd/ExPT.git.
****
ToolkenGPT: Augmenting Frozen Language Models with Massive Tools via Tool Embeddings
Shibo Hao, Tianyang Liu, Zhen Wang, Zhiting Hu
Integrating large language models (LLMs) with various tools has led to increased attention in the field. Existing approaches either involve fine-tuning the LLM, which is both computationally costly and limited to a fixed set of tools, or prompting LLMs by in-context tool demonstrations. Although the latter method offers adaptability to new tools, it struggles with the inherent context length constraint of LLMs when many new tools are presented, and mastering a new set of tools with few-shot examples remains challenging, resulting in suboptimal performance. To address these limitations, we propose a novel solution, named ToolkenGPT, wherein LLMs effectively learn to master tools as predicting tokens through tool embeddings for solving complex tasks. In this framework, each tool is transformed into vector embeddings and plugged into the language model head. Once the function is triggered during text generation, the LLM enters a special function mode to execute the tool calls. Our experiments show that function embeddings effectively help LLMs understand tool use and improve on several tasks, including numerical reasoning, knowledge-based question answering and embodied decision-making.
****
CLIP4HOI: Towards Adapting CLIP for Practical Zero-Shot HOI Detection
Yunyao Mao, Jiajun Deng, Wengang Zhou, Li Li, Yao Fang, Houqiang Li
Zero-shot Human-Object Interaction (HOI) detection aims to identify both seen and unseen HOI categories. A strong zero-shot HOI detector is supposed to be not only capable of discriminating novel interactions but also robust to positional d

istribution discrepancy between seen and unseen categories when locating human-object pairs. However, top-performing zero-shot HOI detectors rely on seen and predefined unseen categories to distill knowledge from CLIP and jointly locate human-object pairs without considering the potential positional distribution discrepancy, leading to impaired transferability. In this paper, we introduce CLIP4HOI, a novel framework for zero-shot HOI detection. CLIP4HOI is developed on the vision-language model CLIP and ameliorates the above issues in the following two aspects. First, to avoid the model from overfitting to the joint positional distribution of seen human-object pairs, we seek to tackle the problem of zero-shot HOI detection in a disentangled two-stage paradigm. To be specific, humans and objects are independently identified and all feasible human-object pairs are processed by Human-Object interactor for pairwise proposal generation. Second, to facilitate better transferability, the CLIP model is elaborately adapted into a fine-grained HOI classifier for proposal discrimination, avoiding data-sensitive knowledge distillation. Finally, experiments on prevalent benchmarks show that our CLIP4HOI outperforms previous approaches on both rare and unseen categories, and sets a series of state-of-the-art records under a variety of zero-shot settings.
****
Transformer-based Planning for Symbolic Regression
Parshin Shojaee, Kazem Meidani, Amir Barati Farimani, Chandan Reddy
Symbolic regression (SR) is a challenging task in machine learning that involves finding a mathematical expression for a function based on its values. Recent advancements in SR have demonstrated the effectiveness of pre-trained transformer models in generating equations as sequences, leveraging large-scale pre-training on synthetic datasets and offering notable advantages in terms of inference time over classical Genetic Programming (GP) methods. However, these models primarily rely on supervised pre-training objectives borrowed from text generation and overlook equation discovery goals like accuracy and complexity. To address this, we propose TPSR, a Transformer-based Planning strategy for Symbolic Regression that incorporates Monte Carlo Tree Search planning algorithm into the transformer decoding process. Unlike conventional decoding strategies, TPSR enables the integration of non-differentiable equation verification feedback, such as fitting accuracy and complexity, as external sources of knowledge into the transformer equation generation process. Extensive experiments on various datasets show that our approach outperforms state-of-the-art methods, enhancing the model's fitting-complexity trade-off, extrapolation abilities, and robustness to noise.
****
Exploring Geometry of Blind Spots in Vision models
Sriram Balasubramanian, Gaurang Sriramanan, Vinu Sankar Sadasivan, Soheil Feizi
Despite the remarkable success of deep neural networks in a myriad of settings, several works have demonstrated their overwhelming sensitivity to near-imperceptible perturbations, known as adversarial attacks. On the other hand, prior works have also observed that deep networks can be under-sensitive, wherein large-magnitude perturbations in input space do not induce appreciable changes to network activations. In this work, we study in detail the phenomenon of under-sensitivity in vision models such as CNNs and Transformers, and present techniques to study the geometry and extent of “equi-confidence” level sets of such networks. We propose a Level Set Traversal algorithm that iteratively explores regions of high confidence with respect to the input space using orthogonal components of the local gradients. Given a source image, we use this algorithm to identify inputs that lie in the same equi-confidence level set as the source image despite being perceptually similar to arbitrary images from other classes. We further observe that the source image is linearly connected by a high-confidence path to these inputs, uncovering a star-like structure for level sets of deep networks. Furthermore, we attempt to identify and estimate the extent of these connected higher-dimensional regions over which the model maintains a high degree of confidence.
****
Provable benefits of annealing for estimating normalizing constants: Importance Sampling, Noise-Contrastive Estimation, and beyond

Omar Chehab, Aapo Hyvarinen, Andrej Risteski
Recent research has developed several Monte Carlo methods for estimating the nor malization constant (partition function) based on the idea of annealing. This me ans sampling successively from a path of distributions which interpolate between a tractable "proposal" distribution and the unnormalized "target" distribution. Prominent estimators in this family include annealed importance sampling and an nealed noise-contrastive estimation (NCE). Such methods hinge on a number of des ign choices: which estimator to use, which path of distributions to use and whet her to use a path at all; so far, there is no definitive theory on which choices are efficient. Here, we evaluate each design choice by the asymptotic estimatio n error it produces. First, we show that using NCE is more efficient than the im portance sampling estimator, but in the limit of infinitesimal path steps, the d ifference vanishes. Second, we find that using the geometric path brings down th e estimation error from an exponential to a polynomial function of the parameter distance between the target and proposal distributions. Third, we find that the arithmetic path, while rarely used, can offer optimality properties over the un iversally-used geometric path. In fact, in a particular limit, the optimal path is arithmetic. Based on this theory, we finally propose a two-step estimator to approximate the optimal path in an efficient way.
****
Strategic Distribution Shift of Interacting Agents via Coupled Gradient Flows
Lauren Conger, Franca Hoffmann, Eric Mazumdar, Lillian Ratliff
We propose a novel framework for analyzing the dynamics of distribution shift in real-world systems that captures the feedback loop between learning algorithms and the distributions on which they are deployed. Prior work largely models feed back-induced distribution shift as adversarial or via an overly simplistic distr ibution-shift structure. In contrast, we propose a coupled partial differential equation model that captures fine-grained changes in the distribution over time by accounting for complex dynamics that arise due to strategic responses to algo rithmic decision-making, non-local endogenous population interactions, and other exogenous sources of distribution shift. We consider two common settings in mac hine learning: cooperative settings with information asymmetries, and competitiv e settings where a learner faces strategic users. For both of these settings, wh en the algorithm retrains via gradient descent, we prove asymptotic convergence of the retraining procedure to a steady-state, both in finite and in infinite di mensions, obtaining explicit rates in terms of the model parameters. To do so we derive new results on the convergence of coupled PDEs that extends what is know n on multi-species systems. Empirically, we show that our approach captures well -documented forms of distribution shifts like polarization and disparate impacts that simpler models cannot capture.
****
Learning Time-Invariant Representations for Individual Neurons from Population D ynamics
Lu Mi, Trung Le, Tianxing He, Eli Shlizerman, Uygar Sümbül
Neurons can display highly variable dynamics. While such variability presumably supports the wide range of behaviors generated by the organism, their gene expre ssions are relatively stable in the adult brain. This suggests that neuronal act ivity is a combination of its time-invariant identity and the inputs the neuron receives from the rest of the circuit. Here, we propose a self-supervised learni ng based method to assign time-invariant representations to individual neurons b ased on permutation-, and population size-invariant summary of population record ings. We fit dynamical models to neuronal activity to learn a representation by considering the activity of both the individual and the neighboring population. Our self-supervised approach and use of implicit representations enable robust i nference against imperfections such as partial overlap of neurons across session s, trial-to-trial variability, and limited availability of molecular (transcript omic) labels for downstream supervised tasks. We demonstrate our method on a pub lic multimodal dataset of mouse cortical neuronal activity and transcriptomic la bels. We report >35\% improvement in predicting the transcriptomic subclass iden tity and >20\% improvement in predicting class identity with respect to the stat

e-of-the-art.
****
GeoTMI: Predicting Quantum Chemical Property with Easy-to-Obtain Geometry via Positional Denoising
Hyeonsu Kim, Jeheon Woo, SEONGHWAN KIM, Seokhyun Moon, Jun Hyeong Kim, Woo Youn Kim
As quantum chemical properties have a dependence on their geometries, graph neural networks (GNNs) using 3D geometric information have achieved high prediction accuracy in many tasks. However, they often require 3D geometries obtained from high-level quantum mechanical calculations, which are practically infeasible, limiting their applicability to real-world problems. To tackle this, we propose a new training framework, GeoTMI, that employs denoising process to predict properties accurately using easy-to-obtain geometries (corrupted versions of correct geometries, such as those obtained from low-level calculations). Our starting point was the idea that the correct geometry is the best description of the target property. Hence, to incorporate information of the correct, GeoTMI aims to maximize mutual information between three variables: the correct and the corrupted geometries and the property. GeoTMI also explicitly updates the corrupted input to approach the correct geometry as it passes through the GNN layers, contributing to more effective denoising. We investigated the performance of the proposed method using 3D GNNs for three prediction tasks: molecular properties, a chemical reaction property, and relaxed energy in a heterogeneous catalytic system. Our results showed consistent improvements in accuracy across various tasks, demonstrating the effectiveness and robustness of GeoTMI.
****
PRED: Pre-training via Semantic Rendering on LiDAR Point Clouds
Hao Yang, Haiyang Wang, Di Dai, Liwei Wang
Pre-training is crucial in 3D-related fields such as autonomous driving where point cloud annotation is costly and challenging. Many recent studies on point cloud pre-training, however, have overlooked the issue of incompleteness, where only a fraction of the points are captured by LiDAR, leading to ambiguity during the training phase. On the other hand, images offer more comprehensive information and richer semantics that can bolster point cloud encoders in addressing the incompleteness issue inherent in point clouds. Yet, incorporating images into point cloud pre-training presents its own challenges due to occlusions, potentially causing misalignments between points and pixels. In this work, we propose PRED, a novel image-assisted pre-training framework for outdoor point clouds in an occlusion-aware manner. The main ingredient of our framework is a Birds-Eye-View (BEV) feature map conditioned semantic rendering, leveraging the semantics of images for supervision through neural rendering. We further enhance our model's performance by incorporating point-wise masking with a high mask ratio (95%). Extensive experiments demonstrate PRED's superiority over prior point cloud pre-training methods, providing significant improvements on various large-scale datasets for 3D perception tasks. Codes will be available at https://github.com/PRED4pc/PRED.
****
Active Observing in Continuous-time Control
Samuel Holt, Alihan Hüyük, Mihaela van der Schaar
The control of continuous-time environments while actively deciding when to take costly observations in time is a crucial yet unexplored problem, particularly relevant to real-world scenarios such as medicine, low-power systems, and resource management. Existing approaches either rely on continuous-time control methods that take regular, expensive observations in time or discrete-time control with costly observation methods, which are inapplicable to continuous-time settings due to the compounding discretization errors introduced by time discretization. In this work, we are the first to formalize the continuous-time control problem with costly observations. Our key theoretical contribution shows that observing at regular time intervals is not optimal in certain environments, while irregular observation policies yield higher expected utility. This perspective paves the way for the development of novel methods that can take irregular observations

in continuous-time control with costly observations. We empirically validate our theoretical findings in various continuous-time environments, including a cancer simulation, by constructing a simple initial method to solve this new problem, with a heuristic threshold on the variance of reward rollouts in an offline continuous-time model-based model predictive control (MPC) planner. Although determining the optimal method remains an open problem, our work offers valuable insights and understanding of this unique problem, laying the foundation for future research in this area.
****

Principled Weight Initialisation for Input-Convex Neural Networks
Pieter-Jan Hoedt, Günter Klambauer
Input-Convex Neural Networks (ICNNs) are networks that guarantee convexity in their input-output mapping. These networks have been successfully applied for energy-based modelling, optimal transport problems and learning invariances.The convexity of ICNNs is achieved by using non-decreasing convex activation functions and non-negative weights. Because of these peculiarities, previous initialisation strategies, which implicitly assume centred weights, are not effective for ICNNs. By studying signal propagation through layers with non-negative weights, we are able to derive a principled weight initialisation for ICNNs. Concretely, we generalise signal propagation theory by removing the assumption that weights are sampled from a centred distribution. In a set of experiments, we demonstrate that our principled initialisation effectively accelerates learning in ICNNs and leads to better generalisation. Moreover, we find that, in contrast to common belief, ICNNs can be trained without skip-connections when initialised correctly. Finally, we apply ICNNs to a real-world drug discovery task and show that they allow for more effective molecular latent space exploration.
****

Automatic Grouping for Efficient Cooperative Multi-Agent Reinforcement Learning
Yifan Zang, Jinmin He, Kai Li, Haobo Fu, Qiang Fu, Junliang Xing, Jian Cheng
Grouping is ubiquitous in natural systems and is essential for promoting efficiency in team coordination. This paper proposes a novel formulation of Group-oriented Multi-Agent Reinforcement Learning (GoMARL), which learns automatic grouping without domain knowledge for efficient cooperation. In contrast to existing approaches that attempt to directly learn the complex relationship between the joint action-values and individual utilities, we empower subgroups as a bridge to model the connection between small sets of agents and encourage cooperation among them, thereby improving the learning efficiency of the whole team. In particular, we factorize the joint action-values as a combination of group-wise values, which guide agents to improve their policies in a fine-grained fashion. We present an automatic grouping mechanism to generate dynamic groups and group action-values. We further introduce a hierarchical control for policy learning that drives the agents in the same group to specialize in similar policies and possess diverse strategies for various groups. Experiments on the StarCraft II micromanagement tasks and Google Research Football scenarios verify our method's effectiveness. Extensive component studies show how grouping works and enhances performance.
****

On the Minimax Regret for Online Learning with Feedback Graphs
Khaled Eldowa, Emmanuel Esposito, Tom Cesari, Nicolò Cesa-Bianchi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

DropPos: Pre-Training Vision Transformers by Reconstructing Dropped Positions
Haochen Wang, Junsong Fan, Yuxi Wang, Kaiyou Song, Tong Wang, ZHAO-XIANG ZHANG
As it is empirically observed that Vision Transformers (ViTs) are quite insensitive to the order of input tokens, the need for an appropriate self-supervised pretext task that enhances the location awareness of ViTs is becoming evident. To address this, we present DropPos, a novel pretext task designed to reconstruct Dropped Positions. The formulation of DropPos is simple: we first drop a large ra

ndom subset of positional embeddings and then the model classifies the actual position for each non-overlapping patch among all possible positions solely based on their visual appearance. To avoid trivial solutions, we increase the difficulty of this task by keeping only a subset of patches visible. Additionally, considering there may be different patches with similar visual appearances, we propose position smoothing and attentive reconstruction strategies to relax this classification problem, since it is not necessary to reconstruct their exact positions in these cases. Empirical evaluations of DropPos show strong capabilities. DropPos outperforms supervised pre-training and achieves competitive results compared with state-of-the-art self-supervised alternatives on a wide range of downstream benchmarks. This suggests that explicitly encouraging spatial reasoning abilities, as DropPos does, indeed contributes to the improved location awareness of ViTs. The code is publicly available at https://github.com/Haochen-Wang409/DropPos.

****

Hierarchical VAEs provide a normative account of motion processing in the primate brain

Hadi Vafaii, Jacob Yates, Daniel Butts

The relationship between perception and inference, as postulated by Helmholtz in the 19th century, is paralleled in modern machine learning by generative models like Variational Autoencoders (VAEs) and their hierarchical variants. Here, we evaluate the role of hierarchical inference and its alignment with brain function in the domain of motion perception. We first introduce a novel synthetic data framework, Retinal Optic Flow Learning (ROFL), which enables control over motion statistics and their causes. We then present a new hierarchical VAE and test it against alternative models on two downstream tasks: (i) predicting ground truth causes of retinal optic flow (e.g., self-motion); and (ii) predicting the responses of neurons in the motion processing pathway of primates. We manipulate the model architectures (hierarchical versus non-hierarchical), loss functions, and the causal structure of the motion stimuli. We find that hierarchical latent structure in the model leads to several improvements. First, it improves the linear decodability of ground truth variables and does so in a sparse and disentangled manner. Second, our hierarchical VAE outperforms previous state-of-the-art models in predicting neuronal responses and exhibits sparse latent-to-neuron relationships. These results depend on the causal structure of the world, indicating that alignment between brains and artificial neural networks depends not only on architecture but also on matching ecologically relevant stimulus statistics. Taken together, our results suggest that hierarchical Bayesian inference underlines the brain's understanding of the world, and hierarchical VAEs can effectively model this understanding.

****

Variational Gaussian Processes with Decoupled Conditionals

Xinran Zhu, Kaiwen Wu, Natalie Maus, Jacob Gardner, David Bindel

Variational Gaussian processes (GPs) approximate exact GP inference by using a small set of inducing points to form a sparse approximation of the true posterior, with the fidelity of the model increasing with additional inducing points. Although the approximation error in principle can be reduced through the use of more inducing points, this leads to scaling optimization challenges and computational complexity. To achieve scalability, inducing point methods typically introduce conditional independencies and then approximations to the training and test conditional distributions. In this paper, we consider an alternative approach to modifying the training and test conditionals, in which we make them more flexible. In particular, we investigate decoupling the parametric form of the predictive mean and covariance in the conditionals, and learn independent parameters for predictive mean and covariance. We derive new evidence lower bounds (ELBO) under these more flexible conditionals, and provide two concrete examples of applying the decoupled conditionals. Empirically, we find this additional flexibility leads to improved model performance on a variety of regression tasks and Bayesian optimization (BO) applications.

****

EgoSchema: A Diagnostic Benchmark for Very Long-form Video Language Understanding

Karttikeya Mangalam, Raiymbek Akshulakov, Jitendra Malik

We introduce EgoSchema, a very long-form video question-answering dataset, and benchmark to evaluate long video understanding capabilities of modern vision and language systems. Derived from Ego4D, EgoSchema consists of over 5000 human curated multiple choice question answer pairs, spanning over 250 hours of real video data, covering a very broad range of natural human activity and behavior. For each question, EgoSchema requires the correct answer to be selected between five given options based on a three-minute-long video clip. While some prior works have proposed video datasets with long clip lengths, we posit that merely the length of the video clip does not truly capture the temporal difficulty of the video task that is being considered. To remedy this, we introduce temporal certificate sets, a general notion for capturing the intrinsic temporal understanding length associated with a broad range of video understanding tasks & datasets. Based on this metric, we find EgoSchema to have intrinsic temporal lengths over 5.7x longer than the second closest dataset and 10x to 100x longer than any other video understanding dataset. Further, our evaluation of several current state-of-the-art video and language models shows them to be severely lacking in long-term video understanding capabilities. Even models with several billions of parameters achieve QA accuracy less than 33% (random is 20%) on the EgoSchema multi-choice question answering task, while humans achieve about 76% accuracy. We posit that EgoSchema, with its long intrinsic temporal structures and diverse complexity, would serve as a valuable evaluation probe for developing effective long-term video understanding systems in the future. Data and Zero-shot model evaluation code will all be open-sourced under the Ego4D license at http://egoschema.github.io.

****

TabMT: Generating tabular data with masked transformers

Manbir Gulati, Paul Roysdon

Autoregressive and Masked Transformers are incredibly effective as generative models and classifiers. While these models are most prevalent in NLP, they also exhibit strong performance in other domains, such as vision. This work contributes to the exploration of transformer-based models in synthetic data generation for diverse application domains. In this paper, we present TabMT, a novel Masked Transformer design for generating synthetic tabular data. TabMT effectively addresses the unique challenges posed by heterogeneous data fields and is natively able to handle missing data. Our design leverages improved masking techniques to allow for generation and demonstrates state-of-the-art performance from extremely small to extremely large tabular datasets. We evaluate TabMT for privacy-focused applications and find that it is able to generate high quality data with superior privacy tradeoffs.

****

Brain Dissection: fMRI-trained Networks Reveal Spatial Selectivity in the Processing of Natural Images

Gabriel Sarch, Michael Tarr, Katerina Fragkiadaki, Leila Wehbe

The alignment between deep neural network (DNN) features and cortical responses currently provides the most accurate quantitative explanation for higher visual areas. At the same time, these model features have been critiqued as uninterpretable explanations, trading one black box (the human brain) for another (a neural network). In this paper, we train networks to directly predict, from scratch, brain responses to images from a large-scale dataset of natural scenes (Allen et. al., 2021). We then use "network dissection" (Bau et. al., 2017), an explainable AI technique used for enhancing neural network interpretability by identifying and localizing the most significant features in images for individual units of a trained network, and which has been used to study category selectivity in the human brain (Khosla & Wehbe, 2022). We adapt this approach to create a hypothesis-neutral model that is then used to explore the tuning properties of specific visual regions beyond category selectivity, which we call "brain dissection". We use brain dissection to examine a range of ecologically important, intermediate properties, including depth, surface normals, curvature, and object relations ac

ross sub-regions of the parietal, lateral, and ventral visual streams, and scene-selective regions. Our findings reveal distinct preferences in brain regions for interpreting visual scenes, with ventro-lateral areas favoring closer and curvier features, medial and parietal areas opting for more varied and flatter 3D elements, and the parietal region uniquely preferring spatial relations. Scene-selective regions exhibit varied preferences, as the retrosplenial complex prefers distant and outdoor features, while the occipital and parahippocampal place areas favor proximity, verticality, and in the case of the OPA, indoor elements. Such findings show the potential of using explainable AI to uncover spatial feature selectivity across the visual cortex, contributing to a deeper, more fine-grained understanding of the functional characteristics of human visual cortex when viewing natural scenes.
****
Action Inference by Maximising Evidence: Zero-Shot Imitation from Observation with World Models
Xingyuan Zhang, Philip Becker-Ehmck, Patrick van der Smagt, Maximilian Karl
Unlike most reinforcement learning agents which require an unrealistic amount of environment interactions to learn a new behaviour, humans excel at learning quickly by merely observing and imitating others. This ability highly depends on the fact that humans have a model of their own embodiment that allows them to infer the most likely actions that led to the observed behaviour. In this paper, we propose Action Inference by Maximising Evidence (AIME) to replicate this behaviour using world models. AIME consists of two distinct phases. In the first phase, the agent learns a world model from its past experience to understand its own body by maximising the ELBO. While in the second phase, the agent is given some observation-only demonstrations of an expert performing a novel task and tries to imitate the expert's behaviour. AIME achieves this by defining a policy as an inference model and maximising the evidence of the demonstration under the policy and world model. Our method is "zero-shot" in the sense that it does not require further training for the world model or online interactions with the environment after given the demonstration. We empirically validate the zero-shot imitation performance of our method on the Walker and Cheetah embodiment of the DeepMind Control Suite and find it outperforms the state-of-the-art baselines. Code is available at: https://github.com/argmax-ai/aime.
****
ProtoDiff: Learning to Learn Prototypical Networks by Task-Guided Diffusion
Yingjun Du, Zehao Xiao, Shengcai Liao, Cees Snoek
Prototype-based meta-learning has emerged as a powerful technique for addressing few-shot learning challenges. However, estimating a deterministic prototype using a simple average function from a limited number of examples remains a fragile process. To overcome this limitation, we introduce ProtoDiff, a novel framework that leverages a task-guided diffusion model during the meta-training phase to gradually generate prototypes, thereby providing efficient class representations. Specifically, a set of prototypes is optimized to achieve per-task prototype overfitting, enabling accurately obtaining the overfitted prototypes for individual tasks.Furthermore, we introduce a task-guided diffusion process within the prototype space, enabling the meta-learning of a generative process that transitions from a vanilla prototype to an overfitted prototype. ProtoDiff gradually generates task-specific prototypes from random noise during the meta-test stage, conditioned on the limited samples available for the new task. Furthermore, to expedite training and enhance ProtoDiff's performance, we propose the utilization of residual prototype learning, which leverages the sparsity of the residual prototype. We conduct thorough ablation studies to demonstrate its ability to accurately capture the underlying prototype distribution and enhance generalization. The new state-of-the-art performance on within-domain, cross-domain, and few-task few-shot classi■cation further substantiates the bene■t of ProtoDiff.
****
Synthetic Experience Replay
Cong Lu, Philip Ball, Yee Whye Teh, Jack Parker-Holder
A key theme in the past decade has been that when large neural networks and larg

e datasets combine they can produce remarkable results. In deep reinforcement learning (RL), this paradigm is commonly made possible through experience replay, whereby a dataset of past experiences is used to train a policy or value function. However, unlike in supervised or self-supervised learning, an RL agent has to collect its own data, which is often limited. Thus, it is challenging to reap the benefits of deep learning, and even small neural networks can overfit at the start of training. In this work, we leverage the tremendous recent progress in generative modeling and propose Synthetic Experience Replay (SynthER), a diffusion-based approach to flexibly upsample an agent's collected experience. We show that SynthER is an effective method for training RL agents across offline and online settings, in both proprioceptive and pixel-based environments. In offline settings, we observe drastic improvements when upsampling small offline datasets and see that additional synthetic data also allows us to effectively train larger networks. Furthermore, SynthER enables online agents to train with a much higher update-to-data ratio than before, leading to a significant increase in sample efficiency, without any algorithmic changes. We believe that synthetic training data could open the door to realizing the full potential of deep learning for replay-based RL algorithms from limited data. Finally, we open-source our code at https://github.com/conglu1997/SynthER.
****

Learning to Tokenize for Generative Retrieval
Weiwei Sun, Lingyong Yan, Zheng Chen, Shuaiqiang Wang, Haichao Zhu, Pengjie Ren, Zhumin Chen, Dawei Yin, Maarten Rijke, Zhaochun Ren
As a new paradigm in information retrieval, generative retrieval directly generates a ranked list of document identifiers (docids) for a given query using generative language models (LMs).How to assign each document a unique docid (denoted as document tokenization) is a critical problem, because it determines whether the generative retrieval model can precisely retrieve any document by simply decoding its docid.Most existing methods adopt rule-based tokenization, which is ad-hoc and does not generalize well.In contrast, in this paper we propose a novel document tokenization learning method, GenRet, which learns to encode the complete document semantics into docids.GenRet learns to tokenize documents into short discrete representations (i.e., docids) via a discrete auto-encoding approach.We develop a progressive training scheme to capture the autoregressive nature of docids and diverse clustering techniques to stabilize the training process.Based on the semantic-embedded docids of any set of documents, the generative retrieval model can learn to generate the most relevant docid only according to the docids' semantic relevance to the queries.We conduct experiments on the NQ320K, MS MARCO, and BEIR datasets.GenRet establishes the new state-of-the-art on the NQ320K dataset.Compared to generative retrieval baselines, GenRet can achieve significant improvements on unseen documents.Moreover, GenRet can also outperform comparable baselines on MS MARCO and BEIR, demonstrating the method's generalizability.
****

A Reduction-based Framework for Sequential Decision Making with Delayed Feedback
Yunchang Yang, Han Zhong, Tianhao Wu, Bin Liu, Liwei Wang, Simon S. Du
We study stochastic delayed feedback in general single-agent and multi-agent sequential decision making, which includes bandits, single-agent Markov decision processes (MDPs), and Markov games (MGs). We propose a novel reduction-based framework, which turns any multi-batched algorithm for sequential decision making with instantaneous feedback into a sample-efficient algorithm that can handle stochastic delays in sequential decision making. By plugging different multi-batched algorithms into our framework, we provide several examples demonstrating that our framework not only matches or improves existing results for bandits, tabular MDPs, and tabular MGs, but also provides the first line of studies on delays in sequential decision making with function approximation. In summary, we provide a complete set of sharp results for single-agent and multi-agent sequential decision making with delayed feedback.
****

Efficient RL with Impaired Observability: Learning to Act with Delayed and Missi

ng State Observations
Minshuo Chen, Yu Bai, H. Vincent Poor, Mengdi Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Unified 3D Segmenter As Prototypical Classifiers
Zheyun Qin, Cheng Han, Qifan Wang, Xiushan Nie, Yilong Yin, Lu Xiankai
The task of point cloud segmentation, comprising semantic, instance, and panoptic segmentation, has been mainly tackled by designing task-specific network architectures, which often lack the flexibility to generalize across tasks, thus resulting in a fragmented research landscape. In this paper, we introduce ProtoSEG, a prototype-based model that unifies semantic, instance, and panoptic segmentation tasks. Our approach treats these three homogeneous tasks as a classification problem with different levels of granularity. By leveraging a Transformer architecture, we extract point embeddings to optimize prototype-class distances and dynamically learn class prototypes to accommodate the end tasks. Our prototypical design enjoys simplicity and transparency, powerful representational learning, and ad-hoc explainability. Empirical results demonstrate that ProtoSEG outperforms concurrent well-known specialized architectures on 3D point cloud benchmarks, achieving 72.3%, 76.4% and 74.2% mIoU for semantic segmentation on S3DIS, ScanNet V2 and SemanticKITTI, 66.8% mCov and 51.2% mAP for instance segmentation on S3DIS and ScanNet V2, 62.4% PQ for panoptic segmentation on SemanticKITTI, validating the strength of our concept and the effectiveness of our algorithm. The code and models are available at https://github.com/zyqin19/PROTOSEG.
****
Cola: A Benchmark for Compositional Text-to-image Retrieval
Arijit Ray, Filip Radenovic, Abhimanyu Dubey, Bryan Plummer, Ranjay Krishna, Kate Saenko
Compositional reasoning is a hallmark of human visual intelligence. Yet, despite the size of large vision-language models, they struggle to represent simple compositions by combining objects with their attributes. To measure this lack of compositional capability, we design Cola, a text-to-image retrieval benchmark to Compose Objects Localized with Attributes. To solve Cola, a model must retrieve images with the correct configuration of attributes and objects and avoid choosing a distractor image with the same objects and attributes but in the wrong configuration. Cola contains about 1.2k composed queries of 168 objects and 197 attributes on around 30K images. Our human evaluation finds that Cola is 83.33% accurate, similar to contemporary compositionality benchmarks. Using Cola as a testbed, we explore empirical modeling designs to adapt pre-trained vision-language models to reason compositionally. We explore 6 adaptation strategies on 2 seminal vision-language models, using compositionality-centric test benchmarks - Cola and CREPE. We find the optimal adaptation strategy is to train a multi-modal attention layer that jointly attends over the frozen pre-trained image and language features. Surprisingly, training multimodal layers on CLIP performs better than tuning a larger FLAVA model with already pre-trained multimodal layers. Furthermore, our adaptation strategy improves CLIP and FLAVA to comparable levels, suggesting that training multimodal layers using contrastive attribute-object data is key, as opposed to using them pre-trained. Lastly, we show that Cola is harder than a closely related contemporary benchmark, CREPE, since simpler fine-tuning strategies without multimodal layers suffice on CREPE, but not on Cola. However, we still see a significant gap between our best adaptation and human accuracy, suggesting considerable room for further research. Project page: https://cs-people.bu.edu/array/research/cola/
****
Estimating Causal Effects Identifiable from a Combination of Observations and Experiments
Yonghan Jung, Ivan Diaz, Jin Tian, Elias Bareinboim
Learning cause and effect relations is arguably one of the central challenges fo

und throughout the data sciences.Formally, determining whether a collection of observational and interventional distributions can be combined to learn a target causal relation is known as the problem of generalized identification (or g-identification) [Lee et al., 2019]. Although g-identification has been well understood and solved in theory, it turns out to be challenging to apply these results in practice, in particular when considering the estimation of the target distribution from finite samples. In this paper, we develop a new, general estimator that exhibits multiply robustness properties for g-identifiable causal functionals. Specifically, we show that any g-identifiable causal effect can be expressed as a function of generalized multi-outcome sequential back-door adjustments that are amenable to estimation. We then construct a corresponding estimator for the g-identification expression that exhibits robustness properties to bias. We analyze the asymptotic convergence properties of the estimator. Finally, we illustrate the use of the proposed estimator in experimental studies. Simulation results corroborate the theory.
****
LMC: Large Model Collaboration with Cross-assessment for Training-Free Open-Set Object Recognition
Haoxuan Qu, Xiaofei Hui, Yujun Cai, Jun Liu
Open-set object recognition aims to identify if an object is from a class that has been encountered during training or not. To perform open-set object recognition accurately, a key challenge is how to reduce the reliance on spurious-discriminative features. In this paper, motivated by that different large models pre-trained through different paradigms can possess very rich while distinct implicit knowledge, we propose a novel framework named Large Model Collaboration (LMC) to tackle the above challenge via collaborating different off-the-shelf large models in a training-free manner. Moreover, we also incorporate the proposed framework with several novel designs to effectively extract implicit knowledge from large models. Extensive experiments demonstrate the efficacy of our proposed framework. Code is available \href{https://github.com/Harryqu123/LMC}{here}.
****
TaskMet: Task-driven Metric Learning for Model Learning
Dishank Bansal, Ricky T. Q. Chen, Mustafa Mukadam, Brandon Amos
Deep learning models are often used with some downstream task. Models solely trained to achieve accurate predictions may struggle to perform well on the desired downstream tasks. We propose using the task loss to learn a metric which parameterizes a loss to train the model. This approach does not alter the optimal prediction model itself, but rather changes the model learning to emphasize the information important for the downstream task. This enables us to achieve the best of both worlds: a prediction model trained in the original prediction space while also being valuable for the desired downstream task. We validate our approach through experiments conducted in two main settings: 1) decision-focused model learning scenarios involving portfolio optimization and budget allocation, and 2) reinforcement learning in noisy environments with distracting states.
****
Pairwise Causality Guided Transformers for Event Sequences
Xiao Shou, Debarun Bhattacharjya, Tian Gao, Dharmashankar Subramanian, Oktie Hassanzadeh, Kristin P Bennett
Although pairwise causal relations have been extensively studied in observational longitudinal analyses across many disciplines, incorporating knowledge of causal pairs into deep learning models for temporal event sequences remains largely unexplored. In this paper, we propose a novel approach for enhancing the performance of transformer-based models in multivariate event sequences by injecting pairwise qualitative causal knowledge such as `event Z amplifies future occurrences of event Y'. We establish a new framework for causal inference in temporal event sequences using a transformer architecture, providing a theoretical justification for our approach, and show how to obtain unbiased estimates of the proposed measure. Experimental results demonstrate that our approach outperforms several state-of-the-art models in terms of prediction accuracy by effectively leveraging knowledge about causal pairs. We also consider a unique application where we

extract knowledge around sequences of societal events by generating them from a large language model, and demonstrate how a causal knowledge graph can help with event prediction in such sequences. Overall, our framework offers a practical means of improving the performance of transformer-based models in multivariate event sequences by explicitly exploiting pairwise causal information.
****
Self-Refine: Iterative Refinement with Self-Feedback
Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegreffe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, Peter Clark
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena
Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, Hao Zhang, Joseph E. Gonzalez, Ion Stoica
Evaluating large language model (LLM) based chat assistants is challenging due to their broad capabilities and the inadequacy of existing benchmarks in measuring human preferences.To address this, we explore using strong LLMs as judges to evaluate these models on more open-ended questions.We examine the usage and limitations of LLM-as-a-judge, including position, verbosity, and self-enhancement biases, as well as limited reasoning ability, and propose solutions to mitigate some of them.We then verify the agreement between LLM judges and human preferences by introducing two benchmarks: MT-bench, a multi-turn question set; and Chatbot Arena, a crowdsourced battle platform.Our results reveal that strong LLM judges like GPT-4 can match both controlled and crowdsourced human preferences well, achieving over 80\% agreement, the same level of agreement between humans.Hence, LLM-as-a-judge is a scalable and explainable way to approximate human preferences, which are otherwise very expensive to obtain.Additionally, we show our benchmark and traditional benchmarks complement each other by evaluating several variants of LLaMA and Vicuna.The MT-bench questions, 3K expert votes, and 30K conversations with human preferences are publicly available at https://github.com/lm-sys/FastChat/tree/main/fastchat/llm_judge.
****
Causal Discovery in Semi-Stationary Time Series
Shanyun Gao, Raghavendra Addanki, Tong Yu, Ryan Rossi, Murat Kocaoglu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Fine-grained Expressivity of Graph Neural Networks
Jan Böker, Ron Levie, Ningyuan Huang, Soledad Villar, Christopher Morris
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CrossCodeEval: A Diverse and Multilingual Benchmark for Cross-File Code Completion
Yangruibo Ding, Zijian Wang, Wasi Ahmad, Hantian Ding, Ming Tan, Nihal Jain, Murali Krishna Ramanathan, Ramesh Nallapati, Parminder Bhatia, Dan Roth, Bing Xiang
Code completion models have made significant progress in recent years, yet current popular evaluation datasets, such as HumanEval and MBPP, predominantly focus on code completion tasks within a single file. This over-simplified setting falls short of representing the real-world software development scenario where repos

itories span multiple files with numerous cross-file dependencies, and accessing and understanding cross-file context is often required to complete the code correctly. To fill in this gap, we propose CrossCodeEval, a diverse and multilingual code completion benchmark that necessitates an in-depth cross-file contextual understanding to complete the code accurately. CrossCodeEval is built on a diverse set of real-world, open-sourced, permissively-licensed repositories in four popular programming languages: Python, Java, TypeScript, and C#. To create examples that strictly require cross-file context for accurate completion, we propose a straightforward yet efficient static-analysis-based approach to pinpoint the use of cross-file context within the current file. Extensive experiments on state-of-the-art code language models like CodeGen and StarCoder demonstrate that CrossCodeEval is extremely challenging when the relevant cross-file context is absent, and we see clear improvements when adding these context into the prompt. However, despite such improvements, the pinnacle of performance remains notably unattained even with the highest-performing model, indicating that CrossCodeEval is also capable of assessing model's capability in leveraging extensive context to make better code completion. Finally, we benchmarked various methods in retrieving cross-file context, and show that CrossCodeEval can also be used to measure the capability of code retrievers.
****
Hierarchical Adaptive Value Estimation for Multi-modal Visual Reinforcement Learning
Yangru Huang, Peixi Peng, Yifan Zhao, Haoran Xu, Mengyue Geng, Yonghong Tian
Integrating RGB frames with alternative modality inputs is gaining increasing traction in many vision-based reinforcement learning (RL) applications. Existing multi-modal vision-based RL methods usually follow a Global Value Estimation (GVE) pipeline, which uses a fused modality feature to obtain a unified global environmental description. However, such a feature-level fusion paradigm with a single critic may fall short in policy learning as it tends to overlook the distinct values of each modality. To remedy this, this paper proposes a Local modality-customized Value Estimation (LVE) paradigm, which dynamically estimates the contribution and adjusts the importance weight of each modality from a value-level perspective. Furthermore, a task-contextual re-fusion process is developed to achieve a task-level re-balance of estimations from both feature and value levels. To this end, a Hierarchical Adaptive Value Estimation (HAVE) framework is formed, which adaptively coordinates the contributions of individual modalities as well as their collective efficacy. Agents trained by HAVE are able to exploit the unique characteristics of various modalities while capturing their intricate interactions, achieving substantially improved performance. We specifically highlight the potency of our approach within the challenging landscape of autonomous driving, utilizing the CARLA benchmark with neuromorphic event and depth data to demonstrate HAVE's capability and the effectiveness of its distinct components.
****
Small Total-Cost Constraints in Contextual Bandits with Knapsacks, with Application to Fairness
Evgenii Chzhen, Christophe Giraud, Zhen LI, Gilles Stoltz
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CAPP-130: A Corpus of Chinese Application Privacy Policy Summarization and Interpretation
pengyun zhu, Long Wen, Jinfei Liu, Feng Xue, Jian Lou, Zhibo Wang, Kui Ren
A privacy policy serves as an online internet protocol crafted by service providers, which details how service providers collect, process, store, manage, and use personal information when users engage with applications. However, these privacy policies are often filled with technobabble and legalese, making them "incomprehensible''. As a result, users often agree to all terms unknowingly, even some terms may conflict with the law, thereby posing a considerable risk to personal

privacy information. One potential solution to alleviate this challenge is to automatically summarize privacy policies using NLP techniques. However, existing techniques primarily focus on extracting key sentences, resulting in comparatively shorter agreements, but failing to address the poor readability caused by the "incomprehensible'' of technobabble and legalese. Moreover, research on Chinese application privacy policy summarization is currently almost nonexistent, and there is a lack of a high-quality corpus suitable for addressing readability issues. To tackle these challenges, we introduce a fine-grained CAPP-130 corpus and a TCSI-pp framework. CAPP-130 contains 130 Chinese privacy policies from popular applications that have been carefully annotated and interpreted by legal experts, resulting in 52,489 annotations and 20,555 rewritten sentences. TCSI-pp first extracts sentences related to the topic specified by users and then uses a generative model to rewrite the sentences into comprehensible summarization. Built upon TSCI-pp, we construct a summarization tool TSCI-pp-zh by selecting RoBERTa from six classification models for sentence extraction and selecting mT5 from five generative models for sentence rewriting. Experimental results show that TCSI-pp-zh outperforms GPT-4 and other baselines in Chinese application privacy policy summarization, demonstrating exceptional readability and reliability. Our data, annotation guidelines, benchmark models, and source code are publicly available at https://github.com/EnlightenedAI/CAPP-130.
****

Neural Oscillators are Universal

Samuel Lanthaler, T. Konstantin Rusch, Siddhartha Mishra

Coupled oscillators are being increasingly used as the basis of machine learning (ML) architectures, for instance in sequence modeling, graph representation learning and in physical neural networks that are used in analog ML devices. We introduce an abstract class of neural oscillators that encompasses these architectures and prove that neural oscillators are universal, i.e, they can approximate any continuous and casual operator mapping between time-varying functions, to desired accuracy. This universality result provides theoretical justification for the use of oscillator based ML systems. The proof builds on a fundamental result of independent interest, which shows that a combination of forced harmonic oscillators with a nonlinear read-out suffices to approximate the underlying operators.
****

PAC-Bayes Generalization Certificates for Learned Inductive Conformal Prediction

Apoorva Sharma, Sushant Veer, Asher Hancock, Heng Yang, Marco Pavone, Anirudha Majumdar

Inductive Conformal Prediction (ICP) provides a practical and effective approach for equipping deep learning models with uncertainty estimates in the form of set-valued predictions which are guaranteed to contain the ground truth with high probability.Despite the appeal of this coverage guarantee, these sets may not be efficient: the size and contents of the prediction sets are not directly controlled, and instead depend on the underlying model and choice of score function.To remedy this, recent work has proposed learning model and score function parameters using data to directly optimize the efficiency of the ICP prediction sets.While appealing, the generalization theory for such an approach is lacking: direct optimization of empirical efficiency may yield prediction sets that are either no longer efficient on test data, or no longer obtain the required coverage on test data.In this work, we use PAC-Bayes theory to obtain generalization bounds on both the coverage and the efficiency of set-valued predictors which can be directly optimized to maximize efficiency while satisfying a desired test coverage. In contrast to prior work, our framework allows us to utilize the entire calibration dataset to learn the parameters of the model and score function, instead of requiring a separate hold-out set for obtaining test-time coverage guarantees.We leverage these theoretical results to provide a practical algorithm for using calibration data to simultaneously fine-tune the parameters of a model and score function while guaranteeing test-time coverage and efficiency of the resulting prediction sets.We evaluate the approach on regression and classification tasks, and outperform baselines calibrated using a Hoeffding bound-based PAC guarantee

on ICP, especially in the low-data regime.
****
Image Captioners Are Scalable Vision Learners Too
Michael Tschannen, Manoj Kumar, Andreas Steiner, Xiaohua Zhai, Neil Houlsby, Lucas Beyer
Contrastive pretraining on image-text pairs from the web is one of the most popular large-scale pretraining strategies for vision backbones, especially in the context of large multimodal models. At the same time, image captioning on this type of data is commonly considered an inferior pretraining strategy. In this paper, we perform a fair comparison of these two pretraining strategies, carefully matching training data, compute, and model capacity. Using a standard encoder-decoder transformer, we find that captioning alone is surprisingly effective: on classification tasks, captioning produces vision encoders competitive with contrastively pretrained encoders, while surpassing them on vision & language tasks. We further analyze the effect of the model architecture and scale, as well as the pretraining data on the representation quality, and find that captioning exhibits the same or better scaling behavior along these axes. Overall our results show that plain image captioning is a more powerful pretraining strategy than was previously believed. Code is available at https://github.com/google-research/big_vision.
****
Diplomat: A Dialogue Dataset for Situated PragMATic Reasoning
Hengli Li, Song-Chun Zhu, Zilong Zheng
The ability to discern and comprehend pragmatic meanings is a cornerstone of social and emotional intelligence, referred to as pragmatic reasoning. Despite the strides made in the development of Large Language Models (LLMs), such as ChatGPT, these models grapple with capturing the nuanced and ambiguous facets of language, falling short of the aspiration to build human-like conversational agents. In this work, we introduce a novel benchmark, the DiPlomat, which delves into the fundamental components of conversational pragmatic reasoning, encompassing situational context reasoning, open-world knowledge acquisition, and unified figurative language understanding. We start by collecting a new human-annotated dialogue dataset, composed of 4,177 multi-turn dialogues and a vocabulary of 48,900 words. Along with the dataset, two tasks are proposed to evaluate machines' pragmatic reasoning capabilities, namely, Pragmatic Reasoning and Identification(PIR) and Conversational Question Answering (CQA). Furthermore, we probe into a zero-shot natural language inference task, where the significance of context in pragmatic reasoning is underscored. Experimental findings illustrate the existing limitations of current prevailing LLMs in the realm of pragmatic reasoning, shedding light on the pressing need for further research to facilitate the emergence of emotional intelligence within human-like conversational agents.
****
CrossGNN: Confronting Noisy Multivariate Time Series Via Cross Interaction Refinement
Qihe Huang, Lei Shen, Ruixin Zhang, Shouhong Ding, Binwu Wang, Zhengyang Zhou, Yang Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Structured Prediction with Stronger Consistency Guarantees
Anqi Mao, Mehryar Mohri, Yutao Zhong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Stanford-ORB: A Real-World 3D Object Inverse Rendering Benchmark
Zhengfei Kuang, Yunzhi Zhang, Hong-Xing Yu, Samir Agarwala, Elliott / Shangzhe W

u, Jiajun Wu
We introduce Stanford-ORB, a new real-world 3D Object inverse Rendering Benchmark. Recent advances in inverse rendering have enabled a wide range of real-world applications in 3D content generation, moving rapidly from research and commercial use cases to consumer devices. While the results continue to improve, there is no real-world benchmark that can quantitatively assess and compare the performance of various inverse rendering methods. Existing real-world datasets typically only consist of the shape and multi-view images of objects, which are not sufficient for evaluating the quality of material recovery and object relighting. Methods capable of recovering material and lighting often resort to synthetic data for quantitative evaluation, which on the other hand does not guarantee generalization to complex real-world environments. We introduce a new dataset of real-world objects captured under a variety of natural scenes with ground-truth 3D scans, multi-view images, and environment lighting. Using this dataset, we establish the first comprehensive real-world evaluation benchmark for object inverse rendering tasks from in-the-wild scenes, and compare the performance of various existing methods. All data, code, and models can be accessed at https://stanfordorb.github.io/
****
Explainable Brain Age Prediction using coVariance Neural Networks
Saurabh Sihag, Gonzalo Mateos, Corey McMillan, Alejandro Ribeiro
In computational neuroscience, there has been an increased interest in developing machine learning algorithms that leverage brain imaging data to provide estimates of "brain age" for an individual. Importantly, the discordance between brain age and chronological age (referred to as "brain age gap") can capture accelerated aging due to adverse health conditions and therefore, can reflect increased vulnerability towards neurological disease or cognitive impairments. However, widespread adoption of brain age for clinical decision support has been hindered due to lack of transparency and methodological justifications in most existing brain age prediction algorithms. In this paper, we leverage coVariance neural networks (VNN) to propose an explanation-driven and anatomically interpretable framework for brain age prediction using cortical thickness features. Specifically, our brain age prediction framework extends beyond the coarse metric of brain age gap in Alzheimer’s disease (AD) and we make two important observations: (i) VNNs can assign anatomical interpretability to elevated brain age gap in AD by identifying contributing brain regions, (ii) the interpretability offered by VNNs is contingent on their ability to exploit specific eigenvectors of the anatomical covariance matrix. Together, these observations facilitate an explainable and anatomically interpretable perspective to the task of brain age prediction.
****
Adversarial Examples Might be Avoidable: The Role of Data Concentration in Adversarial Robustness
Ambar Pal, Jeremias Sulam, Rene Vidal
The susceptibility of modern machine learning classifiers to adversarial examples has motivated theoretical results suggesting that these might be unavoidable. However, these results can be too general to be applicable to natural data distributions. Indeed, humans are quite robust for tasks involving vision. This apparent conflict motivates a deeper dive into the question: Are adversarial examples truly unavoidable? In this work, we theoretically demonstrate that a key property of the data distribution -- concentration on small-volume subsets of the input space -- determines whether a robust classifier exists. We further demonstrate that, for a data distribution concentrated on a union of low-dimensional linear subspaces, utilizing structure in data naturally leads to classifiers that enjoy data-dependent polyhedral robustness guarantees, improving upon methods for provable certification in certain regimes.
****
Structured State Space Models for In-Context Reinforcement Learning
Chris Lu, Yannick Schroecker, Albert Gu, Emilio Parisotto, Jakob Foerster, Satinder Singh, Feryal Behbahani
Structured state space sequence (S4) models have recently achieved state-of-the-

art performance on long-range sequence modeling tasks. These models also have fa st inference speeds and parallelisable training, making them potentially useful in many reinforcement learning settings. We propose a modification to a variant of S4 that enables us to initialise and reset the hidden state in parallel, all owing us to tackle reinforcement learning tasks. We show that our modified archi tecture runs asymptotically faster than Transformers in sequence length and perf orms better than RNN's on a simple memory-based task. We evaluate our modified a rchitecture on a set of partially-observable environments and find that, in prac tice, our model outperforms RNN's while also running over five times faster. The n, by leveraging the model’s ability to handle long-range sequences, we achieve strong performance on a challenging meta-learning task in which the agent is giv en a randomly-sampled continuous control environment, combined with a randomly-s ampled linear projection of the environment's observations and actions. Furtherm ore, we show the resulting model can adapt to out-of-distribution held-out tasks . Overall, the results presented in this paper show that structured state space models are fast and performant for in-context reinforcement learning tasks. We p rovide code at https://github.com/luchris429/s5rl.
****
Sharpness-Aware Minimization Leads to Low-Rank Features
Maksym Andriushchenko, Dara Bahri, Hossein Mobahi, Nicolas Flammarion
Sharpness-aware minimization (SAM) is a recently proposed method that minimizes the sharpness of the training loss of a neural network. While its generalization improvement is well-known and is the primary motivation, we uncover an addition al intriguing effect of SAM: reduction of the feature rank which happens at diff erent layers of a neural network. We show that this low-rank effect occurs very broadly: for different architectures such as fully-connected networks, convoluti onal networks, vision transformers and for different objectives such as regressi on, classification, language-image contrastive training. To better understand th is phenomenon, we provide a mechanistic understanding of how low-rank features a rise in a simple two-layer network. We observe that a significant number of acti vations gets entirely pruned by SAM which directly contributes to the rank reduc tion. We confirm this effect theoretically and check that it can also occur in d eep networks, although the overall rank reduction mechanism can be more complex, especially for deep networks with pre-activation skip connections and self-atte ntion layers.
****
A Spectral Theory of Neural Prediction and Alignment
Abdulkadir Canatar, Jenelle Feather, Albert Wakhloo, SueYeon Chung
The representations of neural networks are often compared to those of biological systems by performing regression between the neural network responses and those measured from biological systems. Many different state-of-the-art deep neural n etworks yield similar neural predictions, but it remains unclear how to differen tiate among models that perform equally well at predicting neural responses. To gain insight into this, we use a recent theoretical framework that relates the g eneralization error from regression to the spectral properties of the model and the target. We apply this theory to the case of regression between model activat ions and neural responses and decompose the neural prediction error in terms of the model eigenspectra, alignment of model eigenvectors and neural responses, an d the training set size. Using this decomposition, we introduce geometrical meas ures to interpret the neural prediction error. We test a large number of deep ne ural networks that predict visual cortical activity and show that there are mult iple types of geometries that result in low neural prediction error as measured via regression. The work demonstrates that carefully decomposing representationa l metrics can provide interpretability of how models are capturing neural activi ty and points the way towards improved models of neural activity.
****
Train Once, Get a Family: State-Adaptive Balances for Offline-to-Online Reinforc ement Learning
Shenzhi Wang, Qisen Yang, Jiawei Gao, Matthieu Lin, HAO CHEN, Liwei Wu, Ning Jia , Shiji Song, Gao Huang

Offline-to-online reinforcement learning (RL) is a training paradigm that combines pre-training on a pre-collected dataset with fine-tuning in an online environment. However, the incorporation of online fine-tuning can intensify the well-known distributional shift problem. Existing solutions tackle this problem by imposing a policy constraint on the policy improvement objective in both offline and online learning. They typically advocate a single balance between policy improvement and constraints across diverse data collections. This one-size-fits-all manner may not optimally leverage each collected sample due to the significant variation in data quality across different states. To this end, we introduce Family Offline-to-Online RL (FamO2O), a simple yet effective framework that empowers existing algorithms to determine state-adaptive improvement-constraint balances. FamO2O utilizes a universal model to train a family of policies with different improvement/constraint intensities, and a balance model to select a suitable policy for each state. Theoretically, we prove that state-adaptive balances are necessary for achieving a higher policy performance upper bound. Empirically, extensive experiments show that FamO2O offers a statistically significant improvement over various existing methods, achieving state-of-the-art performance on the D4RL benchmark. Codes are available at https://github.com/LeapLabTHU/FamO2O.
****
Test-Time Distribution Normalization for Contrastively Learned Visual-language Models
Yifei Zhou, Juntao Ren, Fengyu Li, Ramin Zabih, Ser Nam Lim
Advances in the field of visual-language contrastive learning have made it possible for many downstream applications to be carried out efficiently and accurately by simply taking the dot product between image and text representations. One of the most representative approaches proposed recently known as CLIP has quickly garnered widespread adoption due to its effectiveness. CLIP is trained with an InfoNCE loss that takes into account both positive and negative samples to help learn a much more robust representation space. This paper however reveals that the common downstream practice of taking a dot product is only a zeroth-order approximation of the optimization goal, resulting in a loss of information during test-time. Intuitively, since the model has been optimized based on the InfoNCE loss, test-time procedures should ideally also be in alignment. The question lies in how one can retrieve any semblance of negative samples information during inference in a computationally efficient way. We propose Distribution Normalization (DN), where we approximate the mean representation of a batch of test samples and use such a mean to represent what would be analogous to negative samples in the InfoNCE loss. DN requires no retraining or fine-tuning and can be effortlessly applied during inference. Extensive experiments on a wide variety of downstream tasks exhibit a clear advantage of DN over the dot product on top of other existing test-time augmentation methods.
****
Propagating Knowledge Updates to LMs Through Distillation
Shankar Padmanabhan, Yasumasa Onoe, Michael Zhang, Greg Durrett, Eunsol Choi
Modern language models have the capacity to store and use immense amounts of knowledge about real-world entities, but it remains unclear how to update such knowledge stored in model parameters. While prior methods for updating knowledge in LMs successfully inject atomic facts, updated LMs fail to make inferences based on injected facts. In this work, we demonstrate that a context distillation-based approach can both impart knowledge about entities \emph{and} propagate that knowledge to enable broader inferences. Our approach consists of two stages: transfer set generation and distillation on the transfer set. We first generate a transfer set by prompting a language model to generate continuations from the entity definition. Then, we update the model parameters so that the distribution of the LM (the 'student') matches the distribution of the LM conditioned on the definition (the 'teacher') on the transfer set. Our experiments demonstrate that this approach is more effective at propagating knowledge updates than fine-tuning and other gradient-based knowledge-editing methods. Moreover, it does not compromise performance in other contexts, even when injecting the definitions of up to 150 entities at once.

****
ContiFormer: Continuous-Time Transformer for Irregular Time Series Modeling
Yuqi Chen, Kan Ren, Yansen Wang, Yuchen Fang, Weiwei Sun, Dongsheng Li
Modeling continuous-time dynamics on irregular time series is critical to account for data evolution and correlations that occur continuously. Traditional methods including recurrent neural networks or Transformer models leverage inductive bias via powerful neural architectures to capture complex patterns. However, due to their discrete characteristic, they have limitations in generalizing to continuous-time data paradigms. Though neural ordinary differential equations (Neural ODEs) and their variants have shown promising results in dealing with irregular time series, they often fail to capture the intricate correlations within these sequences. It is challenging yet demanding to concurrently model the relationship between input data points and capture the dynamic changes of the continuous-time system. To tackle this problem, we propose ContiFormer that extends the relation modeling of vanilla Transformer to the continuous-time domain, which explicitly incorporates the modeling abilities of continuous dynamics of Neural ODEs with the attention mechanism of Transformers. We mathematically characterize the expressive power of ContiFormer and illustrate that, by curated designs of function hypothesis, many Transformer variants specialized in irregular time series modeling can be covered as a special case of ContiFormer. A wide range of experiments on both synthetic and real-world datasets have illustrated the superior modeling capacities and prediction performance of ContiFormer on irregular time series data. The project link is https://seqml.github.io/contiformer/.
****
Differentiable Random Partition Models
Thomas Sutter, Alain Ryser, Joram Liebeskind, Julia Vogt
Partitioning a set of elements into an unknown number of mutually exclusive subsets is essential in many machine learning problems.However, assigning elements, such as samples in a dataset or neurons in a network layer, to an unknown and discrete number of subsets is inherently non-differentiable, prohibiting end-to-end gradient-based optimization of parameters.We overcome this limitation by proposing a novel two-step method for inferring partitions, which allows its usage in variational inference tasks.This new approach enables reparameterized gradients with respect to the parameters of the new random partition model.Our method works by inferring the number of elements per subset and, second, by filling these subsets in a learned order.We highlight the versatility of our general-purpose approach on three different challenging experiments: variational clustering, inference of shared and independent generative factors under weak supervision, and multitask learning.
****
Connecting Pre-trained Language Model and Downstream Task via Properties of Representation
Chenwei Wu, Holden Lee, Rong Ge
Recently, researchers have found that representations learned by large-scale pre-trained language models are useful in various downstream tasks. However, there is little theoretical understanding of how pre-training performance is related to downstream task performance. In this paper, we analyze how this performance transfer depends on the properties of the downstream task and the structure of the representations. We consider a log-linear model where a word can be predicted from its context through a network having softmax as its last layer. We show that even if the downstream task is highly structured and depends on a simple function of the hidden representation, there are still cases when a low pre-training loss cannot guarantee good performance on the downstream task. On the other hand, we propose and empirically validate the existence of an ``anchor vector'' in the representation space, and show that this assumption, together with properties of the downstream task, guarantees performance transfer.
****
Generalizable One-shot 3D Neural Head Avatar
Xueting Li, Shalini De Mello, Sifei Liu, Koki Nagano, Umar Iqbal, Jan Kautz
We present a method that reconstructs and animates a 3D head avatar from a singl

e-view portrait image. Existing methods either involve time-consuming optimization for a specific person with multiple images, or they struggle to synthesize intricate appearance details beyond the facial region. To address these limitations, we propose a framework that not only generalizes to unseen identities based on a single-view image without requiring person-specific optimization, but also captures characteristic details within and beyond the face area (e.g. hairstyle, accessories, etc.). At the core of our method are three branches that produce three tri-planes representing the coarse 3D geometry, detailed appearance of a source image, as well as the expression of a target image. By applying volumetric rendering to the combination of the three tri-planes followed by a super-resolution module, our method yields a high fidelity image of the desired identity, expression and pose. Once trained, our model enables efficient 3D head avatar reconstruction and animation via a single forward pass through a network. Experiments show that the proposed approach generalizes well to unseen validation datasets, surpassing SOTA baseline methods by a large margin on head avatar reconstruction and animation.
****
Equivariant Single View Pose Prediction Via Induced and Restriction Representations
Owen Howell, David Klee, Ondrej Biza, Linfeng Zhao, Robin Walters
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Unsupervised Learning for Solving the Travelling Salesman Problem
Yimeng Min, Yiwei Bai, Carla P. Gomes
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ContinuAR: Continuous Autoregression For Infinite-Fidelity Fusion
WEI XING, Yuxin Wang, Zheng Xing
Multi-fidelity fusion has become an important surrogate technique, which provides insights into expensive computer simulations and effectively improves decision-making, e.g., optimization, with less computational cost. Multi-fidelity fusion is much more computationally efficient compared to traditional single-fidelity surrogates. Despite the fast advancement of multi-fidelity fusion techniques, they lack a systematic framework to make use of the fidelity indicator, deal with high-dimensional and arbitrary data structure, and scale well to infinite-fidelity problems. In this work, we first generalize the popular autoregression (AR) to derive a novel linear fidelity differential equation (FiDE), paving the way to tractable infinite-fidelity fusion. We generalize FiDE to a high-dimensional system, which also provides a unifying framework to seemly bridge the gap between many multi- and single-fidelity GP-based models. We then propose ContinuAR, a rank-1 approximation solution to FiDEs, which is tractable to train, compatible with arbitrary multi-fidelity data structure, linearly scalable to the output dimension, and most importantly, delivers consistent SOTA performance with a significant margin over the baseline methods. Compared to the SOTA infinite-fidelity fusion, IFC, ContinuAR achieves up to 4x improvement in accuracy and 62,500x speedup in training time.
****
FOCAL: Contrastive Learning for Multimodal Time-Series Sensing Signals in Factorized Orthogonal Latent Space
Shengzhong Liu, Tomoyoshi Kimura, Dongxin Liu, Ruijie Wang, Jinyang Li, Suhas Diggavi, Mani Srivastava, Tarek Abdelzaher
This paper proposes a novel contrastive learning framework, called FOCAL, for extracting comprehensive features from multimodal time-series sensing signals through self-supervised training. Existing multimodal contrastive frameworks mostly

rely on the shared information between sensory modalities, but do not explicitly consider the exclusive modality information that could be critical to understanding the underlying sensing physics. Besides, contrastive frameworks for time series have not handled the temporal information locality appropriately. FOCAL solves these challenges by making the following contributions: First, given multimodal time series, it encodes each modality into a factorized latent space consisting of shared features and private features that are orthogonal to each other. The shared space emphasizes feature patterns consistent across sensory modalities through a modal-matching objective. In contrast, the private space extracts modality-exclusive information through a transformation-invariant objective. Second, we propose a temporal structural constraint for modality features, such that the average distance between temporally neighboring samples is no larger than that of temporally distant samples. Extensive evaluations are performed on four multimodal sensing datasets with two backbone encoders and two classifiers to demonstrate the superiority of FOCAL. It consistently outperforms the state-of-the-art baselines in downstream tasks with a clear margin, under different ratios of available labels. The code and self-collected dataset are available at https://github.com/tomoyoshki/focal.
****
Assumption violations in causal discovery and the robustness of score matching
Francesco Montagna, Atalanti Mastakouri, Elias Eulig, Nicoletta Noceti, Lorenzo Rosasco, Dominik Janzing, Bryon Aragam, Francesco Locatello
When domain knowledge is limited and experimentation is restricted by ethical, financial, or time constraints, practitioners turn to observational causal discovery methods to recover the causal structure, exploiting the statistical properties of their data. Because causal discovery without further assumptions is an ill-posed problem, each algorithm comes with its own set of usually untestable assumptions, some of which are hard to meet in real datasets. Motivated by these considerations, this paper extensively benchmarks the empirical performance of recent causal discovery methods on observational iid data generated under different background conditions, allowing for violations of the critical assumptions required by each selected approach. Our experimental findings show that score matching-based methods demonstrate surprising performance in the false positive and false negative rate of the inferred graph in these challenging scenarios, and we provide theoretical insights into their performance. This work is also the first effort to benchmark the stability of causal discovery algorithms with respect to the values of their hyperparameters. Finally, we hope this paper will set a new standard for the evaluation of causal discovery methods and can serve as an accessible entry point for practitioners interested in the field, highlighting the empirical implications of different algorithm choices.
****
Normalizing flow neural networks by JKO scheme
Chen Xu, Xiuyuan Cheng, Yao Xie
Normalizing flow is a class of deep generative models for efficient sampling and likelihood estimation, which achieves attractive performance, particularly in high dimensions. The flow is often implemented using a sequence of invertible residual blocks. Existing works adopt special network architectures and regularization of flow trajectories. In this paper, we develop a neural ODE flow network called JKO-iFlow, inspired by the Jordan-Kinderleherer-Otto (JKO) scheme, which unfolds the discrete-time dynamic of the Wasserstein gradient flow. The proposed method stacks residual blocks one after another, allowing efficient block-wise training of the residual blocks, avoiding sampling SDE trajectories and score matching or variational learning, thus reducing the memory load and difficulty in end-to-end training. We also develop adaptive time reparameterization of the flow network with a progressive refinement of the induced trajectory in probability space to improve the model accuracy further. Experiments with synthetic and real data show that the proposed JKO-iFlow network achieves competitive performance compared with existing flow and diffusion models at a significantly reduced computational and memory cost.
****

Stability-penalty-adaptive follow-the-regularized-leader: Sparsity, game-dependency, and best-of-both-worlds
Taira Tsuchiya, Shinji Ito, Junya Honda
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Domain Agnostic Fourier Neural Operators
Ning Liu, Siavash Jafarzadeh, Yue Yu
Fourier neural operators (FNOs) can learn highly nonlinear mappings between function spaces, and have recently become a popular tool for learning responses of complex physical systems. However, to achieve good accuracy and efficiency, FNOs rely on the Fast Fourier transform (FFT), which is restricted to modeling problems on rectangular domains. To lift such a restriction and permit FFT on irregular geometries as well as topology changes, we introduce domain agnostic Fourier neural operator (DAFNO), a novel neural operator architecture for learning surrogates with irregular geometries and evolving domains. The key idea is to incorporate a smoothed characteristic function in the integral layer architecture of FNOs, and leverage FFT to achieve rapid computations, in such a way that the geometric information is explicitly encoded in the architecture. In our empirical evaluation, DAFNO has achieved state-of-the-art accuracy as compared to baseline neural operator models on two benchmark datasets of material modeling and airfoil simulation. To further demonstrate the capability and generalizability of DAFNO in handling complex domains with topology changes, we consider a brittle material fracture evolution problem. With only one training crack simulation sample, DAFNO has achieved generalizability to unseen loading scenarios and substantially different crack patterns from the trained scenario. Our code and data accompanying this paper are available at https://github.com/ningliu-iga/DAFNO.
****
$\textbf{A}^2\textbf{CiD}^2$: Accelerating Asynchronous Communication in Decentralized Deep Learning
Adel Nabli, Eugene Belilovsky, Edouard Oyallon
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
MathNAS: If Blocks Have a Role in Mathematical Architecture Design
Qinsi Wang, Jinghan Ke, Zhi Liang, Sihai Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Block Broyden's Methods for Solving Nonlinear Equations
Chengchang Liu, Cheng Chen, Luo Luo, John C.S. Lui
This paper studies quasi-Newton methods for solving nonlinear equations. We propose block variants of both good and bad Broyden's methods, which enjoy explicit local superlinear convergence rates. Our block good Broyden's method has faster condition-number-free convergence rate than existing Broyden's methods because it takes the advantage of multiple rank modification on the Jacobian estimator. On the other hand, our block bad Broyden's method directly estimates the inverse of the Jacobian provably, which reduces the computational cost of the iteration. Our theoretical results provide some new insights on why good Broyden's method outperforms bad Broyden's method in most of the cases. The empirical results also demonstrate the superiority of our methods and validate our theoretical analysis.
****
Diffusion Hyperfeatures: Searching Through Time and Space for Semantic Correspon

dence
Grace Luo, Lisa Dunlap, Dong Huk Park, Aleksander Holynski, Trevor Darrell
Diffusion models have been shown to be capable of generating high-quality images, suggesting that they could contain meaningful internal representations. Unfortunately, the feature maps that encode a diffusion model's internal information are spread not only over layers of the network, but also over diffusion timesteps, making it challenging to extract useful descriptors. We propose Diffusion Hyperfeatures, a framework for consolidating multi-scale and multi-timestep feature maps into per-pixel feature descriptors that can be used for downstream tasks. These descriptors can be extracted for both synthetic and real images using the generation and inversion processes. We evaluate the utility of our Diffusion Hyperfeatures on the task of semantic keypoint correspondence: our method achieves superior performance on the SPair-71k real image benchmark. We also demonstrate that our method is flexible and transferable: our feature aggregation network trained on the inversion features of real image pairs can be used on the generation features of synthetic image pairs with unseen objects and compositions. Our code is available at https://diffusion-hyperfeatures.github.io.
****
No Change, No Gain: Empowering Graph Neural Networks with Expected Model Change Maximization for Active Learning
Zixing Song, Yifei Zhang, Irwin King
Graph Neural Networks (GNNs) are crucial for machine learning applications with graph-structured data, but their success depends on sufficient labeled data. We present a novel active learning (AL) method for GNNs, extending the Expected Model Change Maximization (EMCM) principle to improve prediction performance on unlabeled data. By presenting a Bayesian interpretation for the node embeddings generated by GNNs under the semi-supervised setting, we efficiently compute the closed-form EMCM acquisition function as the selection criterion for AL without re-training. Our method establishes a direct connection with expected prediction error minimization, offering theoretical guarantees for AL performance. Experiments demonstrate our method's effectiveness compared to existing approaches, in terms of both accuracy and efficiency.
****
Scaling Laws for Hyperparameter Optimization
Arlind Kadra, Maciej Janowski, Martin Wistuba, Josif Grabocka
Hyperparameter optimization is an important subfield of machine learning that focuses on tuning the hyperparameters of a chosen algorithm to achieve peak performance. Recently, there has been a stream of methods that tackle the issue of hyperparameter optimization, however, most of the methods do not exploit the dominant power law nature of learning curves for Bayesian optimization. In this work, we propose Deep Power Laws (DPL), an ensemble of neural network models conditioned to yield predictions that follow a power-law scaling pattern. Our method dynamically decides which configurations to pause and train incrementally by making use of gray-box evaluations. We compare our method against 7 state-of-the-art competitors on 3 benchmarks related to tabular, image, and NLP datasets covering 59 diverse tasks. Our method achieves the best results across all benchmarks by obtaining the best any-time results compared to all competitors.
****
A Robust and Opponent-Aware League Training Method for StarCraft II
Ruozi Huang, Xipeng Wu, Hongsheng Yu, Zhong Fan, Haobo Fu, Qiang Fu, Wei Yang
It is extremely difficult to train a superhuman Artificial Intelligence (AI) for games of similar size to StarCraft II. AlphaStar is the first AI that beat human professionals in the full game of StarCraft II, using a league training framework that is inspired by a game-theoretic approach. In this paper, we improve AlphaStar's league training in two significant aspects. We train goal-conditioned exploiters, whose abilities of spotting weaknesses in the main agent and the entire league are greatly improved compared to the unconditioned exploiters in AlphaStar. In addition, we endow the agents in the league with the new ability of opponent modeling, which makes the agent more responsive to the opponent's real-time strategy. Based on these improvements, we train a better and superhuman AI wi

th orders of magnitude less resources than AlphaStar (see Table 1 for a full com parison). Considering the iconic role of StarCraft II in game AI research, we be lieve our method and results on StarCraft II provide valuable design principles on how one would utilize the general league training framework for obtaining a l east-exploitable strategy in various, large-scale, real-world games.
****
Causal Fairness for Outcome Control
Drago Plecko, Elias Bareinboim
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
DeepPCR: Parallelizing Sequential Operations in Neural Networks
Federico Danieli, Miguel Sarabia, Xavier Suau Cuadros, Pau Rodriguez, Luca Zappe lla
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
DELTA: Diverse Client Sampling for Fasting Federated Learning
Lin Wang, Yongxin Guo, Tao Lin, Xiaoying Tang
Partial client participation has been widely adopted in Federated Learning (FL) to reduce the communication burden efficiently. However, an inadequate client sa mpling scheme can lead to the selection of unrepresentative subsets, resulting i n significant variance in model updates and slowed convergence. Existing samplin g methods are either biased or can be further optimized for faster convergence.I n this paper, we present DELTA, an unbiased sampling scheme designed to alleviat e these issues. DELTA characterizes the effects of client diversity and local va riance, and samples representative clients with valuable information for global model updates. In addition, DELTA is a proven optimal unbiased sampling scheme t hat minimizes variance caused by partial client participation and outperforms ot her unbiased sampling schemes in terms of convergence. Furthermore, to address full-client gradient dependence, we provide a practical version of DELTA dependi ng on the available clients' information, and also analyze its convergence. Our results are validated through experiments on both synthetic and real-world datas ets.
****
OpenAssistant Conversations - Democratizing Large Language Model Alignment
Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi Rui T am, Keith Stevens, Abdullah Barhoum, Duc Nguyen, Oliver Stanley, Richárd Nagyfi, Shahul ES, Sameer Suri, David Glushkov, Arnav Dantuluri, Andrew Maguire, Christ oph Schuhmann, Huu Nguyen, Alexander Mattick
Aligning large language models (LLMs) with human preferences has proven to drast ically improve usability and has driven rapid adoption as demonstrated by ChatGP T.Alignment techniques such as supervised fine-tuning (\textit{SFT}) and reinfo rcement learning from human feedback (\textit{RLHF}) greatly reduce the required skill and domain knowledge to effectively harness the capabilities of LLMs, inc reasing their accessibility and utility across various domains.However, state-of -the-art alignment techniques like \textit{RLHF} rely on high-quality human feed back data, which is expensive to create and often remains proprietary.In an effo rt to democratize research on large-scale alignment, we release OpenAssistant Co nversations, a human-generated, human-annotated assistant-style conversation cor pus consisting of 161,443 messages in 35 different languages, annotated with 461 ,292 quality ratings, resulting in over 10,000 complete and fully annotated conv ersation trees.The corpus is a product of a worldwide crowd-sourcing effort invo lving over 13,500 volunteers.Models trained on OpenAssistant Conversations show consistent improvements on standard benchmarks over respective base models.We re lease our code\footnote{\git} and data\footnote{\data} under a fully permissive

licence.
****
Conformal Meta-learners for Predictive Inference of Individual Treatment Effects
Ahmed M. Alaa, Zaid Ahmad, Mark van der Laan
We investigate the problem of machine learning-based (ML) predictive inference on individual treatment effects (ITEs). Previous work has focused primarily on developing ML-based “meta-learners” that can provide point estimates of the conditional average treatment effect (CATE)—these are model-agnostic approaches for combining intermediate nuisance estimates to produce estimates of CATE. In this paper, we develop conformal meta-learners, a general framework for issuing predictive intervals for ITEs by applying the standard conformal prediction (CP) procedure on top of CATE meta-learners. We focus on a broad class of meta-learners based on two-stage pseudo-outcome regression and develop a stochastic ordering framework to study their validity. We show that inference with conformal meta-learners is marginally valid if their (pseudo-outcome) conformity scores stochastically dominate “oracle” conformity scores evaluated on the unobserved ITEs. Additionally, we prove that commonly used CATE meta-learners, such as the doubly-robust learner, satisfy a model- and distribution-free stochastic (or convex) dominance condition, making their conformal inferences valid for practically-relevant levels of target coverage. Whereas existing procedures conduct inference on nuisance parameters (i.e., potential outcomes) via weighted CP, conformal meta-learners enable direct inference on the target parameter (ITE). Numerical experiments show that conformal meta-learners provide valid intervals with competitive efficiency while retaining the favorable point estimation properties of CATE meta-learners.
****
Simple and Controllable Music Generation
Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, Alexandre Defossez
We tackle the task of conditional music generation. We introduce MusicGen, a single Language Model (LM) that operates over several streams of compressed discrete music representation, i.e., tokens. Unlike prior work, MusicGen is comprised of a single-stage transformer LM together with efficient token interleaving patterns, which eliminates the need for cascading several models, e.g., hierarchically or upsampling. Following this approach, we demonstrate how MusicGen can generate high-quality samples, both mono and stereo, while being conditioned on textual description or melodic features, allowing better controls over the generated output. We conduct extensive empirical evaluation, considering both automatic and human studies, showing the proposed approach is superior to the evaluated baselines on a standard text-to-music benchmark. Through ablation studies, we shed light over the importance of each of the components comprising MusicGen. Music samples, code, and models are available at https://github.com/facebookresearch/audiocraft
****
Temporal Robustness against Data poisoning
Wenxiao Wang, Soheil Feizi
Data poisoning considers cases when an adversary manipulates the behavior of machine learning algorithms through malicious training data. Existing threat models of data poisoning center around a single metric, the number of poisoned samples. In consequence, if attackers can poison more samples than expected with affordable overhead, as in many practical scenarios, they may be able to render existing defenses ineffective in a short time. To address this issue, we leverage timestamps denoting the birth dates of data, which are often available but neglected in the past. Benefiting from these timestamps, we propose a temporal threat model of data poisoning with two novel metrics, earliness and duration, which respectively measure how long an attack started in advance and how long an attack lasted. Using these metrics, we define the notions of temporal robustness against data poisoning, providing a meaningful sense of protection even with unbounded amounts of poisoned samples when the attacks are temporally bounded. We present a benchmark with an evaluation protocol simulating continuous data collection and

periodic deployments of updated models, thus enabling empirical evaluation of temporal robustness. Lastly, we develop and also empirically verify a baseline defense, namely temporal aggregation, offering provable temporal robustness and highlighting the potential of our temporal threat model for data poisoning.
****

Optimal Treatment Regimes for Proximal Causal Learning

Tao Shen, Yifan Cui

A common concern when a policymaker draws causal inferences from and makes decisions based on observational data is that the measured covariates are insufficiently rich to account for all sources of confounding, i.e., the standard no confoundedness assumption fails to hold. The recently proposed proximal causal inference framework shows that proxy variables that abound in real-life scenarios can be leveraged to identify causal effects and therefore facilitate decision-making. Building upon this line of work, we propose a novel optimal individualized treatment regime based on so-called outcome and treatment confounding bridges. We then show that the value function of this new optimal treatment regime is superior to that of existing ones in the literature. Theoretical guarantees, including identification, superiority, excess value bound, and consistency of the estimated regime, are established. Furthermore, we demonstrate the proposed optimal regime via numerical experiments and a real data application.
****

Debias Coarsely, Sample Conditionally: Statistical Downscaling through Optimal Transport and Probabilistic Diffusion Models

Zhong Yi Wan, Ricardo Baptista, Anudhyan Boral, Yi-Fan Chen, John Anderson, Fei Sha, Leonardo Zepeda-Núñez

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Transformers over Directed Acyclic Graphs

Yuankai Luo, Veronika Thost, Lei Shi

Transformer models have recently gained popularity in graph representation learning as they have the potential to learn complex relationships beyond the ones captured by regular graph neural networks.The main research question is how to inject the structural bias of graphs into the transformer architecture,and several proposals have been made for undirected molecular graphs and, recently, also for larger network graphs.In this paper, we study transformers over directed acyclic graphs (DAGs) and propose architecture adaptations tailored to DAGs: (1) An attention mechanism that is considerably more efficient than the regular quadratic complexity of transformers and at the same time faithfully captures the DAG structure, and (2) a positional encoding of the DAG's partial order, complementing the former.We rigorously evaluate our approach over various types of tasks, ranging from classifying source code graphs to nodes in citation networks, and show that it is effective in two important aspects: in making graph transformers generally outperform graph neural networks tailored to DAGs and in improving SOTA graph transformer performance in terms of both quality and efficiency.
****

Understanding and Mitigating Copying in Diffusion Models

Gowthami Somepalli, Vasu Singla, Micah Goldblum, Jonas Geiping, Tom Goldstein

Images generated by diffusion models like Stable Diffusion are increasingly widespread. Recent works and even lawsuits have shown that these models are prone to replicating their training data, unbeknownst to the user. In this paper, we first analyze this memorization problem in text-to-image diffusion models. While it is widely believed that duplicated images in the training set are responsible for content replication at inference time, we observe that the text conditioning of the model plays a similarly important role. In fact, we see in our experiments that data replication often does not happen for unconditional models, while it is common in the text-conditional case. Motivated by our findings, we then propose several techniques for reducing data replication at both training and infer

ence time by randomizing and augmenting image captions in the training set. Code is available at https://github.com/somepago/DCR.
****

Credal Marginal MAP
Radu Marinescu, Debarun Bhattacharjya, Junkyu Lee, Fabio Cozman, Alexander Gray
Credal networks extend Bayesian networks to allow for imprecision in probability values. Marginal MAP is a widely applicable mixed inference task that identifies the most likely assignment for a subset of variables (called MAP variables). However, the task is extremely difficult to solve in credal networks particularly because the evaluation of each complete MAP assignment involves exact likelihood computations (combinatorial sums) over the vertices of a complex joint credal set representing the space of all possible marginal distributions of the MAP variables. In this paper, we explore Credal Marginal MAP inference and develop new exact methods based on variable elimination and depth-first search as well as several approximation schemes based on the mini-bucket partitioning and stochastic local search. An extensive empirical evaluation demonstrates the effectiveness of our new methods on random as well as real-world benchmark problems.
****

Multi-task Representation Learning for Pure Exploration in Bilinear Bandits
Subhojyoti Mukherjee, Qiaomin Xie, Josiah Hanna, Robert Nowak
We study multi-task representation learning for the problem of pure exploration in bilinear bandits. In bilinear bandits, an action takes theform of a pair of arms from two different entity types and the reward is a bilinear function of the known feature vectors of the arms. In the \textit{multi-task bilinear bandit problem}, we aim to find optimal actions for multiple tasks that share a common low-dimensional linear representation. The objective is to leverage this characteristic to expedite the process of identifying the best pair of arms for all tasks. We propose the algorithm GOBLIN that uses an experimental design approach to optimize sample allocations for learning the global representation as well as minimize the number of samples needed to identify the optimal pair of arms in individual tasks. To the best of our knowledge, this is the first study to give sample complexity analysis for pure exploration in bilinear bandits with shared representation. Our results demonstrate that by learning the shared representation across tasks, we achieve significantly improved sample complexity compared to the traditional approach of solving tasks independently.
****

Mechanic: A Learning Rate Tuner
Ashok Cutkosky, Aaron Defazio, Harsh Mehta
We introduce a technique for tuning the learning rate scale factor of any base optimization algorithm and schedule automatically, which we call Mechanic. Our method provides a practical realization of recent theoretical reductions for accomplishing a similar goal in online convex optimization. We rigorously evaluate Mechanic on a range of large scale deep learning tasks with varying batch sizes, schedules, and base optimization algorithms. These experiments demonstrate that depending on the problem, Mechanic either comes very close to, matches or even improves upon manual tuning of learning rates.
****

Compositional Policy Learning in Stochastic Control Systems with Formal Guarantees
■or■e Žikeli■, Mathias Lechner, Abhinav Verma, Krishnendu Chatterjee, Thomas Henzinger
Reinforcement learning has shown promising results in learning neural network policies for complicated control tasks. However, the lack of formal guarantees about the behavior of such policies remains an impediment to their deployment. We propose a novel method for learning a composition of neural network policies in stochastic environments, along with a formal certificate which guarantees that a specification over the policy's behavior is satisfied with the desired probability. Unlike prior work on verifiable RL, our approach leverages the compositional nature of logical specifications provided in SpectRL, to learn over graphs of probabilistic reach-avoid specifications. The formal guarantees are provided by l

earning neural network policies together with reach-avoid supermartingales (RASM) for the graph’s sub-tasks and then composing them into a global policy. We also derive a tighter lower bound compared to previous work on the probability of reach-avoidance implied by a RASM, which is required to find a compositional policy with an acceptable probabilistic threshold for complex tasks with multiple edge policies. We implement a prototype of our approach and evaluate it on a Stochastic Nine Rooms environment.
****
Fast Exact Leverage Score Sampling from Khatri-Rao Products with Applications to Tensor Decomposition
Vivek Bharadwaj, Osman Asif Malik, Riley Murray, Laura Grigori, Aydin Buluc, James Demmel
We present a data structure to randomly sample rows from the Khatri-Rao product of several matrices according to the exact distribution of its leverage scores. Our proposed sampler draws each row in time logarithmic in the height of the Khatri-Rao product and quadratic in its column count, with persistent space overhead at most the size of the input matrices. As a result, it tractably draws samples even when the matrices forming the Khatri-Rao product have tens of millions of rows each. When used to sketch the linear least-squares problems arising in Candecomp / PARAFAC decomposition, our method achieves lower asymptotic complexity per solve than recent state-of-the-art methods. Experiments on billion-scale sparse tensors and synthetic data validate our theoretical claims, with our algorithm achieving higher accuracy than competing methods as the decomposition rank grows.
****
Online Performative Gradient Descent for Learning Nash Equilibria in Decision-Dependent Games
Zihan Zhu, Ethan Fang, Zhuoran Yang
We study the multi-agent game within the innovative framework of decision-dependent games, which establishes a feedback mechanism that population data reacts to agents’ actions and further characterizes the strategic interactions between agents. We focus on finding the Nash equilibrium of decision-dependent games in the bandit feedback setting. However, since agents are strategically coupled, traditional gradient-based methods are infeasible without the gradient oracle. To overcome this challenge, we model the strategic interactions by a general parametric model and propose a novel online algorithm, Online Performative Gradient Descent (OPGD), which leverages the ideas of online stochastic approximation and projected gradient descent to learn the Nash equilibrium in the context of function approximation for the unknown gradient. In particular, under mild assumptions on the function classes defined in the parametric model, we prove that OPGD can find the Nash equilibrium efficiently for strongly monotone decision-dependent games. Synthetic numerical experiments validate our theory.
****
AD-PT: Autonomous Driving Pre-Training with Large-scale Point Cloud Dataset
Jiakang Yuan, Bo Zhang, Xiangchao Yan, Botian Shi, Tao Chen, Yikang LI, Yu Qiao
It is a long-term vision for Autonomous Driving (AD) community that the perception models can learn from a large-scale point cloud dataset, to obtain unified representations that can achieve promising results on different tasks or benchmarks. Previous works mainly focus on the self-supervised pre-training pipeline, meaning that they perform the pre-training and fine-tuning on the same benchmark, which is difficult to attain the performance scalability and cross-dataset application for the pre-training checkpoint. In this paper, for the first time, we are committed to building a large-scale pre-training point-cloud dataset with diverse data distribution, and meanwhile learning generalizable representations from such a diverse pre-training dataset. We formulate the point-cloud pre-training task as a semi-supervised problem, which leverages the few-shot labeled and massive unlabeled point-cloud data to generate the unified backbone representations that can be directly applied to many baseline models and benchmarks, decoupling the AD-related pre-training process and downstream fine-tuning task. During the period of backbone pre-training, by enhancing the scene- and instance-level dist

ribution diversity and exploiting the backbone's ability to learn from unknown instances, we achieve significant performance gains on a series of downstream perception benchmarks including Waymo, nuScenes, and KITTI, under different baseline models like PV-RCNN++, SECOND, CenterPoint.
****
Aging with GRACE: Lifelong Model Editing with Discrete Key-Value Adaptors
Tom Hartvigsen, Swami Sankaranarayanan, Hamid Palangi, Yoon Kim, Marzyeh Ghassemi
Deployed language models decay over time due to shifting inputs, changing user needs, or emergent world-knowledge gaps. When such problems are identified, we want to make targeted edits while avoiding expensive retraining. However, current model editors, which modify such behaviors of pre-trained models, degrade model performance quickly across multiple, sequential edits. We propose GRACE, a \textit{lifelong} model editing method, which implements spot-fixes on streaming errors of a deployed model, ensuring minimal impact on unrelated inputs. GRACE writes new mappings into a pre-trained model's latent space, creating a discrete, local codebook of edits without altering model weights. This is the first method enabling thousands of sequential edits using only streaming errors. Our experiments on T5, BERT, and GPT models show GRACE's state-of-the-art performance in making and retaining edits, while generalizing to unseen inputs. Our code is available at github.com/thartvigsen/grace.
****
On the Identifiability of Sparse ICA without Assuming Non-Gaussianity
Ignavier Ng, Yujia Zheng, Xinshuai Dong, Kun Zhang
Independent component analysis (ICA) is a fundamental statistical tool used to reveal hidden generative processes from observed data. However, traditional ICA approaches struggle with the rotational invariance inherent in Gaussian distributions, often necessitating the assumption of non-Gaussianity in the underlying sources. This may limit their applicability in broader contexts. To accommodate Gaussian sources, we develop an identifiability theory that relies on second-order statistics without imposing further preconditions on the distribution of sources, by introducing novel assumptions on the connective structure from sources to observed variables. Different from recent work that focuses on potentially restrictive connective structures, our proposed assumption of structural variability is both considerably less restrictive and provably necessary. Furthermore, we propose two estimation methods based on second-order statistics and sparsity constraint. Experimental results are provided to validate our identifiability theory and estimation methods.
****
Unbiased Compression Saves Communication in Distributed Optimization: When and How Much?
Yutong He, Xinmeng Huang, Kun Yuan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Pareto Frontiers in Deep Feature Learning: Data, Compute, Width, and Luck
Benjamin Edelman, Surbhi Goel, Sham Kakade, Eran Malach, Cyril Zhang
In modern deep learning, algorithmic choices (such as width, depth, and learning rate) are known to modulate nuanced resource tradeoffs. This work investigates how these complexities necessarily arise for feature learning in the presence of computational-statistical gaps. We begin by considering offline sparse parity learning, a supervised classification problem which admits a statistical query lower bound for gradient-based training of a multilayer perceptron. This lower bound can be interpreted as a multi-resource tradeoff frontier: successful learning can only occur if one is sufficiently rich (large model), knowledgeable (large dataset), patient (many training iterations), or lucky (many random guesses). We show, theoretically and experimentally, that sparse initialization and increasing network width yield significant improvements in sample efficiency in this set

ting. Here, width plays the role of parallel search: it amplifies the probabilit y of finding "lottery ticket" neurons, which learn sparse features more sample-e fficiently. Finally, we show that the synthetic sparse parity task can be useful as a proxy for real problems requiring axis-aligned feature learning. We demons trate improved sample efficiency on tabular classification benchmarks by using w ide, sparsely-initialized MLP models; these networks sometimes outperform tuned random forests.
****
Reliable learning in challenging environments
Maria-Florina F. Balcan, Steve Hanneke, Rattana Pukdee, Dravyansh Sharma
The problem of designing learners that provide guarantees that their predictions are provably correct is of increasing importance in machine learning. However, learning theoretic guarantees have only been considered in very specific setting s. In this work, we consider the design and analysis of reliable learners in ch allenging test-time environments as encountered in modern machine learning probl ems: namely adversarial test-time attacks (in several variations) and natural di stribution shifts. In this work, we provide a reliable learner with provably op timal guarantees in such settings. We discuss computationally feasible implement ations of the learner and further show that our algorithm achieves strong positi ve performance guarantees on several natural examples: for example, linear separ ators under log-concave distributions or smooth boundary classifiers under smoot h probability distributions.
****
Retaining Beneficial Information from Detrimental Data for Neural Network Repair
Long-Kai Huang, Peilin Zhao, Junzhou Huang, Sinno Pan
The performance of deep learning models heavily relies on the quality of the tra ining data. Inadequacies in the training data, such as corrupt input or noisy la bels, can lead to the failure of model generalization. Recent studies propose re pairing the model by identifying the training samples that contribute to the fai lure and removing their influence from the model. However, it is important to no te that the identified data may contain both beneficial and detrimental informat ion. Simply erasing the information of the identified data from the model can ha ve a negative impact on its performance, especially when accurate data is mistak enly identified as detrimental and removed. To overcome this challenge, we propo se a novel approach that leverages the knowledge obtained from a retained clean set. Our method first identifies harmful data by utilizing the clean set, then s eparates the beneficial and detrimental information within the identified data. Finally, we utilize the extracted beneficial information to enhance the model's performance. Through empirical evaluations, we demonstrate that our method outpe rforms baseline approaches in both identifying harmful data and rectifying model failures. Particularly in scenarios where identification is challenging and a s ignificant amount of benign data is involved, our method improves performance wh ile the baselines deteriorate due to the erroneous removal of beneficial informa tion.
****
Unsupervised Optical Flow Estimation with Dynamic Timing Representation for Spik e Camera
Lujie Xia, Ziluo Ding, Rui Zhao, Jiyuan Zhang, Lei Ma, Zhaofei Yu, Tiejun Huang, Ruiqin Xiong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
No-regret Algorithms for Fair Resource Allocation
Abhishek Sinha, Ativ Joshi, Rajarshi Bhattacharjee, Cameron Musco, Mohammad Haji esmaili
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Bypass Exponential Time Preprocessing: Fast Neural Network Training via Weight-D
ata Correlation Preprocessing
Josh Alman, ■■ ■, Zhao Song, Ruizhe Zhang, Danyang Zhuo
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Online PCA in Converging Self-consistent Field Equations
Xihan Li, Xiang Chen, Rasul Tutunov, Haitham Bou Ammar, Lei Wang, Jun Wang
Self-consistent Field (SCF) equation is a type of nonlinear eigenvalue problem i
n which the matrix to be eigen-decomposed is a function of its own eigenvectors.
 It is of great significance in computational science for its connection to the
Schrödinger equation. Traditional fixed-point iteration methods for solving such
 equations suffer from non-convergence issues. In this work, we present a novel
perspective on such SCF equations as a principal component analysis (PCA) for no
n-stationary time series, in which a distribution and its own top principal comp
onents are mutually updated over time, and the equilibrium state of the model co
rresponds to the solution of the SCF equations. By the new perspective, online P
CA techniques are able to engage in so as to enhance the convergence of the mode
l towards the equilibrium state, acting as a new set of tools for converging the
 SCF equations. With several numerical adaptations, we then develop a new algori
thm for converging the SCF equation, and demonstrated its high convergence capac
ity with experiments on both synthesized and real electronic structure scenarios
.
****
DiffPack: A Torsional Diffusion Model for Autoregressive Protein Side-Chain Pack
ing
Yangtian Zhang, Zuobai Zhang, Bozitao Zhong, Sanchit Misra, Jian Tang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
ARTIC3D: Learning Robust Articulated 3D Shapes from Noisy Web Image Collections
Chun-Han Yao, Amit Raj, Wei-Chih Hung, Michael Rubinstein, Yuanzhen Li, Ming-Hsu
an Yang, Varun Jampani
Estimating 3D articulated shapes like animal bodies from monocular images is inh
erently challenging due to the ambiguities of camera viewpoint, pose, texture, l
ighting, etc. We propose ARTIC3D, a self-supervised framework to reconstruct per
-instance 3D shapes from a sparse image collection in-the-wild. Specifically, AR
TIC3D is built upon a skeleton-based surface representation and is further guide
d by 2D diffusion priors from Stable Diffusion. First, we enhance the input imag
es with occlusions/truncation via 2D diffusion to obtain cleaner mask estimates
and semantic features. Second, we perform diffusion-guided 3D optimization to es
timate shape and texture that are of high-fidelity and faithful to input images.
 We also propose a novel technique to calculate more stable image-level gradient
s via diffusion models compared to existing alternatives. Finally, we produce re
alistic animations by fine-tuning the rendered shape and texture under rigid par
t transformations. Extensive evaluations on multiple existing datasets as well a
s newly introduced noisy web image collections with occlusions and truncation de
monstrate that ARTIC3D outputs are more robust to noisy images, higher quality i
n terms of shape and texture details, and more realistic when animated.
****
Noether Embedding: Efficient Learning of Temporal Regularities
Chi Gao, Zidong Zhou, Luping Shi
Learning to detect and encode temporal regularities (TRs) in events is a prerequ
isite for human-like intelligence. These regularities should be formed from limi

ted event samples and stored as easily retrievable representations. Existing eve nt embeddings, however, cannot effectively decode TR validity with well-trained vectors, let alone satisfy the efficiency requirements. We develop Noether Embed ding (NE) as the first efficient TR learner with event embeddings. Specifically, NE possesses the intrinsic time-translation symmetries of TRs indicated as cons erved local energies in the embedding space. This structural bias reduces the ca lculation of each TR validity to embedding each event sample, enabling NE to ach ieve data-efficient TR formation insensitive to sample size and time-efficient T R retrieval in constant time complexity. To comprehensively evaluate the TR lear ning capability of embedding models, we define complementary tasks of TR detecti on and TR query, formulate their evaluation metrics, and assess embeddings on cl assic ICEWS14, ICEWS18, and GDELT datasets. Our experiments demonstrate that NE consistently achieves about double the F1 scores for detecting valid TRs compare d to classic embeddings, and it provides over ten times higher confidence scores for querying TR intervals. Additionally, we showcase NE's potential application s in social event prediction, personal decision-making, and memory-constrained s cenarios.
****
$\texttt{TACO}$: Temporal Latent Action-Driven Contrastive Loss for Visual Reinf orcement Learning
Ruijie Zheng, Xiyao Wang, Yanchao Sun, Shuang Ma, Jieyu Zhao, Huazhe Xu, Hal Dau mé III, Furong Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
On the choice of Perception Loss Function for Learned Video Compression
Sadaf Salehkalaibar, Truong Buu Phan, Jun Chen, Wei Yu, Ashish Khisti
We study causal, low-latency, sequential video compression when the output is su bjected to both a mean squared-error (MSE) distortion loss as well as a percepti on loss to target realism. Motivated by prior approaches, we consider two differ ent perception loss functions (PLFs). The first, PLF-JD, considers the joint di stribution (JD) of all the video frames up to the current one, while the second metric, PLF-FMD, considers the framewise marginal distributions (FMD) between t he source and reconstruction. Using information theoretic analysis and deep-lear ning based experiments, we demonstrate that the choice of PLF can have a signifi cant effect on the reconstruction, especially at low-bit rates. In particular, w hile the reconstruction based on PLF-JD can better preserve the temporal correla tion across frames, it also imposes a significant penalty in distortion compare d to PLF-FMD and further makes it more difficult to recover from errors made in the earlier output frames. Although the choice of PLF decisively affects recon struction quality, we also demonstrate that it may not be essential to commit to a particular PLF during encoding and the choice of PLF can be delegated to the decoder. In particular, encoded representations generated by training a system t o minimize the MSE (without requiring either PLF) can be {\em near universal} and can generate close to optimal reconstructions for either choice of PLF at th e decoder. We validate our results using (one-shot) information-theoretic analy sis, detailed study of the rate-distortion-perception tradeoff of the Gauss-Mark ov source model as well as deep-learning based experiments on moving MNIST and K TH datasets.
****
Imitation Learning from Vague Feedback
Xin-Qiang Cai, Yu-Jie Zhang, Chao-Kai Chiang, Masashi Sugiyama
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Semantic segmentation of sparse irregular point clouds for leaf/wood discriminat

ion
Yuchen BAI, Jean-Baptiste Durand, Grégoire Vincent, Florence Forbes
Lidar (Light Detection and Ranging) has become an essential part of the remote s ensing toolbox used for biosphere monitoring. In particular, Lidar provides the opportunity to map forest leaf area with unprecedented accuracy, while leaf area has remained an important source of uncertainty affecting models of gas exchanges between the vegetation and the atmosphere. Unmanned Aerial Vehicles (UAV) are easy to mobilize and therefore allow frequent revisits to track the response of vegetation to climate change. However, miniature sensors embarked on UAVs usually provide point clouds of limited density, which are further affected by a strong decrease in density from top to bottom of the canopy due to progressively stronger occlusion. In such a context, discriminating leaf points from wood points presents a significant challenge due in particular to strong class imbalance and spatially irregular sampling intensity. Here we introduce a neural network model based on the Pointnet ++ architecture which makes use of point geometry only (excluding any spectral information). To cope with local data sparsity, we propose an innovative sampling scheme which strives to preserve local important geometric information. We also propose a loss function adapted to the severe class imbalance. We show that our model outperforms state-of-the-art alternatives on UAV point clouds. We discuss future possible improvements, particularly regarding much denser point clouds acquired from below the canopy.
****
Max-Margin Token Selection in Attention Mechanism
Davoud Ataee Tarzanagh, Yingcong Li, Xuechen Zhang, Samet Oymak
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Locality-Aware Generalizable Implicit Neural Representation
Doyup Lee, Chiheon Kim, Minsu Cho, WOOK SHIN HAN
Generalizable implicit neural representation (INR) enables a single continuous function, i.e., a coordinate-based neural network, to represent multiple data instances by modulating its weights or intermediate features using latent codes. However, the expressive power of the state-of-the-art modulation is limited due to its inability to localize and capture fine-grained details of data entities such as specific pixels and rays. To address this issue, we propose a novel framework for generalizable INR that combines a transformer encoder with a locality-aware INR decoder. The transformer encoder predicts a set of latent tokens from a data instance to encode local information into each latent token. The locality-aware INR decoder extracts a modulation vector by selectively aggregating the latent tokens via cross-attention for a coordinate input and then predicts the output by progressively decoding with coarse-to-fine modulation through multiple frequency bandwidths. The selective token aggregation and the multi-band feature modulation enable us to learn locality-aware representation in spatial and spectral aspects, respectively. Our framework significantly outperforms previous generalizable INRs and validates the usefulness of the locality-aware latents for downstream tasks such as image generation.
****
StableRep: Synthetic Images from Text-to-Image Models Make Strong Visual Representation Learners
Yonglong Tian, Lijie Fan, Phillip Isola, Huiwen Chang, Dilip Krishnan
We investigate the potential of learning visual representations using synthetic images generated by text-to-image models. This is a natural question in the light of the excellent performance of such models in generating high-quality images. We consider specifically the Stable Diffusion, one of the leading open source text-to-image models. We show that (1) when the generative model is properly configured, training self-supervised methods on synthetic images can match or beat the real image counterpart;(2) by treating the multiple images generated from the same text prompt as positives for each other, we develop a multi-positive contr

astive learning method, which we call StableRep. With solely synthetic images, the representations learned by StableRep surpass the performance of representations learned by SimCLR and CLIP using the same set of text prompts and corresponding real images, on large scale datasets. When we further add language supervision, \name~trained with 20M synthetic images (10M captions) achieves better accuracy than CLIP trained with 50M real images (50M captions).
****
DropCompute: simple and more robust distributed synchronous training via compute variance reduction
Niv Giladi, Shahar Gottlieb, moran shkolnik, Asaf Karnieli, Ron Banner, Elad Hoffer, Kfir Y. Levy, Daniel Soudry
Background: Distributed training is essential for large scale training of deep neural networks (DNNs). The dominant methods for large scale DNN training are synchronous (e.g. All-Reduce), but these require waiting for all workers in each step. Thus, these methods are limited by the delays caused by straggling workers.Results: We study a typical scenario in which workers are straggling due to variability in compute time. We find an analytical relation between compute time properties and scalability limitations, caused by such straggling workers. With these findings, we propose a simple yet effective decentralized method to reduce the variation among workers and thus improve the robustness of synchronous training. This method can be integrated with the widely used All-Reduce. Our findings are validated on large-scale training tasks using 200 Gaudi Accelerators.
****
A Unified Generalization Analysis of Re-Weighting and Logit-Adjustment for Imbalanced Learning
Zitai Wang, Qianqian Xu, Zhiyong Yang, Yuan He, Xiaochun Cao, Qingming Huang
Real-world datasets are typically imbalanced in the sense that only a few classes have numerous samples, while many classes are associated with only a few samples. As a result, a naive ERM learning process will be biased towards the majority classes, making it difficult to generalize to the minority classes. To address this issue, one simple but effective approach is to modify the loss function to emphasize the learning on minority classes, such as re-weighting the losses or adjusting the logits via class-dependent terms. However, existing generalization analysis of such losses is still coarse-grained and fragmented, failing to explain some empirical results. To bridge this gap between theory and practice, we propose a novel technique named data-dependent contraction to capture how these modified losses handle different classes. On top of this technique, a fine-grained generalization bound is established for imbalanced learning, which helps reveal the mystery of re-weighting and logit-adjustment in a unified manner. Furthermore, a principled learning algorithm is developed based on the theoretical insights. Finally, the empirical results on benchmark datasets not only validate the theoretical results but also demonstrate the effectiveness of the proposed method.
****
Sketching Algorithms for Sparse Dictionary Learning: PTAS and Turnstile Streaming
Gregory Dexter, Petros Drineas, David Woodruff, Taisuke Yasuda
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Annotator: A Generic Active Learning Baseline for LiDAR Semantic Segmentation
Binhui Xie, Shuang Li, Qingju Guo, Chi Liu, Xinjing Cheng
Active learning, a label-efficient paradigm, empowers models to interactively query an oracle for labeling new data. In the realm of LiDAR semantic segmentation, the challenges stem from the sheer volume of point clouds, rendering annotation labor-intensive and cost-prohibitive. This paper presents Annotator, a general and efficient active learning baseline, in which a voxel-centric online selection strategy is tailored to efficiently probe and annotate the salient and exempl

ar voxel girds within each LiDAR scan, even under distribution shift. Concretely
, we first execute an in-depth analysis of several common selection strategies s
uch as Random, Entropy, Margin, and then develop voxel confusion degree (VCD) to
 exploit the local topology relations and structures of point clouds. Annotator
excels in diverse settings, with a particular focus on active learning (AL), act
ive source-free domain adaptation (ASFDA), and active domain adaptation (ADA). I
t consistently delivers exceptional performance across LiDAR semantic segmentati
on benchmarks, spanning both simulation-to-real and real-to-real scenarios. Surp
risingly, Annotator exhibits remarkable efficiency, requiring significantly fewe
r annotations, e.g., just labeling five voxels per scan in the SynLiDAR $\rightarrow$ Semant
icKITTI task. This results in impressive performance, achieving 87.8% fully-supe
rvised performance under AL, 88.5% under ASFDA, and 94.4% under ADA. We envision
 that Annotator will offer a simple, general, and efficient solution for label-e
fficient 3D applications.
****
Kissing to Find a Match: Efficient Low-Rank Permutation Representation
Hannah Dröge, Zorah Lähner, Yuval Bahat, Onofre Martorell Nadal, Felix Heide, Mi
chael Moeller
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Creating Multi-Level Skill Hierarchies in Reinforcement Learning
Joshua B. Evans, Özgür ■im■ek
What is a useful skill hierarchy for an autonomous agent? We propose an answer b
ased on a graphical representation of how the interaction between an agent and i
ts environment may unfold. Our approach uses modularity maximisation as a centra
l organising principle to expose the structure of the interaction graph at multi
ple levels of abstraction. The result is a collection of skills that operate at
varying time scales, organised into a hierarchy, where skills that operate over
longer time scales are composed of skills that operate over shorter time scales.
 The entire skill hierarchy is generated automatically, with no human input, inc
luding the skills themselves (their behaviour, when they can be called, and when
 they terminate) as well as the dependency structure between them. In a wide ran
ge of environments, this approach generates skill hierarchies that are intuitive
ly appealing and that considerably improve the learning performance of the agent
.
****
Winner Takes It All: Training Performant RL Populations for Combinatorial Optimi
zation
Nathan Grinsztajn, Daniel Furelos-Blanco, Shikha Surana, Clément Bonnet, Tom Bar
rett
Applying reinforcement learning (RL) to combinatorial optimization problems is a
ttractive as it removes the need for expert knowledge or pre-solved instances. H
owever, it is unrealistic to expect an agent to solve these (often NP-)hard prob
lems in a single shot at inference due to their inherent complexity. Thus, leadi
ng approaches often implement additional search strategies, from stochastic samp
ling and beam-search to explicit fine-tuning. In this paper, we argue for the be
nefits of learning a population of complementary policies, which can be simultan
eously rolled out at inference. To this end, we introduce Poppy, a simple traini
ng procedure for populations. Instead of relying on a predefined or hand-crafted
 notion of diversity, Poppy induces an unsupervised specialization targeted sole
ly at maximizing the performance of the population. We show that Poppy produces
a set of complementary policies, and obtains state-of-the-art RL results on thre
e popular NP-hard problems: traveling salesman, capacitated vehicle routing, and
 job-shop scheduling.
****
Revisiting Scalarization in Multi-Task Learning: A Theoretical Perspective
Yuzheng Hu, Ruicheng Xian, Qilong Wu, Qiuling Fan, Lang Yin, Han Zhao

Linear scalarization, i.e., combining all loss functions by a weighted sum, has
been the default choice in the literature of multi-task learning (MTL) since its
inception. In recent years, there is a surge of interest in developing Speciali
zed Multi-Task Optimizers (SMTOs) that treat MTL as a multi-objective optimizati
on problem. However, it remains open whether there is a fundamental advantage of
SMTOs over scalarization. In fact, heated debates exist in the community compar
ing these two types of algorithms, mostly from an empirical perspective. To appr
oach the above question, in this paper, we revisit scalarization from a theoreti
cal perspective. We focus on linear MTL models and study whether scalarization i
s capable of fully exploring the Pareto front. Our findings reveal that, in cont
rast to recent works that claimed empirical advantages of scalarization, scalari
zation is inherently incapable of full exploration, especially for those Pareto
optimal solutions that strike the balanced trade-offs between multiple tasks. Mo
re concretely, when the model is under-parametrized, we reveal a multi-surface s
tructure of the feasible region and identify necessary and sufficient conditions
for full exploration. This leads to the conclusion that scalarization is in gen
eral incapable of tracing out the Pareto front. Our theoretical results partiall
y answer the open questions in Xin et al. (2021), and provide a more intuitive e
xplanation on why scalarization fails beyond non-convexity. We additionally perf
orm experiments on a real-world dataset using both scalarization and state-of-th
e-art SMTOs. The experimental results not only corroborate our theoretical findi
ngs, but also unveil the potential of SMTOs in finding balanced solutions, which
cannot be achieved by scalarization.
****
Provable Guarantees for Generative Behavior Cloning: Bridging Low-Level Stabilit
y and High-Level Behavior
Adam Block, Ali Jadbabaie, Daniel Pfrommer, Max Simchowitz, Russ Tedrake
We propose a theoretical framework for studying behavior cloning of complex expe
rt demonstrations using generative modeling.Our framework invokes low-level cont
rollers - either learned or implicit in position-command control - to stabilize
imitation around expert demonstrations. We show that with (a) a suitable low-lev
el stability guarantee and (b) a powerful enough generative model as our imitati
on learner, pure supervised behavior cloning can generate trajectories matching
the per-time step distribution of essentially arbitrary expert trajectories in
an optimal transport cost. Our analysis relies on a stochastic continuity proper
ty of the learned policy we call "total variation continuity" (TVC). We then sho
w that TVC can be ensured with minimal degradation of accuracy by combining a po
pular data-augmentation regimen with a novel algorithmic trick: adding augmentat
ion noise at execution time. We instantiate our guarantees for policies paramete
rized by diffusion models and prove that if the learner accurately estimates the
score of the (noise-augmented) expert policy, then the distribution of imitator
trajectories is close to the demonstrator distribution in a natural optimal tra
nsport distance. Our analysis constructs intricate couplings between noise-augme
nted trajectories, a technique that may be of independent interest. We conclude
by empirically validating our algorithmic recommendations, and discussing implic
ations for future research directions for better behavior cloning with generativ
e modeling.
****
Prefix-Tree Decoding for Predicting Mass Spectra from Molecules
Samuel Goldman, John Bradshaw, Jiayi Xin, Connor Coley
Computational predictions of mass spectra from molecules have enabled the discov
ery of clinically relevant metabolites. However, such predictive tools are still
limited as they occupy one of two extremes, either operating (a) by fragmentin
g molecules combinatorially with overly rigid constraints on potential rearrange
ments and poor time complexity or (b) by decoding lossy and nonphysical discreti
zed spectra vectors. In this work, we use a new intermediate strategy for predic
ting mass spectra from molecules by treating mass spectra as sets of molecular f
ormulae, which are themselves multisets of atoms. After first encoding an input
molecular graph, we decode a set of molecular subformulae, each of which specify
a predicted peak in the mass spectrum, the intensities of which are predicted b

y a second model. Our key insight is to overcome the combinatorial possibilities for molecular subformulae by decoding the formula set using a prefix tree structure, atom-type by atom-type, representing a general method for ordered multiset decoding. We show promising empirical results on mass spectra prediction tasks.
****

Knowledge-Augmented Reasoning Distillation for Small Language Models in Knowledge-Intensive Tasks

Minki Kang, Seanie Lee, Jinheon Baek, Kenji Kawaguchi, Sung Ju Hwang

Large Language Models (LLMs) have shown promising performance in knowledge-intensive reasoning tasks that require a compound understanding of knowledge. However, deployment of the LLMs in real-world applications can be challenging due to their high computational requirements and concerns on data privacy.Previous studies have focused on building task-specific small Language Models (LMs) by fine-tuning them with labeled data or distilling LLMs. However, these approaches are ill-suited for knowledge-intensive reasoning tasks due to the limited capacity of small LMs in memorizing the knowledge required.Motivated by our theoretical analysis on memorization, we propose Knowledge-Augmented Reasoning Distillation (KARD), a novel method that fine-tunes small LMs to generate rationales obtained from LLMs with augmented knowledge retrieved from an external knowledge base. Moreover, we further propose a neural reranker to obtain documents relevant to rationale generation. We empirically show that KARD significantly improves the performance of small T5 and GPT models on the challenging knowledge-intensive reasoning datasets, namely MedQA-USMLE, StrategyQA, and OpenbookQA.Notably, our method makes the 250M T5 models achieve superior performance against the fine-tuned 3B models, having 12 times larger parameters, on both MedQA-USMLE and StrategyQA benchmarks.
****

Nonparametric Identifiability of Causal Representations from Unknown Interventions

Julius von Kügelgen, Michel Besserve, Liang Wendong, Luigi Gresele, Armin Keki■, Elias Bareinboim, David Blei, Bernhard Schölkopf

We study causal representation learning, the task of inferring latent causal variables and their causal relations from high-dimensional functions (“mixtures”) of the variables. Prior work relies on weak supervision, in the form of counterfactual pre- and post-intervention views or temporal structure; places restrictive assumptions, such as linearity, on the mixing function or latent causal model; or requires partial knowledge of the generative process, such as the causal graph or intervention targets. We instead consider the general setting in which both the causal model and the mixing function are nonparametric. The learning signal takes the form of multiple datasets, or environments, arising from unknown interventions in the underlying causal model. Our goal is to identify both the ground truth latents and their causal graph up to a set of ambiguities which we show to be irresolvable from interventional data. We study the fundamental setting of two causal variables and prove that the observational distribution and one perfect intervention per node suffice for identifiability, subject to a genericity condition. This condition rules out spurious solutions that involve fine-tuning of the intervened and observational distributions, mirroring similar conditions for nonlinear cause-effect inference. For an arbitrary number of variables, we show that at least one pair of distinct perfect interventional domains per node guarantees identifiability. Further, we demonstrate that the strengths of causal influences among the latent variables are preserved by all equivalent solutions, rendering the inferred representation appropriate for drawing causal conclusions from new data. Our study provides the first identifiability results for the general nonparametric setting with unknown interventions, and elucidates what is possible and impossible for causal representation learning without more direct supervision.
****

Dual Mean-Teacher: An Unbiased Semi-Supervised Framework for Audio-Visual Source Localization

Yuxin Guo, Shijie Ma, Hu Su, Zhiqing Wang, Yuhao Zhao, Wei Zou, Siyang Sun, Yun

Zheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Hierarchical Gaussian Mixture based Task Generative Model for Robust Meta-Learning
Yizhou Zhang, Jingchao Ni, Wei Cheng, Zhengzhang Chen, Liang Tong, Haifeng Chen, Yan Liu
Meta-learning enables quick adaptation of machine learning models to new tasks with limited data. While tasks could come from varying distributions in reality, most of the existing meta-learning methods consider both training and testing tasks as from the same uni-component distribution, overlooking two critical needs of a practical solution: (1) the various sources of tasks may compose a multi-component mixture distribution, and (2) novel tasks may come from a distribution that is unseen during meta-training. In this paper, we demonstrate these two challenges can be solved jointly by modeling the density of task instances. We develop a meta-training framework underlain by a novel Hierarchical Gaussian Mixture based Task Generative Model (HTGM). HTGM extends the widely used empirical process of sampling tasks to a theoretical model, which learns task embeddings, fits the mixture distribution of tasks, and enables density-based scoring of novel tasks. The framework is agnostic to the encoder and scales well with large backbone networks. The model parameters are learned end-to-end by maximum likelihood estimation via an Expectation-Maximization (EM) algorithm. Extensive experiments on benchmark datasets indicate the effectiveness of our method for both sample classification and novel task detection.
****
Temporal Dynamic Quantization for Diffusion Models
Junhyuk So, Jungwon Lee, Daehyun Ahn, Hyungjun Kim, Eunhyeok Park
Diffusion model has gained popularity in vision applications due to its remarkable generative performance and versatility. However, its high storage and computation demands, resulting from the model size and iterative generation, hinder its use on mobile devices. Existing quantization techniques struggle to maintain performance even in 8-bit precision due to the diffusion model's unique property of temporal variation in activation. We introduce a novel quantization method that dynamically adjusts the quantization interval based on time step information, significantly improving output quality. Unlike conventional dynamic quantization techniques, our approach has no computational overhead during inference and is compatible with both post-training quantization (PTQ) and quantization-aware training (QAT). Our extensive experiments demonstrate substantial improvements in output quality with the quantized model across various configurations.
****
Learning Interpretable Low-dimensional Representation via Physical Symmetry
Xuanjie Liu, Daniel Chin, Yichen Huang, Gus Xia
We have recently seen great progress in learning interpretable music representations, ranging from basic factors, such as pitch and timbre, to high-level concepts, such as chord and texture. However, most methods rely heavily on music domain knowledge. It remains an open question what general computational principles give rise to interpretable representations, especially low-dim factors that agree with human perception. In this study, we take inspiration from modern physics and use physical symmetry as a self-consistency constraint for the latent space. Specifically, it requires the prior model that characterises the dynamics of the latent states to be equivariant with respect to certain group transformations. We show that physical symmetry leads the model to learn a linear pitch factor from unlabelled monophonic music audio in a self-supervised fashion. In addition, the same methodology can be applied to computer vision, learning a 3D Cartesian space from videos of a simple moving object without labels. Furthermore, physical symmetry naturally leads to counterfactual representation augmentation, a new technique which improves sample efficiency.

****
ImageBrush: Learning Visual In-Context Instructions for Exemplar-Based Image Manipulation
ya■sheng sun, Yifan Yang, Houwen Peng, Yifei Shen, Yuqing Yang, Han Hu, Lili Qiu, Hideki Koike
While language-guided image manipulation has made remarkable progress, the challenge of how to instruct the manipulation process faithfully reflecting human intentions persists. An accurate and comprehensive description of a manipulation task using natural language is laborious and sometimes even impossible, primarily due to the inherent uncertainty and ambiguity present in linguistic expressions. Is it feasible to accomplish image manipulation without resorting to external cross-modal language information? If this possibility exists, the inherent modality gap would be effortlessly eliminated. In this paper, we propose a novel manipulation methodology, dubbed ImageBrush, that learns visual instructions for more accurate image editing.Our key idea is to employ a pair of transformation images as visual instructions, which not only precisely captures human intention but also facilitates accessibility in real-world scenarios. Capturing visual instructions is particularly challenging because it involves extracting the underlying intentions solely from visual demonstrations and then applying this operation to a new image. To address this challenge, we formulate visual instruction learning as a diffusion-based inpainting problem, where the contextual information is fully exploited through an iterative process of generation. A visual prompting encoder is carefully devised to enhance the model's capacity in uncovering human intent behind the visual instructions. Extensive experiments show that our method generates engaging manipulation results conforming to the transformations entailed in demonstrations. Moreover, our model exhibits robust generalization capabilities on various downstream tasks such as pose transfer, image translation and video inpainting.
****
Meek Separators and Their Applications in Targeted Causal Discovery
Kirankumar Shiragur, Jiaqi Zhang, Caroline Uhler
Learning causal structures from interventional data is a fundamental problem with broad applications across various fields. While many previous works have focused on recovering the entire causal graph, in practice, there are scenarios where learning only part of the causal graph suffices. This is called \emph{targeted} causal discovery. In our work, we focus on two such well-motivated problems: subset search and causal matching. We aim to minimize the number of interventions in both cases.Towards this, we introduce the \emph{Meek separator}, which is a subset of vertices that, when intervened, decomposes the remaining unoriented edges into smaller connected components. We then present an efficient algorithm to find Meek separators that are of small sizes. Such a procedure is helpful in designing various divide-and-conquer-based approaches. In particular, we propose two randomized algorithms that achieve logarithmic approximation for subset search and causal matching, respectively. Our results provide the first known average-case provable guarantees for both problems. We believe that this opens up possibilities to design near-optimal methods for many other targeted causal structure learning problems arising from various applications.
****
CLeAR: Continual Learning on Algorithmic Reasoning for Human-like Intelligence
Bong Gyun Kang, HyunGi Kim, Dahuin Jung, Sungroh Yoon
Continual learning (CL) aims to incrementally learn multiple tasks that are presented sequentially. The significance of CL lies not only in the practical importance but also in studying the learning mechanisms of humans who are excellent continual learners. While most research on CL has been done on structured data such as images, there is a lack of research on CL for abstract logical concepts such as counting, sorting, and arithmetic, which humans learn gradually over time in the real world. In this work, for the first time, we introduce novel algorithmic reasoning (AR) methodology for continual tasks of abstract concepts: CLeAR. Our methodology proposes a one-to-many mapping of input distribution to a shared mapping space, which allows the alignment of various tasks of different dimensio

ns and shared semantics. Our tasks of abstract logical concepts, in the form of formal language, can be classified into Chomsky hierarchies based on their difficulty. In this study, we conducted extensive experiments consisting of 15 tasks with various levels of Chomsky hierarchy, ranging from in-hierarchy to inter-hierarchy scenarios. CLeAR not only achieved near zero forgetting but also improved accuracy during following tasks, a phenomenon known as backward transfer, while previous CL methods designed for image classification drastically failed.
****
SLIBO-Net: Floorplan Reconstruction via Slicing Box Representation with Local Geometry Regularization
Jheng-Wei Su, Kuei-Yu Tung, Chi-Han Peng, Peter Wonka, Hung-Kuo (James) Chu
This paper focuses on improving the reconstruction of 2D floorplans from unstructured 3D point clouds. We identify opportunities for enhancement over the existing methods in three main areas: semantic quality, efficient representation, and local geometric details. To address these, we presents SLIBO-Net, an innovative approach to reconstructing 2D floorplans from unstructured 3D point clouds. We propose a novel transformer-based architecture that employs an efficient floorplan representation, providing improved room shape supervision and allowing for manageable token numbers. By incorporating geometric priors as a regularization mechanism and post-processing step, we enhance the capture of local geometric details. We also propose a scale-independent evaluation metric, correcting the discrepancy in error treatment between varying floorplan sizes. Our approach notably achieves a new state-of-the-art on the Structure3D dataset. The resultant floorplans exhibit enhanced semantic plausibility, substantially improving the overall quality and realism of the reconstructions. Our code and dataset are available online.
****
Fantastic Robustness Measures: The Secrets of Robust Generalization
Hoki Kim, Jinseong Park, Yujin Choi, Jaewook Lee
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Spectral Algorithm for List-Decodable Covariance Estimation in Relative Frobenius Norm
Ilias Diakonikolas, Daniel Kane, Jasper Lee, Ankit Pensia, Thanasis Pittas
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Diff-Foley: Synchronized Video-to-Audio Synthesis with Latent Diffusion Models
Simian Luo, Chuanhao Yan, Chenxu Hu, Hang Zhao
The Video-to-Audio (V2A) model has recently gained attention for its practical application in generating audio directly from silent videos, particularly in video/film production. However, previous methods in V2A have limited generation quality in terms of temporal synchronization and audio-visual relevance. We present Diff-Foley, a synchronized Video-to-Audio synthesis method with a latent diffusion model (LDM) that generates high-quality audio with improved synchronization and audio-visual relevance. We adopt contrastive audio-visual pretraining (CAVP) to learn more temporally and semantically aligned features, then train an LDM with CAVP-aligned visual features on spectrogram latent space. The CAVP-aligned features enable LDM to capture the subtler audio-visual correlation via a cross-attention module. We further significantly improve sample quality with `double guidance'. Diff-Foley achieves state-of-the-art V2A performance on current large scale V2A dataset. Furthermore, we demonstrate Diff-Foley practical applicability and adaptability via customized downstream finetuning. Project Page: https://diff-foley.github.io/
****

The Drunkard’s Odometry: Estimating Camera Motion in Deforming Scenes
David Recasens Lafuente, Martin R. Oswald, Marc Pollefeys, Javier Civera
Estimating camera motion in deformable scenes poses a complex and open research challenge. Most existing non-rigid structure from motion techniques assume to observe also static scene parts besides deforming scene parts in order to establish an anchoring reference. However, this assumption does not hold true in certain relevant application cases such as endoscopies. Deformable odometry and SLAM pipelines, which tackle the most challenging scenario of exploratory trajectories, suffer from a lack of robustness and proper quantitative evaluation methodologies. To tackle this issue with a common benchmark, we introduce the Drunkard's Dataset, a challenging collection of synthetic data targeting visual navigation and reconstruction in deformable environments. This dataset is the first large set of exploratory camera trajectories with ground truth inside 3D scenes where every surface exhibits non-rigid deformations over time. Simulations in realistic 3D buildings lets us obtain a vast amount of data and ground truth labels, including camera poses, RGB images and depth, optical flow and normal maps at high resolution and quality. We further present a novel deformable odometry method, dubbed the Drunkard’s Odometry, which decomposes optical flow estimates into rigid-body camera motion and non-rigid scene deformations. In order to validate our data, our work contains an evaluation of several baselines as well as a novel tracking error metric which does not require ground truth data. Dataset and code: https://davidrecasens.github.io/TheDrunkard'sOdometry/
****
Optimal Treatment Allocation for Efficient Policy Evaluation in Sequential Decision Making
Ting Li, Chengchun Shi, Jianing Wang, Fan Zhou, hongtu zhu
A/B testing is critical for modern technological companies to evaluate the effectiveness of newly developed products against standard baselines. This paper studies optimal designs that aim to maximize the amount of information obtained from online experiments to estimate treatment effects accurately. We propose three optimal allocation strategies in a dynamic setting where treatments are sequentially assigned over time. These strategies are designed to minimize the variance of the treatment effect estimator when data follow a non Markov decision process or a (time-varying) Markov decision process. We further develop estimation procedures based on existing off-policy evaluation (OPE) methods and conduct extensive experiments in various environments to demonstrate the effectiveness of the proposed methodologies. In theory, we prove the optimality of the proposed treatment allocation design and establish upper bounds for the mean squared errors of the resulting treatment effect estimators.
****
Advancing Bayesian Optimization via Learning Correlated Latent Space
Seunghun Lee, Jaewon Chu, Sihyeon Kim, Juyeon Ko, Hyunwoo J. Kim
Bayesian optimization is a powerful method for optimizing black-box functions with limited function evaluations. Recent works have shown that optimization in a latent space through deep generative models such as variational autoencoders leads to effective and efficient Bayesian optimization for structured or discrete data. However, as the optimization does not take place in the input space, it leads to an inherent gap that results in potentially suboptimal solutions. To alleviate the discrepancy, we propose Correlated latent space Bayesian Optimization (CoBO), which focuses on learning correlated latent spaces characterized by a strong correlation between the distances in the latent space and the distances within the objective function. Specifically, our method introduces Lipschitz regularization, loss weighting, and trust region recoordination to minimize the inherent gap around the promising areas. We demonstrate the effectiveness of our approach on several optimization tasks in discrete data, such as molecule design and arithmetic expression fitting, and achieve high performance within a small budget.
****
Generalization bounds for neural ordinary differential equations and deep residual networks

Pierre Marion
Neural ordinary differential equations (neural ODEs) are a popular family of continuous-depth deep learning models. In this work, we consider a large family of parameterized ODEs with continuous-in-time parameters, which include time-dependent neural ODEs. We derive a generalization bound for this class by a Lipschitz-based argument. By leveraging the analogy between neural ODEs and deep residual networks, our approach yields in particular a generalization bound for a class of deep residual networks. The bound involves the magnitude of the difference between successive weight matrices. We illustrate numerically how this quantity affects the generalization capability of neural networks.
****
Global Update Tracking: A Decentralized Learning Algorithm for Heterogeneous Data
Sai Aparna Aketi, Abolfazl Hashemi, Kaushik Roy
Decentralized learning enables the training of deep learning models over large distributed datasets generated at different locations, without the need for a central server. However, in practical scenarios, the data distribution across these devices can be significantly different, leading to a degradation in model performance. In this paper, we focus on designing a decentralized learning algorithm that is less susceptible to variations in data distribution across devices. We propose Global Update Tracking (GUT), a novel tracking-based method that aims to mitigate the impact of heterogeneous data in decentralized learning without introducing any communication overhead. We demonstrate the effectiveness of the proposed technique through an exhaustive set of experiments on various Computer Vision datasets (CIFAR-10, CIFAR-100, Fashion MNIST, and ImageNette), model architectures, and network topologies. Our experiments show that the proposed method achieves state-of-the-art performance for decentralized learning on heterogeneous data via a 1-6% improvement in test accuracy compared to other existing techniques.
****
QuadAttac$K$: A Quadratic Programming Approach to Learning Ordered Top-K Adversarial Attacks
Thomas Paniagua, Ryan Grainger, Tianfu Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Predicting mutational effects on protein-protein binding via a side-chain diffusion probabilistic model
Shiwei Liu, Tian Zhu, Milong Ren, Chungong Yu, Dongbo Bu, Haicang Zhang
Many crucial biological processes rely on networks of protein-protein interactions. Predicting the effect of amino acid mutations on protein-protein binding is important in protein engineering, including therapeutic discovery. However, the scarcity of annotated experimental data on binding energy poses a significant challenge for developing computational approaches, particularly deep learning-based methods. In this work, we propose SidechainDiff, a novel representation learning-based approach that leverages unlabelled experimental protein structures. SidechainDiff utilizes a Riemannian diffusion model to learn the generative process of side-chain conformations and can also give the structural context representations of mutations on the protein-protein interface. Leveraging the learned representations, we achieve state-of-the-art performance in predicting the mutational effects on protein-protein binding. Furthermore, SidechainDiff is the first diffusion-based generative model for side-chains, distinguishing it from prior efforts that have predominantly focused on the generation of protein backbone structures.
****
PETAL: Physics Emulation Through Averaged Linearizations for Solving Inverse Problems
Jihui Jin, Etienne Ollivier, Richard Touret, Matthew McKinley, Karim Sabra, Just

in Romberg
Inverse problems describe the task of recovering an underlying signal of interes t given observables. Typically, the observables are related via some non-linear forward model applied to the underlying unknown signal. Inverting the non-linear forward model can be computationally expensive, as it often involves computing and inverting a linearization at a series of estimates. Rather than inverting th e physics-based model, we instead train a surrogate forward model (emulator) and leverage modern auto-grad libraries to solve for the input within a classical o ptimization framework. Current methods to train emulators are done in a black bo x supervised machine learning fashion and fail to take advantage of any existing knowledge of the forward model. In this article, we propose a simple learned we ighted average model that embeds linearizations of the forward model around vari ous reference points into the model itself, explicitly incorporating known physi cs. Grounding the learned model with physics based linearizations improves the f orward modeling accuracy and provides richer physics based gradient information during the inversion process leading to more accurate signal recovery. We demons trate the efficacy on an ocean acoustic tomography (OAT) example that aims to re cover ocean sound speed profile (SSP) variations from acoustic observations (e.g . eigenray arrival times) within simulation of ocean dynamics in the Gulf of Mex ico.
****
RealTime QA: What's the Answer Right Now?
Jungo Kasai, Keisuke Sakaguchi, yoichi takahashi, Ronan Le Bras, Akari Asai, Xin yan Yu, Dragomir Radev, Noah A. Smith, Yejin Choi, Kentaro Inui
We introduce RealTime QA, a dynamic question answering (QA) platform that announ ces questions and evaluates systems on a regular basis (weekly in this version). RealTime QA inquires about the current world, and QA systems need to answer que stions about novel events or information. It therefore challenges static, conven tional assumptions in open-domain QA datasets and pursues instantaneous applicat ions. We build strong baseline models upon large pretrained language models, inc luding GPT-3 and T5. Our benchmark is an ongoing effort, and this paper presents real-time evaluation results over the past year. Our experimental results show that GPT-3 can often properly update its generation results, based on newly-retr ieved documents, highlighting the importance of up-to-date information retrieval . Nonetheless, we find that GPT-3 tends to return outdated answers when retrieve d documents do not provide sufficient information to find an answer. This sugges ts an important avenue for future research: can an open-domain QA system identif y such unanswerable cases and communicate with the user or even the retrieval mo dule to modify the retrieval results? We hope that RealTime QA will spur progres s in instantaneous applications of question answering and beyond.
****
Learning Transformer Programs
Dan Friedman, Alexander Wettig, Danqi Chen
Recent research in mechanistic interpretability has attempted to reverse-enginee r Transformer models by carefully inspecting network weights and activations. Ho wever, these approaches require considerable manual effort and still fall short of providing complete, faithful descriptions of the underlying algorithms. In th is work, we introduce a procedure for training Transformers that are mechanistic ally interpretable by design. We build on RASP [Weiss et al., 2021], a programmi ng language that can be compiled into Transformer weights. Instead of compiling human-written programs into Transformers, we design a modified Transformer that can be trained using gradient-based optimization and then automatically converte d into a discrete, human-readable program. We refer to these models as Transform er Programs. To validate our approach, we learn Transformer Programs for a varie ty of problems, including an in-context learning task, a suite of algorithmic pr oblems (e.g. sorting, recognizing Dyck languages), and NLP tasks including named entity recognition and text classification. The Transformer Programs can automa tically find reasonable solutions, performing on par with standard Transformers of comparable size; and, more importantly, they are easy to interpret. To demons trate these advantages, we convert Transformers into Python programs and use off

-the-shelf code analysis tools to debug model errors and identify the “circuits” used to solve different sub-problems. We hope that Transformer Programs open a new path toward the goal of intrinsically interpretable machine learning.
****

An Inverse Scaling Law for CLIP Training
Xianhang Li, Zeyu Wang, Cihang Xie
CLIP, one of the pioneering foundation models that connect images and text, has enabled many recent breakthroughs in computer vision. However, its associated training cost is prohibitively high, imposing a significant barrier to its widespread exploration. In this paper, we present a surprising finding that there exists an inverse scaling law for CLIP training, whereby the larger the image/text encoders used, the shorter the sequence length of image/text tokens that can be applied in training. Moreover, we showcase that the strategy for reducing image/text token length plays a crucial role in determining the quality of this scaling law.As a result of this finding, we are able to successfully train CLIP even with limited computational resources. For example, using 8 A100 GPUs, our CLIP models achieve zero-shot top-1 ImageNet-1k accuracies of 63.2% in ~2 days, 67.8% in ~3 days, and 69.3% in ~4 days. Our method also works well when scaling up --- with G/14, we register a new record of 83.0% ImageNet-1k zero-shot accuracy, and meanwhile accelerate the training by ~33x compared to its OpenCLIP counterpart.By reducing the computation barrier associated with CLIP, we hope to inspire more research in this field, particularly from academics. Our code is available at https://github.com/UCSC-VLAA/CLIPA.
****

Sequential Preference Ranking for Efficient Reinforcement Learning from Human Feedback
Minyoung Hwang, Gunmin Lee, Hogun Kee, Chan Woo Kim, Kyungjae Lee, Songhwai Oh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Diffusion-SS3D: Diffusion Model for Semi-supervised 3D Object Detection
Cheng-Ju Ho, Chen-Hsuan Tai, Yen-Yu Lin, Ming-Hsuan Yang, Yi-Hsuan Tsai
Semi-supervised object detection is crucial for 3D scene understanding, efficiently addressing the limitation of acquiring large-scale 3D bounding box annotations. Existing methods typically employ a teacher-student framework with pseudo-labeling to leverage unlabeled point clouds. However, producing reliable pseudo-labels in a diverse 3D space still remains challenging. In this work, we propose Diffusion-SS3D, a new perspective of enhancing the quality of pseudo-labels via the diffusion model for semi-supervised 3D object detection. Specifically, we include noises to produce corrupted 3D object size and class label distributions, and then utilize the diffusion model as a denoising process to obtain bounding box outputs. Moreover, we integrate the diffusion model into the teacher-student framework, so that the denoised bounding boxes can be used to improve pseudo-label generation, as well as the entire semi-supervised learning process. We conduct experiments on the ScanNet and SUN RGB-D benchmark datasets to demonstrate that our approach achieves state-of-the-art performance against existing methods. We also present extensive analysis to understand how our diffusion model design affects performance in semi-supervised learning. The source code will be available at https://github.com/luluho1208/Diffusion-SS3D.
****

Aligning Language Models with Human Preferences via a Bayesian Approach
Jiashuo WANG, Haozhao Wang, Shichao Sun, Wenjie Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

A Smooth Binary Mechanism for Efficient Private Continual Observation

Joel Daniel Andersson, Rasmus Pagh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Training Transformers with 4-bit Integers
Haocheng Xi, ChangHao Li, Jianfei Chen, Jun Zhu
Quantizing the activation, weight, and gradient to 4-bit is promising to accelerate neural network training. However, existing 4-bit training methods require custom numerical formats which are not supported by contemporary hardware. In this work, we propose a training method for transformers with all matrix multiplications implemented with the INT4 arithmetic. Training with an ultra-low INT4 precision is challenging. To achieve this, we carefully analyze the specific structures of activation and gradients in transformers to propose dedicated quantizers for them. For forward propagation, we identify the challenge of outliers and propose a Hadamard quantizer to suppress the outliers. For backpropagation, we leverage the structural sparsity of gradients by proposing bit splitting and leverage score sampling techniques to quantize gradients accurately. Our algorithm achieves competitive accuracy on a wide range of tasks including natural language understanding, machine translation, and image classification. Unlike previous 4-bit training methods, our algorithm can be implemented on the current generation of GPUs. Our prototypical linear operator implementation is up to 2.2 times faster than the FP16 counterparts and speeds up the training by 17.8\% on average for sufficiently large models. Our code is available at https://github.com/xijiu9/Train_Transformers_with_INT4.
****
TD Convergence: An Optimization Perspective
Kavosh Asadi, Shoham Sabach, Yao Liu, Omer Gottesman, Rasool Fakoor
We study the convergence behavior of the celebrated temporal-difference (TD) learning algorithm. By looking at the algorithm through the lens of optimization, we first argue that TD can be viewed as an iterative optimization algorithm where the function to be minimized changes per iteration. By carefully investigating the divergence displayed by TD on a classical counter example, we identify two forces that determine the convergent or divergent behavior of the algorithm. We next formalize our discovery in the linear TD setting with quadratic loss and prove that convergence of TD hinges on the interplay between these two forces. We extend this optimization perspective to prove convergence of TD in a much broader setting than just linear approximation and squared loss. Our results provide a theoretical explanation for the successful application of TD in reinforcement learning.
****
Time Series as Images: Vision Transformer for Irregularly Sampled Time Series
Zekun Li, Shiyang Li, Xifeng Yan
Irregularly sampled time series are increasingly prevalent, particularly in medical domains. While various specialized methods have been developed to handle these irregularities, effectively modeling their complex dynamics and pronounced sparsity remains a challenge. This paper introduces a novel perspective by converting irregularly sampled time series into line graph images, then utilizing powerful pre-trained vision transformers for time series classification in the same way as image classification. This method not only largely simplifies specialized algorithm designs but also presents the potential to serve as a universal framework for time series modeling. Remarkably, despite its simplicity, our approach outperforms state-of-the-art specialized algorithms on several popular healthcare and human activity datasets. Especially in the rigorous leave-sensors-out setting where a portion of variables is omitted during testing, our method exhibits strong robustness against varying degrees of missing observations, achieving an impressive improvement of 42.8% in absolute F1 score points over leading specialized baselines even with half the variables masked. Code and data are available at https://github.com/Leezekun/ViTST.

****
Symbolic Discovery of Optimization Algorithms
Xiangning Chen, Chen Liang, Da Huang, Esteban Real, Kaiyuan Wang, Hieu Pham, Xuanyi Dong, Thang Luong, Cho-Jui Hsieh, Yifeng Lu, Quoc V Le
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Calibrating Diffusion Probabilistic Models
Tianyu Pang, Cheng Lu, Chao Du, Min Lin, Shuicheng Yan, Zhijie Deng
Recently, diffusion probabilistic models (DPMs) have achieved promising results in diverse generative tasks. A typical DPM framework includes a forward process that gradually diffuses the data distribution and a reverse process that recovers the data distribution from time-dependent data scores. In this work, we observe that the stochastic reverse process of data scores is a martingale, from which concentration bounds and the optional stopping theorem for data scores can be derived. Then, we discover a simple way for calibrating an arbitrary pretrained DPM, with which the score matching loss can be reduced and the lower bounds of model likelihood can consequently be increased. We provide general calibration guidelines under various model parametrizations. Our calibration method is performed only once and the resulting models can be used repeatedly for sampling. We conduct experiments on multiple datasets to empirically validate our proposal. Our code is available at https://github.com/thudzj/Calibrated-DPMs.
****
InstructBLIP: Towards General-purpose Vision-Language Models with Instruction Tuning
Wenliang Dai, Junnan Li, DONGXU LI, Anthony Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale N Fung, Steven Hoi
Large-scale pre-training and instruction tuning have been successful at creating general-purpose language models with broad competence. However, building general-purpose vision-language models is challenging due to the rich input distributions and task diversity resulting from the additional visual input. Although vision-language pretraining has been widely studied, vision-language instruction tuning remains under-explored. In this paper, we conduct a systematic and comprehensive study on vision-language instruction tuning based on the pretrained BLIP-2 models. We gather 26 publicly available datasets, covering a wide variety of tasks and capabilities, and transform them into instruction tuning format. Additionally, we introduce an instruction-aware Query Transformer, which extracts informative features tailored to the given instruction. Trained on 13 held-in datasets, InstructBLIP attains state-of-the-art zero-shot performance across all 13 held-out datasets, substantially outperforming BLIP-2 and larger Flamingo models. Our models also lead to state-of-the-art performance when finetuned on individual downstream tasks (e.g., 90.7% accuracy on ScienceQA questions with image contexts). Furthermore, we qualitatively demonstrate the advantages of InstructBLIP over concurrent multimodal models. All InstructBLIP models are open-source.
****
Privacy Auditing with One (1) Training Run
Thomas Steinke, Milad Nasr, Matthew Jagielski
We propose a scheme for auditing differentially private machine learning systems with a single training run. This exploits the parallelism of being able to add or remove multiple training examples independently. We analyze this using the connection between differential privacy and statistical generalization, which avoids the cost of group privacy. Our auditing scheme requires minimal assumptions about the algorithm and can be applied in the black-box or white-box setting. We demonstrate the effectiveness of our framework by applying it to DP-SGD, where we can achieve meaningful empirical privacy lower bounds by training only one model. In contrast, standard methods would require training hundreds of models.
****
Kernel Stein Discrepancy thinning: a theoretical perspective of pathologies and

a practical fix with regularization
Clement Benard, Brian Staber, Sébastien Da Veiga
Stein thinning is a promising algorithm proposed by (Riabiz et al., 2022) for post-processing outputs of Markov chain Monte Carlo (MCMC). The main principle is to greedily minimize the kernelized Stein discrepancy (KSD), which only requires the gradient of the log-target distribution, and is thus well-suited for Bayesian inference. The main advantages of Stein thinning are the automatic remove of the burn-in period, the correction of the bias introduced by recent MCMC algorithms, and the asymptotic properties of convergence towards the target distribution. Nevertheless, Stein thinning suffers from several empirical pathologies, which may result in poor approximations, as observed in the literature. In this article, we conduct a theoretical analysis of these pathologies, to clearly identify the mechanisms at stake, and suggest improved strategies. Then, we introduce the regularized Stein thinning algorithm to alleviate the identified pathologies. Finally, theoretical guarantees and extensive experiments show the high efficiency of the proposed algorithm. An implementation of regularized Stein thinning as the kernax library in python and JAX is available at https://gitlab.com/drti/kernax.
****
Punctuation-level Attack: Single-shot and Single Punctuation Can Fool Text Models
wenqiang wang, Chongyang Du, Tao Wang, Kaihao Zhang, Wenhan Luo, Lin Ma, Wei Liu, Xiaochun Cao
The adversarial attacks have attracted increasing attention in various fields including natural language processing. The current textual attacking models primarily focus on fooling models by adding character-/word-/sentence-level perturbations, ignoring their influence on human perception. In this paper, for the first time in the community, we propose a novel mode of textual attack, punctuation-level attack. With various types of perturbations, including insertion, displacement, deletion, and replacement, the punctuation-level attack achieves promising fooling rates against SOTA models on typical textual tasks and maintains minimal influence on human perception and understanding of the text by mere perturbation of single-shot single punctuation. Furthermore, we propose a search method named Text Position Punctuation Embedding and Paraphrase (TPPEP) to accelerate the pursuit of optimal position to deploy the attack, without exhaustive search, and we present a mathematical interpretation of TPPEP. Thanks to the integrated Text Position Punctuation Embedding (TPPE), the punctuation attack can be applied at a constant cost of time. Experimental results on public datasets and SOTA models demonstrate the effectiveness of the punctuation attack and the proposed TPPE. We additionally apply the single punctuation attack to summarization, semantic-similarity-scoring, and text-to-image tasks, and achieve encouraging results.
****
Towards Hybrid-grained Feature Interaction Selection for Deep Sparse Network
Fuyuan Lyu, Xing Tang, Dugang Liu, Chen Ma, Weihong Luo, Liang Chen, xiuqiang He, Xue (Steve) Liu
Deep sparse networks are widely investigated as a neural network architecture for prediction tasks with high-dimensional sparse features, with which feature interaction selection is a critical component. While previous methods primarily focus on how to search feature interaction in a coarse-grained space, less attention has been given to a finer granularity. In this work, we introduce a hybrid-grained feature interaction selection approach that targets both feature field and feature value for deep sparse networks. To explore such expansive space, we propose a decomposed space which is calculated on the fly. We then develop a selection algorithm called OptFeature, which efficiently selects the feature interaction from both the feature field and the feature value simultaneously. Results from experiments on three large real-world benchmark datasets demonstrate that OptFeature performs well in terms of accuracy and efficiency. Additional studies support the feasibility of our method. All source code are publicly available\footnote{https://anonymous.4open.science/r/OptFeature-Anonymous}.
****

On the Asymptotic Learning Curves of Kernel Ridge Regression under Power-law Dec
ay
Yicheng Li, haobo Zhang, Qian Lin
The widely observed 'benign overfitting phenomenon' in the neural network litera
ture raises the challenge to the `bias-variance trade-off' doctrine in the stati
stical learning theory.Since the generalization ability of the 'lazy trained' ov
er-parametrized neural network can be well approximated by that of the neural ta
ngent kernel regression,the curve of the excess risk (namely, the learning curve
) of kernel ridge regression attracts increasing attention recently.However, mos
t recent arguments on the learning curve are heuristic and are based on the 'Gau
ssian design' assumption.In this paper, under mild and more realistic assumption
s, we rigorously provide a full characterization of the learning curve in the as
ymptotic senseunder a power-law decay condition of the eigenvalues of the kernel
 and also the target function.The learning curve elaborates the effect and the i
nterplay of the choice of the regularization parameter, the source condition and
 the noise.In particular, our results suggest that the 'benign overfitting pheno
menon' exists in over-parametrized neural networks only when the noise level is
small.
****
Mechanism Design for Collaborative Normal Mean Estimation
Yiding Chen, Jerry Zhu, Kirthevasan Kandasamy
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
DiffKendall: A Novel Approach for Few-Shot Learning with Differentiable Kendall'
s Rank Correlation
Kaipeng Zheng, Huishuai Zhang, Weiran Huang
Few-shot learning aims to adapt models trained on the base dataset to novel task
s where the categories were not seen by the model before. This often leads to a
relatively concentrated distribution of feature values across channels on novel
classes, posing challenges in determining channel importance for novel tasks. St
andard few-shot learning methods employ geometric similarity metrics such as cos
ine similarity and negative Euclidean distance to gauge the semantic relatedness
 between two features. However, features with high geometric similarities may ca
rry distinct semantics, especially in the context of few-shot learning. In this
paper, we demonstrate that the importance ranking of feature channels is a more
reliable indicator for few-shot learning than geometric similarity metrics. We o
bserve that replacing the geometric similarity metric with Kendall’s rank correl
ation only during inference is able to improve the performance of few-shot learn
ing across a wide range of methods and datasets with different domains. Furtherm
ore, we propose a carefully designed differentiable loss for meta-training to ad
dress the non-differentiability issue of Kendall’s rank correlation. By replacin
g geometric similarity with differentiable Kendall’s rank correlation, our metho
d can integrate with numerous existing few-shot approaches and is ready for inte
grating with future state-of-the-art methods that rely on geometric similarity m
etrics. Extensive experiments validate the efficacy of the rank-correlation-base
d approach, showcasing a significant improvement in few-shot learning.
****
High-dimensional Contextual Bandit Problem without Sparsity
Junpei Komiyama, Masaaki Imaizumi
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
VidChapters-7M: Video Chapters at Scale
Antoine Yang, Arsha Nagrani, Ivan Laptev, Josef Sivic, Cordelia Schmid
Segmenting untrimmed videos into chapters enables users to quickly navigate to t

he information of their interest. This important topic has been understudied due to the lack of publicly released datasets. To address this issue, we present VidChapters-7M, a dataset of 817K user-chaptered videos including 7M chapters in total. VidChapters-7M is automatically created from videos online in a scalable manner by scraping user-annotated chapters and hence without any additional manual annotation. We introduce the following three tasks based on this data. First, the video chapter generation task consists of temporally segmenting the video and generating a chapter title for each segment. To further dissect the problem, we also define two variants of this task: video chapter generation given ground-truth boundaries, which requires generating a chapter title given an annotated video segment, and video chapter grounding, which requires temporally localizing a chapter given its annotated title. We benchmark both simple baselines as well as state-of-the-art video-language models on these three tasks. We also show that pretraining on VidChapters-7M transfers well to dense video captioning tasks, largely improving the state of the art on the YouCook2 and ViTT benchmarks. Finally, our experiments reveal that downstream performance scales well with the size of the pretraining dataset.
****
Energy-Based Models for Anomaly Detection: A Manifold Diffusion Recovery Approach
Sangwoong Yoon, Young-Uk Jin, Yung-Kyun Noh, Frank Park
We present a new method of training energy-based models (EBMs) for anomaly detection that leverages low-dimensional structures within data. The proposed algorithm, Manifold Projection-Diffusion Recovery (MPDR), first perturbs a data point along a low-dimensional manifold that approximates the training dataset. Then, EBM is trained to maximize the probability of recovering the original data. The training involves the generation of negative samples via MCMC, as in conventional EBM training, but from a different distribution concentrated near the manifold. The resulting near-manifold negative samples are highly informative, reflecting relevant modes of variation in data. An energy function of MPDR effectively learns accurate boundaries of the training data distribution and excels at detecting out-of-distribution samples. Experimental results show that MPDR exhibits strong performance across various anomaly detection tasks involving diverse data types, such as images, vectors, and acoustic signals.
****
Characterizing the Optimal $0-1$ Loss for Multi-class Classification with a Test-time Attacker
Sihui Dai, Wenxin Ding, Arjun Nitin Bhagoji, Daniel Cullina, Heather Zheng, Ben Zhao, Prateek Mittal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Truncated Affinity Maximization: One-class Homophily Modeling for Graph Anomaly Detection
Hezhe Qiao, Guansong Pang
We reveal a one-class homophily phenomenon, which is one prevalent property we find empirically in real-world graph anomaly detection (GAD) datasets, i.e., normal nodes tend to have strong connection/affinity with each other, while the homophily in abnormal nodes is significantly weaker than normal nodes. However, this anomaly-discriminative property is ignored by existing GAD methods that are typically built using a conventional anomaly detection objective, such as data reconstruction.In this work, we explore this property to introduce a novel unsupervised anomaly scoring measure for GAD -- local node affinity-- that assigns a larger anomaly score to nodes that are less affiliated with their neighbors, with the affinity defined as similarity on node attributes/representations. We further propose Truncated Affinity Maximization (TAM) that learns tailored node representations for our anomaly measure by maximizing the local affinity of nodes to their neighbors. Optimizing on the original graph structure can be biased by non-h

omophily edges(i.e., edges connecting normal and abnormal nodes). Thus, TAM is instead optimized on truncated graphs where non-homophily edges are removed iteratively to mitigate this bias. The learned representations result in significantly stronger local affinity for normal nodes than abnormal nodes. Extensive empirical results on 10 real-world GAD datasets show that TAM substantially outperforms seven competing models, achieving over 10% increase in AUROC/AUPRC compared to the best contenders on challenging datasets. Our code is available at https://github.com/mala-lab/TAM-master/.
****
Sample-Conditioned Hypothesis Stability Sharpens Information-Theoretic Generalization Bounds
Ziqiao Wang, Yongyi Mao
We present new information-theoretic generalization guarantees through the a novel construction of the "neighboring-hypothesis" matrix and a new family of stability notions termed sample-conditioned hypothesis (SCH) stability. Our approach yields sharper bounds that improve upon previous information-theoretic bounds in various learning scenarios. Notably, these bounds address the limitations of existing information-theoretic bounds in the context of stochastic convex optimization (SCO) problems, as explored in the recent work by Haghifam et al. (2023).
****
Exploiting Contextual Objects and Relations for 3D Visual Grounding
Li Yang, chunfeng yuan, Ziqi Zhang, Zhongang Qi, Yan Xu, Wei Liu, Ying Shan, Bing Li, Weiping Yang, Peng Li, Yan Wang, Weiming Hu
3D visual grounding, the task of identifying visual objects in 3D scenes based on natural language inputs, plays a critical role in enabling machines to understand and engage with the real-world environment. However, this task is challenging due to the necessity to capture 3D contextual information to distinguish target objects from complex 3D scenes. The absence of annotations for contextual objects and relations further exacerbates the difficulties. In this paper, we propose a novel model, CORE-3DVG, to address these challenges by explicitly learning about contextual objects and relations. Our method accomplishes 3D visual grounding via three sequential modular networks, including a text-guided object detection network, a relation matching network, and a target identification network. During training, we introduce a pseudo-label self-generation strategy and a weakly-supervised method to facilitate the learning of contextual objects and relations, respectively. The proposed techniques allow the networks to focus more effectively on referred objects within 3D scenes by understanding their context better. We validate our model on the challenging Nr3D, Sr3D, and ScanRefer datasets and demonstrate state-of-the-art performance. Our code will be public at https://github.com/yangli18/CORE-3DVG.
****
Learning to Search Feasible and Infeasible Regions of Routing Problems with Flexible Neural k-Opt
Yining Ma, Zhiguang Cao, Yeow Meng Chee
In this paper, we present Neural k-Opt (NeuOpt), a novel learning-to-search (L2S) solver for routing problems. It learns to perform flexible k-opt exchanges based on a tailored action factorization method and a customized recurrent dual-stream decoder. As a pioneering work to circumvent the pure feasibility masking scheme and enable the autonomous exploration of both feasible and infeasible regions, we then propose the Guided Infeasible Region Exploration (GIRE) scheme, which supplements the NeuOpt policy network with feasibility-related features and leverages reward shaping to steer reinforcement learning more effectively. Additionally, we equip NeuOpt with Dynamic Data Augmentation (D2A) for more diverse searches during inference. Extensive experiments on the Traveling Salesman Problem (TSP) and Capacitated Vehicle Routing Problem (CVRP) demonstrate that our NeuOpt not only significantly outstrips existing (masking-based) L2S solvers, but also showcases superiority over the learning-to-construct (L2C) and learning-to-predict (L2P) solvers. Notably, we offer fresh perspectives on how neural solvers can handle VRP constraints. Our code is available: https://github.com/yining043/NeuOpt.

****
Importance Weighted Actor-Critic for Optimal Conservative Offline Reinforcement Learning
Hanlin Zhu, Paria Rashidinejad, Jiantao Jiao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Hierarchical Semi-Implicit Variational Inference with Application to Diffusion Model Acceleration
Longlin Yu, Tianyu Xie, Yu Zhu, Tong Yang, Xiangyu Zhang, Cheng Zhang
Semi-implicit variational inference (SIVI) has been introduced to expand the analytical variational families by defining expressive semi-implicit distributions in a hierarchical manner. However, the single-layer architecture commonly used in current SIVI methods can be insufficient when the target posterior has complicated structures. In this paper, we propose hierarchical semi-implicit variational inference, called HSIVI, which generalizes SIVI to allow more expressive multi-layer construction of semi-implicit distributions. By introducing auxiliary distributions that interpolate between a simple base distribution and the target distribution, the conditional layers can be trained by progressively matching these auxiliary distributions one layer after another. Moreover, given pre-trained score networks, HSIVI can be used to accelerate the sampling process of diffusion models with the score matching objective. We show that HSIVI significantly enhances the expressiveness of SIVI on several Bayesian inference problems with complicated target distributions. When used for diffusion model acceleration, we show that HSIVI can produce high quality samples comparable to or better than the existing fast diffusion model based samplers with a small number of function evaluations on various datasets.
****
Geometry-Aware Adaptation for Pretrained Models
Nicholas Roberts, Xintong Li, Dyah Adila, Sonia Cromp, Tzu-Heng Huang, Jitian Zhao, Frederic Sala
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
JourneyDB: A Benchmark for Generative Image Understanding
Keqiang Sun, Junting Pan, Yuying Ge, Hao Li, Haodong Duan, Xiaoshi Wu, Renrui Zhang, Aojun Zhou, Zipeng Qin, Yi Wang, Jifeng Dai, Yu Qiao, Limin Wang, Hongsheng Li
While recent advancements in vision-language models have had a transformative impact on multi-modal comprehension, the extent to which these models possess the ability to comprehend generated images remains uncertain. Synthetic images, in comparison to real data, encompass a higher level of diversity in terms of both content and style, thereby presenting significant challenges for the models to fully grasp. In light of this challenge, we introduce a comprehensive dataset, referred to as JourneyDB, that caters to the domain of generative images within the context of multi-modal visual understanding. Our meticulously curated dataset comprises 4 million distinct and high-quality generated images, each paired with the corresponding text prompts that were employed in their creation. Furthermore, we additionally introduce an external subset with results of another 22 text-to-image generative models, which makes JourneyDB a comprehensive benchmark for evaluating the comprehension of generated images. On our dataset, we have devised four benchmarks to assess the performance of generated image comprehension in relation to both content and style interpretation. These benchmarks encompass prompt inversion, style retrieval, image captioning, and visual question answering. Lastly, we evaluate the performance of state-of-the-art multi-modal models when applied to the JourneyDB dataset, providing a comprehensive analysis of their s

trengths and limitations in comprehending generated content. We anticipate that the proposed dataset and benchmarks will facilitate further research in the field of generative content understanding. The dataset is publicly available at https://journeydb.github.io.
****

A fast heuristic to optimize time-space tradeoff for large models
Akifumi Imanishi, Zijian Xu, Masayuki Takagi, Sixue Wang, Emilio Castillo
Training large-scale neural networks is heavily constrained by GPU memory. In order to circumvent this limitation, gradient checkpointing, or recomputation is a powerful technique. There is active research in this area with methods such as Checkmake or Moccasin. However, both Checkmate and Moccasin rely on mixed integer linear programming or constraint programming, resulting in limited scalability due to their exponentially large search space.This paper proposes a novel algorithm for recomputation (FastSA) based on a simulated annealing heuristic that achieves comparable or even better solutions than state-of-the-art alternatives. FastSA can optimize computational graphs with thousands of nodes within 3 to 30 seconds, several orders of magnitude faster than current solutions.We applied FastSA to PyTorch models and verified its effectiveness through popular large vision and text models, including recent language models with the transformer architecture. The results demonstrate significant memory reductions by 73% with extra 18% computational overheads on average. Our experiments demonstrate the practicality and effectiveness of our recomputation algorithm, further highlighting its potential for wide application in various deep learning domains.
****
A Unified Conditional Framework for Diffusion-based Image Restoration
Yi Zhang, Xiaoyu Shi, Dasong Li, Xiaogang Wang, Jian Wang, Hongsheng Li
Diffusion Probabilistic Models (DPMs) have recently shown remarkable performance in image generation tasks, which are capable of generating highly realistic images. When adopting DPMs for image restoration tasks, the crucial aspect lies in how to integrate the conditional information to guide the DPMs to generate accurate and natural output, which has been largely overlooked in existing works. In this paper, we present a unified conditional framework based on diffusion models for image restoration. We leverage a lightweight UNet to predict initial guidance and the diffusion model to learn the residual of the guidance. By carefully designing the basic module and integration module for the diffusion model block, we integrate the guidance and other auxiliary conditional information into every block of the diffusion model to achieve spatially-adaptive generation conditioning. To handle high-resolution images, we propose a simple yet effective inter-step patch-splitting strategy to produce arbitrary-resolution images without grid artifacts. We evaluate our conditional framework on three challenging tasks: extreme low-light denoising, deblurring, and JPEG restoration, demonstrating its significant improvements in perceptual quality and the generalization to restoration tasks. The code will be released at https://zhangyi-3.github.io/project/UCDIR/.
****
Environment-Aware Dynamic Graph Learning for Out-of-Distribution Generalization
Haonan Yuan, Qingyun Sun, Xingcheng Fu, Ziwei Zhang, Cheng Ji, Hao Peng, Jianxin Li
Dynamic graph neural networks (DGNNs) are increasingly pervasive in exploiting spatio-temporal patterns on dynamic graphs. However, existing works fail to generalize under distribution shifts, which are common in real-world scenarios. As the generation of dynamic graphs is heavily influenced by latent environments, investigating their impacts on the out-of-distribution (OOD) generalization is critical. However, it remains unexplored with the following two major challenges: (1) How to properly model and infer the complex environments on dynamic graphs with distribution shifts? (2) How to discover invariant patterns given inferred spatio-temporal environments? To solve these challenges, we propose a novel Environment-Aware dynamic Graph LEarning (EAGLE) framework for OOD generalization by modeling complex coupled environments and exploiting spatio-temporal invariant patterns. Specifically, we first design the environment-aware EA-DGNN to model envi

ronments by multi-channel environments disentangling. Then, we propose an environment instantiation mechanism for environment diversification with inferred distributions. Finally, we discriminate spatio-temporal invariant patterns for out-of-distribution prediction by the invariant pattern recognition mechanism and perform fine-grained causal interventions node-wisely with a mixture of instantiated environment samples. Experiments on real-world and synthetic dynamic graph datasets demonstrate the superiority of our method against state-of-the-art baselines under distribution shifts. To the best of our knowledge, we are the first to study OOD generalization on dynamic graphs from the environment learning perspective.
****
Provably Fast Finite Particle Variants of SVGD via Virtual Particle Stochastic Approximation
Aniket Das, Dheeraj Nagaraj
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Flow Factorized Representation Learning
Yue Song, Andy Keller, Nicu Sebe, Max Welling
A prominent goal of representation learning research is to achieve representations which are factorized in a useful manner with respect to the ground truth factors of variation. The fields of disentangled and equivariant representation learning have approached this ideal from a range of complimentary perspectives; however, to date, most approaches have proven to either be ill-specified or insufficiently flexible to effectively separate all realistic factors of interest in a learned latent space. In this work, we propose an alternative viewpoint on such structured representation learning which we call Flow Factorized Representation Learning, and demonstrate it to learn both more efficient and more usefully structured representations than existing frameworks. Specifically, we introduce a generative model which specifies a distinct set of latent probability paths that define different input transformations. Each latent flow is generated by the gradient field of a learned potential following dynamic optimal transport. Our novel setup brings new understandings to both \textit{disentanglement} and \textit{equivariance}. We show that our model achieves higher likelihoods on standard representation learning benchmarks while simultaneously being closer to approximately equivariant models. Furthermore, we demonstrate that the transformations learned by our model are flexibly composable and can also extrapolate to new data, implying a degree of robustness and generalizability approaching the ultimate goal of usefully factorized representation learning.
****
Hierarchical Randomized Smoothing
Yan Scholten, Jan Schuchardt, Aleksandar Bojchevski, Stephan Günnemann
Real-world data is complex and often consists of objects that can be decomposed into multiple entities (e.g. images into pixels, graphs into interconnected nodes). Randomized smoothing is a powerful framework for making models provably robust against small changes to their inputs - by guaranteeing robustness of the majority vote when randomly adding noise before classification. Yet, certifying robustness on such complex data via randomized smoothing is challenging when adversaries do not arbitrarily perturb entire objects (e.g. images) but only a subset of their entities (e.g. pixels). As a solution, we introduce hierarchical randomized smoothing: We partially smooth objects by adding random noise only on a randomly selected subset of their entities. By adding noise in a more targeted manner than existing methods we obtain stronger robustness guarantees while maintaining high accuracy. We initialize hierarchical smoothing using different noising distributions, yielding novel robustness certificates for discrete and continuous domains. We experimentally demonstrate the importance of hierarchical smoothing in image and node classification, where it yields superior robustness-accuracy trade-offs. Overall, hierarchical smoothing is an important contribution toward

s models that are both - certifiably robust to perturbations and accurate.
****
BenchCLAMP: A Benchmark for Evaluating Language Models on Syntactic and Semantic Parsing
Subhro Roy, Samuel Thomson, Tongfei Chen, Richard Shin, Adam Pauls, Jason Eisner, Benjamin Van Durme
Recent work has shown that generation from a prompted or fine-tuned language model can perform well at semantic parsing when the output is constrained to be a valid semantic representation. We introduce BenchCLAMP, a Benchmark to evaluate Constrained LAnguage Model Parsing, that includes context-free grammars for seven semantic parsing datasets and two syntactic parsing datasets with varied output meaning representations, as well as a constrained decoding interface to generate only valid outputs covered by these grammars. We provide low, medium, and high resource splits for each dataset, allowing accurate comparison of various language models under different data regimes. Our benchmark supports evaluation of language models using prompt-based learning as well as fine-tuning. We benchmark seven language models, including two GPT-3 variants available only through an API. Our experiments show that encoder-decoder pretrained language models can achieve similar performance or even surpass state-of-the-art methods for both syntactic and semantic parsing when the model output is constrained to be valid.
****
Stable Nonconvex-Nonconcave Training via Linear Interpolation
Thomas Pethick, Wanyun Xie, Volkan Cevher
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
UniPC: A Unified Predictor-Corrector Framework for Fast Sampling of Diffusion Models
Wenliang Zhao, Lujia Bai, Yongming Rao, Jie Zhou, Jiwen Lu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Coop: Memory is not a Commodity
Jianhao Zhang, Shihan Ma, Peihong Liu, Jinhui Yuan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning with Explanation Constraints
Rattana Pukdee, Dylan Sam, J. Zico Kolter, Maria-Florina F. Balcan, Pradeep Ravikumar
As larger deep learning models are hard to interpret, there has been a recent focus on generating explanations of these black-box models. In contrast, we may have apriori explanations of how models should behave. In this paper, we formalize this notion as learning from explanation constraints and provide a learning theoretic framework to analyze how such explanations can improve the learning of our models. One may naturally ask, "When would these explanations be helpful?"Our first key contribution addresses this question via a class of models that satisfies these explanation constraints in expectation over new data. We provide a characterization of the benefits of these models (in terms of the reduction of their Rademacher complexities) for a canonical class of explanations given by gradient information in the settings of both linear models and two layer neural networks. In addition, we provide an algorithmic solution for our framework, via a variational approximation that achieves better performance and satisfies these constraints more frequently, when compared to simpler augmented Lagrangian methods

to incorporate these explanations. We demonstrate the benefits of our approach over a large array of synthetic and real-world experiments.
****
On the Interplay between Social Welfare and Tractability of Equilibria
Ioannis Anagnostides, Tuomas Sandholm
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Solving Linear Inverse Problems Provably via Posterior Sampling with Latent Diffusion Models
Litu Rout, Negin Raoof, Giannis Daras, Constantine Caramanis, Alex Dimakis, Sanjay Shakkottai
We present the first framework to solve linear inverse problems leveraging pre-trained \textit{latent} diffusion models. Previously proposed algorithms (such as DPS and DDRM) only apply to \textit{pixel-space} diffusion models. We theoretically analyze our algorithm showing provable sample recovery in a linear model setting. The algorithmic insight obtained from our analysis extends to more general settings often considered in practice. Experimentally, we outperform previously proposed posterior sampling algorithms in a wide variety of problems including random inpainting, block inpainting, denoising, deblurring, destriping, and super-resolution.
****
Maximum State Entropy Exploration using Predecessor and Successor Representations
Arnav Kumar Jain, Lucas Lehnert, Irina Rish, Glen Berseth
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
T2T: From Distribution Learning in Training to Gradient Search in Testing for Combinatorial Optimization
Yang Li, Jinpei Guo, Runzhong Wang, Junchi Yan
Extensive experiments have gradually revealed the potential performance bottleneck of modeling Combinatorial Optimization (CO) solving as neural solution prediction tasks. The neural networks, in their pursuit of minimizing the average objective score across the distribution of historical problem instances, diverge from the core target of CO of seeking optimal solutions for every test instance. This calls for an effective search on each problem instance, while the model should serve to provide supporting knowledge that benefits the search. To this end, we propose T2T (Training to Testing) framework that first leverages the generative modeling to estimate the high-quality solution distribution for each instance during training, and then conducts a gradient-based search within the solution space during testing. The proposed neural search paradigm consistently leverages generative modeling, specifically diffusion, for graduated solution improvement. It disrupts the local structure of the given solution by introducing noise and reconstructs a lower-cost solution guided by the optimization objective. Experimental results on Traveling Salesman Problem (TSP) and Maximal Independent Set (MIS) show the significant superiority of T2T, demonstrating an average performance gain of 49.15% for TSP solving and 17.27% for MIS solving compared to the previous state-of-the-art.
****
Learning Curves for Noisy Heterogeneous Feature-Subsampled Ridge Ensembles
Ben Ruben, Cengiz Pehlevan
Feature bagging is a well-established ensembling method which aims to reducepred iction variance by combining predictions of many estimators trained on subsetsor projections of features. Here, we develop a theory of feature-bagging in noisyl east-squares ridge ensembles and simplify the resulting learning curves in the s

pecialcase of equicorrelated data. Using analytical learning curves, we demonstr atethat subsampling shifts the double-descent peak of a linear predictor. This l eadsus to introduce heterogeneous feature ensembling, with estimators built on v aryingnumbers of feature dimensions, as a computationally efficient method to mi tigatedouble-descent. Then, we compare the performance of a feature-subsamplinge nsemble to a single linear predictor, describing a trade-off between noise ampli ficationdue to subsampling and noise reduction due to ensembling. Our qualitativ einsights carry over to linear classifiers applied to image classification tasks withrealistic datasets constructed using a state-of-the-art deep learning featu re map.

****

Neural MMO 2.0: A Massively Multi-task Addition to Massively Multi-agent Learnin g

Joseph Suarez, David Bloomin, Kyoung Whan Choe, Hao Xiang Li, Ryan Sullivan, Nis haanth Kanna, Daniel Scott, Rose Shuman, Herbie Bradley, Louis Castricato, Phill ip Isola, Chenghui Yu, Yuhao Jiang, Qimai Li, Jiaxin Chen, Xiaolong Zhu

Neural MMO 2.0 is a massively multi-agent and multi-task environment for reinfor cement learning research. This version features a novel task-system that broaden s the range of training settings and poses a new challenge in generalization: ev aluation on and against tasks, maps, and opponents never seen during training. M aps are procedurally generated with 128 agents in the standard setting and 1-102 4 supported overall. Version 2.0 is a complete rewrite of its predecessor with t hree-fold improved performance, effectively addressing simulation bottlenecks in online training. Enhancements to compatibility enable training with standard re inforcement learning frameworks designed for much simpler environments. Neural M MO 2.0 is free and open-source with comprehensive documentation available at neu ralmmo.github.io and an active community Discord. To spark initial research on t his new platform, we are concurrently running a competition at NeurIPS 2023.

****

Zero-shot Visual Relation Detection via Composite Visual Cues from Large Languag e Models

Lin Li, Jun Xiao, Guikun Chen, Jian Shao, Yueting Zhuang, Long Chen

Pretrained vision-language models, such as CLIP, have demonstrated strong genera lization capabilities, making them promising tools in the realm of zero-shot vis ual recognition. Visual relation detection (VRD) is a typical task that identifi es relationship (or interaction) types between object pairs within an image. How ever, naively utilizing CLIP with prevalent class-based prompts for zero-shot VR D has several weaknesses, e.g., it struggles to distinguish between different fi ne-grained relation types and it neglects essential spatial information of two o bjects. To this end, we propose a novel method for zero-shot VRD: RECODE, which solves RElation detection via COmposite DEscription prompts. Specifically, RECOD E first decomposes each predicate category into subject, object, and spatial com ponents. Then, it leverages large language models (LLMs) to generate description -based prompts (or visual cues) for each component. Different visual cues enhanc e the discriminability of similar relation categories from different perspective s, which significantly boosts performance in VRD. To dynamically fuse different cues, we further introduce a chain-of-thought method that prompts LLMs to genera te reasonable weights for different visual cues. Extensive experiments on four V RD benchmarks have demonstrated the effectiveness and interpretability of RECODE .

****

ToolQA: A Dataset for LLM Question Answering with External Tools

Yuchen Zhuang, Yue Yu, Kuan Wang, Haotian Sun, Chao Zhang

Large Language Models (LLMs) have demonstrated impressive performance in various NLP tasks, but they still suffer from challenges such as hallucination and weak numerical reasoning. To overcome these challenges, external tools can be used t o enhance LLMs' question-answering abilities. However, current evaluation method s do not distinguish between questions that can be answered using LLMs' internal knowledge and those that require external information through tool use. To addr ess this issue, we introduce a new dataset called ToolQA, which is designed to f

aithfully evaluate LLMs' ability to use external tools for question answering. Our development of ToolQA involved a scalable, automated process for dataset curation, along with 13 specialized tools designed for interaction with external knowledge in order to answer questions. Importantly, we strive to minimize the overlap between our benchmark data and LLMs' pre-training data, enabling a more precise evaluation of LLMs' tool-use reasoning abilities. We conducted an in-depth diagnosis of existing tool-use LLMs to highlight their strengths, weaknesses, and potential improvements. Our findings set a new benchmark for evaluating LLMs and suggest new directions for future advancements. Our data and code are freely available for the broader scientific community on GitHub.
****

BiSLS/SPS: Auto-tune Step Sizes for Stable Bi-level Optimization
Chen Fan, Gaspard Choné-Ducasse, Mark Schmidt, Christos Thrampoulidis
The popularity of bi-level optimization (BO) in deep learning has spurred a growing interest in studying gradient-based BO algorithms.However, existing algorithms involve two coupled learning rates that can be affected by approximation errors when computing hypergradients, making careful fine-tuning necessary to ensure fast convergence. To alleviate this issue, we investigate the use of recently proposed adaptive step-size methods, namely stochastic line search (SLS) and stochastic Polyak step size (SPS), for computing both the upper and lower-level learning rates. First, we revisit the use of SLS and SPS in single-level optimization without the additional interpolation condition that is typically assumed in prior works. For such settings, we investigate new variants of SLS and SPS that improve upon existing suggestions in the literature and are simpler to implement. Importantly, these two variants can be seen as special instances of general family of methods with an envelope-type step-size. This unified envelope strategy allows for the extension of the algorithms and their convergence guarantees to BO settings. Finally, our extensive experiments demonstrate that the new algorithms, which are available in both SGD and Adam versions, can find large learning rates with minimal tuning and converge faster than corresponding vanilla SGD or Adam BO algorithms that require fine-tuning.
****

Compositional Abilities Emerge Multiplicatively: Exploring Diffusion Models on a Synthetic Task
Maya Okawa, Ekdeep S Lubana, Robert Dick, Hidenori Tanaka
Modern generative models exhibit unprecedented capabilities to generate extremely realistic data. However, given the inherent compositionality of the real world, reliable use of these models in practical applications requires that they exhibit the capability to compose a novel set of concepts to generate outputs not seen in the training data set. Prior work demonstrates that recent diffusion models do exhibit intriguing compositional generalization abilities, but also fail unpredictably. Motivated by this, we perform a controlled study for understanding compositional generalization in conditional diffusion models in a synthetic setting, varying different attributes of the training data and measuring the model's ability to generate samples out-of-distribution. Our results show: (i) the order in which the ability to generate samples from a concept and compose them emerges is governed by the structure of the underlying data-generating process; (ii) performance on compositional tasks exhibits a sudden "emergence" due to multiplicative reliance on the performance of constituent tasks, partially explaining emergent phenomena seen in generative models; and (iii) composing concepts with lower frequency in the training data to generate out-of-distribution samples requires considerably more optimization steps compared to generating in-distribution samples. Overall, our study lays a foundation for understanding emergent capabilities and compositionality in generative models from a data-centric perspective.
****

Extracting Reward Functions from Diffusion Models
Felipe Nuti, Tim Franzmeyer, João F. Henriques
Diffusion models have achieved remarkable results in image generation, and have similarly been used to learn high-performing policies in sequential decision-making tasks. Decision-making diffusion models can be trained on lower-quality data

, and then be steered with a reward function to generate near-optimal trajectori
es.We consider the problem of extracting a reward function by comparing a decisi
on-making diffusion model that models low-reward behavior and one that models hi
gh-reward behavior; a setting related to inverse reinforcement learning. We firs
t define the notion of a \emph{relative reward function of two diffusion models}
 and show conditions under which it exists and is unique. We then devise a pract
ical learning algorithm for extracting it by aligning the gradients of a reward
function -- parametrized by a neural network -- to the difference in outputs of
both diffusion models.Our method finds correct reward functions in navigation en
vironments, and we demonstrate that steering the base model with the learned rew
ard functions results in significantly increased performance in standard locomot
ion benchmarks.Finally, we demonstrate that our approach generalizes beyond sequ
ential decision-making by learning a reward-like function from two large-scale i
mage generation diffusion models. The extracted reward function successfully ass
igns lower rewards to harmful images.
****
Disentangling Voice and Content with Self-Supervision for Speaker Recognition
TIANCHI LIU, Kong Aik Lee, Qiongqiong Wang, Haizhou Li
For speaker recognition, it is difficult to extract an accurate speaker represe
ntation from speech because of its mixture of speaker traits and content. This p
aper proposes a disentanglement framework that simultaneously models speaker tra
its and content variability in speech. It is realized with the use of three Gaus
sian inference layers, each consisting of a learnable transition model that extr
acts distinct speech components. Notably, a strengthened transition model is spe
cifically designed to model complex speech dynamics. We also propose a self-supe
rvision method to dynamically disentangle content without the use of labels othe
r than speaker identities. The efficacy of the proposed framework is validated
via experiments conducted on the VoxCeleb and SITW datasets with 9.56\% and 8.24
\% average reductions in EER and minDCF, respectively. Since neither additional
model training nor data is specifically needed, it is easily applicable in pract
ical use.
****
Automatic Integration for Spatiotemporal Neural Point Processes
Zihao Zhou, Rose Yu
Learning continuous-time point processes is essential to many discrete event for
ecasting tasks. However, integration poses a major challenge, particularly for s
patiotemporal point processes (STPPs), as it involves calculating the likelihood
 through triple integrals over space and time. Existing methods for integrating
STPP either assume a parametric form of the intensity function, which lacks flex
ibility; or approximating the intensity with Monte Carlo sampling, which introdu
ces numerical errors. Recent work by Omi et al. proposes a dual network approach
 for efficient integration of flexible intensity function. However, their method
 only focuses on the 1D temporal point process. In this paper, we introduce a no
vel paradigm: Auto-STPP (Automatic Integration for Spatiotemporal Neural Point P
rocesses) that extends the dual network approach to 3D STPP. While previous work
 provides a foundation, its direct extension overly restricts the intensity func
tion and leads to computational challenges. In response, we introduce a decompos
able parametrization for the integral network using ProdNet. This approach, leve
raging the product of simplified univariate graphs, effectively sidesteps the co
mputational complexities inherent in multivariate computational graphs. We prove
 the consistency of Auto-STPP and validate it on synthetic data and benchmark re
al-world datasets. Auto-STPP shows a significant advantage in recovering complex
 intensity functions from irregular spatiotemporal events, particularly when the
 intensity is sharply localized. Our code is open-source at https://github.com/R
ose-STL-Lab/AutoSTPP.
****
Identifiability Guarantees for Causal Disentanglement from Soft Interventions
Jiaqi Zhang, Kristjan Greenewald, Chandler Squires, Akash Srivastava, Karthikeya
n Shanmugam, Caroline Uhler
Causal disentanglement aims to uncover a representation of data using latent var

iables that are interrelated through a causal model. Such a representation is identifiable if the latent model that explains the data is unique. In this paper, we focus on the scenario where unpaired observational and interventional data are available, with each intervention changing the mechanism of a latent variable. When the causal variables are fully observed, statistically consistent algorithms have been developed to identify the causal model under faithfulness assumptions. We here show that identifiability can still be achieved with unobserved causal variables, given a generalized notion of faithfulness. Our results guarantee that we can recover the latent causal model up to an equivalence class and predict the effect of unseen combinations of interventions, in the limit of infinite data. We implement our causal disentanglement framework by developing an autoencoding variational Bayes algorithm and apply it to the problem of predicting combinatorial perturbation effects in genomics.
****
Equivariant Adaptation of Large Pretrained Models
Arnab Kumar Mondal, Siba Smarak Panigrahi, Oumar Kaba, Sai Rajeswar Mudumba, Siamak Ravanbakhsh
Equivariant networks are specifically designed to ensure consistent behavior with respect to a set of input transformations, leading to higher sample efficiency and more accurate and robust predictions. However, redesigning each component of prevalent deep neural network architectures to achieve chosen equivariance is a difficult problem and can result in a computationally expensive network during both training and inference. A recently proposed alternative towards equivariance that removes the architectural constraints is to use a simple canonicalization network that transforms the input to a canonical form before feeding it to an unconstrained prediction network. We show here that this approach can effectively be used to make a large pretrained network equivariant. However, we observe that the produced canonical orientations can be misaligned with those of the training distribution, hindering performance. Using dataset-dependent priors to inform the canonicalization function, we are able to make large pretrained models equivariant while maintaining their performance. This significantly improves the robustness of these models to deterministic transformations of the data, such as rotations. We believe this equivariant adaptation of large pretrained models can help their domain-specific applications with known symmetry priors.
****
HT-Step: Aligning Instructional Articles with How-To Videos
Triantafyllos Afouras, Effrosyni Mavroudi, Tushar Nagarajan, Huiyu Wang, Lorenzo Torresani
We introduce HT-Step, a large-scale dataset containing temporal annotations of instructional article steps in cooking videos. It includes 122k segment-level annotations over 20k narrated videos (approximately 2.3k hours) of the HowTo100M dataset.Each annotation provides a temporal interval, and a categorical step label from a taxonomy of 4,958 unique steps automatically mined from wikiHow articles which include rich descriptions of each step.Our dataset significantly surpasses existing labeled step datasets in terms of scale, number of tasks, and richness of natural language step descriptions. Based on these annotations, we introduce a strongly supervised benchmark for aligning instructional articles with how-to videos and present a comprehensive evaluation of baseline methods for this task.By publicly releasing these annotations and defining rigorous evaluation protocols and metrics,we hope to significantly accelerate research in the field of procedural activity understanding.
****
Provable Training for Graph Contrastive Learning
Yue Yu, Xiao Wang, Mengmei Zhang, Nian Liu, Chuan Shi
Graph Contrastive Learning (GCL) has emerged as a popular training approach for learning node embeddings from augmented graphs without labels. Despite the key principle that maximizing the similarity between positive node pairs while minimizing it between negative node pairs is well established, some fundamental problems are still unclear. Considering the complex graph structure, are some nodes consistently well-trained and following this principle even with different graph a

ugmentations? Or are there some nodes more likely to be untrained across graph augmentations and violate the principle? How to distinguish these nodes and further guide the training of GCL? To answer these questions, we first present experimental evidence showing that the training of GCL is indeed imbalanced across all nodes. To address this problem, we propose the metric "node compactness", which is the lower bound of how a node follows the GCL principle related to the range of augmentations. We further derive the form of node compactness theoretically through bound propagation, which can be integrated into binary cross-entropy as a regularization. To this end, we propose the PrOvable Training (POT) for GCL, which regularizes the training of GCL to encode node embeddings that follows the GCL principle better. Through extensive experiments on various benchmarks, POT consistently improves the existing GCL approaches, serving as a friendly plugin.
****

Generalized Weighted Path Consistency for Mastering Atari Games

Dengwei Zhao, Shikui Tu, Lei Xu

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Scaling Data-Constrained Language Models

Niklas Muennighoff, Alexander Rush, Boaz Barak, Teven Le Scao, Nouamane Tazi, Aleksandra Piktus, Sampo Pyysalo, Thomas Wolf, Colin A. Raffel

The current trend of scaling language models involves increasing both parameter count and training dataset size. Extrapolating this trend suggests that training dataset size may soon be limited by the amount of text data available on the internet. Motivated by this limit, we investigate scaling language models in data-constrained regimes. Specifically, we run a large set of experiments varying the extent of data repetition and compute budget, ranging up to 900 billion training tokens and 9 billion parameter models. We find that with constrained data for a fixed compute budget, training with up to 4 epochs of repeated data yields negligible changes to loss compared to having unique data. However, with more repetition, the value of adding compute eventually decays to zero. We propose and empirically validate a scaling law for compute optimality that accounts for the decreasing value of repeated tokens and excess parameters. Finally, we experiment with approaches mitigating data scarcity, including augmenting the training dataset with code data or removing commonly used filters. Models and datasets from our 400 training runs are freely available at https://github.com/huggingface/datablations.
****

A Definition of Continual Reinforcement Learning

David Abel, Andre Barreto, Benjamin Van Roy, Doina Precup, Hado P. van Hasselt, Satinder Singh

In a standard view of the reinforcement learning problem, an agent’s goal is to efficiently identify a policy that maximizes long-term reward. However, this perspective is based on a restricted view of learning as finding a solution, rather than treating learning as endless adaptation. In contrast, continual reinforcement learning refers to the setting in which the best agents never stop learning. Despite the importance of continual reinforcement learning, the community lacks a simple definition of the problem that highlights its commitments and makes its primary concepts precise and clear. To this end, this paper is dedicated to carefully defining the continual reinforcement learning problem. We formalize the notion of agents that “never stop learning” through a new mathematical language for analyzing and cataloging agents. Using this new language, we define a continual learning agent as one that can be understood as carrying out an implicit search process indefinitely, and continual reinforcement learning as the setting in which the best agents are all continual learning agents. We provide two motivating examples, illustrating that traditional views of multi-task reinforcement learning and continual supervised learning are special cases of our definition. Collectively, these definitions and perspectives formalize many intuitive concepts

at the heart of learning, and open new research pathways surrounding continual learning agents.
****
A Dual-Stream Neural Network Explains the Functional Segregation of Dorsal and Ventral Visual Pathways in Human Brains
Minkyu Choi, Kuan Han, Xiaokai Wang, Yizhen Zhang, Zhongming Liu
The human visual system uses two parallel pathways for spatial processing and object recognition. In contrast, computer vision systems tend to use a single feedforward pathway, rendering them less robust, adaptive, or efficient than human vision. To bridge this gap, we developed a dual-stream vision model inspired by the human eyes and brain. At the input level, the model samples two complementary visual patterns to mimic how the human eyes use magnocellular and parvocellular retinal ganglion cells to separate retinal inputs to the brain. At the backend, the model processes the separate input patterns through two branches of convolutional neural networks (CNN) to mimic how the human brain uses the dorsal and ventral cortical pathways for parallel visual processing. The first branch (WhereCNN) samples a global view to learn spatial attention and control eye movements. The second branch (WhatCNN) samples a local view to represent the object around the fixation. Over time, the two branches interact recurrently to build a scene representation from moving fixations. We compared this model with the human brains processing the same movie and evaluated their functional alignment by linear transformation. The WhereCNN and WhatCNN branches were found to differentially match the dorsal and ventral pathways of the visual cortex, respectively, primarily due to their different learning objectives, rather than their distinctions in retinal sampling or sensitivity to attention-driven eye movements. These model-based results lead us to speculate that the distinct responses and representations of the ventral and dorsal streams are more influenced by their distinct goals in visual attention and object recognition than by their specific bias or selectivity in retinal inputs. This dual-stream model takes a further step in brain-inspired computer vision, enabling parallel neural networks to actively explore and understand the visual surroundings.
****
When Do Transformers Shine in RL? Decoupling Memory from Credit Assignment
Tianwei Ni, Michel Ma, Benjamin Eysenbach, Pierre-Luc Bacon
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Hypothesis Selection with Memory Constraints
Maryam Aliakbarpour, Mark Bun, Adam Smith
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Optimization or Architecture: How to Hack Kalman Filtering
Ido Greenberg, Netanel Yannay, Shie Mannor
In non-linear filtering, it is traditional to compare non-linear architectures such as neural networks to the standard linear Kalman Filter (KF). We observe that this mixes the evaluation of two separate components: the non-linear architecture, and the parameters optimization method. In particular, the non-linear model is often optimized, whereas the reference KF model is not. We argue that both should be optimized similarly, and to that end present the Optimized KF (OKF). We demonstrate that the KF may become competitive to neural models – if optimized using OKF. This implies that experimental conclusions of certain previous studies were derived from a flawed process. The advantage of OKF over the standard KF is further studied theoretically and empirically, in a variety of problems. Conveniently, OKF can replace the KF in real-world systems by merely updating the parameters.

****
Online robust non-stationary estimation
Abishek Sankararaman, Balakrishnan Narayanaswamy
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
POP-3D: Open-Vocabulary 3D Occupancy Prediction from Images
Antonin Vobecky, Oriane Siméoni, David Hurych, Spyridon Gidaris, Andrei Bursuc, Patrick Pérez, Josef Sivic
We describe an approach to predict open-vocabulary 3D semantic voxel occupancy map from input 2D images with the objective of enabling 3D grounding, segmentation and retrieval of free-form language queries. This is a challenging problem because of the 2D-3D ambiguity and the open-vocabulary nature of the target tasks, where obtaining annotated training data in 3D is difficult. The contributions of this work are three-fold. First, we design a new model architecture for open-vocabulary 3D semantic occupancy prediction. The architecture consists of a 2D-3D encoder together with occupancy prediction and 3D-language heads. The output is a dense voxel map of 3D grounded language embeddings enabling a range of open-vocabulary tasks. Second, we develop a tri-modal self-supervised learning algorithm that leverages three modalities: (i) images, (ii) language and (iii) LiDAR point clouds, and enables training the proposed architecture using a strong pre-trained vision-language model without the need for any 3D manual language annotations. Finally, we demonstrate quantitatively the strengths of the proposed model on several open-vocabulary tasks:Zero-shot 3D semantic segmentation using existing datasets; 3D grounding and retrieval of free-form language queries, using a small dataset that we propose as an extension of nuScenes. You can find the project page here https://vobecant.github.io/POP3D.
****
Faster Relative Entropy Coding with Greedy Rejection Coding
Gergely Flamich, Stratis Markou, José Miguel Hernández-Lobato
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Sparse Parameterization for Epitomic Dataset Distillation
Xing Wei, Anjia Cao, Funing Yang, Zhiheng Ma
The success of deep learning relies heavily on large and diverse datasets, but the storage, preprocessing, and training of such data present significant challenges. To address these challenges, dataset distillation techniques have been proposed to obtain smaller synthetic datasets that capture the essential information of the originals. In this paper, we introduce a Sparse Parameterization for Epitomic datasEt Distillation (SPEED) framework, which leverages the concept of dictionary learning and sparse coding to distill epitomes that represent pivotal information of the dataset. SPEED prioritizes proper parameterization of the synthetic dataset and introduces techniques to capture spatial redundancy within and between synthetic images. We propose Spatial-Agnostic Epitomic Tokens (SAETs) and Sparse Coding Matrices (SCMs) to efficiently represent and select significant features. Additionally, we build a Feature-Recurrent Network (FReeNet) to generate hierarchical features with high compression and storage efficiency. Experimental results demonstrate the superiority of SPEED in handling high-resolution datasets, achieving state-of-the-art performance on multiple benchmarks and downstream applications. Our framework is compatible with a variety of dataset matching approaches, generally enhancing their performance. This work highlights the importance of proper parameterization in epitomic dataset distillation and opens avenues for efficient representation learning. Source code is available at https://github.com/MIV-XJTU/SPEED.
****

HAP: Structure-Aware Masked Image Modeling for Human-Centric Perception
Junkun Yuan, Xinyu Zhang, Hao Zhou, Jian Wang, Zhongwei Qiu, Zhiyin Shao, Shaofeng Zhang, Sifan Long, Kun Kuang, Kun Yao, Junyu Han, Errui Ding, Lanfen Lin, Fei Wu, Jingdong Wang
Model pre-training is essential in human-centric perception. In this paper, we first introduce masked image modeling (MIM) as a pre-training approach for this task. Upon revisiting the MIM training strategy, we reveal that human structure priors offer significant potential. Motivated by this insight, we further incorporate an intuitive human structure prior - human parts - into pre-training. Specifically, we employ this prior to guide the mask sampling process. Image patches, corresponding to human part regions, have high priority to be masked out. This encourages the model to concentrate more on body structure information during pre-training, yielding substantial benefits across a range of human-centric perception tasks. To further capture human characteristics, we propose a structure-invariant alignment loss that enforces different masked views, guided by the human part prior, to be closely aligned for the same image. We term the entire method as HAP. HAP simply uses a plain ViT as the encoder yet establishes new state-of-the-art performance on 11 human-centric benchmarks, and on-par result on one dataset. For example, HAP achieves 78.1% mAP on MSMT17 for person re-identification, 86.54% mA on PA-100K for pedestrian attribute recognition, 78.2% AP on MS COCO for 2D pose estimation, and 56.0 PA-MPJPE on 3DPW for 3D pose and shape estimation.
****
Trust Your ∇: Gradient-based Intervention Targeting for Causal Discovery
Mateusz Olko, Micha■ Zaj■c, Aleksandra Nowak, Nino Scherrer, Yashas Annadani, Stefan Bauer, ■ukasz Kuci■ski, Piotr Mi■o■
Inferring causal structure from data is a challenging task of fundamental importance in science. Often, observational data alone is not enough to uniquely identify a system’s causal structure. The use of interventional data can address this issue, however, acquiring these samples typically demands a considerable investment of time and physical or financial resources. In this work, we are concerned with the acquisition of interventional data in a targeted manner to minimize the number of required experiments. We propose a novel Gradient-based Intervention Targeting method, abbreviated GIT, that ’trusts’ the gradient estimator of a gradient-based causal discovery framework to provide signals for the intervention targeting function. We provide extensive experiments in simulated and real-world datasets and demonstrate that GIT performs on par with competitive baselines, surpassing them in the low-data regime.
****
SyncDiffusion: Coherent Montage via Synchronized Joint Diffusions
Yuseung Lee, Kunho Kim, Hyunjin Kim, Minhyuk Sung
The remarkable capabilities of pretrained image diffusion models have been utilized not only for generating fixed-size images but also for creating panoramas. However, naive stitching of multiple images often results in visible seams. Recent techniques have attempted to address this issue by performing joint diffusions in multiple windows and averaging latent features in overlapping regions. However, these approaches, which focus on seamless montage generation, often yield incoherent outputs by blending different scenes within a single image. To overcome this limitation, we propose SyncDiffusion, a plug-and-play module that synchronizes multiple diffusions through gradient descent from a perceptual similarity loss. Specifically, we compute the gradient of the perceptual loss using the predicted denoised images at each denoising step, providing meaningful guidance for achieving coherent montages. Our experimental results demonstrate that our method produces significantly more coherent outputs compared to previous methods (66.35% vs. 33.65% in our user study) while still maintaining fidelity (as assessed by GIQA) and compatibility with the input prompt (as measured by CLIP score). We further demonstrate the versatility of our method across three plug-and-play applications: layout-guided image generation, conditional image generation and 360-degree panorama generation. Our project page is at https://syncdiffusion.github.io.

****
Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the Mediterranean
Spyridon Kondylatos, Ioannis Prapas, Gustau Camps-Valls, Ioannis Papoutsis
We introduce Mesogeos, a large-scale multi-purpose dataset for wildfire modeling in the Mediterranean. Mesogeos integrates variables representing wildfire drivers (meteorology, vegetation, human activity) and historical records of wildfire ignitions and burned areas for 17 years (2006-2022). It is designed as a cloud-friendly spatio-temporal dataset, namely a datacube, harmonizing all variables in a grid of 1km x 1km x 1-day resolution. The datacube structure offers opportunities to assess machine learning (ML) usage in various wildfire modeling tasks. We extract two ML-ready datasets that establish distinct tracks to demonstrate this potential: (1) short-term wildfire danger forecasting and (2) final burned area estimation given the point of ignition. We define appropriate metrics and baselines to evaluate the performance of models in each track. By publishing the datacube, along with the code to create the ML datasets and models, we encourage the community to foster the implementation of additional tracks for mitigating the increasing threat of wildfires in the Mediterranean.
****
Deep learning with kernels through RKHM and the Perron-Frobenius operator
Yuka Hashimoto, Masahiro Ikeda, Hachem Kadri
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SmoothHess: ReLU Network Feature Interactions via Stein's Lemma
Max Torop, Aria Masoomi, Davin Hill, Kivanc Kose, Stratis Ioannidis, Jennifer Dy
Several recent methods for interpretability model feature interactions by looking at the Hessian of a neural network. This poses a challenge for ReLU networks, which are piecewise-linear and thus have a zero Hessian almost everywhere. We propose SmoothHess, a method of estimating second-order interactions through Stein's Lemma. In particular, we estimate the Hessian of the network convolved with a Gaussian through an efficient sampling algorithm, requiring only network gradient calls. SmoothHess is applied post-hoc, requires no modifications to the ReLU network architecture, and the extent of smoothing can be controlled explicitly. We provide a non-asymptotic bound on the sample complexity of our estimation procedure. We validate the superior ability of SmoothHess to capture interactions on benchmark datasets and a real-world medical spirometry dataset.
****
MLFMF: Data Sets for Machine Learning for Mathematical Formalization
Andrej Bauer, Matej Petkovi■, Ljupco Todorovski
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DreamSim: Learning New Dimensions of Human Visual Similarity using Synthetic Data
Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, Phillip Isola
Current perceptual similarity metrics operate at the level of pixels and patches. These metrics compare images in terms of their low-level colors and textures, but fail to capture mid-level similarities and differences in image layout, object pose, and semantic content. In this paper, we develop a perceptual metric that assesses images holistically. Our first step is to collect a new dataset of human similarity judgments over image pairs that are alike in diverse ways. Critical to this dataset is that judgments are nearly automatic and shared by all observers. To achieve this we use recent text-to-image models to create synthetic pairs that are perturbed along various dimensions. We observe that popular percept

ual metrics fall short of explaining our new data, and we introduce a new metric , DreamSim, tuned to better align with human perception. We analyze how our metric is affected by different visual attributes, and find that it focuses heavily on foreground objects and semantic content while also being sensitive to color and layout. Notably, despite being trained on synthetic data, our metric generalizes to real images, giving strong results on retrieval and reconstruction tasks. Furthermore, our metric outperforms both prior learned metrics and recent large vision models on these tasks. Our project page: https://dreamsim-nights.github.io/
****
Explaining the Uncertain: Stochastic Shapley Values for Gaussian Process Models
Siu Lun Chau, Krikamol Muandet, Dino Sejdinovic
We present a novel approach for explaining Gaussian processes (GPs) that can utilize the full analytical covariance structure present in GPs. Our method is based on the popular solution concept of Shapley values extended to stochastic cooperative games, resulting in explanations that are random variables. The GP explanations generated using our approach satisfy similar favorable axioms to standard Shapley values and possess a tractable covariance function across features and data observations. This covariance allows for quantifying explanation uncertainties and studying the statistical dependencies between explanations. We further extend our framework to the problem of predictive explanation, and propose a Shapley prior over the explanation function to predict Shapley values for new data based on previously computed ones. Our extensive illustrations demonstrate the effectiveness of the proposed approach.
****
WBCAtt: A White Blood Cell Dataset Annotated with Detailed Morphological Attributes
Satoshi Tsutsui, Winnie Pang, Bihan Wen
The examination of blood samples at a microscopic level plays a fundamental role in clinical diagnostics. For instance, an in-depth study of White Blood Cells (WBCs), a crucial component of our blood, is essential for diagnosing blood-related diseases such as leukemia and anemia. While multiple datasets containing WBC images have been proposed, they mostly focus on cell categorization, often lacking the necessary morphological details to explain such categorizations, despite the importance of explainable artificial intelligence (XAI) in medical domains. This paper seeks to address this limitation by introducing comprehensive annotations for WBC images. Through collaboration with pathologists, a thorough literature review, and manual inspection of microscopic images, we have identified 11 morphological attributes associated with the cell and its components (nucleus, cytoplasm, and granules). We then annotated ten thousand WBC images with these attributes, resulting in 113k labels (11 attributes x 10.3k images). Annotating at this level of detail and scale is unprecedented, offering unique value to AI in pathology. Moreover, we conduct experiments to predict these attributes from cell images, and also demonstrate specific applications that can benefit from our detailed annotations. Overall, our dataset paves the way for interpreting WBC recognition models, further advancing XAI in the fields of pathology and hematology.
****
Graph Mixture of Experts: Learning on Large-Scale Graphs with Explicit Diversity Modeling
Haotao Wang, Ziyu Jiang, Yuning You, Yan Han, Gaowen Liu, Jayanth Srinivasa, Ramana Kompella, Zhangyang "Atlas" Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Interpretable and Explainable Logical Policies via Neurally Guided Symbolic Abstraction
Quentin Delfosse, Hikaru Shindo, Devendra Dhami, Kristian Kersting

The limited priors required by neural networks make them the dominating choice to encode and learn policies using reinforcement learning (RL). However, they are also black-boxes, making it hard to understand the agent's behavior, especially when working on the image level. Therefore, neuro-symbolic RL aims at creating policies that are interpretable in the first place.Unfortunately, interpretability is not explainability. To achieve both, we introduce Neurally gUided Differentiable loGic policiEs (NUDGE). NUDGE exploits trained neural network-based agents to guide the search of candidate-weighted logic rules, then uses differentiable logic to train the logic agents. Our experimental evaluation demonstrates that NUDGE agents can induce interpretable and explainable policies while outperforming purely neural ones and showing good flexibility to environments of different initial states and problem sizes.
****

Personalized Dictionary Learning for Heterogeneous Datasets

Geyu Liang, Naichen Shi, Raed AL Kontar, Salar Fattahi

We introduce a relevant yet challenging problem named Personalized Dictionary Learning (PerDL), where the goal is to learn sparse linear representations from heterogeneous datasets that share some commonality. In PerDL, we model each dataset's shared and unique features as global and local dictionaries. Challenges for PerDL not only are inherited from classical dictionary learning(DL), but also arise due to the unknown nature of the shared and unique features. In this paper, we rigorously formulate this problem and provide conditions under which the global and local dictionaries can be provably disentangled. Under these conditions, we provide a meta-algorithm called Personalized Matching and Averaging (PerMA) that can recover both global and local dictionaries from heterogeneous datasets. PerMA is highly efficient; it converges to the ground truth at a linear rate under suitable conditions. Moreover, it automatically borrows strength from strong learners to improve the prediction of weak learners. As a general framework for extracting global and local dictionaries, we show the application of PerDL in different learning tasks, such as training with imbalanced datasets and video surveillance.
****

Graph-Structured Gaussian Processes for Transferable Graph Learning

Jun Wu, Lisa Ainsworth, Andrew Leakey, Haixun Wang, Jingrui He

Transferable graph learning involves knowledge transferability from a source graph to a relevant target graph. The major challenge of transferable graph learning is the distribution shift between source and target graphs induced by individual node attributes and complex graph structures. To solve this problem, in this paper, we propose a generic graph-structured Gaussian process framework (GraphGP) for adaptively transferring knowledge across graphs with either homophily or heterophily assumptions. Specifically, GraphGP is derived from a novel graph structure-aware neural network in the limit on the layer width. The generalization analysis of GraphGP explicitly investigates the connection between knowledge transferability and graph domain similarity. Extensive experiments on several transferable graph learning benchmarks demonstrate the efficacy of GraphGP over state-of-the-art Gaussian process baselines.
****

Language Models are Weak Learners

Hariharan Manikandan, Yiding Jiang, J. Zico Kolter

A central notion in practical and theoretical machine learning is that of a weak learner, classifiers that achieve better-than-random performance (on any given distribution over data), even by a small margin. Such weak learners form the practical basis for canonical machine learning methods such as boosting. In this work, we illustrate that prompt-based large language models can operate effectively as said weak learners. Specifically, we illustrate the use of a large language model (LLM) as a weak learner in a boosting algorithm applied to tabular data. We show that by providing (properly sampled according to the distribution of interest) text descriptions of tabular data samples, LLMs can produce a summary of the samples that serves as a template for classification, and achieves the aim of acting as a weak learner on this task. We incorporate these models into a

boosting approach, which in many settings can leverage the knowledge within the LLM to outperform traditional tree-based boosting. The model outperforms both few-shot learning and occasionally even more involved fine-tuning procedures, particularly for some tasks involving small numbers of data points. The results illustrate the potential for prompt-based LLMs to function not just as few-shot learners themselves, but as components of larger machine learning models.
****
SlotDiffusion: Object-Centric Generative Modeling with Diffusion Models
Ziyi Wu, Jingyu Hu, Wuyue Lu, Igor Gilitschenski, Animesh Garg
Object-centric learning aims to represent visual data with a set of object entities (a.k.a. slots), providing structured representations that enable systematic generalization.Leveraging advanced architectures like Transformers, recent approaches have made significant progress in unsupervised object discovery.In addition, slot-based representations hold great potential for generative modeling, such as controllable image generation and object manipulation in image editing.However, current slot-based methods often produce blurry images and distorted objects, exhibiting poor generative modeling capabilities.In this paper, we focus on improving slot-to-image decoding, a crucial aspect for high-quality visual generation.We introduce SlotDiffusion -- an object-centric Latent Diffusion Model (LDM) designed for both image and video data.Thanks to the powerful modeling capacity of LDMs, SlotDiffusion surpasses previous slot models in unsupervised object segmentation and visual generation across six datasets.Furthermore, our learned object features can be utilized by existing object-centric dynamics models, improving video prediction quality and downstream temporal reasoning tasks.Finally, we demonstrate the scalability of SlotDiffusion to unconstrained real-world datasets such as PASCAL VOC and COCO, when integrated with self-supervised pre-trained image encoders.
****
ZoomTrack: Target-aware Non-uniform Resizing for Efficient Visual Tracking
Yutong Kou, Jin Gao, Bing Li, Gang Wang, Weiming Hu, Yizheng Wang, Liang Li
Recently, the transformer has enabled the speed-oriented trackers to approach state-of-the-art (SOTA) performance with high-speed thanks to the smaller input size or the lighter feature extraction backbone, though they still substantially lag behind their corresponding performance-oriented versions. In this paper, we demonstrate that it is possible to narrow or even close this gap while achieving high tracking speed based on the smaller input size. To this end, we non-uniformly resize the cropped image to have a smaller input size while the resolution of the area where the target is more likely to appear is higher and vice versa. This enables us to solve the dilemma of attending to a larger visual field while retaining more raw information for the target despite a smaller input size. Our formulation for the non-uniform resizing can be efficiently solved through quadratic programming (QP) and naturally integrated into most of the crop-based local trackers. Comprehensive experiments on five challenging datasets based on two kinds of transformer trackers, \ie, OSTrack and TransT, demonstrate consistent improvements over them. In particular, applying our method to the speed-oriented version of OSTrack even outperforms its performance-oriented counterpart by 0.6\% AUC on TNL2K, while running 50\% faster and saving over 55\% MACs. Codes and models are available at https://github.com/Kou-99/ZoomTrack.
****
Improving neural network representations using human similarity judgments
Lukas Muttenthaler, Lorenz Linhardt, Jonas Dippel, Robert A. Vandermeulen, Katherine Hermann, Andrew Lampinen, Simon Kornblith
Deep neural networks have reached human-level performance on many computer vision tasks. However, the objectives used to train these networks enforce only that similar images are embedded at similar locations in the representation space, and do not directly constrain the global structure of the resulting space. Here, we explore the impact of supervising this global structure by linearly aligning it with human similarity judgments. We find that a naive approach leads to large changes in local representational structure that harm downstream performance. Thus, we propose a novel method that aligns the global structure of representation

s while preserving their local structure. This global-local transform considerably improves accuracy across a variety of few-shot learning and anomaly detection tasks. Our results indicate that human visual representations are globally organized in a way that facilitates learning from few examples, and incorporating this global structure into neural network representations improves performance on downstream tasks.
****
Hard Prompts Made Easy: Gradient-Based Discrete Optimization for Prompt Tuning and Discovery
Yuxin Wen, Neel Jain, John Kirchenbauer, Micah Goldblum, Jonas Geiping, Tom Goldstein
The strength of modern generative models lies in their ability to be controlled through prompts. Hard prompts comprise interpretable words and tokens, and are typically hand-crafted by humans. Soft prompts, on the other hand, consist of continuous feature vectors. These can be discovered using powerful optimization methods, but they cannot be easily edited, re-used across models, or plugged into a text-based interface. We describe an easy-to-use approach to automatically optimize hard text prompts through efficient gradient-based optimization. Our approach can be readily applied to text-to-image and text-only applications alike. This method allows API users to easily generate, discover, and mix and match image concepts without prior knowledge of how to prompt the model. Furthermore, using our method, we can bypass token-level content filters imposed by Midjourney by optimizing through the open-sourced text encoder.
****
Bilevel Coreset Selection in Continual Learning: A New Formulation and Algorithm
Jie Hao, Kaiyi Ji, Mingrui Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
HyP-NeRF: Learning Improved NeRF Priors using a HyperNetwork
Bipasha Sen, Gaurav Singh, Aditya Agarwal, Rohith Agaram, Madhava Krishna, Srinath Sridhar
Neural Radiance Fields (NeRF) have become an increasingly popular representation to capture high-quality appearance and shape of scenes and objects. However, learning generalizable NeRF priors over categories of scenes or objects has been challenging due to the high dimensionality of network weight space. To address the limitations of existing work on generalization, multi-view consistency and to improve quality, we propose HyP-NeRF, a latent conditioning method for learning generalizable category-level NeRF priors using hypernetworks. Rather than using hypernetworks to estimate only the weights of a NeRF, we estimate both the weights and the multi-resolution hash encodings resulting in significant quality gains. To improve quality even further, we incorporate a denoise and finetune strategy that denoises images rendered from NeRFs estimated by the hypernetwork and finetunes it while retaining multiview consistency. These improvements enable us to use HyP-NeRF as a generalizable prior for multiple downstream tasks including NeRF reconstruction from single-view or cluttered scenes and text-to-NeRF. We provide qualitative comparisons and evaluate HyP-NeRF on three tasks: generalization, compression, and retrieval, demonstrating our state-of-the-art results.
****
MultiVENT: Multilingual Videos of Events and Aligned Natural Text
Kate Sanders, David Etter, Reno Kriz, Benjamin Van Durme
Everyday news coverage has shifted from traditional broadcasts towards a wide range of presentation formats such as first-hand, unedited video footage. Datasets that reflect the diverse array of multimodal, multilingual news sources available online could be used to teach models to benefit from this shift, but existing news video datasets focus on traditional news broadcasts produced for English-speaking audiences. We address this limitation by constructing MultiVENT, a dataset of multilingual, event-centric videos grounded in text documents across five

target languages. MultiVENT includes both news broadcast videos and non-professi onal event footage, which we use to analyze the state of online news videos and how they can be leveraged to build robust, factually accurate models. Finally, w e provide a model for complex, multilingual video retrieval to serve as a basel ine for information retrieval using MultiVENT.
****
GEO-Bench: Toward Foundation Models for Earth Monitoring
Alexandre Lacoste, Nils Lehmann, Pau Rodriguez, Evan Sherwin, Hannah Kerner, Bjö rn Lütjens, Jeremy Irvin, David Dao, Hamed Alemohammad, Alexandre Drouin, Mehmet Gunturkun, Gabriel Huang, David Vazquez, Dava Newman, Yoshua Bengio, Stefano Er mon, Xiaoxiang Zhu
Recent progress in self-supervision has shown that pre-training large neural net works on vast amounts of unsupervised data can lead to substantial increases in generalization to downstream tasks. Such models, recently coined foundation mode ls, have been transformational to the field of natural language processing.Varia nts have also been proposed for image data, but their applicability to remote se nsing tasks is limited.To stimulate the development of foundation models for Ear th monitoring, we propose a benchmark comprised of six classification and six se gmentation tasks, which were carefully curated and adapted to be both relevant t o the field and well-suited for model evaluation. We accompany this benchmark wi th a robust methodology for evaluating models and reporting aggregated results t o enable a reliable assessment of progress. Finally, we report results for 20 ba selines to gain information about the performance of existing models.We believe that this benchmark will be a driver of progress across a variety of Earth monit oring tasks.
****
Gold-YOLO: Efficient Object Detector via Gather-and-Distribute Mechanism
Chengcheng Wang, Wei He, Ying Nie, Jianyuan Guo, Chuanjian Liu, Yunhe Wang, Kai Han
In the past years, YOLO-series models have emerged as the leading approaches in the area of real-time object detection. Many studies pushed up the baseline to a higher level by modifying the architecture, augmenting data and designing new l osses. However, we find previous models still suffer from information fusion pro blem, although Feature Pyramid Network (FPN) and Path Aggregation Network (PANet) have alleviated this. Therefore, this study provides an advanced Gatherand-Dis tribute mechanism (GD) mechanism, which is realized with convolution and self-at tention operations. This new designed model named as Gold-YOLO, which boosts the multi-scale feature fusion capabilities and achieves an ideal balance between l atency and accuracy across all model scales. Additionally, we implement MAE-styl e pretraining in the YOLO-series for the first time, allowing YOLOseries models could be to benefit from unsupervised pretraining. Gold-YOLO-N attains an outsta nding 39.9% AP on the COCO val2017 datasets and 1030 FPS on a T4 GPU, which outp erforms the previous SOTA model YOLOv6-3.0-N with similar FPS by +2.4%. The PyTo rch code is available at https://github.com/huawei-noah/Efficient-Computing/tree /master/Detection/Gold-YOLO, and the MindSpore code is available at https://gite e.com/mindspore/models/tree/master/research/cv/Gold_YOLO.
****
Curriculum Learning for Graph Neural Networks: Which Edges Should We Learn First
Zheng Zhang, Junxiang Wang, Liang Zhao
Graph Neural Networks (GNNs) have achieved great success in representing data wi th dependencies by recursively propagating and aggregating messages along the ed ges. However, edges in real-world graphs often have varying degrees of difficult y, and some edges may even be noisy to the downstream tasks. Therefore, existing GNNs may lead to suboptimal learned representations because they usually treat every edge in the graph equally. On the other hand, Curriculum Learning (CL), wh ich mimics the human learning principle of learning data samples in a meaningful order, has been shown to be effective in improving the generalization ability a nd robustness of representation learners by gradually proceeding from easy to mo re difficult samples during training. Unfortunately, existing CL strategies are designed for independent data samples and cannot trivially generalize to handle

data dependencies. To address these issues, we propose a novel CL strategy to gradually incorporate more edges into training according to their difficulty from easy to hard, where the degree of difficulty is measured by how well the edges are expected given the model training status. We demonstrate the strength of our proposed method in improving the generalization ability and robustness of learned representations through extensive experiments on nine synthetic datasets and nine real-world datasets. The code for our proposed method is available at https://github.com/rollingstonezz/Curriculumlearningfor_GNNs
****
Unified Lower Bounds for Interactive High-dimensional Estimation under Information Constraints
Jayadev Acharya, Clément L Canonne, Ziteng Sun, Himanshu Tyagi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Differentiable Registration of Images and LiDAR Point Clouds with VoxelPoint-to-Pixel Matching
Junsheng Zhou, Baorui Ma, Wenyuan Zhang, Yi Fang, Yu-Shen Liu, Zhizhong Han
Cross-modality registration between 2D images captured by cameras and 3D point clouds from LiDARs is a crucial task in computer vision and robotic. Previous methods estimate 2D-3D correspondences by matching point and pixel patterns learned by neural networks, and use Perspective-n-Points (PnP) to estimate rigid transformation during post-processing. However, these methods struggle to map points and pixels to a shared latent space robustly since points and pixels have very different characteristics with patterns learned in different manners (MLP and CNN), and they also fail to construct supervision directly on the transformation since the PnP is non-differentiable, which leads to unstable registration results. To address these problems, we propose to learn a structured cross-modality latent space to represent pixel features and 3D features via a differentiable probabilistic PnP solver. Specifically, we design a triplet network to learn VoxelPoint-to-Pixel matching, where we represent 3D elements using both voxels and points to learn the cross-modality latent space with pixels. We design both the voxel and pixel branch based on CNNs to operate convolutions on voxels/pixels represented in grids, and integrate an additional point branch to regain the information lost during voxelization. We train our framework end-to-end by imposing supervisions directly on the predicted pose distribution with a probabilistic PnP solver. To explore distinctive patterns of cross-modality features, we design a novel loss with adaptive-weighted optimization for cross-modality feature description. The experimental results on KITTI and nuScenes datasets show significant improvements over the state-of-the-art methods.
****
Ecosystem-level Analysis of Deployed Machine Learning Reveals Homogeneous Outcomes
Connor Toups, Rishi Bommasani, Kathleen Creel, Sarah Bana, Dan Jurafsky, Percy S. Liang
Machine learning is traditionally studied at the model level: researchers measure and improve the accuracy, robustness, bias, efficiency, and other dimensions of specific models. In practice, however, the societal impact of any machine learning model is partially determined by the context into which it is deployed. To capture this, we introduce ecosystem-level analysis: rather than analyzing a single model, we consider the collection of models that are deployed in a given context. For example, ecosystem-level analysis in hiring recognizes that a job candidate’s outcomes are determined not only by a single hiring algorithm or firm but instead by the collective decisions of all the firms to which the candidate applied. Across three modalities (text, images, speech) and 11 datasets, we establish a clear trend: deployed machine learning is prone to systemic failure, meaning some users are exclusively misclassified by all models available. Even when individual models improve at the population level over time, we find these improv

ements rarely reduce the prevalence of systemic failure. Instead, the benefits of these improvements predominantly accrue to individuals who are already correctly classified by other models. In light of these trends, we analyze medical imaging for dermatology, a setting where the costs of systemic failure are especially high. While traditional analyses reveal that both models and humans exhibit racial performance disparities, ecosystem-level analysis reveals new forms of racial disparity in model predictions that do not present in human predictions. These examples demonstrate that ecosystem-level analysis has unique strengths in characterizing the societal impact of machine learning.
****

MVDiffusion: Enabling Holistic Multi-view Image Generation with Correspondence-Aware Diffusion

Shitao Tang, Fuyang Zhang, Jiacheng Chen, Peng Wang, Yasutaka Furukawa

This paper introduces MVDiffusion, a simple yet effective method for generating consistent multi-view images from text prompts given pixel-to-pixel correspondences (e.g., perspective crops from a panorama or multi-view images given depth maps and poses). Unlike prior methods that rely on iterative image warping and inpainting, MVDiffusion simultaneously generates all images with a global awareness, effectively addressing the prevalent error accumulation issue. At its core, MVDiffusion processes perspective images in parallel with a pre-trained text-to-image diffusion model, while integrating novel correspondence-aware attention layers to facilitate cross-view interactions. For panorama generation, while only trained with 10k panoramas, MVDiffusion is able to generate high-resolution photorealistic images for arbitrary texts or extrapolate one perspective image to a 360-degree view. For multi-view depth-to-image generation, MVDiffusion demonstrates state-of-the-art performance for texturing a scene mesh. The project page is at https://mvdiffusion.github.io/.
****

The geometry of hidden representations of large transformer models

Lucrezia Valeriani, Diego Doimo, Francesca Cuturello, Alessandro Laio, Alessio Ansuini, Alberto Cazzaniga

Large transformers are powerful architectures used for self-supervised data analysis across various data types, including protein sequences, images, and text. In these models, the semantic structure of the dataset emerges from a sequence of transformations between one representation and the next. We characterize the geometric and statistical properties of these representations and how they change as we move through the layers.By analyzing the intrinsic dimension (ID) and neighbor composition, we find that the representations evolve similarly in transformers trained on protein language taskand image reconstruction tasks. In the first layers, the data manifold expands, becoming high-dimensional, and then contracts significantly in the intermediate layers. In the last part of the model, the ID remains approximately constant or forms a second shallow peak. We show that the semantic information of the dataset is better expressed at the end of the first peak, and this phenomenon can be observed across many models trained on diverse datasets.Based on our findings, we point out an explicit strategy to identify, without supervision, the layers that maximize semantic content: representations at intermediate layers corresponding to a relative minimum of the ID profile are more suitable for downstream learning tasks.
****

Django: Detecting Trojans in Object Detection Models via Gaussian Focus Calibration

Guangyu Shen, Siyuan Cheng, Guanhong Tao, Kaiyuan Zhang, Yingqi Liu, Shengwei An, Shiqing Ma, Xiangyu Zhang

Object detection models are vulnerable to backdoor or trojan attacks, where an attacker can inject malicious triggers into the model, leading to altered behavior during inference. As a defense mechanism, trigger inversion leverages optimization to reverse-engineer triggers and identify compromised models. While existing trigger inversion methods assume that each instance from the support set is equally affected by the injected trigger, we observe that the poison effect can vary significantly across bounding boxes in object detection models due to its den

se prediction nature, leading to an undesired optimization objective misalignmen t issue for existing trigger reverse-engineering methods. To address this challe nge, we propose the first object detection backdoor detection framework Django (Detecting Trojans in Object Detection Models via Gaussian Focus Calibration). It leverages a dynamic Gaussian weighting scheme that prioritizes more vulnerable victim boxes and assigns appropriate coefficients to calibrate the optimization objective during trigger inversion. In addition, we combine Django with a novel label proposal pre-processing technique to enhance its efficiency. We evaluate D jango on 3 object detection image datasets, 3 model architectures, and 2 types o f attacks, with a total of 168 models. Our experimental results show that Django outperforms 6 state-of-the-art baselines, with up to 38% accuracy improvement a nd 10x reduced overhead. The code is available at https://github.com/PurduePAML/ DJGO.
****
CORNN: Convex optimization of recurrent neural networks for rapid inference of n eural dynamics
Fatih Dinc, Adam Shai, Mark Schnitzer, Hidenori Tanaka
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
A Unified Framework for Rank-based Loss Minimization
Rufeng Xiao, Yuze Ge, Rujun Jiang, Yifan Yan
The empirical loss, commonly referred to as the average loss, is extensively uti lized for training machine learning models. However, in order to address the div erse performance requirements of machine learning models, the use of the rank-ba sed loss is prevalent, replacing the empirical loss in many cases. The rank-base d loss comprises a weighted sum of sorted individual losses, encompassing both c onvex losses like the spectral risk, which includes the empirical risk and condi tional value-at-risk, and nonconvex losses such as the human-aligned risk and th e sum of the ranked range loss. In this paper, we introduce a unified framework for the optimization of the rank-based loss through the utilization of a proxima l alternating direction method of multipliers. We demonstrate the convergence an d convergence rate of the proposed algorithm under mild conditions. Experiments conducted on synthetic and real datasets illustrate the effectiveness and effici ency of the proposed algorithm.
****
LambdaBeam: Neural Program Search with Higher-Order Functions and Lambdas
Kensen Shi, Hanjun Dai, Wen-Ding Li, Kevin Ellis, Charles Sutton
Search is an important technique in program synthesis that allows for adaptive s trategies such as focusing on particular search directions based on execution re sults. Several prior works have demonstrated that neural models are effective at guiding program synthesis searches. However, a common drawback of those approac hes is the inability to handle iterative loops, higher-order functions, or lambd a functions, thus limiting prior neural searches from synthesizing longer and mo re general programs. We address this gap by designing a search algorithm called LambdaBeam that can construct arbitrary lambda functions that compose operations within a given DSL. We create semantic vector representations of the execution behavior of the lambda functions and train a neural policy network to choose whi ch lambdas to construct during search, and pass them as arguments to higher-orde r functions to perform looping computations. Our experiments show that LambdaBea m outperforms neural, symbolic, and LLM-based techniques in an integer list mani pulation domain.
****
HQA-Attack: Toward High Quality Black-Box Hard-Label Adversarial Attack on Text
Han Liu, Zhi Xu, Xiaotong Zhang, Feng Zhang, Fenglong Ma, Hongyang Chen, Hong Yu , Xianchao Zhang
Black-box hard-label adversarial attack on text is a practical and challenging t ask, as the text data space is inherently discrete and non-differentiable, and o

nly the predicted label is accessible. Research on this problem is still in the embryonic stage and only a few methods are available. Nevertheless, existing methods rely on the complex heuristic algorithm or unreliable gradient estimation strategy, which probably fall into the local optimum and inevitably consume numerous queries, thus are difficult to craft satisfactory adversarial examples with high semantic similarity and low perturbation rate in a limited query budget. To alleviate above issues, we propose a simple yet effective framework to generate high quality textual adversarial examples under the black-box hard-label attack scenarios, named HQA-Attack. Specifically, after initializing an adversarial example randomly, HQA-attack first constantly substitutes original words back as many as possible, thus shrinking the perturbation rate. Then it leverages the synonym set of the remaining changed words to further optimize the adversarial example with the direction which can improve the semantic similarity and satisfy the adversarial condition simultaneously. In addition, during the optimizing procedure, it searches a transition synonym word for each changed word, thus avoiding traversing the whole synonym set and reducing the query number to some extent. Extensive experimental results on five text classification datasets, three natural language inference datasets and two real-world APIs have shown that the proposed HQA-Attack method outperforms other strong baselines significantly.
****

Augmentation-Free Dense Contrastive Knowledge Distillation for Efficient Semantic Segmentation
Jiawei Fan, Chao Li, Xiaolong Liu, Meina Song, Anbang Yao
In recent years, knowledge distillation methods based on contrastive learning have achieved promising results on image classification and object detection tasks. However, in this line of research, we note that less attention is paid to semantic segmentation. Existing methods heavily rely on data augmentation and memory buffer, which entail high computational resource demands when applying them to handle semantic segmentation that requires to preserve high-resolution feature maps for making dense pixel-wise predictions. In order to address this problem, we present Augmentation-free Dense Contrastive Knowledge Distillation (Af-DCD), a new contrastive distillation learning paradigm to train compact and accurate deep neural networks for semantic segmentation applications. Af-DCD leverages a masked feature mimicking strategy, and formulates a novel contrastive learning loss via taking advantage of tactful feature partitions across both channel and spatial dimensions, allowing to effectively transfer dense and structured local knowledge learnt by the teacher model to a target student model while maintaining training efficiency. Extensive experiments on five mainstream benchmarks with various teacher-student network pairs demonstrate the effectiveness of our approach. For instance, DeepLabV3-Res18|DeepLabV3-MBV2 model trained by Af-DCD reaches 77.03\%|76.38\% mIOU on Cityscapes dataset when choosing DeepLabV3-Res101 as the teacher, setting new performance records. Besides that, Af-DCD achieves an absolute mIOU improvement of 3.26\%|3.04\%|2.75\%|2.30\%|1.42\% compared with individually trained counterpart on Cityscapes|Pascal VOC|Camvid|ADE20K|COCO-Stuff-164K. Code is available at https://github.com/OSVAI/Af-DCD.
****

On the Need for a Language Describing Distribution Shifts: Illustrations on Tabular Datasets
Jiashuo Liu, Tianyu Wang, Peng Cui, Hongseok Namkoong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Diverse Community Data for Benchmarking Data Privacy Algorithms
Aniruddha Sen, Christine Task, Dhruv Kapur, Gary Howarth, Karan Bhagat
The Collaborative Research Cycle (CRC) is a National Institute of Standards and Technology (NIST) benchmarking program intended to strengthen understanding of tabular data deidentification technologies. Deidentification algorithms are vulnerable to the same bias and privacy issues that impact other data analytics and m

achine learning applications, and it can even amplify those issues by contaminat ing downstream applications. This paper summarizes four CRC contributions: theor etical work on the relationship between diverse populations and challenges for e quitable deidentification; public benchmark data focused on diverse populations and challenging features; a comprehensive open source suite of evaluation metrol ogy for deidentified datasets; and an archive of more than 450 deidentified data samples from a broad range of techniques. The initial set of evaluation results demonstrate the value of the CRC tools for investigations in this field.
****
Coneheads: Hierarchy Aware Attention
Albert Tseng, Tao Yu, Toni Liu, Christopher M. De Sa
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Benchmark of Machine Learning Force Fields for Semiconductor Simulations: Datase ts, Metrics, and Comparative Analysis
Geonu Kim, Byunggook Na, Gunhee Kim, Hyuntae Cho, Seungjin Kang, Hee Sun Lee, Sa erom Choi, Heejae Kim, Seungwon Lee, Yongdeok Kim
As semiconductor devices become miniaturized and their structures become more co mplex, there is a growing need for large-scale atomic-level simulations as a les s costly alternative to the trial-and-error approach during development.Although machine learning force fields (MLFFs) can meet the accuracy and scale requireme nts for such simulations, there are no open-access benchmarks for semiconductor materials.Hence, this study presents a comprehensive benchmark suite that consis ts of two semiconductor material datasets and ten MLFF models with six evaluatio n metrics. We select two important semiconductor thin-film materials silicon nit ride and hafnium oxide, and generate their datasets using computationally expens ive density functional theory simulations under various scenarios at a cost of 2 .6k GPU days.Additionally, we provide a variety of architectures as baselines: d escriptor-based fully connected neural networks and graph neural networks with r otational invariant or equivariant features.We assess not only the accuracy of e nergy and force predictions but also five additional simulation indicators to de termine the practical applicability of MLFF models in molecular dynamics simulat ions.To facilitate further research, our benchmark suite is available at https:/ /github.com/SAITPublic/MLFF-Framework.
****
Vulnerabilities in Video Quality Assessment Models: The Challenge of Adversarial Attacks
Aoxiang Zhang, Yu Ran, Weixuan Tang, Yuan-Gen Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Unsupervised Behavior Extraction via Random Intent Priors
Hao Hu, Yiqin Yang, Jianing Ye, Ziqing Mai, Chongjie Zhang
Reward-free data is abundant and contains rich prior knowledge of human behavior s, but it is not well exploited by offline reinforcement learning (RL) algorithm s. In this paper, we propose UBER, an unsupervised approach to extract useful be haviors from offline reward-free datasets via diversified rewards. UBER assigns different pseudo-rewards sampled from a given prior distribution to different ag ents to extract a diverse set of behaviors, and reuse them as candidate policies to facilitate the learning of new tasks. Perhaps surprisingly, we show that rew ards generated from random neural networks are sufficient to extract diverse and useful behaviors, some even close to expert ones. We provide both empirical and theoretical evidences to justify the use of random priors for the reward functi on. Experiments on multiple benchmarks showcase UBER's ability to learn effectiv e and diverse behavior sets that enhance sample efficiency for online RL, outper

forming existing baselines. By reducing reliance on human supervision, UBER broadens the applicability of RL to real-world scenarios with abundant reward-free data.
****
Deconstructing Data Reconstruction: Multiclass, Weight Decay and General Losses
Gon Buzaglo, Niv Haim, Gilad Yehudai, Gal Vardi, Yakir Oz, Yaniv Nikankin, Michal Irani
Memorization of training data is an active research area, yet our understanding of the inner workings of neural networks is still in its infancy.Recently, Haim et al. 2022 proposed a scheme to reconstruct training samples from multilayer perceptron binary classifiers, effectively demonstrating that a large portion of training samples are encoded in the parameters of such networks.In this work, we extend their findings in several directions, including reconstruction from multiclass and convolutional neural networks. We derive a more general reconstruction scheme which is applicable to a wider range of loss functions such as regression losses. Moreover, we study the various factors that contribute to networks' susceptibility to such reconstruction schemes. Intriguingly, we observe that using weight decay during training increases reconstructability both in terms of quantity and quality. Additionally, we examine the influence of the number of neurons relative to the number of training samples on the reconstructability.Code: https://github.com/gonbuzaglo/decoreco
****
Information Maximizing Curriculum: A Curriculum-Based Approach for Learning Versatile Skills
Denis Blessing, Onur Celik, Xiaogang Jia, Moritz Reuss, Maximilian Li, Rudolf Lioutikov, Gerhard Neumann
Imitation learning uses data for training policies to solve complex tasks. However,when the training data is collected from human demonstrators, it often leadsto multimodal distributions because of the variability in human actions. Mostimitation learning methods rely on a maximum likelihood (ML) objective to learna parameterized policy, but this can result in suboptimal or unsafe behavior dueto the mode-averaging property of the ML objective. In this work, we proposeInformation Maximizing Curriculum, a curriculum-based approach that assignsa weight to each data point and encourages the model to specialize in the data itcan represent, effectively mitigating the mode-averaging problem by allowing themodel to ignore data from modes it cannot represent. To cover all modes and thus,enable versatile behavior, we extend our approach to a mixture of experts (MoE)policy, where each mixture component selects its own subset of the training datafor learning. A novel, maximum entropy-based objective is proposed to achievefull coverage of the dataset, thereby enabling the policy to encompass all modeswithin the data distribution. We demonstrate the effectiveness of our approach oncomplex simulated control tasks using versatile human demonstrations, achievingsuperior performance compared to state-of-the-art methods.
****
Unleash the Potential of Image Branch for Cross-modal 3D Object Detection
Yifan Zhang, Qijian Zhang, Junhui Hou, Yixuan Yuan, Guoliang Xing
To achieve reliable and precise scene understanding, autonomous vehicles typically incorporate multiple sensing modalities to capitalize on their complementary attributes. However, existing cross-modal 3D detectors do not fully utilize the image domain information to address the bottleneck issues of the LiDAR-based detectors. This paper presents a new cross-modal 3D object detector, namely UPIDet, which aims to unleash the potential of the image branch from two aspects. First, UPIDet introduces a new 2D auxiliary task called normalized local coordinate map estimation. This approach enables the learning of local spatial-aware features from the image modality to supplement sparse point clouds. Second, we discover that the representational capability of the point cloud backbone can be enhanced through the gradients backpropagated from the training objectives of the image branch, utilizing a succinct and effective point-to-pixel module. Extensive experiments and ablation studies validate the effectiveness of our method. Notably, we achieved the top rank in the highly competitive cyclist class of the KITTI b

enchmark at the time of submission. The source code is available at https://gith
ub.com/Eaphan/UPIDet.
****
Model Sparsity Can Simplify Machine Unlearning
jinghan jia, Jiancheng Liu, Parikshit Ram, Yuguang Yao, Gaowen Liu, Yang Liu, PR
ANAY SHARMA, Sijia Liu
In response to recent data regulation requirements, machine unlearning (MU) has
emerged as a critical process to remove the influence of specific examples from
a given model. Although exact unlearning can be achieved through complete model
retraining using the remaining dataset, the associated computational costs have
driven the development of efficient, approximate unlearning techniques. Moving b
eyond data-centric MU approaches, our study introduces a novel model-based persp
ective: model sparsification via weight pruning, which is capable of reducing th
e gap between exact unlearning and approximate unlearning. We show in both theor
y and practice that model sparsity can boost the multi-criteria unlearning perfo
rmance of an approximate unlearner, closing the approximation gap, while continu
ing to be efficient. This leads to a new MU paradigm,   termed prune first, the
n unlearn, which infuses a sparse prior to the unlearning process. Building on t
his insight, we also develop a sparsity-aware unlearning method that utilizes sp
arsity regularization to enhance the training process of approximate unlearning.
 Extensive experiments show that our proposals consistently benefit MU in variou
s unlearning scenarios. A notable highlight is the 77% unlearning efficacy gain
of fine-tuning (one of the simplest approximate unlearning methods) when using o
ur proposed sparsity-aware unlearning method. Furthermore, we showcase the pract
ical impact of our proposed MU methods through two specific use cases: defending
 against backdoor attacks, and enhancing transfer learning through source class
removal. These applications demonstrate the versatility and effectiveness of our
 approaches in addressing a variety of machine learning challenges beyond unlear
ning for data privacy. Codes are available at https://github.com/OPTML-Group/Unl
earn-Sparse.
****
IDRNet: Intervention-Driven Relation Network for Semantic Segmentation
Zhenchao Jin, Xiaowei Hu, Lingting Zhu, Luchuan Song, Li Yuan, Lequan Yu
Co-occurrent visual patterns suggest that pixel relation modeling facilitates de
nse prediction tasks, which inspires the development of numerous context modelin
g paradigms, \emph{e.g.}, multi-scale-driven and similarity-driven context schem
es. Despite the impressive results, these existing paradigms often suffer from i
nadequate or ineffective contextual information aggregation due to reliance on l
arge amounts of predetermined priors. To alleviate the issues, we propose a nove
l \textbf{I}ntervention-\textbf{D}riven \textbf{R}elation \textbf{Net}work (\tex
tbf{IDRNet}), which leverages a deletion diagnostics procedure to guide the mode
ling of contextual relations among different pixels. Specifically, we first grou
p pixel-level representations into semantic-level representations with the guida
nce of pseudo labels and further improve the distinguishability of the grouped r
epresentations with a feature enhancement module. Next, a deletion diagnostics p
rocedure is conducted to model relations of these semantic-level representations
 via perceiving the network outputs and the extracted relations are utilized to
guide the semantic-level representations to interact with each other. Finally,
the interacted representations are utilized to augment original pixel-level repr
esentations for final predictions. Extensive experiments are conducted to valida
te the effectiveness of IDRNet quantitatively and qualitatively. Notably, our in
tervention-driven context scheme brings consistent performance improvements to s
tate-of-the-art segmentation frameworks and achieves competitive results on popu
lar benchmark datasets, including ADE20K, COCO-Stuff, PASCAL-Context, LIP, and C
ityscapes.
****
Phase diagram of early training dynamics in deep neural networks: effect of the
learning rate, depth, and width
Dayal Singh Kalra, Maissam Barkeshli
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Neural Algorithmic Reasoning Without Intermediate Supervision
Gleb Rodionov, Liudmila Prokhorenkova
Neural algorithmic reasoning is an emerging area of machine learning focusing on building models that can imitate the execution of classic algorithms, such as sorting, shortest paths, etc. One of the main challenges is to learn algorithms that are able to generalize to out-of-distribution data, in particular with significantly larger input sizes. Recent work on this problem has demonstrated the advantages of learning algorithms step-by-step, giving models access to all intermediate steps of the original algorithm. In this work, we instead focus on learning neural algorithmic reasoning only from the input-output pairs without appealing to the intermediate supervision. We propose simple but effective architectural improvements and also build a self-supervised objective that can regularise intermediate computations of the model without access to the algorithm trajectory. We demonstrate that our approach is competitive to its trajectory-supervised counterpart on tasks from the CLRS Algorithmic Reasoning Benchmark and achieves new state-of-the-art results for several problems, including sorting, where we obtain significant improvements. Thus, learning without intermediate supervision is a promising direction for further research on neural reasoners.
****

On the Powerfulness of Textual Outlier Exposure for Visual OoD Detection
Sangha Park, Jisoo Mok, Dahuin Jung, Saehyung Lee, Sungroh Yoon
Successful detection of Out-of-Distribution (OoD) data is becoming increasingly important to ensure safe deployment of neural networks. One of the main challenges in OoD detection is that neural networks output overconfident predictions on OoD data, make it difficult to determine OoD-ness of data solely based on their predictions. Outlier exposure addresses this issue by introducing an additional loss that encourages low-confidence predictions on OoD data during training. While outlier exposure has shown promising potential in improving OoD detection performance, all previous studies on outlier exposure have been limited to utilizing visual outliers. Drawing inspiration from the recent advancements in vision-language pre-training, this paper venture out to the uncharted territory of textual outlier exposure. First, we uncover the benefits of using textual outliers by replacing real or virtual outliers in the image-domain with textual equivalents. Then, we propose various ways of generating preferable textual outliers. Our extensive experiments demonstrate that generated textual outliers achieve competitive performance on large-scale OoD and hard OoD benchmarks. Furthermore, we conduct empirical analyses of textual outliers to provide primary criteria for designing advantageous textual outliers: near-distribution, descriptiveness, and inclusion of visual semantics.
****

Estimating Propensity for Causality-based Recommendation without Exposure Data
Zhongzhou Liu, Yuan Fang, Min Wu
Causality-based recommendation systems focus on the causal effects of user-item interactions resulting from item exposure (i.e., which items are recommended or exposed to the user), as opposed to conventional correlation-based recommendation. They are gaining popularity due to their multi-sided benefits to users, sellers and platforms alike. However, existing causality-based recommendation methods require additional input in the form of exposure data and/or propensity scores (i.e., the probability of exposure) for training. Such data, crucial for modeling causality in recommendation, are often not available in real-world situations due to technical or privacy constraints. In this paper, we bridge the gap by proposing a new framework, called Propensity Estimation for Causality-based Recommendation (PropCare). It can estimate the propensity and exposure from a more practical setup, where only interaction data are available without any ground truth on exposure or propensity in training and inference. We demonstrate that, by relating the pairwise characteristics between propensity and item popularity, PropC

are enables competitive causality-based recommendation given only the convention
al interaction data. We further present a theoretical analysis on the bias of th
e causal effect under our model estimation. Finally, we empirically evaluate Pr
opCare through both quantitative and qualitative experiments.
****
A Robust Exact Algorithm for the Euclidean Bipartite Matching Problem
Akshaykumar Gattani, Sharath Raghvendra, Pouyan Shirzadian
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Content-based Unrestricted Adversarial Attack
Zhaoyu Chen, Bo Li, Shuang Wu, Kaixun Jiang, Shouhong Ding, Wenqiang Zhang
Unrestricted adversarial attacks typically manipulate the semantic content of an
 image (e.g., color or texture) to create adversarial examples that are both eff
ective and photorealistic, demonstrating their ability to deceive human percepti
on and deep neural networks with stealth and success. However, current works usu
ally sacrifice unrestricted degrees and subjectively select some image content t
o guarantee the photorealism of unrestricted adversarial examples, which limits
its attack performance. To ensure the photorealism of adversarial examples and b
oost attack performance, we propose a novel unrestricted attack framework called
 Content-based Unrestricted Adversarial Attack. By leveraging a low-dimensional
manifold that represents natural images, we map the images onto the manifold and
 optimize them along its adversarial direction. Therefore, within this framework
, we implement Adversarial Content Attack (ACA) based on Stable Diffusion and ca
n generate high transferable unrestricted adversarial examples with various adve
rsarial contents. Extensive experimentation and visualization demonstrate the ef
ficacy of ACA, particularly in surpassing state-of-the-art attacks by an average
 of 13.3-50.4\% and 16.8-48.0\% in normally trained models and defense methods,
respectively.
****
On Dynamic Programming Decompositions of Static Risk Measures in Markov Decision
 Processes
Jia Lin Hau, Erick Delage, Mohammad Ghavamzadeh, Marek Petrik
Optimizing static risk-averse objectives in Markov decision processes is difficu
lt because they do not admit standard dynamic programming equations common in Re
inforcement Learning (RL) algorithms. Dynamic programming decompositions that au
gment the state space with discrete risk levels have recently gained popularity
in the RL community. Prior work has shown that these decompositions are optimal
when the risk level is discretized sufficiently. However, we show that these pop
ular decompositions for Conditional-Value-at-Risk (CVaR) and Entropic-Value-at-R
isk (EVaR) are inherently suboptimal regardless of the discretization level. In
particular, we show that a saddle point property assumed to hold in prior litera
ture may be violated. However, a decomposition does hold for Value-at-Risk and o
ur proof demonstrates how this risk measure differs from CVaR and EVaR. Our find
ings are significant because risk-averse algorithms are used in high-stake envir
onments, making their correctness much more critical.
****
Benchmarking Robustness of Adaptation Methods on Pre-trained Vision-Language Mod
els
Shuo Chen, Jindong Gu, Zhen Han, Yunpu Ma, Philip Torr, Volker Tresp
Various adaptation methods, such as LoRA, prompts, and adapters, have been propo
sed to enhance the performance of pre-trained vision-language models in specific
 domains. As test samples in real-world applications usually differ from adaptat
ion data, the robustness of these adaptation methods against distribution shifts
 are essential. In this study, we assess the robustness of 11 widely-used adapta
tion methods across 4 vision-language datasets under multimodal corruptions. Con
cretely, we introduce 7 benchmark datasets, including 96 visual and 87 textual c
orruptions, to investigate the robustness of different adaptation methods, the i

mpact of available adaptation examples, and the influence of trainable parameter size during adaptation. Our analysis reveals that: 1) Adaptation methods are more sensitive to text corruptions than visual corruptions. 2) Full fine-tuning does not consistently provide the highest robustness; instead, adapters can achieve better robustness with comparable clean performance. 3) Contrary to expectations, our findings indicate that increasing the number of adaptation data and parameters does not guarantee enhanced robustness; instead, it results in even lower robustness. We hope this study could benefit future research in the development of robust multimodal adaptation methods. The benchmark, code, and dataset used in this study can be accessed at https://adarobustness.github.io.
****
Evaluating the Moral Beliefs Encoded in LLMs
Nino Scherrer, Claudia Shi, Amir Feder, David Blei
This paper presents a case study on the design, administration, post-processing, and evaluation of surveys on large language models (LLMs). It comprises two components:(1) A statistical method for eliciting beliefs encoded in LLMs. We introduce statistical measures and evaluation metrics that quantify the probability of an LLM "making a choice", the associated uncertainty, and the consistency of that choice.(2) We apply this method to study what moral beliefs are encoded in different LLMs, especially in ambiguous cases where the right choice is not obvious.We design a large-scale survey comprising 680 high-ambiguity moral scenarios (e.g., "Should I tell a white lie?") and 687 low-ambiguity moral scenarios (e.g., "Should I stop for a pedestrian on the road?"). Each scenario includes a description, two possible actions, and auxiliary labels indicating violated rules (e.g., "do not kill"). We administer the survey to 28 open- and closed-source LLMs.We find that (a) in unambiguous scenarios, most models ``choose" actions that align with commonsense. In ambiguous cases, most models express uncertainty.(b) Some models are uncertain about choosing the commonsense action because their responses are sensitive to the question-wording.(c) Some models reflect clear preferences in ambiguous scenarios. Specifically, closed-source models tend to agree with each other.
****
Enhancing Adversarial Robustness via Score-Based Optimization
Boya Zhang, Weijian Luo, Zhihua Zhang
Adversarial attacks have the potential to mislead deep neural network classifiers by introducing slight perturbations. Developing algorithms that can mitigate the effects of these attacks is crucial for ensuring the safe use of artificial intelligence. Recent studies have suggested that score-based diffusion models are effective in adversarial defenses. However, existing diffusion-based defenses rely on the sequential simulation of the reversed stochastic differential equations of diffusion models, which are computationally inefficient and yield suboptimal results. In this paper, we introduce a novel adversarial defense scheme named ScoreOpt, which optimizes adversarial samples at test-time, towards original clean data in the direction guided by score-based priors. We conduct comprehensive experiments on multiple datasets, including CIFAR10, CIFAR100 and ImageNet. Our experimental results demonstrate that our approach outperforms existing adversarial defenses in terms of both robustness performance and inference speed.
****
Aligning Optimization Trajectories with Diffusion Models for Constrained Design Generation
Giorgio Giannone, Akash Srivastava, Ole Winther, Faez Ahmed
Generative models have significantly influenced both vision and language domains, ushering in innovative multimodal applications. Although these achievements have motivated exploration in scientific and engineering fields, challenges emerge, particularly in constrained settings with limited data where precision is crucial. Traditional engineering optimization methods rooted in physics often surpass generative models in these contexts. To address these challenges, we introduce Diffusion Optimization Models (DOM) and Trajectory Alignment (TA), a learning framework that demonstrates the efficacy of aligning the sampling trajectory of diffusion models with the trajectory derived from physics-based iterative optimiz

ation methods. This alignment ensures that the sampling process remains grounded in the underlying physical principles. This alignment eliminates the need for costly preprocessing, external surrogate models, or extra labeled data, generating feasible and high-performance designs efficiently. We apply our framework to structural topology optimization, a fundamental problem in mechanical design, evaluating its performance on in- and out-of-distribution configurations. Our results demonstrate that TA outperforms state-of-the-art deep generative models on in-distribution configurations and halves the inference computational cost. When coupled with a few steps of optimization, it also improves manufacturability for out-of-distribution conditions. DOM's efficiency and performance improvements significantly expedite design processes and steer them toward optimal and manufacturable outcomes, highlighting the potential of generative models in data-driven design.
****

Optimal cross-learning for contextual bandits with unknown context distributions
Jon Schneider, Julian Zimmert
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Conservative Offline Policy Adaptation in Multi-Agent Games
Chengjie Wu, Pingzhong Tang, Jun Yang, Yujing Hu, Tangjie Lv, Changjie Fan, Chongjie Zhang
Prior research on policy adaptation in multi-agent games has often relied on online interaction with the target agent in training, which can be expensive and impractical in real-world scenarios. Inspired by recent progress in offline reinforcement learn- ing, this paper studies offline policy adaptation, which aims to utilize the target agent’s behavior data to exploit its weakness or enable effective cooperation. We investigate its distinct challenges of distributional shift and risk-free deviation, and propose a novel learning objective, conservative offline adaptation, that optimizes the worst-case performance against any dataset consistent proxy models. We pro- pose an efficient algorithm called Constrained Self-Play (CSP) that incorporates dataset information into regularized policy learning. We prove that CSP learns a near-optimal risk-free offline adaptation policy upon convergence. Empirical results demonstrate that CSP outperforms non-conservative baselines in various environments, including Maze, predator-prey, MuJoCo, and Google Football.
****

Bounding the Invertibility of Privacy-preserving Instance Encoding using Fisher Information
Kiwan Maeng, Chuan Guo, Sanjay Kariyappa, G. Edward Suh
Privacy-preserving instance encoding aims to encode raw data into feature vectors without revealing their privacy-sensitive information. When designed properly, these encodings can be used for downstream ML applications such as training and inference with limited privacy risk. However, the vast majority of existing schemes do not theoretically justify that their encoding is non-invertible, and their privacy-enhancing properties are only validated empirically against a limited set of attacks. In this paper, we propose a theoretically-principled measure for the invertibility of instance encoding based on Fisher information that is broadly applicable to a wide range of popular encoders. We show that dFIL can be used to bound the invertibility of encodings both theoretically and empirically, providing an intuitive interpretation of the privacy of instance encoding.
****

Adjustable Robust Reinforcement Learning for Online 3D Bin Packing
Yuxin Pan, Yize Chen, Fangzhen Lin
Designing effective policies for the online 3D bin packing problem (3D-BPP) has been a long-standing challenge, primarily due to the unpredictable nature of incoming box sequences and stringent physical constraints. While current deep reinforcement learning (DRL) methods for online 3D-BPP have shown promising results

in optimizing average performance over an underlying box sequence distribution, they often fail in real-world settings where some worst-case scenarios can materialize. Standard robust DRL algorithms tend to overly prioritize optimizing the worst-case performance at the expense of performance under normal problem instance distribution. To address these issues, we first introduce a permutation-based attacker to investigate the practical robustness of both DRL-based and heuristic methods proposed for solving online 3D-BPP. Then, we propose an adjustable robust reinforcement learning (AR2L) framework that allows efficient adjustment of robustness weights to achieve the desired balance of the policy's performance in average and worst-case environments. Specifically, we formulate the objective function as a weighted sum of expected and worst-case returns, and derive the lower performance bound by relating to the return under a mixture dynamics. To realize this lower bound, we adopt an iterative procedure that searches for the associated mixture dynamics and improves the corresponding policy. We integrate this procedure into two popular robust adversarial algorithms to develop the exact and approximate AR2L algorithms. Experiments demonstrate that AR2L is versatile in the sense that it improves policy robustness while maintaining an acceptable level of performance for the nominal case.
****
Promises and Pitfalls of Threshold-based Auto-labeling
Harit Vishwakarma, Heguang Lin, Frederic Sala, Ramya Korlakai Vinayak
Creating large-scale high-quality labeled datasets is a major bottleneck in supervised machine learning workflows. Threshold-based auto-labeling (TBAL), where validation data obtained from humans is used to find a confidence threshold above which the data is machine-labeled, reduces reliance on manual annotation. TBAL is emerging as a widely-used solution in practice. Given the long shelf-life and diverse usage of the resulting datasets, understanding when the data obtained by such auto-labeling systems can be relied on is crucial. This is the first work to analyze TBAL systems and derive sample complexity bounds on the amount of human-labeled validation data required for guaranteeing the quality of machine-labeled data. Our results provide two crucial insights. First, reasonable chunks of unlabeled data can be automatically and accurately labeled by seemingly bad models. Second, a hidden downside of TBAL systems is potentially prohibitive validation data usage. Together, these insights describe the promise and pitfalls of using such systems. We validate our theoretical guarantees with extensive experiments on synthetic and real datasets.
****
CAMEL: Communicative Agents for "Mind" Exploration of Large Language Model Society
Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, Bernard Ghanem
The rapid advancement of chat-based language models has led to remarkable progress in complex task-solving. However, their success heavily relies on human input to guide the conversation, which can be challenging and time-consuming. This paper explores the potential of building scalable techniques to facilitate autonomous cooperation among communicative agents, and provides insight into their “cognitive” processes. To address the challenges of achieving autonomous cooperation, we propose a novel communicative agent framework named role-playing . Our approach involves using inception prompting to guide chat agents toward task completion while maintaining consistency with human intentions. We showcase how role-playing can be used to generate conversational data for studying the behaviors and capabilities of a society of agents, providing a valuable resource for investigating conversational language models. In particular, we conduct comprehensive studies on instruction-following cooperation in multi-agent settings. Our contributions include introducing a novel communicative agent framework, offering a scalable approach for studying the cooperative behaviors and capabilities of multi-agent systems, and open-sourcing our library to support research on communicative agents and beyond: https://github.com/camel-ai/camel.
****
Graph Neural Networks for Road Safety Modeling: Datasets and Evaluations for Accident Analysis

Abhinav Nippani, Dongyue Li, Haotian Ju, Haris Koutsopoulos, Hongyang Zhang
We consider the problem of traffic accident analysis on a road network based on road network connections and traffic volume. Previous works have designed various deep-learning methods using historical records to predict traffic accident occurrences. However, there is a lack of consensus on how accurate existing methods are, and a fundamental issue is the lack of public accident datasets for comprehensive evaluations. This paper constructs a large-scale, unified dataset of traffic accident records from official reports of various states in the US, totaling 9 million records, accompanied by road networks and traffic volume reports. Using this new dataset, we evaluate existing deep-learning methods for predicting the occurrence of accidents on road networks. Our main finding is that graph neural networks such as GraphSAGE can accurately predict the number of accidents on roads with less than 22% mean absolute error (relative to the actual count) and whether an accident will occur or not with over 87% AUROC, averaged over states. We achieve these results by using multitask learning to account for cross-state variabilities (e.g., availability of accident labels) and transfer learning to combine traffic volume with accident prediction. Ablation studies highlight the importance of road graph-structural features, amongst other features. Lastly, we discuss the implications of the analysis and develop a package for easily using our new dataset.
****
SUBP: Soft Uniform Block Pruning for 1$\times$N Sparse CNNs Multithreading Acceleration
JINGYANG XIANG, Siqi Li, Jun Chen, Guang Dai, Shipeng Bai, Yukai Ma, Yong Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Adaptive Linear Estimating Equations
Mufang Ying, Koulik Khamaru, Cun-Hui Zhang
Sequential data collection has emerged as a widely adopted technique for enhancing the efficiency of data gathering processes. Despite its advantages, such data collection mechanism often introduces complexities to the statistical inference procedure. For instance, the ordinary least squares (OLS) estimator in an adaptive linear regression model can exhibit non-normal asymptotic behavior, posing challenges for accurate inference and interpretation. In this paper, we propose a general method for constructing debiased estimator which remedies this issue. It makes use of the idea of adaptive linear estimating equations, and we establish theoretical guarantees of asymptotic normality, supplemented by discussions on achieving near-optimal asymptotic variance. A salient feature of our estimator is that in the context of multi-armed bandits, our estimator retains the non-asymptotic performance of the least squares estimator while obtaining asymptotic normality property. Consequently, this work helps connect two fruitful paradigms of adaptive inference: a) non-asymptotic inference using concentration inequalities and b) asymptotic inference via asymptotic normality.
****
Robust Knowledge Transfer in Tiered Reinforcement Learning
Jiawei Huang, Niao He
In this paper, we study the Tiered Reinforcement Learning setting, a parallel transfer learning framework, where the goal is to transfer knowledge from the low-tier (source) task to the high-tier (target) task to reduce the exploration risk of the latter while solving the two tasks in parallel. Unlike previous work, we do not assume the low-tier and high-tier tasks share the same dynamics or reward functions, and focus on robust knowledge transfer without prior knowledge on the task similarity. We identify a natural and necessary condition called the ``Optimal Value Dominance'' for our objective. Under this condition, we propose novel online learning algorithms such that, for the high-tier task, it can achieve constant regret on partial states depending on the task similarity and retain near-optimal regret when the two tasks are dissimilar, while for the low-tier task

, it can keep near-optimal without making sacrifice. Moreover, we further study the setting with multiple low-tier tasks, and propose a novel transfer source se lection mechanism, which can ensemble the information from all low-tier tasks an d allow provable benefits on a much larger state-action space.
****

Bypassing the Simulator: Near-Optimal Adversarial Linear Contextual Bandits
Haolin Liu, Chen-Yu Wei, Julian Zimmert
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

GenEval: An object-focused framework for evaluating text-to-image alignment
Dhruba Ghosh, Hannaneh Hajishirzi, Ludwig Schmidt
Recent breakthroughs in diffusion models, multimodal pretraining, and efficient finetuning have led to an explosion of text-to-image generative models. Given hu man evaluation is expensive and difficult to scale, automated methods are critic al for evaluating the increasingly large number of new models. However, most cur rent automated evaluation metrics like FID or CLIPScore only offer a distributio n-level measure of image quality or image-text alignment, and are unsuited for f ine-grained or instance-level analysis. In this paper, we introduce GenEval, an object-focused framework to evaluate compositional image properties such as obje ct co-occurrence, position, count, and color. We show that current object detect ion models can be leveraged to evaluate text-to-image models on a variety of gen eration tasks with strong human agreement, and that other discriminative vision models can be linked to this pipeline to further verify properties like object c olor. We then evaluate several open-source text-to-image models and analyze thei r relative reasoning capabilities on our benchmark. We find that recent models d emonstrate significant improvement on these tasks, though they are still lacking in complex capabilities such as spatial relations and attribute binding. Finall y, we demonstrate how GenEval might be used to help discover existing failure mo des, in order to inform development of the next generation of text-to-image mode ls. Our code to run the GenEval framework will be made publicly available at htt ps://github.com/djghosh13/geneval.
****

Generalization in the Face of Adaptivity: A Bayesian Perspective
Moshe Shenfeld, Katrina Ligett
Repeated use of a data sample via adaptively chosen queries can rapidly lead to overfitting, wherein the empirical evaluation of queries on the sample significa ntly deviates from their mean with respect to the underlying data distribution. It turns out that simple noise addition algorithms suffice to prevent this issue , and differential privacy-based analysis of these algorithms shows that they ca n handle an asymptotically optimal number of queries. However, differential pri vacy's worst-case nature entails scaling such noise to the range of the queries even for highly-concentrated queries, or introducing more complex algorithms.In this paper, we prove that straightforward noise-addition algorithms already prov ide variance-dependent guarantees that also extend to unbounded queries. This im provement stems from a novel characterization that illuminates the core problem of adaptive data analysis. We show that the harm of adaptivity results from the covariance between the new query and a Bayes factor-based measure of how much in formation about the data sample was encoded in the responses given to past queri es. We then leverage this characterization to introduce a new data-dependent sta bility notion that can bound this covariance.
****

Convergence of Adam Under Relaxed Assumptions
Haochuan Li, Alexander Rakhlin, Ali Jadbabaie
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.

****
On the Convergence of Encoder-only Shallow Transformers
Yongtao Wu, Fanghui Liu, Grigorios Chrysos, Volkan Cevher
In this paper, we aim to build the global convergence theory of encoder-only shallow Transformers under a realistic setting from the perspective of architectures, initialization, and scaling under a finite width regime. The difficulty lies in how to tackle the softmax in self-attention mechanism, the core ingredient of Transformer. In particular, we diagnose the scaling scheme, carefully tackle the input/output of softmax, and prove that quadratic overparameterization is sufficient for global convergence of our shallow Transformers under commonly-used He/LeCun initialization in practice. Besides, neural tangent kernel (NTK) based analysis is also given, which facilitates a comprehensive comparison. Our theory demonstrates the separation on the importance of different scaling schemes and initialization. We believe our results can pave the way for a better understanding of modern Transformers, particularly on training dynamics.
****
SoundCam: A Dataset for Finding Humans Using Room Acoustics
Mason Wang, Samuel Clarke, Jui-Hsien Wang, Ruohan Gao, Jiajun Wu
A room’s acoustic properties are a product of the room’s geometry, the objects within the room, and their specific positions. A room’s acoustic properties can be characterized by its impulse response (RIR) between a source and listener location, or roughly inferred from recordings of natural signals present in the room. Variations in the positions of objects in a room can effect measurable changes in the room’s acoustic properties, as characterized by the RIR. Existing datasets of RIRs either do not systematically vary positions of objects in an environment, or they consist of only simulated RIRs. We present SoundCam, the largest dataset of unique RIRs from in-the-wild rooms publicly released to date. It includes 5,000 10-channel real-world measurements of room impulse responses and 2,000 10-channel recordings of music in three different rooms, including a controlled acoustic lab, an in-the-wild living room, and a conference room, with different humans in positions throughout each room. We show that these measurements can be used for interesting tasks, such as detecting and identifying humans, and tracking their positions.
****
Accelerated On-Device Forward Neural Network Training with Module-Wise Descending Asynchronism
Xiaohan Zhao, Hualin Zhang, Zhouyuan Huo, Bin Gu
On-device learning faces memory constraints when optimizing or fine-tuning on edge devices with limited resources. Current techniques for training deep models on edge devices rely heavily on backpropagation. However, its high memory usage calls for a reassessment of its dominance.In this paper, we propose forward gradient descent (FGD) as a potential solution to overcome the memory capacity limitation in on-device learning. However, FGD's dependencies across layers hinder parallel computation and can lead to inefficient resource utilization.To mitigate this limitation, we propose AsyncFGD, an asynchronous framework that decouples dependencies, utilizes module-wise stale parameters, and maximizes parallel computation. We demonstrate its convergence to critical points through rigorous theoretical analysis.Empirical evaluations conducted on NVIDIA's AGX Orin, a popular embedded device, show that AsyncFGD reduces memory consumption and enhances hardware efficiency, offering a novel approach to on-device learning.
****
Optimal Parameter and Neuron Pruning for Out-of-Distribution Detection
Chao Chen, Zhihang Fu, Kai Liu, Ze Chen, Mingyuan Tao, Jieping Ye
For a machine learning model deployed in real world scenarios, the ability of detecting out-of-distribution (OOD) samples is indispensable and challenging. Most existing OOD detection methods focused on exploring advanced training skills or training-free tricks to prevent the model from yielding overconfident confidence score for unknown samples. The training-based methods require expensive training cost and rely on OOD samples which are not always available, while most training-free methods can not efficiently utilize the prior information from the trai

ning data. In this work, we propose an \textbf{O}ptimal \textbf{P}arameter and \textbf{N}euron \textbf{P}runing (\textbf{OPNP}) approach, which aims to identify and remove those parameters and neurons that lead to over-fitting. The main method is divided into two steps. In the first step, we evaluate the sensitivity of the model parameters and neurons by averaging gradients over all training samples. In the second step, the parameters and neurons with exceptionally large or close to zero sensitivities are removed for prediction. Our proposal is training-free, compatible with other post-hoc methods, and exploring the information from all training data. Extensive experiments are performed on multiple OOD detection tasks and model architectures, showing that our proposed OPNP consistently outperforms the existing methods by a large margin.
****
Unbalanced Low-rank Optimal Transport Solvers
Meyer Scetbon, Michal Klein, Giovanni Palla, Marco Cuturi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Geodesic Multi-Modal Mixup for Robust Fine-Tuning
Changdae Oh, Junhyuk So, Hoyoon Byun, YongTaek Lim, Minchul Shin, Jong-June Jeon, Kyungwoo Song
Pre-trained multi-modal models, such as CLIP, provide transferable embeddings and show promising results in diverse applications. However, the analysis of learned multi-modal embeddings is relatively unexplored, and the embedding transferability can be improved. In this work, we observe that CLIP holds separated embedding subspaces for two different modalities, and then we investigate it through the lens of \textit{uniformity-alignment} to measure the quality of learned representation. Both theoretically and empirically, we show that CLIP retains poor uniformity and alignment even after fine-tuning. Such a lack of alignment and uniformity might restrict the transferability and robustness of embeddings. To this end, we devise a new fine-tuning method for robust representation equipping better alignment and uniformity. First, we propose a \textit{Geodesic Multi-Modal Mixup} that mixes the embeddings of image and text to generate hard negative samples on the hypersphere. Then, we fine-tune the model on hard negatives as well as original negatives and positives with contrastive loss. Based on the theoretical analysis about hardness guarantee and limiting behavior, we justify the use of our method. Extensive experiments on retrieval, calibration, few- or zero-shot classification (under distribution shift), embedding arithmetic, and image captioning further show that our method provides transferable representations, enabling robust model adaptation on diverse tasks.
****
Scissorhands: Exploiting the Persistence of Importance Hypothesis for LLM KV Cache Compression at Test Time
Zichang Liu, Aditya Desai, Fangshuo Liao, Weitao Wang, Victor Xie, Zhaozhuo Xu, Anastasios Kyrillidis, Anshumali Shrivastava
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Asymmetric Certified Robustness via Feature-Convex Neural Networks
Samuel Pfrommer, Brendon Anderson, Julien Piet, Somayeh Sojoudi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Unified Fast Gradient Clipping Framework for DP-SGD
Weiwei Kong, Andres Munoz Medina

A well-known numerical bottleneck in the differentially-private stochastic gradient descent (DP-SGD) algorithm is the computation of the gradient norm for each example in a large input batch. When the loss function in DP-SGD is consists of an intermediate linear operation, existing methods in the literature have proposed decompositions of gradients that are amenable to fast norm computations. In this paper, we present a framework that generalizes the above approach to arbitrary (possibly nonlinear) intermediate operations. Moreover, we show that for certain operations, such as fully-connected and embedding layer computations, further improvements to the runtime and storage costs of existing decompositions can be deduced using certain components of our framework. Finally, preliminary numerical experiments are given to demonstrate the substantial effects of the aforementioned improvements.
****

Offline Multi-Agent Reinforcement Learning with Implicit Global-to-Local Value Regularization

Xiangsen Wang, Haoran Xu, Yinan Zheng, Xianyuan Zhan

Offline reinforcement learning (RL) has received considerable attention in recent years due to its attractive capability of learning policies from offline datasets without environmental interactions. Despite some success in the single-agent setting, offline multi-agent RL (MARL) remains to be a challenge. The large joint state-action space and the coupled multi-agent behaviors pose extra complexities for offline policy optimization. Most existing offline MARL studies simply apply offline data-related regularizations on individual agents, without fully considering the multi-agent system at the global level. In this work, we present OMIGA, a new offline multi-agent RL algorithm with implicit global-to-local value regularization. OMIGA provides a principled framework to convert global-level value regularization into equivalent implicit local value regularizations and simultaneously enables in-sample learning, thus elegantly bridging multi-agent value decomposition and policy learning with offline regularizations. Based on comprehensive experiments on the offline multi-agent MuJoCo and StarCraft II micro-management tasks, we show that OMIGA achieves superior performance over the state-of-the-art offline MARL methods in almost all tasks.
****

Benchmarking Large Language Models on CMExam - A comprehensive Chinese Medical Exam Dataset

Junling Liu, Peilin Zhou, Yining Hua, Dading Chong, Zhongyu Tian, Andrew Liu, Helin Wang, Chenyu You, Zhenhua Guo, LEI ZHU, Michael Lingzhi Li

Recent advancements in large language models (LLMs) have transformed the field of question answering (QA). However, evaluating LLMs in the medical field is challenging due to the lack of standardized and comprehensive datasets. To address this gap, we introduce CMExam, sourced from the Chinese National Medical Licensing Examination. CMExam consists of 60K+ multiple-choice questions for standardized and objective evaluations, as well as solution explanations for model reasoning evaluation in an open-ended manner. For in-depth analyses of LLMs, we invited medical professionals to label five additional question-wise annotations, including disease groups, clinical departments, medical disciplines, areas of competency, and question difficulty levels. Alongside the dataset, we further conducted thorough experiments with representative LLMs and QA algorithms on CMExam. The results show that GPT-4 had the best accuracy of 61.6% and a weighted F1 score of 0.617. These results highlight a great disparity when compared to human accuracy, which stood at 71.6%. For explanation tasks, while LLMs could generate relevant reasoning and demonstrate improved performance after finetuning, they fall short of a desired standard, indicating ample room for improvement. To the best of our knowledge, CMExam is the first Chinese medical exam dataset to provide comprehensive medical annotations. The experiments and findings of LLM evaluation also provide valuable insights into the challenges and potential solutions in developing Chinese medical QA systems and LLM evaluation pipelines.
****

A Logic for Expressing Log-Precision Transformers

William Merrill, Ashish Sabharwal

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Universal Prompt Tuning for Graph Neural Networks

Taoran Fang, Yunchao Zhang, YANG YANG, Chunping Wang, Lei Chen

In recent years, prompt tuning has sparked a research surge in adapting pre-trained models. Unlike the unified pre-training strategy employed in the language field, the graph field exhibits diverse pre-training strategies, posing challenges in designing appropriate prompt-based tuning methods for graph neural networks. While some pioneering work has devised specialized prompting functions for models that employ edge prediction as their pre-training tasks, these methods are limited to specific pre-trained GNN models and lack broader applicability. In this paper, we introduce a universal prompt-based tuning method called Graph Prompt Feature (GPF) for pre-trained GNN models under any pre-training strategy. GPF operates on the input graph's feature space and can theoretically achieve an equivalent effect to any form of prompting function. Consequently, we no longer need to illustrate the prompting function corresponding to each pre-training strategy explicitly. Instead, we employ GPF to obtain the prompted graph for the downstream task in an adaptive manner. We provide rigorous derivations to demonstrate the universality of GPF and make guarantee of its effectiveness. The experimental results under various pre-training strategies indicate that our method performs better than fine-tuning, with an average improvement of about 1.4% in full-shot scenarios and about 3.2% in few-shot scenarios. Moreover, our method significantly outperforms existing specialized prompt-based tuning methods when applied to models utilizing the pre-training strategy they specialize in. These numerous advantages position our method as a compelling alternative to fine-tuning for downstream adaptations.
****

Stochastic Approximation Approaches to Group Distributionally Robust Optimization

Lijun Zhang, Peng Zhao, Zhen-Hua Zhuang, Tianbao Yang, Zhi-Hua Zhou

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Learning Efficient Surrogate Dynamic Models with Graph Spline Networks

Chuanbo Hua, Federico Berto, Michael Poli, Stefano Massaroli, Jinkyoo Park

While complex simulations of physical systems have been widely used in engineering and scientific computing, lowering their often prohibitive computational requirements has only recently been tackled by deep learning approaches. In this paper, we present GraphSplineNets, a novel deep-learning method to speed up the forecasting of physical systems by reducing the grid size and number of iteration steps of deep surrogate models. Our method uses two differentiable orthogonal spline collocation methods to efficiently predict response at any location in time and space. Additionally, we introduce an adaptive collocation strategy in space to prioritize sampling from the most important regions. GraphSplineNets improve the accuracy-speedup tradeoff in forecasting various dynamical systems with increasing complexity, including the heat equation, damped wave propagation, Navier-Stokes equations, and real-world ocean currents in both regular and irregular domains.
****

Efficient Adaptation of Large Vision Transformer via Adapter Re-Composing

Wei Dong, Dawei Yan, Zhijun Lin, Peng Wang

The advent of high-capacity pre-trained models has revolutionized problem-solving in computer vision, shifting the focus from training task-specific models to adapting pre-trained models. Consequently, effectively adapting large pre-trained models to downstream tasks in an efficient manner has become a prominent resear

ch area. Existing solutions primarily concentrate on designing lightweight adapters and their interaction with pre-trained models, with the goal of minimizing the number of parameters requiring updates. In this study, we propose a novel Adapter Re-Composing (ARC) strategy that addresses efficient pre-trained model adaptation from a fresh perspective. Our approach considers the reusability of adaptation parameters and introduces a parameter-sharing scheme. Specifically, we leverage symmetric down-/up-projections to construct bottleneck operations, which are shared across layers. By learning low-dimensional re-scaling coefficients, we can effectively re-compose layer-adaptive adapters. This parameter-sharing strategy in adapter design allows us to further reduce the number of new parameters while maintaining satisfactory performance, thereby offering a promising approach to compress the adaptation cost. We conduct experiments on 24 downstream image classification tasks using various Vision Transformer variants to evaluate our method. The results demonstrate that our approach achieves compelling transfer learning performance with a reduced parameter count. Our code is available at https://github.com/DavidYanAnDe/ARC.
****

Hardness of Low Rank Approximation of Entrywise Transformed Matrix Products
Tamas Sarlos, Xingyou Song, David Woodruff, Richard Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Efficient Training of Energy-Based Models Using Jarzynski Equality
Davide Carbone, Mengjian Hua, Simon Coste, Eric Vanden-Eijnden
Energy-based models (EBMs) are generative models inspired by statistical physics with a wide range of applications in unsupervised learning. Their performance is well measured by the cross-entropy (CE) of the model distribution relative to the data distribution. Using the CE as the objective for training is however challenging because the computation of its gradient with respect to the model parameters requires sampling the model distribution. Here we show how results for nonequilibrium thermodynamics based on Jarzynski equality together with tools from sequential Monte-Carlo sampling can be used to perform this computation efficiently and avoid the uncontrolled approximations made using the standard contrastive divergence algorithm. Specifically, we introduce a modification of the unadjusted Langevin algorithm (ULA) in which each walker acquires a weight that enables the estimation of the gradient of the cross-entropy at any step during GD, thereby bypassing sampling biases induced by slow mixing of ULA. We illustrate these results with numerical experiments on Gaussian mixture distributions as well as the MNIST and CIFAR-10 datasets. We show that the proposed approach outperforms methods based on the contrastive divergence algorithm in all the considered situations.
****

High Precision Causal Model Evaluation with Conditional Randomization
Chao Ma, Cheng Zhang
The gold standard for causal model evaluation involves comparing model predictions with true effects estimated from randomized controlled trials (RCT). However, RCTs are not always feasible or ethical to perform. In contrast, conditionally randomized experiments based on inverse probability weighting (IPW) offer a more realistic approach but may suffer from high estimation variance. To tackle this challenge and enhance causal model evaluation in real-world conditional randomization settings, we introduce a novel low-variance estimator for causal error, dubbed as the pairs estimator. By applying the same IPW estimator to both the model and true experimental effects, our estimator effectively cancels out the variance due to IPW and achieves a smaller asymptotic variance. Empirical studies demonstrate the improved of our estimator, highlighting its potential on achieving near-RCT performance. Our method offers a simple yet powerful solution to evaluate causal inference models in conditional randomization settings without complicated modification of the IPW estimator itself, paving the way for more robust a

nd reliable model assessments.
****
Reducing Blackwell and Average Optimality to Discounted MDPs via the Blackwell Discount Factor
Julien Grand-Clément, Marek Petrik
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Marginal Density Ratio for Off-Policy Evaluation in Contextual Bandits
Muhammad Faaiz Taufiq, Arnaud Doucet, Rob Cornish, Jean-Francois Ton
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SPAE: Semantic Pyramid AutoEncoder for Multimodal Generation with Frozen LLMs
Lijun Yu, Yong Cheng, Zhiruo Wang, Vivek Kumar, Wolfgang Macherey, Yanping Huang, David Ross, Irfan Essa, Yonatan Bisk, Ming-Hsuan Yang, Kevin P. Murphy, Alexander Hauptmann, Lu Jiang
In this work, we introduce Semantic Pyramid AutoEncoder (SPAE) for enabling frozen LLMs to perform both understanding and generation tasks involving non-linguistic modalities such as images or videos. SPAE converts between raw pixels and interpretable lexical tokens (or words) extracted from the LLM's vocabulary. The resulting tokens capture both the rich semantic meaning and the fine-grained details needed for visual reconstruction, effectively translating the visual content into a language comprehensible to the LLM, and empowering it to perform a wide array of multimodal tasks. Our approach is validated through in-context learning experiments with frozen PaLM 2 and GPT 3.5 on a diverse set of image understanding and generation tasks.Our method marks the first successful attempt to enable a frozen LLM to generate image content while surpassing state-of-the-art performance in image understanding tasks, under the same setting, by over 25%.
****
Energy-based learning algorithms for analog computing: a comparative study
Benjamin Scellier, Maxence Ernoult, Jack Kendall, Suhas Kumar
Energy-based learning algorithms have recently gained a surge of interest due to their compatibility with analog (post-digital) hardware. Existing algorithms include contrastive learning (CL), equilibrium propagation (EP) and coupled learning (CpL), all consisting in contrasting two states, and differing in the type of perturbation used to obtain the second state from the first one. However, these algorithms have never been explicitly compared on equal footing with same models and datasets, making it difficult to assess their scalability and decide which one to select in practice. In this work, we carry out a comparison of seven learning algorithms, namely CL and different variants of EP and CpL depending on the signs of the perturbations. Specifically, using these learning algorithms, we train deep convolutional Hopfield networks (DCHNs) on five vision tasks (MNIST, F-MNIST, SVHN, CIFAR-10 and CIFAR-100). We find that, while all algorithms yield comparable performance on MNIST, important differences in performance arise as the difficulty of the task increases. Our key findings reveal that negative perturbations are better than positive ones, and highlight the centered variant of EP (which uses two perturbations of opposite sign) as the best-performing algorithm. We also endorse these findings with theoretical arguments. Additionally, we establish new SOTA results with DCHNs on all five datasets, both in performance and speed. In particular, our DCHN simulations are 13.5 times faster with respect to Laborieux et al. (2021), which we achieve thanks to the use of a novel energy minimisation algorithm based on asynchronous updates, combined with reduced precision (16 bits).
****
Distribution Learnability and Robustness

Shai Ben-David, Alex Bie, Gautam Kamath, Tosca Lechner
We examine the relationship between learnability and robust learnability for the problem of distribution learning.We show that learnability implies robust learnability if the adversary can only perform additive contamination (and consequently, under Huber contamination), but not if the adversary is allowed to perform subtractive contamination. Thus, contrary to other learning settings (e.g., PAC learning of function classes), realizable learnability does not imply agnostic learnability. We also explore related implications in the context of compression schemes and differentially private learnability.
****
Behavior Alignment via Reward Function Optimization
Dhawal Gupta, Yash Chandak, Scott Jordan, Philip S. Thomas, Bruno C. da Silva
Designing reward functions for efficiently guiding reinforcement learning (RL) agents toward specific behaviors is a complex task.This is challenging since it requires the identification of reward structures that are not sparse and that avoid inadvertently inducing undesirable behaviors. Naively modifying the reward structure to offer denser and more frequent feedback can lead to unintended outcomes and promote behaviors that are not aligned with the designer's intended goal. Although potential-based reward shaping is often suggested as a remedy, we systematically investigate settings where deploying it often significantly impairs performance. To address these issues, we introduce a new framework that uses a bi-level objective to learn \emph{behavior alignment reward functions}. These functions integrate auxiliary rewards reflecting a designer's heuristics and domain knowledge with the environment's primary rewards. Our approach automatically determines the most effective way to blend these types of feedback, thereby enhancing robustness against heuristic reward misspecification. Remarkably, it can also adapt an agent's policy optimization process to mitigate suboptimalities resulting from limitations and biases inherent in the underlying RL algorithms. We evaluate our method's efficacy on a diverse set of tasks, from small-scale experiments to high-dimensional control challenges. We investigate heuristic auxiliary rewards of varying quality---some of which are beneficial and others detrimental to the learning process. Our results show that our framework offers a robust and principled way to integrate designer-specified heuristics. It not only addresses key shortcomings of existing approaches but also consistently leads to high-performing solutions, even when given misaligned or poorly-specified auxiliary reward functions.
****
Tuning Multi-mode Token-level Prompt Alignment across Modalities
Dongsheng Wang, Miaoge Li, Xinyang Liu, MingSheng Xu, Bo Chen, Hanwang Zhang
Advancements in prompt tuning of vision-language models have underscored their potential in enhancing open-world visual concept comprehension. However, prior works only primarily focus on single-mode (only one prompt for each modality) and holistic level (image or sentence) semantic alignment, which fails to capture the sample diversity, leading to sub-optimal prompt discovery. To address the limitation, we propose a multi-mode token-level tuning framework that leverages the optimal transportation to learn and align a set of prompt tokens across modalities. Specifically, we rely on two essential factors: 1) multi-mode prompts discovery, which guarantees diverse semantic representations, and 2) token-level alignment, which helps explore fine-grained similarity. Consequently, the similarity can be calculated as a hierarchical transportation problem between the modality-specific sets. Extensive experiments on popular image recognition benchmarks show the superior generalization and few-shot abilities of our approach. The qualitative analysis demonstrates that the learned prompt tokens have the ability to capture diverse visual concepts.
****
Censored Sampling of Diffusion Models Using 3 Minutes of Human Feedback
TaeHo Yoon, Kibeom Myoung, Keon Lee, Jaewoong Cho, Albert No, Ernest Ryu
Diffusion models have recently shown remarkable success in high-quality image generation. Sometimes, however, a pre-trained diffusion model exhibits partial misalignment in the sense that the model can generate good images, but it sometimes

outputs undesirable images. If so, we simply need to prevent the generation of the bad images, and we call this task censoring. In this work, we present censored generation with a pre-trained diffusion model using a reward model trained on minimal human feedback. We show that censoring can be accomplished with extreme human feedback efficiency and that labels generated with a mere few minutes of human feedback are sufficient.
****
Timewarp: Transferable Acceleration of Molecular Dynamics by Learning Time-Coarsened Dynamics
Leon Klein, Andrew Foong, Tor Fjelde, Bruno Mlodozeniec, Marc Brockschmidt, Sebastian Nowozin, Frank Noe, Ryota Tomioka
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Harnessing Hard Mixed Samples with Decoupled Regularizer
Zicheng Liu, Siyuan Li, Ge Wang, Lirong Wu, Cheng Tan, Stan Z. Li
Mixup is an efficient data augmentation approach that improves the generalization of neural networks by smoothing the decision boundary with mixed data. Recently, dynamic mixup methods have improved previous \textit{static} policies effectively (e.g., linear interpolation) by maximizing target-related salient regions in mixed samples, but excessive additional time costs are not acceptable. These additional computational overheads mainly come from optimizing the mixed samples according to the mixed labels. However, we found that the extra optimizing step may be redundant because label-mismatched mixed samples are informative hard mixed samples for deep models to localize discriminative features. In this paper, we thus are not trying to propose a more complicated dynamic mixup policy but rather an efficient mixup objective function with decoupled regularizer, named decoupled mixup (DM). The primary effect is that DM can adaptively utilize those hard mixed samples to mine discriminative features without losing the original smoothness of mixup. As a result, DM enables static mixup methods to achieve comparable or even exceed the performance of dynamic methods without any extra computation. This also leads to an interesting objective design problem for mixup training that we need to focus on both smoothing the decision boundaries and identifying discriminative features. Extensive experiments on supervised and semi-supervised learning benchmarks across seven datasets validate the effectiveness of DM.
****
The Utility of “Even if” Semifactual Explanation to Optimise Positive Outcomes
Eoin Kenny, Weipeng Huang
When users receive either a positive or negative outcome from an automated system, Explainable AI (XAI) has almost exclusively focused on how to mutate negative outcomes into positive ones by crossing a decision boundary using counterfactuals (e.g., "If you earn 2k more, we will accept your loan application"). Here, we instead focus on positive outcomes, and take the novel step of using XAI to optimise them (e.g., "Even if you wish to half your down-payment, we will still accept your loan application"). Explanations such as these that employ "even if..." reasoning, and do not cross a decision boundary, are known as semifactuals. To instantiate semifactuals in this context, we introduce the concept of Gain (i.e., how much a user stands to benefit from the explanation), and consider the first causal formalisation of semifactuals. Tests on benchmark datasets show our algorithms are better at maximising gain compared to prior work, and that causality is important in the process. Most importantly however, a user study supports our main hypothesis by showing people find semifactual explanations more useful than counterfactuals when they receive the positive outcome of a loan acceptance.
****
VLATTACK: Multimodal Adversarial Attacks on Vision-Language Tasks via Pre-trained Models
Ziyi Yin, Muchao Ye, Tianrong Zhang, Tianyu Du, Jinguo Zhu, Han Liu, Jinghui Chen, Ting Wang, Fenglong Ma

Vision-Language (VL) pre-trained models have shown their superiority on many multimodal tasks. However, the adversarial robustness of such models has not been fully explored. Existing approaches mainly focus on exploring the adversarial robustness under the white-box setting, which is unrealistic. In this paper, we aim to investigate a new yet practical task to craft image and text perturbations using pre-trained VL models to attack black-box fine-tuned models on different downstream tasks. Towards this end, we propose VLATTACK to generate adversarial samples by fusing perturbations of images and texts from both single-modal and multi-modal levels. At the single-modal level, we propose a new block-wise similarity attack (BSA) strategy to learn image perturbations for disrupting universal representations. Besides, we adopt an existing text attack strategy to generate text perturbations independent of the image-modal attack. At the multi-modal level, we design a novel iterative cross-search attack (ICSA) method to update adversarial image-text pairs periodically, starting with the outputs from the single-modal level. We conduct extensive experiments to attack three widely-used VL pretrained models for six tasks on eight datasets. Experimental results show that the proposed VLATTACK framework achieves the highest attack success rates on all tasks compared with state-of-the-art baselines, which reveals a significant blind spot in the deployment of pre-trained VL models.
****

Mode Connectivity in Auction Design

Christoph Hertrich, Yixin Tao, László A. Végh

Optimal auction design is a fundamental problem in algorithmic game theory. This problem is notoriously difficult already in very simple settings. Recent work in differentiable economics showed that neural networks can efficiently learn known optimal auction mechanisms and discover interesting new ones. In an attempt to theoretically justify their empirical success, we focus on one of the first such networks, RochetNet, and a generalized version for affine maximizer auctions. We prove that they satisfy mode connectivity, i.e., locally optimal solutions are connected by a simple, piecewise linear path such that every solution on the path is almost as good as one of the two local optima. Mode connectivity has been recently investigated as an intriguing empirical and theoretically justifiable property of neural networks used for prediction problems. Our results give the first such analysis in the context of differentiable economics, where neural networks are used directly for solving non-convex optimization problems.
****

Katakomba: Tools and Benchmarks for Data-Driven NetHack

Vladislav Kurenkov, Alexander Nikulin, Denis Tarasov, Sergey Kolesnikov

NetHack is known as the frontier of reinforcement learning research where learning-based methods still need to catch up to rule-based solutions. One of the promising directions for a breakthrough is using pre-collected datasets similar to recent developments in robotics, recommender systems, and more under the umbrella of offline reinforcement learning (ORL). Recently, a large-scale NetHack dataset was released; while it was a necessary step forward, it has yet to gain wide adoption in the ORL community. In this work, we argue that there are three major obstacles for adoption: tool-wise, implementation-wise, and benchmark-wise. To address them, we develop an open-source library that provides workflow fundamentals familiar to the ORL community: pre-defined D4RL-style tasks, uncluttered baseline implementations, and reliable evaluation tools with accompanying configs and logs synced to the cloud.
****

Deep Neural Collapse Is Provably Optimal for the Deep Unconstrained Features Model

Peter Súkeník, Marco Mondelli, Christoph H. Lampert

Neural collapse (NC) refers to the surprising structure of the last layer of deep neural networks in the terminal phase of gradient descent training. Recently, an increasing amount of experimental evidence has pointed to the propagation of NC to earlier layers of neural networks. However, while the NC in the last layer is well studied theoretically, much less is known about its multi-layered counterpart - deep neural collapse (DNC). In particular, existing work focuses either

on linear layers or only on the last two layers at the price of an extra assump tion. Our work fills this gap by generalizing the established analytical framewo rk for NC - the unconstrained features model - to multiple non-linear layers. Ou r key technical contribution is to show that, in a deep unconstrained features m odel, the unique global optimum for binary classification exhibits all the prope rties typical of DNC. This explains the existing experimental evidence of DNC. W e also empirically show that (i) by optimizing deep unconstrained features model s via gradient descent, the resulting solution agrees well with our theory, and (ii) trained networks recover the unconstrained features suitable for the occurr ence of DNC, thus supporting the validity of this modeling principle.
****
IEBins: Iterative Elastic Bins for Monocular Depth Estimation
Shuwei Shao, Zhongcai Pei, Xingming Wu, Zhong Liu, Weihai Chen, Zhengguo Li
Monocular depth estimation (MDE) is a fundamental topic of geometric computer vi sion and a core technique for many downstream applications. Recently, several me thods reframe the MDE as a classification-regression problem where a linear comb ination of probabilistic distribution and bin centers is used to predict depth. In this paper, we propose a novel concept of iterative elastic bins (IEBins) for the classification-regression-based MDE. The proposed IEBins aims to search for high-quality depth by progressively optimizing the search range, which involves multiple stages and each stage performs a finer-grained depth search in the tar get bin on top of its previous stage. To alleviate the possible error accumulati on during the iterative process, we utilize a novel elastic target bin to replac e the original target bin, the width of which is adjusted elastically based on t he depth uncertainty. Furthermore, we develop a dedicated framework composed of a feature extractor and an iterative optimizer that has powerful temporal contex t modeling capabilities benefiting from the GRU-based architecture. Extensive ex periments on the KITTI, NYU-Depth-v2 and SUN RGB-D datasets demonstrate that the proposed method surpasses prior state-of-the-art competitors. The source code i s publicly available at https://github.com/ShuweiShao/IEBins.
****
Fine-Tuning Language Models with Just Forward Passes
Sadhika Malladi, Tianyu Gao, Eshaan Nichani, Alex Damian, Jason D. Lee, Danqi Ch en, Sanjeev Arora
Fine-tuning language models (LMs) has yielded success on diverse downstream task s, but as LMs grow in size, backpropagation requires a prohibitively large amoun t of memory. Zeroth-order (ZO) methods can in principle estimate gradients using only two forward passes but are theorized to be catastrophically slow for optim izing large models. In this work, we propose a memory-efficient zerothorder opti mizer (MeZO), adapting the classical ZO-SGD method to operate in-place, thereby fine-tuning LMs with the same memory footprint as inference. For example, with a single A100 80GB GPU, MeZO can train a 30-billion parameter model, whereas fine -tuning with backpropagation can train only a 2.7B LM with the same budget. We c onduct comprehensive experiments across model types (masked and autoregressive L Ms), model scales (up to 66B), and downstream tasks (classification, multiple-ch oice, and generation). Our results demonstrate that (1) MeZO significantly outpe rforms in-context learning and linear probing; (2) MeZO achieves comparable perf ormance to fine-tuning with backpropagation across multiple tasks, with up to 12 × memory reduction and up to 2× GPU-hour reduction in our implementation; (3) Me ZO is compatible with both full-parameter and parameter-efficient tuning techniq ues such as LoRA and prefix tuning; (4) MeZO can effectively optimize non-differ entiable objectives (e.g., maximizing accuracy or F1). We support our empirical findings with theoretical insights, highlighting how adequate pre-training and t ask prompts enable MeZO to fine-tune huge models, despite classical ZO analyses suggesting otherwise.
****
HEDNet: A Hierarchical Encoder-Decoder Network for 3D Object Detection in Point Clouds
Gang Zhang, Chen Junnan, Guohuan Gao, Jianmin Li, Xiaolin Hu
3D object detection in point clouds is important for autonomous driving systems.

A primary challenge in 3D object detection stems from the sparse distribution o f points within the 3D scene. Existing high-performance methods typically employ 3D sparse convolutional neural networks with small kernels to extract features. To reduce computational costs, these methods resort to submanifold sparse convo lutions, which prevent the information exchange among spatially disconnected fea tures. Some recent approaches have attempted to address this problem by introduc ing large-kernel convolutions or self-attention mechanisms, but they either achi eve limited accuracy improvements or incur excessive computational costs. We pro pose HEDNet, a hierarchical encoder-decoder network for 3D object detection, whi ch leverages encoder-decoder blocks to capture long-range dependencies among fea tures in the spatial space, particularly for large and distant objects. We condu cted extensive experiments on the Waymo Open and nuScenes datasets. HEDNet achie ved superior detection accuracy on both datasets than previous state-of-the-art methods with competitive efficiency. The code is available at https://github.com /zhanggang001/HEDNet.
****
FedGame: A Game-Theoretic Defense against Backdoor Attacks in Federated Learning
Jinyuan Jia, Zhuowen Yuan, Dinuka Sahabandu, Luyao Niu, Arezoo Rajabi, Bhaskar R amasubramanian, Bo Li, Radha Poovendran
Federated learning (FL) provides a distributed training paradigm where multiple clients can jointly train a global model without sharing their local data. Howev er, recent studies have shown that FL offers an additional surface for backdoor attacks. For instance, an attacker can compromise a subset of clients and thus c orrupt the global model to misclassify an input with a backdoor trigger as the a dversarial target. Existing defenses for FL against backdoor attacks usually det ect and exclude the corrupted information from the compromised clients based on a static attacker model. However, such defenses are inadequate against dynamic a ttackers who strategically adapt their attack strategies. To bridge this gap, we model the strategic interactions between the defender and dynamic attackers as a minimax game. Based on the analysis of the game, we design an interactive defe nse mechanism FedGame. We prove that under mild assumptions, the global model tr ained with FedGame under backdoor attacks is close to that trained without attac ks. Empirically, we compare FedGame with multiple state-of-the-art baselines on several benchmark datasets under various attacks. We show that FedGame can effec tively defend against strategic attackers and achieves significantly higher robu stness than baselines. Our code is available at: https://github.com/AI-secure/Fe dGame.
****
Ensemble-based Deep Reinforcement Learning for Vehicle Routing Problems under Di stribution Shift
YUAN JIANG, Zhiguang Cao, Yaoxin Wu, Wen Song, Jie Zhang
While performing favourably on the independent and identically distributed (i.i. d.) instances, most of the existing neural methods for vehicle routing problems (VRPs) struggle to generalize in the presence of a distribution shift. To tackle this issue, we propose an ensemble-based deep reinforcement learning method for VRPs, which learns a group of diverse sub-policies to cope with various instanc e distributions. In particular, to prevent convergence of the parameters to the same one, we enforce diversity across sub-policies by leveraging Bootstrap with random initialization. Moreover, we also explicitly pursue inequality between su b-policies by exploiting regularization terms during training to further enhance diversity. Experimental results show that our method is able to outperform the state-of-the-art neural baselines on randomly generated instances of various dis tributions, and also generalizes favourably on the benchmark instances from TSPL ib and CVRPLib, which confirmed the effectiveness of the whole method and the re spective designs.
****
Module-wise Training of Neural Networks via the Minimizing Movement Scheme
Skander Karkar, Ibrahim Ayed, Emmanuel de Bézenac, Patrick Gallinari
Greedy layer-wise or module-wise training of neural networks is compelling in co nstrained and on-device settings where memory is limited, as it circumvents a nu

mber of problems of end-to-end back-propagation. However, it suffers from a stag nation problem, whereby early layers overfit and deeper layers stop increasing t he test accuracy after a certain depth. We propose to solve this issue by introd ucing a simple module-wise regularization inspired by the minimizing movement sc heme for gradient flows in distribution space. We call the method TRGL for Trans port Regularized Greedy Learning and study it theoretically, proving that it lea ds to greedy modules that are regular and that progressively solve the task. Exp erimentally, we show improved accuracy of module-wise training of various archit ectures such as ResNets, Transformers and VGG, when our regularization is added, superior to that of other module-wise training methods and often to end-to-end training, with as much as 60% less memory usage.
****
POMDP Planning for Object Search in Partially Unknown Environment
Yongbo Chen, Hanna Kurniawati
Efficiently searching for target objects in complex environments that contain va rious types of furniture, such as shelves, tables, and beds, is crucial for mobi le robots, but it poses significant challenges due to various factors such as lo calization errors, limited field of view, and visual occlusion. To address this problem, we propose a Partially Observable Markov Decision Process (POMDP) formu lation with a growing state space for object search in a 3D region. We solve thi s POMDP by carefully designing a perception module and developing a planning alg orithm, called Growing Partially Observable Monte-Carlo Planning (GPOMCP), based on online Monte-Carlo tree search and belief tree reuse with a novel upper conf idence bound. We have demonstrated that belief tree reuse is reasonable and achi eves good performance when the belief differences are limited. Additionally, we introduce a guessed target object with an updating grid world to guide the searc h in the information-less and reward-less cases, like the absence of any detecte d objects. We tested our approach using Gazebo simulations on four scenarios of target finding in a realistic indoor living environment with the Fetch robot sim ulator. Compared to the baseline approaches, which are based on POMCP, our resul ts indicate that our approach enables the robot to find the target object with a higher success rate faster while using the same computational requirements.
****
On the Statistical Consistency of Risk-Sensitive Bayesian Decision-Making
Prateek Jaiswal, Harsha Honnappa, Vinayak Rao
We study data-driven decision-making problems in the Bayesian framework, where t he expectation in the Bayes risk is replaced by a risk-sensitive entropic risk m easure with respect to the posterior distribution. We focus on problems where ca lculating the posterior distribution is intractable, a typical situation in mode rn applications with large datasets and complex data generating models. We lever age a dual representation of the entropic risk measure to introduce a novel risk -sensitive variational Bayesian (RSVB) framework for jointly computing a risk-se nsitive posterior approximation and the corresponding decision rule. Our general framework includes \textit{loss-calibrated} VB (Lacoste-Julien et al. [2011]) as a special case. We also study the impact of these computational approximation s on the predictive performance of the inferred decision rules. We compute the c onvergence rates of the RSVB approximate posterior and the corresponding optimal value. We illustrate our theoretical findings in parametric and nonparametric s ettings with the help of three examples.
****
Does Graph Distillation See Like Vision Dataset Counterpart?
Beining Yang, Kai Wang, Qingyun Sun, Cheng Ji, Xingcheng Fu, Hao Tang, Yang You, Jianxin Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Online Learning under Adversarial Nonlinear Constraints
Pavel Kolev, Georg Martius, Michael Muehlebach

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Sample-Efficient and Safe Deep Reinforcement Learning via Reset Deep Ensemble Agents
Woojun Kim, Yongjae Shin, Jongeui Park, Youngchul Sung
Deep reinforcement learning (RL) has achieved remarkable success in solving complex tasks through its integration with deep neural networks (DNNs) as function approximators. However, the reliance on DNNs has introduced a new challenge called primacy bias, whereby these function approximators tend to prioritize early experiences, leading to overfitting. To alleviate this bias, a reset method has been proposed, which involves periodic resets of a portion or the entirety of a deep RL agent while preserving the replay buffer. However, the use of this method can result in performance collapses after executing the reset, raising concerns from the perspective of safe RL and regret minimization. In this paper, we propose a novel reset-based method that leverages deep ensemble learning to address the limitations of the vanilla reset method and enhance sample efficiency. The effectiveness of the proposed method is validated through various experiments including those in the domain of safe RL. Numerical results demonstrate its potential for real-world applications requiring high sample efficiency and safety considerations.
****
Im-Promptu: In-Context Composition from Image Prompts
Bhishma Dedhia, Michael Chang, Jake Snell, Tom Griffiths, Niraj Jha
Large language models are few-shot learners that can solve diverse tasks from a handful of demonstrations. This implicit understanding of tasks suggests that the attention mechanisms over word tokens may play a role in analogical reasoning. In this work, we investigate whether analogical reasoning can enable in-context composition over composable elements of visual stimuli. First, we introduce a suite of three benchmarks to test the generalization properties of a visual in-context learner. We formalize the notion of an analogy-based in-context learner and use it to design a meta-learning framework called Im-Promptu. Whereas the requisite token granularity for language is well established, the appropriate compositional granularity for enabling in-context generalization in visual stimuli is usually unspecified. To this end, we use Im-Promptu to train multiple agents with different levels of compositionality, including vector representations, patch representations, and object slots. Our experiments reveal tradeoffs between extrapolation abilities and the degree of compositionality, with non-compositional representations extending learned composition rules to unseen domains but performing poorly on combinatorial tasks. Patch-based representations require patches to contain entire objects for robust extrapolation. At the same time, object-centric tokenizers coupled with a cross-attention module generate consistent and high-fidelity solutions, with these inductive biases being particularly crucial for compositional generalization. Lastly, we demonstrate a use case of Im-Promptu as an intuitive programming interface for image generation.
****
Sequential Predictive Two-Sample and Independence Testing
Aleksandr Podkopaev, Aaditya Ramdas
We study the problems of sequential nonparametric two-sample and independence testing. Sequential tests process data online and allow using observed data to decide whether to stop and reject the null hypothesis or to collect more data, while maintaining type I error control. We build upon the principle of (nonparametric) testing by betting, where a gambler places bets on future observations and their wealth measures evidence against the null hypothesis. While recently developed kernel-based betting strategies often work well on simple distributions, selecting a suitable kernel for high-dimensional or structured data, such as images, is often nontrivial. To address this drawback, we design prediction-based betting strategies that rely on the following fact: if a sequentially updated predict

or starts to consistently determine (a) which distribution an instance is drawn from, or (b) whether an instance is drawn from the joint distribution or the product of the marginal distributions (the latter produced by external randomization), it provides evidence against the two-sample or independence nulls respectively. We empirically demonstrate the superiority of our tests over kernel-based approaches under structured settings. Our tests can be applied beyond the case of independent and identically distributed data, remaining valid and powerful even when the data distribution drifts over time.
****

Towards Symmetry-Aware Generation of Periodic Materials
Youzhi Luo, Chengkai Liu, Shuiwang Ji
We consider the problem of generating periodic materials with deep models. While symmetry-aware molecule generation has been studied extensively, periodic materials possess different symmetries, which have not been completely captured by existing methods.In this work, we propose SyMat, a novel material generation approach that can capture physical symmetries of periodic material structures. SyMat generates atom types and lattices of materials through generating atom type sets, lattice lengths and lattice angles with a variational auto-encoder model. In addition, SyMat employs a score-based diffusion model to generate atom coordinates of materials, in which a novel symmetry-aware probabilistic model is used in the coordinate diffusion process. We show that SyMat is theoretically invariant to all symmetry transformations on materials and demonstrate that SyMat achieves promising performance on random generation and property optimization tasks. Our code is publicly available as part of the AIRS library (https://github.com/divelab/AIRS).
****

DICES Dataset: Diversity in Conversational AI Evaluation for Safety
Lora Aroyo, Alex Taylor, Mark Díaz, Christopher Homan, Alicia Parrish, Gregory Serapio-García, Vinodkumar Prabhakaran, Ding Wang
Machine learning approaches often require training and evaluation datasets with a clear separation between positive and negative examples. This requirement overly simplifies the natural subjectivity present in many tasks, and obscures the inherent diversity in human perceptions and opinions about many content items. Preserving the variance in content and diversity in human perceptions in datasets is often quite expensive and laborious. This is especially troubling when building safety datasets for conversational AI systems, as safety is socio-culturally situated in this context. To demonstrate this crucial aspect of conversational AI safety, and to facilitate in-depth model performance analyses, we introduce the DICES (Diversity In Conversational AI Evaluation for Safety) dataset that contains fine-grained demographics information about raters, high replication of ratings per item to ensure statistical power for analyses, and encodes rater votes as distributions across different demographics to allow for in-depth explorations of different aggregation strategies. The DICES dataset enables the observation and measurement of variance, ambiguity, and diversity in the context of safety for conversational AI. We further describe a set of metrics that show how rater diversity influences safety perception across different geographic regions, ethnicity groups, age groups, and genders. The goal of the DICES dataset is to be used as a shared resource and benchmark that respects diverse perspectives during safety evaluation of conversational AI systems.
****

SALSA VERDE: a machine learning attack on LWE with sparse small secrets
Cathy Li, Emily Wenger, Zeyuan Allen-Zhu, Francois Charton, Kristin E. Lauter
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Designing Robust Transformers using Robust Kernel Density Estimation
Xing Han, Tongzheng Ren, Tan Nguyen, Khai Nguyen, Joydeep Ghosh, Nhat Ho
Transformer-based architectures have recently exhibited remarkable successes acr

oss different domains beyond just powering large language models. However, existing approaches typically focus on predictive accuracy and computational cost, largely ignoring certain other practical issues such as robustness to contaminated samples. In this paper, by re-interpreting the self-attention mechanism as a non-parametric kernel density estimator, we adapt classical robust kernel density estimation methods to develop novel classes of transformers that are resistant to adversarial attacks and data contamination. We first propose methods that down-weight outliers in RKHS when computing the self-attention operations. We empirically show that these methods produce improved performance over existing state-of-the-art methods, particularly on image data under adversarial attacks. Then we leverage the median-of-means principle to obtain another efficient approach that results in noticeably enhanced performance and robustness on language modeling and time series classification tasks. Our methods can be combined with existing transformers to augment their robust properties, thus promising to impact a wide variety of applications.
****
Benchmarking Distribution Shift in Tabular Data with TableShift
Josh Gardner, Zoran Popovic, Ludwig Schmidt
Robustness to distribution shift has become a growing concern for text and image models as they transition from research subjects to deployment in the real world. However, high-quality benchmarks for distribution shift in tabular machine learning tasks are still lacking despite the widespread real-world use of tabular data and differences in the models used for tabular data in comparison to text and images. As a consequence, the robustness of tabular models to distribution shift is poorly understood. To address this issue, we introduce TableShift, a distribution shift benchmark for tabular data. TableShift contains 15 binary classification tasks in total, each with an associated shift, and includes a diverse set of data sources, prediction targets, and distribution shifts. The benchmark covers domains including finance, education, public policy, healthcare, and civic participation, and is accessible using only a few lines of Python code via the TableShift API. We conduct a large-scale study comparing several state-of-the-art tabular data models alongside robust learning and domain generalization methods on the benchmark tasks. Our study demonstrates (1) a linear trend between in-distribution (ID) and out-of-distribution (OOD) accuracy; (2) domain robustness methods can reduce shift gaps but at the cost of reduced ID accuracy; (3) a strong relationship between shift gap (difference between ID and OOD performance) and shifts in the label distribution. The benchmark data, Python package, model implementations, and more information about TableShift are available at https://github.com/mlfoundations/tableshift and https://tableshift.org .
****
Weakly Supervised 3D Open-vocabulary Segmentation
Kunhao Liu, Fangneng Zhan, Jiahui Zhang, MUYU XU, Yingchen Yu, Abdulmotaleb El Saddik, Christian Theobalt, Eric Xing, Shijian Lu
Open-vocabulary segmentation of 3D scenes is a fundamental function of human perception and thus a crucial objective in computer vision research. However, this task is heavily impeded by the lack of large-scale and diverse 3D open-vocabulary segmentation datasets for training robust and generalizable models. Distilling knowledge from pre-trained 2D open-vocabulary segmentation models helps but it compromises the open-vocabulary feature as the 2D models are mostly finetuned with close-vocabulary datasets. We tackle the challenges in 3D open-vocabulary segmentation by exploiting pre-trained foundation models CLIP and DINO in a weakly supervised manner. Specifically, given only the open-vocabulary text descriptions of the objects in a scene, we distill the open-vocabulary multimodal knowledge and object reasoning capability of CLIP and DINO into a neural radiance field (NeRF), which effectively lifts 2D features into view-consistent 3D segmentation. A notable aspect of our approach is that it does not require any manual segmentation annotations for either the foundation models or the distillation process. Extensive experiments show that our method even outperforms fully supervised models trained with segmentation annotations in certain scenes, suggesting that 3D open-vocabulary segmentation can be effectively learned from 2D images and text-

image pairs. Code is available at https://github.com/Kunhao-Liu/3D-OVS.
****

Better Private Linear Regression Through Better Private Feature Selection
Travis Dick, Jennifer Gillenwater, Matthew Joseph
Existing work on differentially private linear regression typically assumes that end users can precisely set data bounds or algorithmic hyperparameters. End users often struggle to meet these requirements without directly examining the data (and violating privacy). Recent work has attempted to develop solutions that shift these burdens from users to algorithms, but they struggle to provide utility as the feature dimension grows. This work extends these algorithms to higher-dimensional problems by introducing a differentially private feature selection method based on Kendall rank correlation. We prove a utility guarantee for the setting where features are normally distributed and conduct experiments across 25 datasets. We find that adding this private feature selection step before regression significantly broadens the applicability of ``plug-and-play'' private linear regression algorithms at little additional cost to privacy, computation, or decision-making by the end user.
****

Geometric Neural Diffusion Processes
Emile Mathieu, Vincent Dutordoir, Michael Hutchinson, Valentin De Bortoli, Yee Whye Teh, Richard Turner
Denoising diffusion models have proven to be a flexible and effective paradigm for generative modelling.Their recent extension to infinite dimensional Euclidean spaces has allowed for the modelling of stochastic processes.However, many problems in the natural sciences incorporate symmetries and involve data living in non-Euclidean spaces.In this work, we extend the framework of diffusion models to incorporate a series of geometric priors in infinite-dimension modelling.We do so by a) constructing a noising process which admits, as limiting distribution, a geometric Gaussian process that transforms under the symmetry group of interest, and b) approximating the score with a neural network that is equivariant w.r.t. this group.We show that with these conditions, the generative functional model admits the same symmetry.We demonstrate scalability and capacity of the model, using a novel Langevin-based conditional sampler, to fit complex scalar and vector fields, with Euclidean and spherical codomain, on synthetic and real-world weather data.
****

Online Adaptive Policy Selection in Time-Varying Systems: No-Regret via Contractive Perturbations
Yiheng Lin, James A. Preiss, Emile Anand, Yingying Li, Yisong Yue, Adam Wierman
We study online adaptive policy selection in systems with time-varying costs and dynamics. We develop the Gradient-based Adaptive Policy Selection (GAPS) algorithm together with a general analytical framework for online policy selection via online optimization. Under our proposed notion of contractive policy classes, we show that GAPS approximates the behavior of an ideal online gradient descent algorithm on the policy parameters while requiring less information and computation. When convexity holds, our algorithm is the first to achieve optimal policy regret. When convexity does not hold, we provide the first local regret bound for online policy selection. Our numerical experiments show that GAPS can adapt to changing environments more quickly than existing benchmarks.
****

IMP-MARL: a Suite of Environments for Large-scale Infrastructure Management Planning via MARL
Pascal Leroy, Pablo G. Morato, Jonathan Pisane, Athanasios Kolios, Damien Ernst
We introduce IMP-MARL, an open-source suite of multi-agent reinforcement learning (MARL) environments for large-scale Infrastructure Management Planning (IMP), offering a platform for benchmarking the scalability of cooperative MARL methods in real-world engineering applications.In IMP, a multi-component engineering system is subject to a risk of failure due to its components' damage condition.Specifically, each agent plans inspections and repairs for a specific system component, aiming to minimise maintenance costs while cooperating to minimise system f

ailure risk.With IMP-MARL, we release several environments including one related to offshore wind structural systems, in an effort to meet today's needs to improve management strategies to support sustainable and reliable energy systems.Supported by IMP practical engineering environments featuring up to 100 agents, we conduct a benchmark campaign, where the scalability and performance of state-of-the-art cooperative MARL methods are compared against expert-based heuristic policies. The results reveal that centralised training with decentralised execution methods scale better with the number of agents than fully centralised or decentralised RL approaches, while also outperforming expert-based heuristic policies in most IMP environments.Based on our findings, we additionally outline remaining cooperation and scalability challenges that future MARL methods should still address.Through IMP-MARL, we encourage the implementation of new environments and the further development of MARL methods.
****
Learning Multi-agent Behaviors from Distributed and Streaming Demonstrations
Shicheng Liu, Minghui Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CAPro: Webly Supervised Learning with Cross-modality Aligned Prototypes
Yulei Qin, Xingyu Chen, Yunhang Shen, Chaoyou Fu, Yun Gu, Ke Li, Xing Sun, Rongrong Ji
Webly supervised learning has attracted increasing attention for its effectiveness in exploring publicly accessible data at scale without manual annotation. However, most existing methods of learning with web datasets are faced with challenges from label noise, and they have limited assumptions on clean samples under various noise. For instance, web images retrieved with queries of ”tiger cat“ (a cat species) and ”drumstick“ (a musical instrument) are almost dominated by images of tigers and chickens, which exacerbates the challenge of fine-grained visual concept learning. In this case, exploiting both web images and their associated texts is a requisite solution to combat real-world noise. In this paper, we propose Cross-modality Aligned Prototypes (CAPro), a unified prototypical contrastive learning framework to learn visual representations with correct semantics. For one thing, we leverage textual prototypes, which stem from the distinct concept definition of classes, to select clean images by text matching and thus disambiguate the formation of visual prototypes. For another, to handle missing and mismatched noisy texts, we resort to the visual feature space to complete and enhance individual texts and thereafter improve text matching. Such semantically aligned visual prototypes are further polished up with high-quality samples, and engaged in both cluster regularization and noise removal. Besides, we propose collective bootstrapping to encourage smoother and wiser label reference from appearance-similar instances in a manner of dictionary look-up. Extensive experiments on WebVision1k and NUS-WIDE (Web) demonstrate that CAPro well handles realistic noise under both single-label and multi-label scenarios. CAPro achieves new state-of-the-art performance and exhibits robustness to open-set recognition. Codes are available at https://github.com/yuleiqin/capro.
****
Diversify \& Conquer: Outcome-directed Curriculum RL via Out-of-Distribution Disagreement
Daesol Cho, Seungjae Lee, H. Jin Kim
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Distribution-Free Model-Agnostic Regression Calibration via Nonparametric Methods
Shang Liu, Zhongze Cai, Xiaocheng Li

In this paper, we consider the uncertainty quantification problem for regression models. Specifically, we consider an individual calibration objective for characterizing the quantiles of the prediction model. While such an objective is well-motivated from downstream tasks such as newsvendor cost, the existing methods have been largely heuristic and lack of statistical guarantee in terms of individual calibration. We show via simple examples that the existing methods focusing on population-level calibration guarantees such as average calibration or sharpness can lead to harmful and unexpected results. We propose simple nonparametric calibration methods that are agnostic of the underlying prediction model and enjoy both computational efficiency and statistical consistency. Our approach enables a better understanding of the possibility of individual calibration, and we establish matching upper and lower bounds for the calibration error of our proposed methods. Technically, our analysis combines the nonparametric analysis with a covering number argument for parametric analysis, which advances the existing theoretical analyses in the literature of nonparametric density estimation and quantile bandit problems. Importantly, the nonparametric perspective sheds new theoretical insights into regression calibration in terms of the curse of dimensionality and reconciles the existing results on the impossibility of individual calibration. To our knowledge, we make the first effort to reach both individual calibration and finite-sample guarantee with minimal assumptions in terms of conformal prediction. Numerical experiments show the advantage of such a simple approach under various metrics, and also under covariates shift. We hope our work provides a simple benchmark and a starting point of theoretical ground for future research on regression calibration.
****
Synthetic-to-Real Pose Estimation with Geometric Reconstruction
Qiuxia Lin, Kerui Gu, Linlin Yang, Angela Yao
Pose estimation is remarkably successful under supervised learning, but obtaining annotations, especially for new deployments, is costly and time-consuming. This work tackles adapting models trained on synthetic data to real-world target domains with only unlabelled data. A common approach is model fine-tuning with pseudo-labels from the target domain; yet many pseudo-labelling strategies cannot provide sufficient high-quality pose labels. This work proposes a reconstruction-based strategy as a complement to pseudo-labelling for synthetic-to-real domain adaptation. We generate the driving image by geometrically transforming a base image according to the predicted keypoints and enforce a reconstruction loss to refine the predictions. It provides a novel solution to effectively correct confident yet inaccurate keypoint locations through image reconstruction in domain adaptation. Our approach outperforms the previous state-of-the-arts by 8% for PCK on four large-scale hand and human real-world datasets. In particular, we excel on endpoints such as fingertips and head, with 7.2% and 29.9% improvements in PCK.
****
Parallel Spiking Neurons with High Efficiency and Ability to Learn Long-term Dependencies
Wei Fang, Zhaofei Yu, Zhaokun Zhou, Ding Chen, Yanqi Chen, Zhengyu Ma, Timothée Masquelier, Yonghong Tian
Vanilla spiking neurons in Spiking Neural Networks (SNNs) use charge-fire-reset neuronal dynamics, which can only be simulated serially and can hardly learn long-time dependencies. We find that when removing reset, the neuronal dynamics can be reformulated in a non-iterative form and parallelized. By rewriting neuronal dynamics without reset to a general formulation, we propose the Parallel Spiking Neuron (PSN), which generates hidden states that are independent of their predecessors, resulting in parallelizable neuronal dynamics and extremely high simulation speed. The weights of inputs in the PSN are fully connected, which maximizes the utilization of temporal information. To avoid the use of future inputs for step-by-step inference, the weights of the PSN can be masked, resulting in the masked PSN. By sharing weights across time-steps based on the masked PSN, the sliding PSN is proposed to handle sequences of varying lengths. We evaluate the PSN family on simulation speed and temporal/static data classification, and the r

esults show the overwhelming advantage of the PSN family in efficiency and accuracy. To the best of our knowledge, this is the first study about parallelizing spiking neurons and can be a cornerstone for the spiking deep learning research. Our codes are available at https://github.com/fangwei123456/Parallel-Spiking-Neuron.
****
Learning Fine-grained View-Invariant Representations from Unpaired Ego-Exo Videos via Temporal Alignment
Zihui (Sherry) Xue, Kristen Grauman
The egocentric and exocentric viewpoints of a human activity look dramatically different, yet invariant representations to link them are essential for many potential applications in robotics and augmented reality. Prior work is limited to learning view-invariant features from paired synchronized viewpoints. We relax that strong data assumption and propose to learn fine-grained action features that are invariant to the viewpoints by aligning egocentric and exocentric videos in time, even when not captured simultaneously or in the same environment. To this end, we propose AE2, a self-supervised embedding approach with two key designs: (1) an object-centric encoder that explicitly focuses on regions corresponding to hands and active objects; (2) a contrastive-based alignment objective that leverages temporally reversed frames as negative samples. For evaluation, we establish a benchmark for fine-grained video understanding in the ego-exo context, comprising four datasets---including an ego tennis forehand dataset we collected, along with dense per-frame labels we annotated for each dataset. On the four datasets, our AE2 method strongly outperforms prior work in a variety of fine-grained downstream tasks, both in regular and cross-view settings.
****
Training Private Models That Know What They Don’t Know
Stephan Rabanser, Anvith Thudi, Abhradeep Guha Thakurta, Krishnamurthy Dvijotham, Nicolas Papernot
Training reliable deep learning models which avoid making overconfident but incorrect predictions is a longstanding challenge. This challenge is further exacerbated when learning has to be differentially private: protection provided to sensitive data comes at the price of injecting additional randomness into the learning process. In this work, we conduct a thorough empirical investigation of selective classifiers---that can abstain under uncertainty---under a differential privacy constraint. We find that some popular selective prediction approaches are ineffective in a differentially private setting because they increase the risk of privacy leakage. At the same time, we identify that a recent approach that only uses checkpoints produced by an off-the-shelf private learning algorithm stands out as particularly suitable under DP. Further, we show that differential privacy does not just harm utility but also degrades selective classification performance. To analyze this effect across privacy levels, we propose a novel evaluation mechanism which isolates selective prediction performance across model utility levels at full coverage. Our experimental results show that recovering the performance level attainable by non-private models is possible but comes at a considerable coverage cost as the privacy budget decreases.
****
Direct Preference Optimization: Your Language Model is Secretly a Reward Model
Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, Chelsea Finn
While large-scale unsupervised language models (LMs) learn broad world knowledge and some reasoning skills, achieving precise control of their behavior is difficult due to the completely unsupervised nature of their training. Existing methods for gaining such steerability collect human labels of the relative quality of model generations and fine-tune the unsupervised LM to align with these preferences, often with reinforcement learning from human feedback (RLHF). However, RLHF is a complex and often unstable procedure, first fitting a reward model that reflects the human preferences, and then fine-tuning the large unsupervised LM using reinforcement learning to maximize this estimated reward without drifting too far from the original model. In this paper, we leverage a mapping between rewa

rd functions and optimal policies to show that this constrained reward maximization problem can be optimized exactly with a single stage of policy training, essentially solving a classification problem on the human preference data. The resulting algorithm, which we call Direct Preference Optimization (DPO), is stable, performant, and computationally lightweight, eliminating the need for fitting a reward model, sampling from the LM during fine-tuning, or performing significant hyperparameter tuning. Our experiments show that DPO can fine-tune LMs to align with human preferences as well as or better than existing methods. Notably, fine-tuning with DPO exceeds RLHF's ability to control sentiment of generations and improves response quality in summarization and single-turn dialogue while being substantially simpler to implement and train.
****
Diffusion Representation for Asymmetric Kernels via Magnetic Transform
Mingzhen He, FAN He, Ruikai Yang, Xiaolin Huang
As a nonlinear dimension reduction technique, the diffusion map (DM) has been widely used. In DM, kernels play an important role for capturing the nonlinear relationship of data. However, only symmetric kernels can be used now, which prevents the use of DM in directed graphs, trophic networks, and other real-world scenarios where the intrinsic and extrinsic geometries in data are asymmetric. A promising technique is the magnetic transform which converts an asymmetric matrix to a Hermitian one. However, we are facing essential problems, including how diffusion distance could be preserved and how divergence could be avoided during diffusion process. Via theoretical proof, we successfully establish a diffusion representation framework with the magnetic transform, named MagDM. The effectiveness and robustness for dealing data endowed with asymmetric proximity are demonstrated on three synthetic datasets and two trophic networks.
****
Uncovering the Hidden Dynamics of Video Self-supervised Learning under Distribution Shifts
Pritam Sarkar, Ahmad Beirami, Ali Etemad
Video self-supervised learning (VSSL) has made significant progress in recent years. However, the exact behavior and dynamics of these models under different forms of distribution shift are not yet known. In this paper, we comprehensively study the behavior of six popular self-supervised methods (v-SimCLR, v-MoCo, v-BYOL, v-SimSiam, v-DINO, v-MAE) in response to various forms of natural distribution shift, i.e., (i) context shift, (ii) viewpoint shift, (iii) actor shift, (iv) source shift, (v) generalizability to unknown classes (zero-shot), and (vi) open-set recognition. To perform this extensive study, we carefully craft a test bed consisting of 17 in-distribution and out-of-distribution benchmark pairs using available public datasets and a series of evaluation protocols to stress-test the different methods under the intended shifts. Our study uncovers a series of intriguing findings and interesting behaviors of VSSL methods. For instance, we observe that while video models generally struggle with context shifts, v-MAE and supervised learning exhibit more robustness. Moreover, our study shows that v-MAE is a strong temporal learner, whereas contrastive methods, v-SimCLR and v-MoCo, exhibit strong performances against viewpoint shifts. When studying the notion of open-set recognition, we notice a trade-off between closed-set and open-set recognition performance if the pretrained VSSL encoders are used without finetuning. We hope that our work will contribute to the development of robust video representation learning frameworks for various real-world scenarios. The project page and code are available at: https://pritamqu.github.io/OOD-VSSL.
****
M^{2}SODAI: Multi-Modal Maritime Object Detection Dataset With RGB and Hyperspectral Image Sensors
Jonggyu Jang, Sangwoo Oh, Youjin Kim, Dongmin Seo, Youngchol Choi, Hyun Jong Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****
Projection-Free Methods for Solving Nonconvex-Concave Saddle Point Problems
Morteza Boroun, Erfan Yazdandoost Hamedani, Afrooz Jalilzadeh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Better Correlation and Robustness: A Distribution-Balanced Self-Supervised Learning Framework for Automatic Dialogue Evaluation
Peiwen Yuan, Xinglin Wang, Jiayi Shi, Bin Sun, Yiwei Li, Prof. Kan
Turn-level dialogue evaluation models (TDEMs), using self-supervised learning (SSL) framework, have achieved state-of-the-art performance in open-domain dialogue evaluation. However, these models inevitably face two potential problems. First, they have low correlations with humans on medium coherence samples as the SSL framework often brings training data with unbalanced coherence distribution. Second, the SSL framework leads TDEM to nonuniform score distribution. There is a danger that the nonuniform score distribution will weaken the robustness of TDEM through our theoretical analysis. To tackle these problems, we propose Better Correlation and Robustness (BCR), a distribution-balanced self-supervised learning framework for TDEM. Given a dialogue dataset, BCR offers an effective training set reconstructing method to provide coherence-balanced training signals and further facilitate balanced evaluating abilities of TDEM. To get a uniform score distribution, a novel loss function is proposed, which can adjust adaptively according to the uniformity of score distribution estimated by kernel density estimation. Comprehensive experiments on 17 benchmark datasets show that vanilla BERT-base using BCR outperforms SOTA methods significantly by 11.3% on average. BCR also demonstrates strong generalization ability as it can lead multiple SOTA methods to attain better correlation and robustness.
****
Learning Topology-Agnostic EEG Representations with Geometry-Aware Modeling
Ke Yi, Yansen Wang, Kan Ren, Dongsheng Li
Large-scale pre-training has shown great potential to enhance models on downstream tasks in vision and language. Developing similar techniques for scalp electroencephalogram (EEG) is suitable since unlabelled data is plentiful. Meanwhile, various sampling channel selections and inherent structural and spatial information bring challenges and avenues to improve existing pre-training strategies further. In order to break boundaries between different EEG resources and facilitate cross-dataset EEG pre-training, we propose to map all kinds of channel selections to a unified topology. We further introduce MMM, a pre-training framework with Multi-dimensional position encoding, Multi-level channel hierarchy, and Multi-stage pre-training strategy built on the unified topology to obtain topology-agnostic representations. Experiments demonstrate that our approach yields impressive improvements over previous state-of-the-art techniques on emotional recognition benchmark datasets.
****
Correlation Aware Sparsified Mean Estimation Using Random Projection
Shuli Jiang, PRANAY SHARMA, Gauri Joshi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Accelerating Value Iteration with Anchoring
Jongmin Lee, Ernest Ryu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Echoes Beyond Points: Unleashing the Power of Raw Radar Data in Multi-modality F
usion
Yang Liu, Feng Wang, Naiyan Wang, ZHAO-XIANG ZHANG
Radar is ubiquitous in autonomous driving systems due to its low cost and good adaptability to bad weather. Nevertheless, the radar detection performance is usually inferior because its point cloud is sparse and not accurate due to the poor azimuth and elevation resolution. Moreover, point cloud generation algorithms already drop weak signals to reduce the false targets which may be suboptimal for the use of deep fusion. In this paper, we propose a novel method named EchoFusion to skip the existing radar signal processing pipeline and then incorporate the radar raw data with other sensors. Specifically, we first generate the Bird's Eye View (BEV) queries and then take corresponding spectrum features from radar to fuse with other sensors. By this approach, our method could utilize both rich and lossless distance and speed clues from radar echoes and rich semantic clues from images, making our method surpass all existing methods on the RADIal dataset, and approach the performance of LiDAR. The code will be released on https://github.com/tusen-ai/EchoFusion.
****
D4: Improving LLM Pretraining via Document De-Duplication and Diversification
Kushal Tirumala, Daniel Simig, Armen Aghajanyan, Ari Morcos
Over recent years, an increasing amount of compute and data has been poured into training large language models (LLMs), usually by doing one-pass learning on as many tokens as possible randomly selected from large-scale web corpora. While training on ever-larger portions of the internet leads to consistent performance improvements, the size of these improvements diminishes with scale, and there has been little work exploring the effect of data selection on pre-training and downstream performance beyond simple de-duplication methods such as MinHash. Here, we show that careful data selection (on top of de-duplicated data) via pre-trained model embeddings can speed up training (20% efficiency gains) and improves average downstream accuracy on 16 NLP tasks (up to 2%) at the 6.7B model scale. Furthermore, we show that repeating data intelligently consistently outperforms baseline training (while repeating random data performs worse than baseline training). Our results indicate that clever data selection can significantly improve LLM pre-training, calls into question the common practice of training for a single epoch on as much data as possible, and demonstrates a path to keep improving our models past the limits of randomly sampling web data.
****
Effective Bayesian Heteroscedastic Regression with Deep Neural Networks
Alexander Immer, Emanuele Palumbo, Alexander Marx, Julia Vogt
Flexibly quantifying both irreducible aleatoric and model-dependent epistemic uncertainties plays an important role for complex regression problems. While deep neural networks in principle can provide this flexibility and learn heteroscedastic aleatoric uncertainties through non-linear functions, recent works highlight that maximizing the log likelihood objective parameterized by mean and variance can lead to compromised mean fits since the gradient are scaled by the predictive variance, and propose adjustments in line with this premise. We instead propose to use the natural parametrization of the Gaussian, which has been shown to be more stable for heteroscedastic regression based on non-linear feature maps and Gaussian processes. Further, we emphasize the significance of principled regularization of the network parameters and prediction. We therefore propose an efficient Laplace approximation for heteroscedastic neural networks that allows automatic regularization through empirical Bayes and provides epistemic uncertainties, both of which improve generalization.We showcase on a range of regression problems—including a new heteroscedastic image regression benchmark—that our methods are scalable, improve over previous approaches for heteroscedastic regression, and provide epistemic uncertainty without requiring hyperparameter tuning.
****
Multi-task learning with summary statistics
Parker Knight, Rui Duan
Multi-task learning has emerged as a powerful machine learning paradigm for inte

grating data from multiple sources, leveraging similarities between tasks to improve overall model performance. However, the application of multi-task learning to real-world settings is hindered by data-sharing constraints, especially in healthcare settings. To address this challenge, we propose a flexible multi-task learning framework utilizing summary statistics from various sources. Additionally, we present an adaptive parameter selection approach based on a variant of Lepski's method, allowing for data-driven tuning parameter selection when only summary statistics are accessible. Our systematic non-asymptotic analysis characterizes the performance of the proposed methods under various regimes of the source datasets' sample complexity and overlap. We demonstrate our theoretical findings and the performance of the method through extensive simulations. This work offers a more flexible tool for training related models across various domains, with practical implications in genetic risk prediction and many other fields.
****
Estimating Noise Correlations Across Continuous Conditions With Wishart Processes
Amin Nejatbakhsh, Isabel Garon, Alex Williams
The signaling capacity of a neural population depends on the scale and orientation of its covariance across trials. Estimating this "noise" covariance is challenging and is thought to require a large number of stereotyped trials. New approaches are therefore needed to interrogate the structure of neural noise across rich, naturalistic behaviors and sensory experiences, with few trials per condition. Here, we exploit the fact that conditions are smoothly parameterized in many experiments and leverage Wishart process models to pool statistical power from trials in neighboring conditions. We demonstrate that these models perform favorably on experimental data from the mouse visual cortex and monkey motor cortex relative to standard covariance estimators. Moreover, they produce smooth estimates of covariance as a function of stimulus parameters, enabling estimates of noise correlations in entirely unseen conditions as well as continuous estimates of Fisher information—a commonly used measure of signal fidelity. Together, our results suggest that Wishart processes are broadly applicable tools for quantification and uncertainty estimation of noise correlations in trial-limited regimes, paving the way toward understanding the role of noise in complex neural computations and behavior.
****
Toward Understanding Generative Data Augmentation
Chenyu Zheng, Guoqiang Wu, Chongxuan LI
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
TOA: Task-oriented Active VQA
xiaoying xing, Mingfu Liang, Ying Wu
Knowledge-based visual question answering (VQA) requires external knowledge to answer the question about an image. Early methods explicitly retrieve knowledge from external knowledge bases, which often introduce noisy information. Recently large language models like GPT-3 have shown encouraging performance as implicit knowledge source and revealed planning abilities. However, current large language models can not effectively understand image inputs, thus it remains an open problem to extract the image information and input to large language models. Prior works have used image captioning and object descriptions to represent the image. However, they may either drop the essential visual information to answer the question correctly or involve irrelevant objects to the task-of-interest. To address this problem, we propose to let large language models make an initial hypothesis according to their knowledge, then actively collect the visual evidence required to verify the hypothesis. In this way, the model can attend to the essential visual information in a task-oriented manner. We leverage several vision modules from the perspectives of spatial attention (i.e., Where to look) and attribute attention (i.e., What to look), which is similar to human cognition. The expe

riments show that our proposed method outperforms the baselines on open-ended knowledge-based VQA datasets and presents clear reasoning procedure with better interpretability.
****
Universal Gradient Descent Ascent Method for Nonconvex-Nonconcave Minimax Optimization
Taoli Zheng, Linglingzhi Zhu, Anthony Man-Cho So, Jose Blanchet, Jiajin Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Evaluating Adversarial Robustness of Large Vision-Language Models
Yunqing Zhao, Tianyu Pang, Chao Du, Xiao Yang, Chongxuan LI, Ngai-Man (Man) Cheung, Min Lin
Large vision-language models (VLMs) such as GPT-4 have achieved unprecedented performance in response generation, especially with visual inputs, enabling more creative and adaptable interaction than large language models such as ChatGPT. Nonetheless, multimodal generation exacerbates safety concerns, since adversaries may successfully evade the entire system by subtly manipulating the most vulnerable modality (e.g., vision). To this end, we propose evaluating the robustness of open-source large VLMs in the most realistic and high-risk setting, where adversaries have only black-box system access and seek to deceive the model into returning the targeted responses. In particular, we first craft targeted adversarial examples against pretrained models such as CLIP and BLIP, and then transfer these adversarial examples to other VLMs such as MiniGPT-4, LLaVA, UniDiffuser, BLIP-2, and Img2Prompt. In addition, we observe that black-box queries on these VLMs can further improve the effectiveness of targeted evasion, resulting in a surprisingly high success rate for generating targeted responses. Our findings provide a quantitative understanding regarding the adversarial vulnerability of large VLMs and call for a more thorough examination of their potential security flaws before deployment in practice. Our project page: https://yunqing-me.github.io/AttackVLM/.
****
Generator Born from Classifier
Runpeng Yu, Xinchao Wang
In this paper, we make a bold attempt toward an ambitious task: given a pre-trained classifier, we aim to reconstruct an image generator, without relying on any data samples. From a black-box perspective, this challenge seems intractable, since it inevitably involves identifying the inverse function for a classifier, which is, by nature, an information extraction process. As such, we resort to leveraging the knowledge encapsulated within the parameters of the neural network. Grounded on the theory of Maximum-Margin Bias of gradient descent, we propose a novel learning paradigm, in which the generator is trained to ensure that the convergence conditions of the network parameters are satisfied over the generated distribution of the samples. Empirical validation from various image generation tasks substantiates the efficacy of our strategy.
****
Scattering Vision Transformer: Spectral Mixing Matters
Badri Patro, Vijay Agneeswaran
Vision transformers have gained significant attention and achieved state-of-the-art performance in various computer vision tasks, including image classification, instance segmentation, and object detection. However, challenges remain in addressing attention complexity and effectively capturing fine-grained information within images. Existing solutions often resort to down-sampling operations, such as pooling, to reduce computational cost. Unfortunately, such operations are non-invertible and can result in information loss. In this paper, we present a novel approach called Scattering Vision Transformer (SVT) to tackle these challenges. SVT incorporates a spectrally scattering network that enables the capture of intricate image details. SVT overcomes the invertibility issue associated with d

own-sampling operations by separating low-frequency and high-frequency components. Furthermore, SVT introduces a unique spectral gating network utilizing Einstein multiplication for token and channel mixing, effectively reducing complexity. We show that SVT achieves state-of-the-art performance on the ImageNet dataset with a significant reduction in a number of parameters and FLOPS. SVT shows 2\% improvement over LiTv2 and iFormer. SVT-H-S reaches 84.2\% top-1 accuracy, while SVT-H-B reaches 85.2\% (state-of-art for base versions) and SVT-H-L reaches 85.7\% (again state-of-art for large versions). SVT also shows comparable results in other vision tasks such as instance segmentation. SVT also outperforms other transformers in transfer learning on standard datasets such as CIFAR10, CIFAR100, Oxford Flower, and Stanford Car datasets. The project page is available on this webpage.\url{https://badripatro.github.io/svt/}.
****
Task-aware world model learning with meta weighting via bi-level optimization
Huining Yuan, Hongkun Dou, Xingyu Jiang, Yue Deng
Aligning the world model with the environment for the agent’s specific task is crucial in model-based reinforcement learning. While value-equivalent models may achieve better task awareness than maximum-likelihood models, they sacrifice a large amount of semantic information and face implementation issues. To combine the benefits of both types of models, we propose Task-aware Environment Modeling Pipeline with bi-level Optimization (TEMPO), a bi-level model learning framework that introduces an additional level of optimization on top of a maximum-likelihood model by incorporating a meta weighter network that weights each training sample. The meta weighter in the upper level learns to generate novel sample weights by minimizing a proposed task-aware model loss. The model in the lower level focuses on important samples while maintaining rich semantic information in state representations. We evaluate TEMPO on a variety of continuous and discrete control tasks from the DeepMind Control Suite and Atari video games. Our results demonstrate that TEMPO achieves state-of-the-art performance regarding asymptotic performance, training stability, and convergence speed.
****
LEPARD: Learning Explicit Part Discovery for 3D Articulated Shape Reconstruction
Di Liu, Anastasis Stathopoulos, Qilong Zhangli, Yunhe Gao, Dimitris Metaxas
Reconstructing the 3D articulated shape of an animal from a single in-the-wild image is a challenging task. We propose LEPARD, a learning-based framework that discovers semantically meaningful 3D parts and reconstructs 3D shapes in a part-based manner. This is advantageous as 3D parts are robust to pose variations due to articulations and their shape is typically simpler than the overall shape of the object. In our framework, the parts are explicitly represented as parameterized primitive surfaces with global and local deformations in 3D that deform to match the image evidence. We propose a kinematics-inspired optimization to guide each transformation of the primitive deformation given 2D evidence. Similar to recent approaches, LEPARD is only trained using off-the-shelf deep features from DINO and does not require any form of 2D or 3D annotations. Experiments on 3D animal shape reconstruction, demonstrate significant improvement over existing alternatives in terms of both the overall reconstruction performance as well as the ability to discover semantically meaningful and consistent parts.
****
Curriculum Learning With Infant Egocentric Videos
Saber Sheybani, Himanshu Hansaria, Justin Wood, Linda Smith, Zoran Tiganj
Infants possess a remarkable ability to rapidly learn and process visual inputs. As an infant's mobility increases, so does the variety and dynamics of their visual inputs. Is this change in the properties of the visual inputs beneficial or even critical for the proper development of the visual system? To address this question, we used video recordings from infants wearing head-mounted cameras to train a variety of self-supervised learning models. Critically, we separated the infant data by age group and evaluated the importance of training with a curriculum aligned with developmental order. We found that initiating learning with the data from the youngest age group provided the strongest learning signal and led to the best learning outcomes in terms of downstream task performance. We then

showed that the benefits of the data from the youngest age group are due to the slowness and simplicity of the visual experience. The results provide strong empirical evidence for the importance of the properties of the early infant experience and developmental progression in training. More broadly, our approach and findings take a noteworthy step towards reverse engineering the learning mechanisms in newborn brains using image-computable models from artificial intelligence.
****
MarioGPT: Open-Ended Text2Level Generation through Large Language Models
Shyam Sudhakaran, Miguel González-Duque, Matthias Freiberger, Claire Glanois, Elias Najarro, Sebastian Risi
Procedural Content Generation (PCG) is a technique to generate complex and diverse environments in an automated way. However, while generating content with PCG methods is often straightforward, generating meaningful content that reflects specific intentions and constraints remains challenging. Furthermore, many PCG algorithms lack the ability to generate content in an open-ended manner. Recently, Large Language Models (LLMs) have shown to be incredibly effective in many diverse domains. These trained LLMs can be fine-tuned, re-using information and accelerating training for new tasks. Here, we introduce MarioGPT, a fine-tuned GPT2 model trained to generate tile-based game levels, in our case Super Mario Bros levels. MarioGPT can not only generate diverse levels, but can be text-prompted for controllable level generation, addressing one of the key challenges of current PCG techniques. As far as we know, MarioGPT is the first text-to-level model and combined with novelty search it enables the generation of diverse levels with varying play-style dynamics (i.e. player paths) and the open-ended discovery of an increasingly diverse range of content. Code available at https://github.com/shyamsn97/mario-gpt.
****
Is Learning in Games Good for the Learners?
William Brown, Jon Schneider, Kiran Vodrahalli
We consider a number of questions related to tradeoffs between reward and regret in repeated gameplay between two agents. To facilitate this, we introduce a notion of generalized equilibrium which allows for asymmetric regret constraints, and yields polytopes of feasible values for each agent and pair of regret constraints, where we show that any such equilibrium is reachable by a pair of algorithms which maintain their regret guarantees against arbitrary opponents. As a central example, we highlight the case one agent is no-swap and the other's regret is unconstrained. We show that this captures an extension of Stackelberg equilibria with a matching optimal value, and that there exists a wide class of games where a player can significantly increase their utility by deviating from a no-swap-regret algorithm against a no-swap learner (in fact, almost any game without pure Nash equilibria is of this form). Additionally, we make use of generalized equilibria to consider tradeoffs in terms of the opponent's algorithm choice. We give a tight characterization for the maximal reward obtainable against some no-regret learner, yet we also show a class of games in which this is bounded away from the value obtainable against the class of common "mean-based" no-regret algorithms. Finally, we consider the question of learning reward-optimal strategies via repeated play with a no-regret agent when the game is initially unknown. Again we show tradeoffs depending on the opponent's learning algorithm: the Stackelberg strategy is learnable in exponential time with any no-regret agent (and in polynomial time with any no-adaptive-regret agent) for any game where it is learnable via queries, and there are games where it is learnable in polynomial time against any no-swap-regret agent but requires exponential time against a mean-based no-regret agent.
****
The Shaped Transformer: Attention Models in the Infinite Depth-and-Width Limit
Lorenzo Noci, Chuning Li, Mufan Li, Bobby He, Thomas Hofmann, Chris J. Maddison, Dan Roy
In deep learning theory, the covariance matrix of the representations serves as aproxy to examine the network’s trainability. Motivated by the success of Transform-ers, we study the covariance matrix of a modified Softmax-based attention mo

delwith skip connections in the proportional limit of infinite-depth-and-width. Weshow that at initialization the limiting distribution can be described by a stochasticdifferential equation (SDE) indexed by the depth-to-width ratio. To achieve awell-defined stochastic limit, the Transformer's attention mechanism is modifiedby centering the Softmax output at identity, and scaling the Softmax logits by awidth-dependent temperature parameter. We examine the stability of the networkthrough the corresponding SDE, showing how the scale of both the drift and diffu-sion can be elegantly controlled with the aid of residual connections. The existenceof a stable SDE implies that the covariance structure is well-behaved, even for verylarge depth and width, thus preventing the notorious issues of rank degeneracyin deep attention models. Finally, we show, through simulations, that the SDEprovides a surprisingly good description of the corresponding finite-size model.We coin the name shaped Transformer for these architectural modifications.

****

Decompose Novel into Known: Part Concept Learning For 3D Novel Class Discovery

Tingyu Weng, Jun Xiao, Haiyong Jiang

In this work, we address 3D novel class discovery (NCD) that discovers novel classes from an unlabeled dataset by leveraging the knowledge of disjoint known classes. The key challenge of 3D NCD is that learned features by known class recognition are heavily biased and hinder generalization to novel classes. Since geometric parts are more generalizable across different classes, we propose to decompose novel into known parts, coined DNIK, to mitigate the above problems. DNIK learns a part concept bank encoding rich part geometric patterns from known classes so that novel 3D shapes can be represented as part concept compositions to facilitate cross-category generalization. Moreover, we formulate three constraints on part concepts to ensure diverse part concepts without collapsing. A part relation encoding module (PRE) is also developed to leverage part-wise spatial relations for better recognition. We construct three 3D NCD tasks for evaluation and extensive experiments show that our method achieves significantly superior results than SOTA baselines (+11.7%, +14.1%, and +16.3% improvements on average for three tasks, respectively). Code and data will be released.

****

Long Sequence Hopfield Memory

Hamza Chaudhry, Jacob Zavatone-Veth, Dmitry Krotov, Cengiz Pehlevan

Sequence memory is an essential attribute of natural and artificial intelligence that enables agents to encode, store, and retrieve complex sequences of stimuli and actions. Computational models of sequence memory have been proposed where recurrent Hopfield-like neural networks are trained with temporally asymmetric Hebbian rules. However, these networks suffer from limited sequence capacity (maximal length of the stored sequence) due to interference between the memories. Inspired by recent work on Dense Associative Memories, we expand the sequence capacity of these models by introducing a nonlinear interaction term, enhancing separation between the patterns. We derive novel scaling laws for sequence capacity with respect to network size, significantly outperforming existing scaling laws for models based on traditional Hopfield networks, and verify these theoretical results with numerical simulation. Moreover, we introduce a generalized pseudoinverse rule to recall sequences of highly correlated patterns. Finally, we extend this model to store sequences with variable timing between states' transitions and describe a biologically-plausible implementation, with connections to motor neuroscience.

****

Provably Safe Reinforcement Learning with Step-wise Violation Constraints

Nuoya Xiong, Yihan Du, Longbo Huang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

Human spatiotemporal pattern learning as probabilistic program synthesis

Tracey Mills, Josh Tenenbaum, Samuel Cheyette
People are adept at learning a wide variety of structured patterns from small am
ounts of data, presenting a conundrum from the standpoint of the bias-variance t
radeoff: what kinds of representations and algorithms support the joint flexibil
ity and data-paucity of human learning? One possibility is that people "learn by
 programming": inducing probabilistic models to fit observed data. Here, we expe
rimentally test human learning in the domain of structured 2-dimensional pattern
s, using a task in which participants repeatedly predicted where a dot would mov
e based on its previous trajectory. We evaluate human performance against standa
rd parametric and non-parametric time-series models, as well as two Bayesian pro
gram synthesis models whose hypotheses vary in their degree of structure: a comp
ositional Gaussian Process model and a structured "Language of Thought" (LoT) mo
del. We find that signatures of human pattern learning are best explained by the
 LoT model, supporting the idea that the flexibility and data-efficiency of huma
n structure learning can be understood as probabilistic inference over an expres
sive space of programs.
****
Fair Allocation of Indivisible Chores: Beyond Additive Costs
Bo Li, Fangxiao Wang, Yu Zhou
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Robust Second-Order Nonconvex Optimization and Its Application to Low Rank Matri
x Sensing
Shuyao Li, Yu Cheng, Ilias Diakonikolas, Jelena Diakonikolas, Rong Ge, Stephen W
right
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Incentivized Communication for Federated Bandits
Zhepei Wei, Chuanhao Li, Haifeng Xu, Hongning Wang
Most existing works on federated bandits take it for granted that all clients ar
e altruistic about sharing their data with the server for the collective good wh
enever needed. Despite their compelling theoretical guarantee on performance and
 communication efficiency, this assumption is overly idealistic and oftentimes v
iolated in practice, especially when the algorithm is operated over self-interes
ted clients, who are reluctant to share data without explicit benefits. Negligen
ce of such self-interested behaviors can significantly affect the learning effic
iency and even the practical operability of federated bandit learning. In light
of this, we aim to spark new insights into this under-explored research area by
formally introducing an incentivized communication problem for federated bandits
, where the server shall motivate clients to share data by providing incentives.
 Without loss of generality, we instantiate this bandit problem with the context
ual linear setting and propose the first incentivized communication protocol, na
mely, Inc-FedUCB, that achieves near-optimal regret with provable communication
and incentive cost guarantees. Extensive empirical experiments on both synthetic
 and real-world datasets further validate the effectiveness of the proposed meth
od across various environments.
****
Domain Watermark: Effective and Harmless Dataset Copyright Protection is Closed
at Hand
Junfeng Guo, Yiming Li, Lixu Wang, Shu-Tao Xia, Heng Huang, Cong Liu, Bo Li
The prosperity of deep neural networks (DNNs) is largely benefited from open-sou
rce datasets, based on which users can evaluate and improve their methods. In th
is paper, we revisit backdoor-based dataset ownership verification (DOV), which
is currently the only feasible approach to protect the copyright of open-source

datasets. We reveal that these methods are fundamentally harmful given that they could introduce malicious misclassification behaviors to watermarked DNNs by the adversaries. In this paper, we design DOV from another perspective by making watermarked models (trained on the protected dataset) correctly classify some `hard' samples that will be misclassified by the benign model. Our method is inspired by the generalization property of DNNs, where we find a \emph{hardly-generalized domain} for the original dataset (as its \emph{domain watermark}). It can be easily learned with the protected dataset containing modified samples. Specifically, we formulate the domain generation as a bi-level optimization and propose to optimize a set of visually-indistinguishable clean-label modified data with similar effects to domain-watermarked samples from the hardly-generalized domain to ensure watermark stealthiness. We also design a hypothesis-test-guided ownership verification via our domain watermark and provide the theoretical analyses of our method. Extensive experiments on three benchmark datasets are conducted, which verify the effectiveness of our method and its resistance to potential adaptive methods.
****

DISCO-10M: A Large-Scale Music Dataset

Luca Lanzendörfer, Florian Grötschla, Emil Funke, Roger Wattenhofer

Music datasets play a crucial role in advancing research in machine learning for music. However, existing music datasets suffer from limited size, accessibility, and lack of audio resources. To address these shortcomings, we present DISCO-10M, a novel and extensive music dataset that surpasses the largest previously available music dataset by an order of magnitude. To ensure high-quality data, we implement a multi-stage filtering process. This process incorporates similarities based on textual descriptions and audio embeddings. Moreover, we provide precomputed CLAP embeddings alongside DISCO-10M, facilitating direct application on various downstream tasks. These embeddings enable efficient exploration of machine learning applications on the provided data. With DISCO-10M, we aim to democratize and facilitate new research to help advance the development of novel machine learning models for music: https://huggingface.co/DISCOX
****

Guide Your Agent with Adaptive Multimodal Rewards

Changyeon Kim, Younggyo Seo, Hao Liu, Lisa Lee, Jinwoo Shin, Honglak Lee, Kimin Lee

Developing an agent capable of adapting to unseen environments remains a difficult challenge in imitation learning. This work presents Adaptive Return-conditioned Policy (ARP), an efficient framework designed to enhance the agent's generalization ability using natural language task descriptions and pre-trained multimodal encoders. Our key idea is to calculate a similarity between visual observations and natural language instructions in the pre-trained multimodal embedding space (such as CLIP) and use it as a reward signal. We then train a return-conditioned policy using expert demonstrations labeled with multimodal rewards. Because the multimodal rewards provide adaptive signals at each timestep, our ARP effectively mitigates the goal misgeneralization. This results in superior generalization performances even when faced with unseen text instructions, compared to existing text-conditioned policies. To improve the quality of rewards, we also introduce a fine-tuning method for pre-trained multimodal encoders, further enhancing the performance. Video demonstrations and source code are available on the project website: \url{https://sites.google.com/view/2023arp}.
****

Fine-Grained Theoretical Analysis of Federated Zeroth-Order Optimization

Jun Chen, Hong Chen, Bin Gu, Hao Deng

Federated zeroth-order optimization (FedZO) algorithm enjoys the advantages of both zeroth-order optimization and federated learning, and has shown exceptional performance on black-box attack and softmax regression tasks. However, there is no generalization analysis for FedZO, and its analysis on computing convergence rate is slower than the corresponding first-order optimization setting. This paper aims to establish systematic theoretical assessments of FedZO by developing the analysis technique of on-average model stability. We establish the first gene

ralization error bound of FedZO under the Lipschitz continuity and smoothness conditions. Then, refined generalization and optimization bounds are provided by replacing bounded gradient with heavy-tailed gradient noise and utilizing the second-order Taylor expansion for gradient approximation. With the help of a new error decomposition strategy, our theoretical analysis is also extended to the asynchronous case. For FedZO, our fine-grained analysis fills the theoretical gap on the generalization guarantees and polishes the convergence characterization of the computing algorithm.
****
Sparse Deep Learning for Time Series Data: Theory and Applications
Mingxuan Zhang, Yan Sun, Faming Liang
Sparse deep learning has become a popular technique for improving the performance of deep neural networks in areas such as uncertainty quantification, variable selection, and large-scale network compression. However, most existing research has focused on problems where the observations are independent and identically distributed (i.i.d.), and there has been little work on the problems where the observations are dependent, such as time series data and sequential data in natural language processing. This paper aims to address this gap by studying the theory for sparse deep learning with dependent data. We show that sparse recurrent neural networks (RNNs) can be consistently estimated, and their predictions are asymptotically normally distributed under appropriate assumptions, enabling the prediction uncertainty to be correctly quantified. Our numerical results show that sparse deep learning outperforms state-of-the-art methods, such as conformal predictions, in prediction uncertainty quantification for time series data. Furthermore, our results indicate that the proposed method can consistently identify the autoregressive order for time series data and outperform existing methods in large-scale model compression. Our proposed method has important practical implications in fields such as finance, healthcare, and energy, where both accurate point estimates and prediction uncertainty quantification are of concern.
****
Imbalanced Mixed Linear Regression
Pini Zilber, Boaz Nadler
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
GAN You See Me? Enhanced Data Reconstruction Attacks against Split Inference
Ziang Li, Mengda Yang, Yaxin Liu, Juan Wang, Hongxin Hu, Wenzhe Yi, Xiaoyang Xu
Split Inference (SI) is an emerging deep learning paradigm that addresses computational constraints on edge devices and preserves data privacy through collaborative edge-cloud approaches. However, SI is vulnerable to Data Reconstruction Attacks (DRA), which aim to reconstruct users' private prediction instances. Existing attack methods suffer from various limitations. Optimization-based DRAs do not leverage public data effectively, while Learning-based DRAs depend heavily on auxiliary data quantity and distribution similarity. Consequently, these approaches yield unsatisfactory attack results and are sensitive to defense mechanisms. To overcome these challenges, we propose a GAN-based LAtent Space Search attack (GLASS) that harnesses abundant prior knowledge from public data using advanced StyleGAN technologies. Additionally, we introduce GLASS++ to enhance reconstruction stability. Our approach represents the first GAN-based DRA against SI, and extensive evaluation across different split points and adversary setups demonstrates its state-of-the-art performance. Moreover, we thoroughly examine seven defense mechanisms, highlighting our method's capability to reveal private information even in the presence of these defenses.
****
Random-Access Infinite Context Length for Transformers
Amirkeivan Mohtashami, Martin Jaggi
While Transformers have shown remarkable success in natural language processing, their attention mechanism's large memory requirements have limited their abilit

y to handle longer contexts. Prior approaches, such as recurrent memory or retri
eval-based augmentation, have either compromised the random-access flexibility o
f attention (i.e., the capability to select any token in the entire context) or
relied on separate mechanisms for relevant context retrieval, which may not be c
ompatible with the model's attention. In this paper, we present a novel approach
 that allows access to the complete context while retaining random-access flexib
ility, closely resembling running attention on the entire context. Our method us
es a landmark token to represent each block of the input and trains the attentio
n to use it for selecting relevant blocks, enabling retrieval of blocks directly
 through the attention mechanism instead of by relying on a separate mechanism.
Our approach seamlessly integrates with specialized data structures and the syst
em's memory hierarchy, enabling processing of arbitrarily long context lengths.
We demonstrate that our method can obtain comparable performance with Transforme
r-XL while significantly reducing the number of retrieved tokens in each step. F
inally, we show that fine-tuning LLaMA 7B with our method successfully extends i
ts context length capacity to over 32k tokens, allowing for inference at the con
text lengths of GPT-4. We release the implementation of landmark attention and t
he code to reproduce our experiments at https://github.com/epfml/landmark-attent
ion/.
****
Egocentric Planning for Scalable Embodied Task Achievement
Xiatoian Liu, Hector Palacios, Christian Muise
Embodied agents face significant challenges when tasked with performing actions
in diverse environments, particularly in generalizing across object types and ex
ecuting suitable actions to accomplish tasks. Furthermore, agents should exhibit
 robustness, minimizing the execution of illegal actions. In this work, we prese
nt Egocentric Planning, an innovative approach that combines symbolic planning a
nd Object-oriented POMDPs to solve tasks in complex environments, harnessing exi
sting models for visual perception and natural language processing. We evaluated
 our approach in ALFRED, a simulated environment designed for domestic tasks, an
d demonstrated its high scalability, achieving an impressive 36.07\% unseen succ
ess rate in the ALFRED benchmark and winning the ALFRED challenge at CVPR Embodi
ed AI workshop. Our method requires reliable perception and the specification or
 learning of a symbolic description of the preconditions and effects of the agen
t's actions, as well as what object types reveal information about others. It ca
n naturally scale to solve new tasks beyond ALFRED, as long as they can be solve
d using the available skills. This work offers a solid baseline for studying end
-to-end and hybrid methods that aim to generalize to new tasks, including recent
 approaches relying on LLMs, but often struggle to scale to long sequences of ac
tions or produce robust plans for novel tasks.
****
Removing Hidden Confounding in Recommendation: A Unified Multi-Task Learning App
roach
Haoxuan Li, Kunhan Wu, Chunyuan Zheng, Yanghao Xiao, Hao Wang, Zhi Geng, Fuli Fe
ng, Xiangnan He, Peng Wu
In recommender systems, the collected data used for training is always subject t
o selection bias, which poses a great challenge for unbiased learning. Previous
studies proposed various debiasing methods based on observed user and item featu
res, but ignored the effect of hidden confounding. To address this problem, rece
nt works suggest the use of sensitivity analysis for worst-case control of the u
nknown true propensity, but only valid when the true propensity is near to the n
ominal propensity within a finite bound. In this paper, we first perform theoret
ical analysis to reveal the possible failure of previous approaches, including p
ropensity-based, multi-task learning, and bi-level optimization methods, in achi
eving unbiased learning when hidden confounding is present. Then, we propose a u
nified multi-task learning approach to remove hidden confounding, which uses a f
ew unbiased ratings to calibrate the learned nominal propensities and nominal er
ror imputations from biased data. We conduct extensive experiments on three publ
icly available benchmark datasets containing a fully exposed large-scale industr
ial dataset, validating the effectiveness of the proposed methods in removing hi

dden confounding.
****
Generative Category-level Object Pose Estimation via Diffusion Models
Jiyao Zhang, Mingdong Wu, Hao Dong
Object pose estimation plays a vital role in embodied AI and computer vision, en abling intelligent agents to comprehend and interact with their surroundings. De spite the practicality of category-level pose estimation, current approaches enc ounter challenges with partially observed point clouds, known as the multihypoth esis issue. In this study, we propose a novel solution by reframing categoryleve l object pose estimation as conditional generative modeling, departing from trad itional point-to-point regression. Leveraging score-based diffusion models, we e stimate object poses by sampling candidates from the diffusion model and aggrega ting them through a two-step process: filtering out outliers via likelihood esti mation and subsequently mean-pooling the remaining candidates. To avoid the cost ly integration process when estimating the likelihood, we introduce an alternati ve method that distils an energy-based model from the original score-based model , enabling end-to-end likelihood estimation. Our approach achieves state-of-the-art performance on the REAL275 dataset, surpassing 50% and 60% on strict 5 ■ 2cm and 5 ■ 5cm metrics, respectively. Furthermore, our method demonstrates strong generalization to novel categories without the need for fine-tuning and can read ily adapt to object pose tracking tasks, yielding comparable results to the curr ent state-of-the-art baselines. Our checkpoints and demonstrations can be found at https://sites.google.com/view/genpose.
****
On the explainable properties of 1-Lipschitz Neural Networks: An Optimal Transpo rt Perspective
Mathieu Serrurier, Franck Mamalet, Thomas FEL, Louis Béthune, Thibaut Boissin
Input gradients have a pivotal role in a variety of applications, including adve rsarial attack algorithms for evaluating model robustness, explainable AI techni ques for generating saliency maps, and counterfactual explanations. However, sal iency maps generated by traditional neural networks are often noisy and provide limited insights. In this paper, we demonstrate that, on the contrary, the salie ncy maps of 1-Lipschitz neural networks, learnt with the dual loss of an optimal transportation problem, exhibit desirable XAI properties:They are highly concen trated on the essential parts of the image with low noise, significantly outperf orming state-of-the-art explanation approaches across various models and metrics . We also prove that these maps align unprecedentedly well with human explanatio ns on ImageNet. To explain the particularly beneficial properties of the salienc y map for such models, we prove this gradient encodes both the direction of the transportation plan and the direction towards the nearest adversarial attack. F ollowing the gradient down to the decision boundary is no longer considered an a dversarial attack, but rather a counterfactual explanation that explicitly trans ports the input from one class to another. Thus, Learning with such a loss join tly optimizes the classification objective and the alignment of the gradient , i .e. the saliency map, to the transportation plan direction. These networks were previously known to be certifiably robust by design, and we demonstrate that the y scale well for large problems and models, and are tailored for explainability using a fast and straightforward method.
****
DatasetDM: Synthesizing Data with Perception Annotations Using Diffusion Models
Weijia Wu, Yuzhong Zhao, Hao Chen, Yuchao Gu, Rui Zhao, Yefei He, Hong Zhou, Mik e Zheng Shou, Chunhua Shen
Current deep networks are very data-hungry and benefit from training on large-sc ale datasets, which are often time-consuming to collect and annotate. By contras t, synthetic data can be generated infinitely using generative models such as DA LL-E and diffusion models, with minimal effort and cost. In this paper, we prese nt DatasetDM, a generic dataset generation model that can produce diverse synthe ticimages and the corresponding high-quality perception annotations (e.g., segme ntation masks, and depth). Our method builds upon the pre-trained diffusion mode l and extends text-guided image synthesis to perception data generation. We show

 that the rich latent code of the diffusion model can be effectively decoded as accurate perception annotations using a decoder module. Training the decoder onl y needs less than 1% (around 100 images) of manually labeled images, enabling th e generation of an infinitely large annotated dataset. Then these synthetic data can be used for training various perception models on downstream tasks. To show case the power of the proposed approach, we generate datasets with rich dense pi xel-wise labels for a wide range of downstream tasks, including semantic15segmen tation, instance segmentation, and depth estimation. Notably, it achieves 1) sta te-of-the-art results on semantic segmentation and instance segmentation; 2) sig nificantly more efficient and robust in domain generalization than the real data ; 3) state-of-the-art results in zero-shot segmentation setting; and 4) flexibil ity for efficient application and novel task composition (e.g., image editing)
****
No-Regret Online Prediction with Strategic Experts
Omid Sadeghi, Maryam Fazel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Learning Unseen Modality Interaction
Yunhua Zhang, Hazel Doughty, Cees Snoek
Multimodal learning assumes all modality combinations of interest are available during training to learn cross-modal correspondences. In this paper, we challeng e this modality-complete assumption for multimodal learning and instead strive f or generalization to unseen modality combinations during inference. We pose the problem of unseen modality interaction and introduce a first solution. It exploi ts a module that projects the multidimensional features of different modalities into a common space with rich information preserved. This allows the information to be accumulated with a simple summation operation across available modalities . To reduce overfitting to less discriminative modality combinations during trai ning, we further improve the model learning with pseudo-supervision indicating t he reliability of a modality’s prediction. We demonstrate that our approach is e ffective for diverse tasks and modalities by evaluating it for multimodal video classification, robot state regression, and multimedia retrieval. Project websit e: https://xiaobai1217.github.io/Unseen-Modality-Interaction/.
****
Autonomous Capability Assessment of Sequential Decision-Making Systems in Stocha stic Settings
Pulkit Verma, Rushang Karia, Siddharth Srivastava
It is essential for users to understand what their AI systems can and can't do i n order to use them safely. However, the problem of enabling users to assess AI systems with sequential decision-making (SDM) capabilities is relatively underst udied. This paper presents a new approach for modeling the capabilities of black -box AI systems that can plan and act, along with the possible effects and requi rements for executing those capabilities in stochastic settings. We present an a ctive-learning approach that can effectively interact with a black-box SDM syste m and learn an interpretable probabilistic model describing its capabilities. Th eoretical analysis of the approach identifies the conditions under which the lea rning process is guaranteed to converge to the correct model of the agent; empir ical evaluations on different agents and simulated scenarios show that this appr oach is few-shot generalizable and can effectively describe the capabilities of arbitrary black-box SDM agents in a sample-efficient manner.
****
Model-Free Active Exploration in Reinforcement Learning
Alessio Russo, Alexandre Proutiere
We study the problem of exploration in Reinforcement Learning and present a nove l model-free solution. We adopt an information-theoretical viewpoint and start f rom the instance-specific lower bound of the number of samples that have to be collected to identify a nearly-optimal policy. Deriving this lower bound along w

ith the optimal exploration strategy entails solving an intricate optimization problem and requires a model of the system. In turn, most existing sample optimal exploration algorithms rely on estimating the model. We derive an approximation of the instance-specific lower bound that only involves quantities that can be inferred using model-free approaches. Leveraging this approximation, we devise an ensemble-based model-free exploration strategy applicable to both tabular and continuous Markov decision processes. Numerical results demonstrate that our strategy is able to identify efficient policies faster than state-of-the-art exploration approaches.
****

Universality laws for Gaussian mixtures in generalized linear models
Yatin Dandi, Ludovic Stephan, Florent Krzakala, Bruno Loureiro, Lenka Zdeborová
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

ALGO: Synthesizing Algorithmic Programs with Generated Oracle Verifiers
Kexun Zhang, Danqing Wang, Jingtao Xia, William Yang Wang, Lei Li
Large language models (LLMs) excel at implementing code from functionality descriptions but struggle with algorithmic problems that require not only implementation but also identification of the suitable algorithm. Moreover, LLM-generated programs lack guaranteed correctness and require human verification. To address these challenges, we propose ALGO, a framework that synthesizes Algorithmic programs with LLM-Generated Oracles to guide the generation and verify their correctness. ALGO first generates a reference oracle by prompting an LLM to exhaustively enumerate all the combinations of relevant variables. This oracle is then utilized to guide an arbitrary search strategy in exploring the algorithm space and to verify the synthesized algorithms. Our study shows that the LLM-generatedoracles are correct for 88% of the cases. With the oracles as verifiers, ALGO can be integrated with any existing code generation model in a model-agnostic manner to enhance its performance. Experiments show that when equipped with ALGO, we achieve an 8× better one-submission pass rate over the Codex model and a 2.6× better one-submission pass rate over CodeT, the current state-of-the-art model on CodeContests. We can also get 1.3× better pass rate over the ChatGPT Code Interpreter on unseen problems. The problem set we used for testing, the prompts we used, the verifier and solution programs, and the test cases generated by ALGOare available at https://github.com/zkx06111/ALGO.
****

Private Everlasting Prediction
Moni Naor, Kobbi Nissim, Uri Stemmer, Chao Yan
A private learner is trained on a sample of labeled points and generates a hypothesis that can be used for predicting the labels of newly sampled points while protecting the privacy of the training set [Kasiviswannathan et al., FOCS 2008]. Past research uncovered that private learners may need to exhibit significantly higher sample complexity than non-private learners as is the case of learning of one-dimensional threshold functions [Bun et al., FOCS 2015, Alon et al., STOC 2019].We explore prediction as an alternative to learning. A predictor answers a stream of classification queries instead of outputting a hypothesis. Earlier work has considered a private prediction model with a single classification query [Dwork and Feldman, COLT 2018]. We observe that when answering a stream of queries, a predictor must modify the hypothesis it uses over time, and in a manner that cannot rely solely on the training set.We introduce {\em private everlasting prediction} taking into account the privacy of both the training set {\em and} the (adaptively chosen) queries made to the predictor. We then present a generic construction of private everlasting predictors in the PAC model.The sample complexity of the initial training sample in our construction is quadratic (up to polylog factors) in the VC dimension of the concept class. Our construction allows prediction for all concept classes with finite VC dimension, and in particular threshold functions over infinite domains, for which (traditional) private learni

ng is known to be impossible.
****
A Holistic Approach to Unifying Automatic Concept Extraction and Concept Importa nce Estimation
Thomas FEL, Victor Boutin, Louis Béthune, Remi Cadene, Mazda Moayeri, Léo Andéol , Mathieu Chalvidal, Thomas Serre
In recent years, concept-based approaches have emerged as some of the most promi sing explainability methods to help us interpret the decisions of Artificial Neu ral Networks (ANNs). These methods seek to discover intelligible visual ``concep ts'' buried within the complex patterns of ANN activations in two key steps: (1) concept extraction followed by (2) importance estimation. While these two steps are shared across methods, they all differ in their specific implementations. H ere, we introduce a unifying theoretical framework that recast the first step -- concept extraction problem -- as a special case of dictionary learning, and we formalize the second step -- concept importance estimation -- as a more general form of attribution method.This framework offers several advantages as it allows us: (i) to propose new evaluation metrics for comparing different concept extra ction approaches; (ii) to leverage modern attribution methods and evaluation met rics to extend and systematically evaluate state-of-the-art concept-based approa ches and importance estimation techniques; (iii) to derive theoretical guarante es regarding the optimality of such methods. We further leverage our framework t o try to tackle a crucial question in explainability: how to efficiently identif y clusters of data points that are classified based on a similar shared strategy .To illustrate these findings and to highlight the main strategies of a model, w e introduce a visual representation called the strategic cluster graph. Finally, we present Lens, a dedicated website that offers a complete compilation of thes e visualizations for all classes of the ImageNet dataset.
****
Lung250M-4B: A Combined 3D Dataset for CT- and Point Cloud-Based Intra-Patient L ung Registration
Fenja Falta, Christoph Großbröhmer, Alessa Hering, Alexander Bigalke, Mattias He inrich
A popular benchmark for intra-patient lung registration is provided by the DIR-L AB COPDgene dataset consisting of large-motion in- and expiratory breath-hold CT pairs. This dataset alone, however, does not provide enough samples to properly train state-of-the-art deep learning methods. Other public datasets often also provide only small sample sizes or include primarily small motions between scans that do not translate well to larger deformations. For point-based geometric re gistration, the PVT1010 dataset provides a large number of vessel point clouds w ithout any correspondences and a labeled test set corresponding to the COPDgene cases. However, the absence of correspondences for supervision complicates train ing, and a fair comparison with image-based algorithms is infeasible, since CT s cans for the training data are not publicly available.We here provide a combined benchmark for image- and point-based registration approaches. We curated a tota l of 248 public multi-centric in- and expiratory lung CT scans from 124 patients , which show large motion between scans, processed them to ensure sufficient hom ogeneity between the data and generated vessel point clouds that are well distri buted even deeper inside the lungs. For supervised training, we provide vein and artery segmentations of the vessels and multiple thousand image-derived keypoin t correspondences for each pair. For validation, we provide multiple scan pairs with manual landmark annotations. Finally, as first baselines on our new benchma rk, we evaluate several image and point cloud registration methods on the datase t.
****
An Iterative Self-Learning Framework for Medical Domain Generalization
Zhenbang Wu, Huaxiu Yao, David Liebovitz, Jimeng Sun
Deep learning models have been widely used to assist doctors with clinical decis ion-making. However, these models often encounter a significant performance drop when applied to data that differs from the distribution they were trained on. T his challenge is known as the domain shift problem. Existing domain generalizati

on algorithms attempt to address this problem by assuming the availability of do main IDs and training a single model to handle all domains. However, in healthca re settings, patients can be classified into numerous latent domains, where the actual domain categorizations are unknown. Furthermore, each patient domain exhi bits distinct clinical characteristics, making it sub-optimal to train a single model for all domains. To overcome these limitations, we propose SLGD, a self-le arning framework that iteratively discovers decoupled domains and trains persona lized classifiers for each decoupled domain. We evaluate the generalizability of SLGD across spatial and temporal data distribution shifts on two real-world pub lic EHR datasets: eICU and MIMIC-IV. Our results show that SLGD achieves up to 1 1% improvement in the AUPRC score over the best baseline.
****
A benchmark of categorical encoders for binary classification
Federico Matteucci, Vadim Arzamasov, Klemens Böhm
Categorical encoders transform categorical features into numerical representatio ns that are indispensable for a wide range of machine learning models.Existing e ncoder benchmark studies lack generalizability because of their limited choice o f (1) encoders, (2) experimental factors, and (3) datasets. Additionally, incons istencies arise from the adoption of varying aggregation strategies.This paper i s the most comprehensive benchmark of categorical encoders to date, including an extensive evaluation of 32 configurations of encoders from diverse families, wi th 36 combinations of experimental factors, and on 50 datasets.The study shows t he profound influence of dataset selection, experimental factors, and aggregatio n strategies on the benchmark's conclusions~----~aspects disregarded in previous encoder benchmarks.Our code is available at \url{https://github.com/DrCohomology /EncoderBenchmarking}.
****
Structure of universal formulas
Dmitry Yarotsky
By universal formulas we understand parameterized analytic expressions that have a fixed complexity, but nevertheless can approximate any continuous function on a compact set. There exist various examples of such formulas, including some in the form of neural networks. In this paper we analyze the essential structural elements of these highly expressive models. We introduce a hierarchy of expressi veness classes connecting the global approximability property to the weaker prop erty of infinite VC dimension, and prove a series of classification results for several increasingly complex functional families. In particular, we introduce a general family of polynomially-exponentially-algebraic functions that, as we pro ve, is subject to polynomial constraints. As a consequence, we show that fixed-s ize neural networks with not more than one layer of neurons having transcendenta l activations (e.g., sine or standard sigmoid) cannot in general approximate fun ctions on arbitrary finite sets. On the other hand, we give examples of function al families, including two-hidden-layer neural networks, that approximate functi ons on arbitrary finite sets, but fail to do that on the whole domain of defini tion.
****
Model-enhanced Vector Index
Hailin Zhang, Yujing Wang, Qi Chen, Ruiheng Chang, Ting Zhang, Ziming Miao, Ying yan Hou, Yang Ding, Xupeng Miao, Haonan Wang, Bochen Pang, Yuefeng Zhan, Hao Sun , Weiwei Deng, Qi Zhang, Fan Yang, Xing Xie, Mao Yang, Bin CUI
Embedding-based retrieval methods construct vector indices to search for documen t representations that are most similar to the query representations. They are w idely used in document retrieval due to low latency and decent recall performanc e. Recent research indicates that deep retrieval solutions offer better model qu ality, but are hindered by unacceptable serving latency and the inability to sup port document updates. In this paper, we aim to enhance the vector index with en d-to-end deep generative models, leveraging the differentiable advantages of dee p retrieval models while maintaining desirable serving efficiency. We propose Mo del-enhanced Vector Index (MEVI), a differentiable model-enhanced index empowere d by a twin-tower representation model. MEVI leverages a Residual Quantization (

RQ) codebook to bridge the sequence-to-sequence deep retrieval and embedding-based models. To substantially reduce the inference time, instead of decoding the unique document ids in long sequential steps, we first generate some semantic virtual cluster ids of candidate documents in a small number of steps, and then leverage the well-adapted embedding vectors to further perform a fine-grained search for the relevant documents in the candidate virtual clusters. We empirically show that our model achieves better performance on the commonly used academic benchmarks MSMARCO Passage and Natural Questions, with comparable serving latency to dense retrieval solutions.
****
Wide Neural Networks as Gaussian Processes: Lessons from Deep Equilibrium Models
Tianxiang Gao, Xiaokai Huo, Hailiang Liu, Hongyang Gao
Neural networks with wide layers have attracted significant attention due to their equivalence to Gaussian processes, enabling perfect fitting of training data while maintaining generalization performance, known as benign overfitting. However, existing results mainly focus on shallow or finite-depth networks, necessitating a comprehensive analysis of wide neural networks with infinite-depth layers, such as neural ordinary differential equations (ODEs) and deep equilibrium models (DEQs). In this paper, we specifically investigate the deep equilibrium model (DEQ), an infinite-depth neural network with shared weight matrices across layers. Our analysis reveals that as the width of DEQ layers approaches infinity, it converges to a Gaussian process, establishing what is known as the Neural Network and Gaussian Process (NNGP) correspondence. Remarkably, this convergence holds even when the limits of depth and width are interchanged, which is not observed in typical infinite-depth Multilayer Perceptron (MLP) networks. Furthermore, we demonstrate that the associated Gaussian vector remains non-degenerate for any pairwise distinct input data, ensuring a strictly positive smallest eigenvalue of the corresponding kernel matrix using the NNGP kernel. These findings serve as fundamental elements for studying the training and generalization of DEQs, laying the groundwork for future research in this area.
****
Tree Variational Autoencoders
Laura Manduchi, Moritz Vandenhirtz, Alain Ryser, Julia Vogt
We propose Tree Variational Autoencoder (TreeVAE), a new generative hierarchical clustering model that learns a flexible tree-based posterior distribution over latent variables. TreeVAE hierarchically divides samples according to their intrinsic characteristics, shedding light on hidden structures in the data. It adapts its architecture to discover the optimal tree for encoding dependencies between latent variables. The proposed tree-based generative architecture enables lightweight conditional inference and improves generative performance by utilizing specialized leaf decoders.   We show that TreeVAE uncovers underlying clusters in the data and finds meaningful hierarchical relations between the different groups on a variety of datasets, including real-world imaging data.   We present empirically that TreeVAE provides a more competitive log-likelihood lower bound than the sequential counterparts.   Finally, due to its generative nature, TreeVAE is able to generate new samples from the discovered clusters via conditional sampling.
****
Dynamo-Depth: Fixing Unsupervised Depth Estimation for Dynamical Scenes
Yihong Sun, Bharath Hariharan
Unsupervised monocular depth estimation techniques have demonstrated encouraging results but typically assume that the scene is static. These techniques suffer when trained on dynamical scenes, where apparent object motion can equally be explained by hypothesizing the object's independent motion, or by altering its depth. This ambiguity causes depth estimators to predict erroneous depth for moving objects. To resolve this issue, we introduce Dynamo-Depth, an unifying approach that disambiguates dynamical motion by jointly learning monocular depth, 3D independent flow field, and motion segmentation from unlabeled monocular videos. Specifically, we offer our key insight that a good initial estimation of motion segmentation is sufficient for jointly learning depth and independent motion despi

te the fundamental underlying ambiguity. Our proposed method achieves state-of-the-art performance on monocular depth estimation on Waymo Open and nuScenes Dataset with significant improvement in the depth of moving objects. Code and additional results are available at https://dynamo-depth.github.io.
****
LIMA: Less Is More for Alignment
Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, LILI YU, Susan Zhang, Gargi Ghosh, Mike Lewis, Luke Zettlemoyer, Omer Levy
Large language models are trained in two stages: (1) unsupervised pretraining from raw text, to learn general-purpose representations, and (2) large scale instruction tuning and reinforcement learning, to better align to end tasks and user preferences. We measure the relative importance of these two stages by training LIMA, a 65B parameter LLaMa language model fine-tuned with the standard supervised loss on only 1,000 carefully curated prompts and responses, without any reinforcement learning or human preference modeling.LIMA demonstrates remarkably strong performance, learning to follow specific response formats from only a handful of examples in the training data, including complex queries that range from planning trip itineraries to speculating about alternate history.Moreover, the model tends to generalize well to unseen tasks that did not appear in the training data.In a controlled human study, responses from LIMA are either equivalent or strictly preferred to GPT-4 in 43\% of cases; this statistic is as high as 58\% when compared to Bard and 65\% versus DaVinci003, which was trained with human feedback.Taken together, these results strongly suggest that almost all knowledge in large language models is learned during pretraining, and only limited instruction tuning data is necessary to teach models to produce high quality output.
****
Bi-Level Offline Policy Optimization with Limited Exploration
Wenzhuo Zhou
We study offline reinforcement learning (RL) which seeks to learn a good policy based on a fixed, pre-collected dataset. A fundamental challenge behind this task is the distributional shift due to the dataset lacking sufficient exploration, especially under function approximation. To tackle this issue, we propose a bi-level structured policy optimization algorithm that models a hierarchical interaction between the policy (upper-level) and the value function (lower-level). The lower level focuses on constructing a confidence set of value estimates that maintain sufficiently small weighted average Bellman errors, while controlling uncertainty arising from distribution mismatch. Subsequently, at the upper level, the policy aims to maximize a conservative value estimate from the confidence set formed at the lower level. This novel formulation preserves the maximum flexibility of the implicitly induced exploratory data distribution, enabling the power of model extrapolation. In practice, it can be solved through a computationally efficient, penalized adversarial estimation procedure. Our theoretical regret guarantees do not rely on any data-coverage and completeness-type assumptions, only requiring realizability. These guarantees also demonstrate that the learned policy represents the ``best effort'' among all policies, as no other policies can outperform it. We evaluate our model using a blend of synthetic, benchmark, and real-world datasets for offline RL, showing that it performs competitively with state-of-the-art methods.
****
Generating QM1B with PySCF$_{\text{IPU}}$
Alexander Mathiasen, Hatem Helal, Kerstin Klaser, Paul Balanca, Josef Dean, Carlo Luschi, Dominique Beaini, Andrew Fitzgibbon, Dominic Masters
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Unified Enhancement of Privacy Bounds for Mixture Mechanisms via f-Differential Privacy

Chendi Wang, Buxin Su, Jiayuan Ye, Reza Shokri, Weijie Su
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On the Role of Entanglement and Statistics in Learning
Srinivasan Arunachalam, Vojtech Havlicek, Louis Schatzki
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Molecule Joint Auto-Encoding: Trajectory Pretraining with 2D and 3D Diffusion
weitao Du, Jiujiu Chen, Xuecang Zhang, Zhi-Ming Ma, Shengchao Liu
Recently, artificial intelligence for drug discovery has raised increasing interest in both machine learning and chemistry domains. The fundamental building block for drug discovery is molecule geometry and thus, the molecule's geometrical representation is the main bottleneck to better utilize machine learning techniques for drug discovery. In this work, we propose a pretraining method for molecule joint auto-encoding (MoleculeJAE). MoleculeJAE can learn both the 2D bond (topology) and 3D conformation (geometry) information, and a diffusion process model is applied to mimic the augmented trajectories of such two modalities, based on which, MoleculeJAE will learn the inherent chemical structure in a self-supervised manner. Thus, the pretrained geometrical representation in MoleculeJAE is expected to benefit downstream geometry-related tasks. Empirically, MoleculeJAE proves its effectiveness by reaching state-of-the-art performance on 15 out of 20 tasks by comparing it with 12 competitive baselines.
****
Federated Learning with Manifold Regularization and Normalized Update Reaggregation
Xuming An, Li Shen, Han Hu, Yong Luo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Long-Term Fairness with Unknown Dynamics
Tongxin Yin, Reilly Raab, Mingyan Liu, Yang Liu
While machine learning can myopically reinforce social inequalities, it may also be used to dynamically seek equitable outcomes. In this paper, we formalize long-term fairness as an online reinforcement learning problem for a policy affecting human populations. This formulation accommodates dynamical control objectives, such as achieving equitable population states, that cannot be incorporated into static formulations of fairness. We demonstrate that algorithmic solutions to the proposed fairness problem can adapt to unknown dynamics and, by sacrificing short-term incentives, drive the policy-population system towards more desirable equilibria. For the proposed setting, we develop an algorithm that adapts recent work in online learning and prove that this algorithm achieves simultaneous probabilistic bounds on cumulative loss and cumulative violations of fairness. In the classification setting subject to group fairness, we compare our proposed algorithm to several baselines, including the repeated retraining of myopic or distributionally robust classifiers, and to a deep reinforcement learning algorithm that lacks fairness guarantees. Our experiments model human populations according to evolutionary game theory and integrate real-world datasets.
****
YouTubePD: A Multimodal Benchmark for Parkinson’s Disease Analysis
Andy Zhou, Samuel Li, Pranav Sriram, Xiang Li, Jiahua Dong, Ansh Sharma, Yuanyi Zhong, Shirui Luo, Volodymyr Kindratenko, George Heintz, Christopher Zallek, Yu-Xiong Wang

The healthcare and AI communities have witnessed a growing interest in the devel opment of AI-assisted systems for automated diagnosis of Parkinson's Disease (PD), one of the most prevalent neurodegenerative disorders. However, the progress in this area has been significantly impeded by the absence of a unified, publicl y available benchmark, which prevents comprehensive evaluation of existing PD an alysis methods and the development of advanced models. This work overcomes these challenges by introducing YouTubePD -- the first publicly available multimodal benchmark designed for PD analysis. We crowd-source existing videos featured wit h PD from YouTube, exploit multimodal information including in-the-wild videos, audio data, and facial landmarks across 200+ subject videos, and provide dense a nd diverse annotations from clinical expert. Based on our benchmark, we propose three challenging and complementary tasks encompassing both discriminative and g enerative tasks, along with a comprehensive set of corresponding baselines. Expe rimental evaluation showcases the potential of modern deep learning and computer vision techniques, in particular the generalizability of the models developed o n YouTubePD to real-world clinical settings, while revealing their limitations. We hope our work paves the way for future research in this direction.
****
Fantastic Weights and How to Find Them: Where to Prune in Dynamic Sparse Trainin g
Aleksandra Nowak, Bram Grooten, Decebal Constantin Mocanu, Jacek Tabor
Dynamic Sparse Training (DST) is a rapidly evolving area of research that seeks to optimize the sparse initialization of a neural network by adapting its topolo gy during training. It has been shown that under specific conditions, DST is ab le to outperform dense models. The key components of this framework are the prun ing and growing criteria, which are repeatedly applied during the training proce ss to adjust the network’s sparse connectivity. While the growing criterion's im pact on DST performance is relatively well studied, the influence of the pruning criterion remains overlooked. To address this issue, we design and perform an e xtensive empirical analysis of various pruning criteria to better understand the ir impact on the dynamics of DST solutions. Surprisingly, we find that most of t he studied methods yield similar results. The differences become more significan t in the low-density regime, where the best performance is predominantly given b y the simplest technique: magnitude-based pruning.
****
Tree-Based Diffusion Schrödinger Bridge with Applications to Wasserstein Barycen ters
Maxence Noble, Valentin De Bortoli, Arnaud Doucet, Alain Durmus
Multi-marginal Optimal Transport (mOT), a generalization of OT, aims at minimizi ng the integral of a cost function with respect to a distribution with some pres cribed marginals. In this paper, we consider an entropic version of mOT with a tree-structured quadratic cost, i.e., a function that can be written as a sum o f pairwise cost functions between the nodes of a tree. To address this problem, we develop Tree-based Diffusion Schr\"odinger Bridge (TreeDSB), an extension o f the Diffusion Schr\"odinger Bridge (DSB) algorithm. TreeDSB corresponds to a dynamic and continuous state-space counterpart of the multimarginal Sinkhorn al gorithm. A notable use case of our methodology is to compute Wasserstein baryce nters which can be recast as the solution of a mOT problem on a star-shaped tre e. We demonstrate that our methodology can be applied in high-dimensional settin gs such as image interpolation and Bayesian fusion.
****
Bucks for Buckets (B4B): Active Defenses Against Stealing Encoders
Jan Dubi■ski, Stanis■aw Pawlak, Franziska Boenisch, Tomasz Trzcinski, Adam Dzied zic
Machine Learning as a Service (MLaaS) APIs provide ready-to-use and high-utility encoders that generate vector representations for given inputs. Since these enc oders are very costly to train, they become lucrative targets for model stealing attacks during which an adversary leverages query access to the API to replicat e the encoder locally at a fraction of the original training costs. We propose B ucks for Buckets (B4B), the first active defense that prevents stealing while th

e attack is happening without degrading representation quality for legitimate AP
I users. Our defense relies on the observation that the representations returned
 to adversaries who try to steal the encoder's functionality cover a significant
ly larger fraction of the embedding space than representations of legitimate use
rs who utilize the encoder to solve a particular downstream task. B4B leverages
this to adaptively adjust the utility of the returned representations according
to a user's coverage of the embedding space. To prevent adaptive adversaries fro
m eluding our defense by simply creating multiple user accounts (sybils), B4B al
so individually transforms each user's representations. This prevents the advers
ary from directly aggregating representations over multiple accounts to create t
heir stolen encoder copy. Our active defense opens a new path towards securely s
haring and democratizing encoders over public APIs.
****
A General Framework for Equivariant Neural Networks on Reductive Lie Groups
Ilyes Batatia, Mario Geiger, Jose Munoz, Tess Smidt, Lior Silberman, Christoph O
rtner
Reductive Lie Groups, such as the orthogonal groups, the Lorentz group, or the u
nitary groups, play essential roles across scientific fields as diverse as high
energy physics, quantum mechanics, quantum chromodynamics, molecular dynamics, c
omputer vision, and imaging. In this paper, we present a general Equivariant Neu
ral Network architecture capable of respecting the symmetries of the finite-dime
nsional representations of any reductive Lie Group. Our approach generalizes the
 successful ACE and MACE architectures for atomistic point clouds to any data eq
uivariant to a reductive Lie group action. We also introduce the lie-nn software
 library, which provides all the necessary tools to develop and implement such g
eneral G-equivariant neural networks. It implements routines for the reduction o
f generic tensor products of representations into irreducible representations, m
aking it easy to apply our architecture to a wide range of problems and groups.
The generality and performance of our approach are demonstrated by applying it t
o the tasks of top quark decay tagging (Lorentz group) and shape recognition (or
thogonal group).
****
Context-PIPs: Persistent Independent Particles Demands Spatial Context Features
Weikang Bian, Zhaoyang Huang, Xiaoyu Shi, Yitong Dong, Yijin Li, Hongsheng Li
We tackle the problem of Persistent Independent Particles (PIPs), also called Tr
acking Any Point (TAP), in videos, which specifically aims at estimating persist
ent long-term trajectories of query points in videos. Previous methods attempted
 to estimate these trajectories independently to incorporate longer image sequen
ces, therefore, ignoring the potential benefits of incorporating spatial context
 features. We argue that independent video point tracking also demands spatial c
ontext features. To this end, we propose a novel framework Context-PIPs, which e
ffectively improves point trajectory accuracy by aggregating spatial context fea
tures in videos. Context-PIPs contains two main modules: 1) a SOurse Feature Enh
ancement (SOFE) module, and 2) a TArget Feature Aggregation (TAFA) module. Conte
xt-PIPs significantly improves PIPs all-sided, reducing 11.4\% Average Trajector
y Error of Occluded Points (ATE-Occ) on CroHD and increasing 11.8\% Average Perc
entage of Correct Keypoint (A-PCK) on TAP-Vid-Kinetics. Demos are available at \
url{https://wkbian.github.io/Projects/Context-PIPs/}.
****
GloptiNets: Scalable Non-Convex Optimization with Certificates
Gaspard Beugnot, Julien Mairal, Alessandro Rudi
We present a novel approach to non-convex optimization with certificates, which
handles smooth functions on the hypercube or on the torus. Unlike traditional me
thods that rely on algebraic properties, our algorithm exploits the regularity o
f the target function intrinsic in the decay of its Fourier spectrum. By definin
g a tractable family of models, we allow {\em at the same time} to obtain precis
e certificates and to leverage the advanced and powerful computational technique
s developed to optimize neural networks. In this way the scalability of our appr
oach is naturally enhanced by parallel computing with GPUs. Our approach, when a
pplied to the case of polynomials of moderate dimensions but with thousands of c

oefficients, outperforms the state-of-the-art optimization methods with certificates, as the ones based on Lasserre's hierarchy, addressing problems intractable for the competitors.
****
Ethical Considerations for Responsible Data Curation
Jerone Andrews, Dora Zhao, William Thong, Apostolos Modas, Orestis Papakyriakopoulos, Alice Xiang
Human-centric computer vision (HCCV) data curation practices often neglect privacy and bias concerns, leading to dataset retractions and unfair models. HCCV datasets constructed through nonconsensual web scraping lack crucial metadata for comprehensive fairness and robustness evaluations. Current remedies are post hoc, lack persuasive justification for adoption, or fail to provide proper contextualization for appropriate application. Our research focuses on proactive, domain-specific recommendations, covering purpose, privacy and consent, and diversity, for curating HCCV evaluation datasets, addressing privacy and bias concerns. We adopt an ante hoc reflective perspective, drawing from current practices, guidelines, dataset withdrawals, and audits, to inform our considerations and recommendations.
****
Meta-Adapter: An Online Few-shot Learner for Vision-Language Model
cheng cheng, Lin Song, Ruoyi Xue, Hang Wang, Hongbin Sun, Yixiao Ge, Ying Shan
The contrastive vision-language pre-training, known as CLIP, demonstrates remarkable potential in perceiving open-world visual concepts, enabling effective zero-shot image recognition. Nevertheless, few-shot learning methods based on CLIP typically require offline fine-tuning of the parameters on few-shot samples, resulting in longer inference time and the risk of overfitting in certain domains. To tackle these challenges, we propose the Meta-Adapter, a lightweight residual-style adapter, to refine the CLIP features guided by the few-shot samples in an online manner. With a few training samples, our method can enable effective few-shot learning capabilities and generalize to unseen data or tasks without additional fine-tuning, achieving competitive performance and high efficiency. Without bells and whistles, our approach outperforms the state-of-the-art online few-shot learning method by an average of 3.6\% on eight image classification datasets with higher inference speed. Furthermore, our model is simple and flexible, serving as a plug-and-play module directly applicable to downstream tasks. Without further fine-tuning, Meta-Adapter obtains notable performance improvements in open-vocabulary object detection and segmentation tasks.
****
Taming Local Effects in Graph-based Spatiotemporal Forecasting
Andrea Cini, Ivan Marisca, Daniele Zambon, Cesare Alippi
Spatiotemporal graph neural networks have shown to be effective in time series forecasting applications, achieving better performance than standard univariate predictors in several settings. These architectures take advantage of a graph structure and relational inductive biases to learn a single (global) inductive model to predict any number of the input time series, each associated with a graph node. Despite the gain achieved in computational and data efficiency w.r.t. fitting a set of local models, relying on a single global model can be a limitation whenever some of the time series are generated by a different spatiotemporal stochastic process. The main objective of this paper is to understand the interplay between globality and locality in graph-based spatiotemporal forecasting, while contextually proposing a methodological framework to rationalize the practice of including trainable node embeddings in such architectures. We ascribe to trainable node embeddings the role of amortizing the learning of specialized components. Moreover, embeddings allow for 1) effectively combining the advantages of shared message-passing layers with node-specific parameters and 2) efficiently transferring the learned model to new node sets. Supported by strong empirical evidence, we provide insights and guidelines for specializing graph-based models to the dynamics of each time series and show how this aspect plays a crucial role in obtaining accurate predictions.
****

Latent Space Translation via Semantic Alignment
Valentino Maiorca, Luca Moschella, Antonio Norelli, Marco Fumero, Francesco Locatello, Emanuele Rodolà
While different neural models often exhibit latent spaces that are alike when exposed to semantically related data, this intrinsic similarity is not always immediately discernible. Towards a better understanding of this phenomenon, our work shows how representations learned from these neural modules can be translated between different pre-trained networks via simpler transformations than previously thought. An advantage of this approach is the ability to estimate these transformations using standard, well-understood algebraic procedures that have closed-form solutions. Our method directly estimates a transformation between two given latent spaces, thereby enabling effective stitching of encoders and decoders without additional training. We extensively validate the adaptability of this translation procedure in different experimental settings: across various trainings, domains, architectures (e.g., ResNet, CNN, ViT), and in multiple downstream tasks (classification, reconstruction). Notably, we show how it is possible to zero-shot stitch text encoders and vision decoders, or vice-versa, yielding surprisingly good classification performance in this multimodal setting.
****
A case for reframing automated medical image classification as segmentation
Sarah Hooper, Mayee Chen, Khaled Saab, Kush Bhatia, Curtis Langlotz, Christopher Ré
Image classification and segmentation are common applications of deep learning to radiology. While many tasks can be framed using either classification or segmentation, classification has historically been cheaper to label and more widely used. However, recent work has drastically reduced the cost of training segmentation networks. In light of this recent work, we reexamine the choice of training classification vs. segmentation models. First, we use an information theoretic approach to analyze why segmentation vs. classification models may achieve different performance on the same dataset and overarching task. We then implement multiple methods for using segmentation models to classify medical images, which we call segmentation-for-classification, and compare these methods against traditional classification on three retrospective datasets. We use our analysis and experiments to summarize the benefits of switching from segmentation to classification, including: improved sample efficiency, enabling improved performance with fewer labeled images (up to an order of magnitude lower), on low-prevalence classes, and on certain rare subgroups (up to 161.1\% improved recall); improved robustness to spurious correlations (up to 44.8\% improved robust AUROC); and improved model interpretability, evaluation, and error analysis.
****
Efficient Policy Adaptation with Contrastive Prompt Ensemble for Embodied Agents
wonje choi, Woo Kyung Kim, SeungHyun Kim, Honguk Woo
For embodied reinforcement learning (RL) agents interacting with the environment, it is desirable to have rapid policy adaptation to unseen visual observations, but achieving zero-shot adaptation capability is considered as a challenging problem in the RL context. To address the problem, we present a novel contrastive prompt ensemble (ConPE) framework which utilizes a pretrained vision-language model and a set of visual prompts, thus enables efficient policy learning and adaptation upon a wide range of environmental and physical changes encountered by embodied agents. Specifically, we devise a guided-attention-based ensemble approach with multiple visual prompts on the vision-language model to construct robust state representations. Each prompt is contrastively learned in terms of an individual domain factors that significantly affects the agent's egocentric perception and observation. For a given task, the attention-based ensemble and policy are jointly learned so that the resulting state representations not only generalize to various domains but are also optimized for learning the task. Through experiments, we show that ConPE outperforms other state-of-the-art algorithms for several embodied agent tasks including navigation in AI2THOR, manipulation in Metaworld, and autonomous driving in CARLA, while also improving the sample efficiency of policy learning and adaptation.

****
Improvements on Uncertainty Quantification for Node Classification via Distance Based Regularization
Russell Hart, Linlin Yu, Yifei Lou, Feng Chen
Deep neural networks have achieved significant success in the last decades, but they are not well-calibrated and often produce unreliable predictions. A large number of literature relies on uncertainty quantification to evaluate the reliability of a learning model, which is particularly important for applications of out-of-distribution (OOD) detection and misclassification detection. We are interested in uncertainty quantification for interdependent node-level classification. We start our analysis based on graph posterior networks (GPNs) that optimize the uncertainty cross-entropy (UCE)-based loss function. We describe the theoretical limitations of the widely-used UCE loss. To alleviate the identified drawbacks, we propose a distance-based regularization that encourages clustered OOD nodes to remain clustered in the latent space. We conduct extensive comparison experiments on eight standard datasets and demonstrate that the proposed regularization outperforms the state-of-the-art in both OOD detection and misclassification detection.
****
Efficient Learning of Linear Graph Neural Networks via Node Subsampling
Seiyun Shin, Ilan Shomorony, Han Zhao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DPM-Solver-v3: Improved Diffusion ODE Solver with Empirical Model Statistics
Kaiwen Zheng, Cheng Lu, Jianfei Chen, Jun Zhu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Active Bipartite Ranking
James Cheshire, Vincent Laurent, Stephan Clémençon
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Are Emergent Abilities of Large Language Models a Mirage?
Rylan Schaeffer, Brando Miranda, Sanmi Koyejo
Recent work claims that large language models display \textit{emergent abilities}, abilities not present in smaller-scale models that are present in larger-scale models.What makes emergent abilities intriguing is two-fold: their \textit{sharpness}, transitioning seemingly instantaneously from not present to present, and their \textit{unpredictability}, appearing at seemingly unforeseeable model scales.Here, we present an alternative explanation for emergent abilities: that for a particular task and model family, when analyzing fixed model outputs, emergent abilities appear due the researcher’s choice of metric rather than due to fundamental changes in model behavior with scale. Specifically, nonlinear or discontinuous metrics produce apparent emergent abilities, whereas linear or continuous metrics produce smooth, continuous, predictable changes in model performance.We present our alternative explanation in a simple mathematical model, then test it in three complementary ways: we (1) make, test and confirm three predictions on the effect of metric choice using the InstructGPT/GPT-3 family on tasks with claimed emergent abilities, (2) make, test and confirm two predictions about metric choices in a meta-analysis of emergent abilities on BIG-Bench; and (3) show how to choose metrics to produce never-before-seen seemingly emergent abilities in multiple vision tasks across diverse deep networks.Via all three analyses, we

provide evidence that alleged emergent abilities evaporate with different metrics or with better statistics, and may not be a fundamental property of scaling AI models.
****
Reward-agnostic Fine-tuning: Provable Statistical Benefits of Hybrid Reinforcement Learning
Gen Li, Wenhao Zhan, Jason D. Lee, Yuejie Chi, Yuxin Chen
This paper studies tabular reinforcement learning (RL) in the hybrid setting, which assumes access to both an offline dataset and online interactions with the unknown environment. A central question boils down to how to efficiently utilize online data to strengthen and complement the offline dataset and enable effective policy fine-tuning. Leveraging recent advances in reward-agnostic exploration and offline RL, we design a three-stage hybrid RL algorithm that beats the best of both worlds --- pure offline RL and pure online RL --- in terms of sample complexities. The proposed algorithm does not require any reward information during data collection. Our theory is developed based on a new notion called single-policy partial concentrability, which captures the trade-off between distribution mismatch and miscoverage and guides the interplay between offline and online data.
****
The Exact Sample Complexity Gain from Invariances for Kernel Regression
Behrooz Tahmasebi, Stefanie Jegelka
In practice, encoding invariances into models improves sample complexity. In this work, we study this phenomenon from a theoretical perspective. In particular, we provide minimax optimal rates for kernel ridge regression on compact manifolds, with a target function that is invariant to a group action on the manifold. Our results hold for any smooth compact Lie group action, even groups of positive dimension. For a finite group, the gain effectively multiplies the number of samples by the group size. For groups of positive dimension, the gain is observed by a reduction in the manifold's dimension, in addition to a factor proportional to the volume of the quotient space. Our proof takes the viewpoint of differential geometry, in contrast to the more common strategy of using invariant polynomials. This new geometric viewpoint on learning with invariances may be of independent interest.
****
FaceDNeRF: Semantics-Driven Face Reconstruction, Prompt Editing and Relighting with Diffusion Models
Hao ZHANG, Tianyuan DAI, Yanbo Xu, Yu-Wing Tai, Chi-Keung Tang
The ability to create high-quality 3D faces from a single image has become increasingly important with wide applications in video conferencing, AR/VR, and advanced video editing in movie industries. In this paper, we propose Face Diffusion NeRF (FaceDNeRF), a new generative method to reconstruct high-quality Face NeRFs from single images, complete with semantic editing and relighting capabilities. FaceDNeRF utilizes high-resolution 3D GAN inversion and expertly trained 2D latent-diffusion model, allowing users to manipulate and construct Face NeRFs in zero-shot learning without the need for explicit 3D data. With carefully designed illumination and identity preserving loss, as well as multi-modal pre-training, FaceDNeRF offers users unparalleled control over the editing process enabling them to create and edit face NeRFs using just single-view images, text prompts, and explicit target lighting. The advanced features of FaceDNeRF have been designed to produce more impressive results than existing 2D editing approaches that rely on 2D segmentation maps for editable attributes. Experiments show that our FaceDNeRF achieves exceptionally realistic results and unprecedented flexibility in editing compared with state-of-the-art 3D face reconstruction and editing methods. Our code will be available at https://github.com/BillyXYB/FaceDNeRF.
****
Chanakya: Learning Runtime Decisions for Adaptive Real-Time Perception
Anurag Ghosh, Vaibhav Balloli, Akshay Nambi, Aditya Singh, Tanuja Ganu
Real-time perception requires planned resource utilization. Computational planning in real-time perception is governed by two considerations -- accuracy and lat

ency. There exist run-time decisions (e.g. choice of input resolution) that induce tradeoffs affecting performance on a given hardware, arising from intrinsic (content, e.g. scene clutter) and extrinsic (system, e.g. resource contention) characteristics. Earlier runtime execution frameworks employed rule-based decision algorithms and operated with a fixed algorithm latency budget to balance these concerns, which is sub-optimal and inflexible. We propose Chanakya, a learned approximate execution framework that naturally derives from the streaming perception paradigm, to automatically learn decisions induced by these tradeoffs instead. Chanakya is trained via novel rewards balancing accuracy and latency implicitly, without approximating either objectives. Chanakya simultaneously considers intrinsic and extrinsic context, and predicts decisions in a flexible manner. Chanakya, designed with low overhead in mind, outperforms state-of-the-art static and dynamic execution policies on public datasets on both server GPUs and edge devices.
****

RoboCLIP: One Demonstration is Enough to Learn Robot Policies

Sumedh Sontakke, Jesse Zhang, Séb Arnold, Karl Pertsch, Erdem B■y■k, Dorsa Sadigh, Chelsea Finn, Laurent Itti

Reward specification is a notoriously difficult problem in reinforcement learning, requiring extensive expert supervision to design robust reward functions. Imitation learning (IL) methods attempt to circumvent these problems by utilizing expert demonstrations instead of using an extrinsic reward function but typically require a large number of in-domain expert demonstrations. Inspired by advances in the field of Video-and-Language Models (VLMs), we present RoboCLIP, an online imitation learning method that uses a single demonstration (overcoming the large data requirement) in the form of a video demonstration or a textual description of the task to generate rewards without manual reward function design. Additionally, RoboCLIP can also utilize out-of-domain demonstrations, like videos of humans solving the task for reward generation, circumventing the need to have the same demonstration and deployment domains. RoboCLIP utilizes pretrained VLMs without any finetuning for reward generation. Reinforcement learning agents trained with RoboCLIP rewards demonstrate 2-3 times higher zero-shot performance than competing imitation learning methods on downstream robot manipulation tasks, doing so using only one video/text demonstration. Visit our website at https://sites.google.com/view/roboclip/home for experiment videos.
****

Contrast Everything: A Hierarchical Contrastive Framework for Medical Time-Series

Yihe Wang, Yu Han, Haishuai Wang, Xiang Zhang

Contrastive representation learning is crucial in medical time series analysis as it alleviates dependency on labor-intensive, domain-specific, and scarce expert annotations. However, existing contrastive learning methods primarily focus on one single data level, which fails to fully exploit the intricate nature of medical time series. To address this issue, we present COMET, an innovative hierarchical framework that leverages data consistencies at all inherent levels in medical time series. Our meticulously designed model systematically captures data consistency from four potential levels: observation, sample, trial, and patient levels. By developing contrastive loss at multiple levels, we can learn effective representations that preserve comprehensive data consistency, maximizing information utilization in a self-supervised manner. We conduct experiments in the challenging patient-independent setting. We compare COMET against six baselines using three diverse datasets, which include ECG signals for myocardial infarction and EEG signals for Alzheimer’s and Parkinson’s diseases. The results demonstrate that COMET consistently outperforms all baselines, particularly in setup with 10% and 1% labeled data fractions across all datasets. These results underscore the significant impact of our framework in advancing contrastive representation learning techniques for medical time series. The source code is available at https://github.com/DL4mHealth/COMET.
****

Importance-aware Co-teaching for Offline Model-based Optimization

Ye Yuan, Can (Sam) Chen, Zixuan Liu, Willie Neiswanger, Xue (Steve) Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Large Language Model as Attributed Training Data Generator: A Tale of Diversity and Bias
Yue Yu, Yuchen Zhuang, Jieyu Zhang, Yu Meng, Alexander J. Ratner, Ranjay Krishna, Jiaming Shen, Chao Zhang
Large language models (LLMs) have been recently leveraged as training data generators for various natural language processing (NLP) tasks. While previous research has explored different approaches to training models using generated data, they generally rely on simple class-conditional prompts, which may limit the diversity of the generated data and inherit systematic biases of LLM. Thus, we investigate training data generation with diversely attributed prompts (e.g., specifying attributes like length and style), which have the potential to yield diverse and attributed generated data. Our investigation focuses on datasets with high cardinality and diverse domains, wherein we demonstrate that attributed prompts outperform simple class-conditional prompts in terms of the resulting model's performance. Additionally, we present a comprehensive empirical study on data generation encompassing vital aspects like bias, diversity, and efficiency, and highlight three key observations: firstly, synthetic datasets generated by simple prompts exhibit significant biases, such as regional bias; secondly, attribute diversity plays a pivotal role in enhancing model performance; lastly, attributed prompts achieve the performance of simple class-conditional prompts while utilizing only 5\% of the querying cost of ChatGPT associated with the latter. The data and code are available on {\url{https://github.com/yueyu1030/AttrPrompt}}.
****
GraphPatcher: Mitigating Degree Bias for Graph Neural Networks via Test-time Augmentation
Mingxuan Ju, Tong Zhao, Wenhao Yu, Neil Shah, Yanfang Ye
Recent studies have shown that graph neural networks (GNNs) exhibit strong biases towards the node degree: they usually perform satisfactorily on high-degree nodes with rich neighbor information but struggle with low-degree nodes. Existing works tackle this problem by deriving either designated GNN architectures or training strategies specifically for low-degree nodes. Though effective, these approaches unintentionally create an artificial out-of-distribution scenario, where models mainly or even only observe low-degree nodes during the training, leading to a downgraded performance for high-degree nodes that GNNs originally perform well at. In light of this, we propose a test-time augmentation framework, namely GraphPatcher, to enhance test-time generalization of any GNNs on low-degree nodes. Specifically, GraphPatcher iteratively generates virtual nodes to patch artificially created low-degree nodes via corruptions, aiming at progressively reconstructing target GNN's predictions over a sequence of increasingly corrupted nodes. Through this scheme, GraphPatcher not only learns how to enhance low-degree nodes (when the neighborhoods are heavily corrupted) but also preserves the original superior performance of GNNs on high-degree nodes (when lightly corrupted). Additionally, GraphPatcher is model-agnostic and can also mitigate the degree bias for either self-supervised or supervised GNNs. Comprehensive experiments are conducted over seven benchmark datasets and GraphPatcher consistently enhances common GNNs' overall performance by up to 3.6% and low-degree performance by up to 6.5%, significantly outperforming state-of-the-art baselines. The source code is publicly available at https://github.com/jumxglhf/GraphPatcher.
****
Optimal privacy guarantees for a relaxed threat model: Addressing sub-optimal adversaries in differentially private machine learning
Georgios Kaissis, Alexander Ziller, Stefan Kolek, Anneliese Riess, Daniel Rueckert
Differentially private mechanisms restrict the membership inference capabilities

 of powerful (optimal) adversaries against machine learning models. Such adversaries are rarely encountered in practice. In this work, we examine a more realistic threat model relaxation, where (sub-optimal) adversaries lack access to the exact model training database, but may possess related or partial data. We then formally characterise and experimentally validate adversarial membership inference capabilities in this setting in terms of hypothesis testing errors. Our work helps users to interpret the privacy properties of sensitive data processing systems under realistic threat model relaxations and choose appropriate noise levels for their use-case.
****
Stochastic Approximation Algorithms for Systems of Interacting Particles
Mohammad Reza Karimi Jaghargh, Ya-Ping Hsieh, Andreas Krause
Interacting particle systems have proven highly successful in various machinelearning tasks, including approximate Bayesian inference and neural network optimization. However, the analysis of thesesystems often relies on the simplifying assumption of the \emph{mean-field} limit, where particlenumbers approach infinity and infinitesimal step sizes are used. In practice, discrete time steps,finite particle numbers, and complex integration schemes are employed, creating a theoretical gapbetween continuous-time and discrete-time processes. In this paper, we present a novel frameworkthat establishes a precise connection between these discrete-time schemes and their correspondingmean-field limits in terms of convergence properties and asymptotic behavior. By adopting a dynamical system perspective, our framework seamlessly integrates various numerical schemes that are typically analyzed independently. For example, our framework provides a unified treatment of optimizing an infinite-width two-layer neural network and sampling via Stein Variational Gradient descent, which were previously studied in isolation.
****
Provable Guarantees for Neural Networks via Gradient Feature Learning
Zhenmei Shi, Junyi Wei, Yingyu Liang
Neural networks have achieved remarkable empirical performance, while the current theoretical analysis is not adequate for understanding their success, e.g., the Neural Tangent Kernel approach fails to capture their key feature learning ability, while recent analyses on feature learning are typically problem-specific. This work proposes a unified analysis framework for two-layer networks trained by gradient descent. The framework is centered around the principle of feature learning from gradients, and its effectiveness is demonstrated by applications in several prototypical problems, such as mixtures of Gaussians and parity functions.The framework also sheds light on interesting network learning phenomena such as feature learning beyond kernels and the lottery ticket hypothesis.
****
Binary Radiance Fields
Seungjoo Shin, Jaesik Park
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A U-turn on Double Descent: Rethinking Parameter Counting in Statistical Learning
Alicia Curth, Alan Jeffares, Mihaela van der Schaar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
FABind: Fast and Accurate Protein-Ligand Binding
Qizhi Pei, Kaiyuan Gao, Lijun Wu, Jinhua Zhu, Yingce Xia, Shufang Xie, Tao Qin, Kun He, Tie-Yan Liu, Rui Yan
Modeling the interaction between proteins and ligands and accurately predicting their binding structures is a critical yet challenging task in drug discovery. R

ecent advancements in deep learning have shown promise in addressing this challenge, with sampling-based and regression-based methods emerging as two prominent approaches. However, these methods have notable limitations. Sampling-based methods often suffer from low efficiency due to the need for generating multiple candidate structures for selection. On the other hand, regression-based methods offer fast predictions but may experience decreased accuracy. Additionally, the variation in protein sizes often requires external modules for selecting suitable binding pockets, further impacting efficiency. In this work, we propose FABind, an end-to-end model that combines pocket prediction and docking to achieve accurate and fast protein-ligand binding. FABind incorporates a unique ligand-informed pocket prediction module, which is also leveraged for docking pose estimation. The model further enhances the docking process by incrementally integrating the predicted pocket to optimize protein-ligand binding, reducing discrepancies between training and inference. Through extensive experiments on benchmark datasets, our proposed FABind demonstrates strong advantages in terms of effectiveness and efficiency compared to existing methods. Our code is available at https://github.com/QizhiPei/FABind.
****
Geometric Transformer with Interatomic Positional Encoding
Yusong Wang, Shaoning Li, Tong Wang, Bin Shao, Nanning Zheng, Tie-Yan Liu
The widespread adoption of Transformer architectures in various data modalities has opened new avenues for the applications in molecular modeling. Nevertheless, it remains elusive that whether the Transformer-based architecture can do molecular modeling as good as equivariant GNNs. In this paper, by designing Interatomic Positional Encoding (IPE) thatparameterizes atomic environments as Transformer's positional encodings,we propose Geoformer, a novel geometric Transformer to effectively model molecular structures for various molecular property prediction. We evaluate Geoformer on several benchmarks, including the QM9 dataset and the recently proposed Molecule3D dataset. Compared with both Transformers and equivariant GNN models, Geoformer outperforms the state-of-the-art (SoTA) algorithms on QM9, and achieves the best performance on Molecule3D for both random and scaffold splits.By introducing IPE, Geoformer paves the way for molecular geometric modeling based on Transformer architecture.Codes are available at https://github.com/microsoft/AI2BMD/tree/Geoformer.
****
A Diffusion-Model of Joint Interactive Navigation
Matthew Niedoba, Jonathan Lavington, Yunpeng Liu, Vasileios Lioutas, Justice Sefas, Xiaoxuan Liang, Dylan Green, Setareh Dabiri, Berend Zwartsenberg, Adam Scibior, Frank Wood
Simulation of autonomous vehicle systems requires that simulated traffic participants exhibit diverse and realistic behaviors. The use of prerecorded real-world traffic scenarios in simulation ensures realism but the rarity of safety critical events makes large scale collection of driving scenarios expensive. In this paper, we present DJINN -- a diffusion based method of generating traffic scenarios. Our approach jointly diffuses the trajectories of all agents, conditioned on a flexible set of state observations from the past, present, or future. On popular trajectory forecasting datasets, we report state of the art performance on joint trajectory metrics. In addition, we demonstrate how DJINN flexibly enables direct test-time sampling from a variety of valuable conditional distributions including goal-based sampling, behavior-class sampling, and scenario editing.
****
Diversifying Spatial-Temporal Perception for Video Domain Generalization
Kun-Yu Lin, Jia-Run Du, Yipeng Gao, Jiaming Zhou, Wei-Shi Zheng
Video domain generalization aims to learn generalizable video classification models for unseen target domains by training in a source domain.A critical challenge of video domain generalization is to defend against the heavy reliance on domain-specific cues extracted from the source domain when recognizing target videos. To this end, we propose to perceive diverse spatial-temporal cues in videos, aiming to discover potential domain-invariant cues in addition to domain-specific cues. We contribute a novel model named Spatial-Temporal Diversification Networ

k (STDN), which improves the diversity from both space and time dimensions of video data. First, our STDN proposes to discover various types of spatial cues within individual frames by spatial grouping. Then, our STDN proposes to explicitly model spatial-temporal dependencies between video contents at multiple space-time scales by spatial-temporal relation modeling. Extensive experiments on three benchmarks of different types demonstrate the effectiveness and versatility of our approach.
****
Conformal Prediction for Time Series with Modern Hopfield Networks
Andreas Auer, Martin Gauch, Daniel Klotz, Sepp Hochreiter
To quantify uncertainty, conformal prediction methods are gaining continuously more interest and have already been successfully applied to various domains. However, they are difficult to apply to time series as the autocorrelative structure of time series violates basic assumptions required by conformal prediction. We propose HopCPT, a novel conformal prediction approach for time series that not only copes with temporal structures but leverages them. We show that our approach is theoretically well justified for time series where temporal dependencies are present. In experiments, we demonstrate that our new approach outperforms state-of-the-art conformal prediction methods on multiple real-world time series datasets from four different domains.
****
Cross-modal Prompts: Adapting Large Pre-trained Models for Audio-Visual Downstream Tasks
Haoyi Duan, Yan Xia, Zhou Mingze, Li Tang, Jieming Zhu, Zhou Zhao
In recent years, the deployment of large-scale pre-trained models in audio-visual downstream tasks has yielded remarkable outcomes. However, these models, primarily trained on single-modality unconstrained datasets, still encounter challenges in feature extraction for multi-modal tasks, leading to suboptimal performance. This limitation arises due to the introduction of irrelevant modality-specific information during encoding, which adversely affects the performance of downstream tasks. To address this challenge, this paper proposes a novel Dual-Guided Spatial-Channel-Temporal (DG-SCT) attention mechanism. This mechanism leverages audio and visual modalities as soft prompts to dynamically adjust the parameters of pre-trained models based on the current multi-modal input features. Specifically, the DG-SCT module incorporates trainable cross-modal interaction layers into pre-trained audio-visual encoders, allowing adaptive extraction of crucial information from the current modality across spatial, channel, and temporal dimensions, while preserving the frozen parameters of large-scale pre-trained models. Experimental evaluations demonstrate that our proposed model achieves state-of-the-art results across multiple downstream tasks, including AVE, AVVP, AVS, and AVQA. Furthermore, our model exhibits promising performance in challenging few-shot and zero-shot scenarios. The source code and pre-trained models are available at https://github.com/haoyi-duan/DG-SCT.
****
Causes and Effects of Unanticipated Numerical Deviations in Neural Network Inference Frameworks
Alex Schlögl, Nora Hofer, Rainer Böhme
Hardware-specific optimizations in machine learning (ML) frameworks can cause numerical deviations of inference results. Quite surprisingly, despite using a fixed trained model and fixed input data, inference results are not consistent across platforms, and sometimes not even deterministic on the same platform. We study the causes of these numerical deviations for convolutional neural networks (CNN) on realistic end-to-end inference pipelines and in isolated experiments. Results from 75 distinct platforms suggest that the main causes of deviations on CPUs are differences in SIMD use, and the selection of convolution algorithms at runtime on GPUs. We link the causes and propagation effects to properties of the ML model and evaluate potential mitigations. We make our research code publicly available.
****
Learning via Wasserstein-Based High Probability Generalisation Bounds

Paul Viallard, Maxime Haddouche, Umut Simsekli, Benjamin Guedj
Minimising upper bounds on the population risk or the generalisation gap has bee
n widely used in structural risk minimisation (SRM) -- this is in particular at
the core of PAC-Bayesian learning. Despite its successes and unfailing surge of
interest in recent years, a limitation of the PAC-Bayesian framework is that mos
t bounds involve a Kullback-Leibler (KL) divergence term (or its variations), wh
ich might exhibit erratic behavior and fail to capture the underlying geometric
structure of the learning problem -- hence restricting its use in practical appl
ications.As a remedy, recent studies have attempted to replace the KL divergence
 in the PAC-Bayesian bounds with the Wasserstein distance. Even though these bou
nds alleviated the aforementioned issues to a certain extent, they either hold i
n expectation, are for bounded losses, or are nontrivial to minimize in an SRM f
ramework. In this work, we contribute to this line of research and prove novel
Wasserstein distance-based PAC-Bayesian generalisation bounds for both batch lea
rning with independent and identically distributed (i.i.d.) data, and online lea
rning with potentially non-i.i.d. data. Contrary to previous art, our bounds are
 stronger in the sense that (i) they hold with high probability, (ii) they apply
 to unbounded (potentially heavy-tailed) losses, and (iii) they lead to optimiza
ble training objectives that can be used in SRM. As a result we derive novel Was
serstein-based PAC-Bayesian learning algorithms and we illustrate their empirica
l advantage on a variety of experiments.
****
Towards Anytime Classification in Early-Exit Architectures by Enforcing Conditio
nal Monotonicity
Metod Jazbec, James Allingham, Dan Zhang, Eric Nalisnick
Modern predictive models are often deployed to environments in which computation
al budgets are dynamic. Anytime algorithms are well-suited to such environments
 as, at any point during computation, they can output a prediction whose quality
 is a function of computation time. Early-exit neural networks have garnered att
ention in the context of anytime computation due to their capability to provide
intermediate predictions at various stages throughout the network. However, we d
emonstrate that current early-exit networks are not directly applicable to anyti
me settings, as the quality of predictions for individual data points is not gua
ranteed to improve with longer computation. To address this shortcoming, we prop
ose an elegant post-hoc modification, based on the Product-of-Experts, that enco
urages an early-exit network to become gradually confident. This gives our deep
models the property of conditional monotonicity in the prediction quality---an e
ssential building block towards truly anytime predictive modeling using early-ex
it architectures. Our empirical results on standard image-classification tasks d
emonstrate that such behaviors can be achieved while preserving competitive accu
racy on average.
****
A Scalable Neural Network for DSIC Affine Maximizer Auction Design
Zhijian Duan, Haoran Sun, Yurong Chen, Xiaotie Deng
Automated auction design aims to find empirically high-revenue mechanisms throug
h machine learning. Existing works on multi item auction scenarios can be roughl
y divided into RegretNet-like and affine maximizer auctions (AMAs) approaches. H
owever, the former cannot strictly ensure dominant strategy incentive compatibil
ity (DSIC), while the latter faces scalability issue due to the large number of
allocation candidates. To address these limitations, we propose AMenuNet, a scal
able neural network that constructs the AMA parameters (even including the alloc
ation menu) from bidder and item representations. AMenuNet is always DSIC and in
dividually rational (IR) due to the properties of AMAs, and it enhances scalabil
ity by generating candidate allocations through a neural network. Additionally,
AMenuNet is permutation equivariant, and its number of parameters is independent
 of auction scale. We conduct extensive experiments to demonstrate that AMenuNet
 outperforms strong baselines in both contextual and non-contextual multi-item a
uctions, scales well to larger auctions, generalizes well to different settings,
 and identifies useful deterministic allocations. Overall, our proposed approach
 offers an effective solution to automated DSIC auction design, with improved sc

alability and strong revenue performance in various settings.
****
Med-UniC: Unifying Cross-Lingual Medical Vision-Language Pre-Training by Diminishing Bias
Zhongwei Wan, Che Liu, Mi Zhang, Jie Fu, Benyou Wang, Sibo Cheng, Lei Ma, César Quilodrán-Casas, Rossella Arcucci
The scarcity of data presents a critical obstacle to the efficacy of medical vision-language pre-training (VLP). A potential solution lies in the combination of datasets from various language communities.Nevertheless, the main challenge stems from the complexity of integrating diverse syntax and semantics, language-specific medical terminology, and culture-specific implicit knowledge. Therefore, one crucial aspect to consider is the presence of community bias caused by different languages.This paper presents a novel framework named Unifying Cross-Lingual Medical Vision-Language Pre-Training (\textbf{Med-UniC}), designed to integrate multi-modal medical data from the two most prevalent languages, English and Spanish. Specifically, we propose \textbf{C}ross-lingual \textbf{T}ext Alignment \textbf{R}egularization (\textbf{CTR}) to explicitly unify cross-lingual semantic representations of medical reports originating from diverse language communities. \textbf{CTR} is optimized through latent language disentanglement, rendering our optimization objective to not depend on negative samples, thereby significantly mitigating the bias from determining positive-negative sample pairs within analogous medical reports. Furthermore, it ensures that the cross-lingual representation is not biased toward any specific language community.\textbf{Med-UniC} reaches superior performance across 5 medical image tasks and 10 datasets encompassing over 30 diseases, offering a versatile framework for unifying multi-modal medical data within diverse linguistic communities.The experimental outcomes highlight the presence of community bias in cross-lingual VLP. Reducing this bias enhances the performance not only in vision-language tasks but also in uni-modal visual tasks.
****
Coordinating Distributed Example Orders for Provably Accelerated Training
A. Feder Cooper, Wentao Guo, Duc Khiem Pham, Tiancheng Yuan, Charlie Ruan, Yucheng Lu, Christopher M. De Sa
Recent research on online Gradient Balancing (GraB) has revealed that there exist permutation-based example orderings for SGD that are guaranteed to outperform random reshuffling (RR). Whereas RR arbitrarily permutes training examples, GraB leverages stale gradients from prior epochs to order examples -- achieving a provably faster convergence rate than RR. However, GraB is limited by design: while it demonstrates an impressive ability to scale-up training on centralized data, it does not naturally extend to modern distributed ML workloads. We therefore propose Coordinated Distributed GraB (CD-GraB), which uses insights from prior work on kernel thinning to translate the benefits of provably faster permutation-based example ordering to distributed settings. With negligible overhead, CD-GraB exhibits a linear speedup in convergence rate over centralized GraB and outperforms distributed RR on a variety of benchmark tasks.
****
Individualized Dosing Dynamics via Neural Eigen Decomposition
Stav Belogolovsky, Ido Greenberg, Danny Eytan, Shie Mannor
Dosing models often use differential equations to model biological dynamics. Neural differential equations in particular can learn to predict the derivative of a process, which permits predictions at irregular points of time. However, this temporal flexibility often comes with a high sensitivity to noise, whereas medical problems often present high noise and limited data. Moreover, medical dosing models must generalize reliably over individual patients and changing treatment policies. To address these challenges, we introduce the Neural Eigen Stochastic Differential Equation algorithm (NESDE). NESDE provides individualized modeling (using a hypernetwork over patient-level parameters); generalization to new treatment policies (using decoupled control); tunable expressiveness according to the noise level (using piecewise linearity); and fast, continuous, closed-form prediction (using spectral representation). We demonstrate the robustness of NESDE

in both synthetic and real medical problems, and use the learned dynamics to publish simulated medical gym environments.
****

Unified Embedding: Battle-Tested Feature Representations for Web-Scale ML Systems

Benjamin Coleman, Wang-Cheng Kang, Matthew Fahrbach, Ruoxi Wang, Lichan Hong, Ed Chi, Derek Cheng

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Counterfactual Generation with Identifiability Guarantees

Hanqi Yan, Lingjing Kong, Lin Gui, Yuejie Chi, Eric Xing, Yulan He, Kun Zhang

Counterfactual generation lies at the core of various machine learning tasks, including image translation and controllable text generation. This generation process usually requires the identification of the disentangled latent representations, such as content and style, that underlie the observed data. However, it becomes more challenging when faced with a scarcity of paired data and labelling information. Existing disentangled methods crucially rely on oversimplified assumptions, such as assuming independent content and style variables, to identify the latent variables, even though such assumptions may not hold for complex data distributions. For instance, food reviews tend to involve words like “tasty”, whereas movie reviews commonly contain words such as “thrilling” for the same positive sentiment. This problem is exacerbated when data are sampled from multiple domains since the dependence between content and style may vary significantly over domains. In this work, we tackle the domain-varying dependence between the content and the style variables inherent in the counterfactual generation task. We provide identification guarantees for such latent-variable models by leveraging the relative sparsity of the influences from different latent variables. Our theoretical insights enable the development of a doMain AdapTive counTerfactual gEneration model, called (MATTE). Our theoretically grounded framework achieves state-of-the-art performance in unsupervised style transfer tasks, where neither paired data nor style labels are utilized, across four large-scale datasets.
****

A Batch-to-Online Transformation under Random-Order Model

Jing Dong, Yuichi Yoshida

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

The CLIP Model is Secretly an Image-to-Prompt Converter

Yuxuan Ding, Chunna Tian, Haoxuan Ding, Lingqiao Liu

The Stable Diffusion model is a prominent text-to-image generation model that relies on a text prompt as its input, which is encoded using the Contrastive Language-Image Pre-Training (CLIP). However, text prompts have limitations when it comes to incorporating implicit information from reference images. Existing methods have attempted to address this limitation by employing expensive training procedures involving millions of training samples for image-to-image generation. In contrast, this paper demonstrates that the CLIP model, as utilized in Stable Diffusion, inherently possesses the ability to instantaneously convert images into text prompts. Such an image-to-prompt conversion can be achieved by utilizing a linear projection matrix that is calculated in a closed form. Moreover, the paper showcases that this capability can be further enhanced by either utilizing a small amount of similar-domain training data (approximately 100 images) or incorporating several online training steps (around 30 iterations) on the reference images. By leveraging these approaches, the proposed method offers a simple and flexible solution to bridge the gap between images and text prompts. This methodology can be applied to various tasks such as image variation and image editing, f

acilitating more effective and seamless interaction between images and textual prompts.
****
Invariant Anomaly Detection under Distribution Shifts: A Causal Perspective
João Carvalho, Mengtao Zhang, Robin Geyer, Carlos Cotrini, Joachim M Buhmann
Anomaly detection (AD) is the machine learning task of identifying highly discrepant abnormal samples by solely relying on the consistency of the normal training samples. Under the constraints of a distribution shift, the assumption that training samples and test samples are drawn from the same distribution breaks down. In this work, by leveraging tools from causal inference we attempt to increase the resilience of anomaly detection models to different kinds of distribution shifts. We begin by elucidating a simple yet necessary statistical property that ensures invariant representations, which is critical for robust AD under both domain and covariate shifts. From this property, we derive a regularization term which, when minimized, leads to partial distribution invariance across environments. Through extensive experimental evaluation on both synthetic and real-world tasks, covering a range of six different AD methods, we demonstrated significant improvements in out-of-distribution performance. Under both covariate and domain shift, models regularized with our proposed term showed marked increased robustness. Code is available at: https://github.com/JoaoCarv/invariant-anomaly-detection
****
Stable Bias: Evaluating Societal Representations in Diffusion Models
Sasha Luccioni, Christopher Akiki, Margaret Mitchell, Yacine Jernite
As machine learning-enabled Text-to-Image (TTI) systems are becoming increasingly prevalent and seeing growing adoption as commercial services, characterizing the social biases they exhibit is a necessary first step to lowering their risk of discriminatory outcomes. This evaluation, however, is made more difficult by the synthetic nature of these systems’ outputs: common definitions of diversity are grounded in social categories of people living in the world, whereas the artificial depictions of fictive humans created by these systems have no inherent gender or ethnicity. To address this need, we propose a new method for exploring the social biases in TTI systems. Our approach relies on characterizing the variation in generated images triggered by enumerating gender and ethnicity markers in the prompts, and comparing it to the variation engendered by spanning different professions. This allows us to (1) identify specific bias trends, (2) provide targeted scores to directly compare models in terms of diversity and representation, and (3) jointly model interdependent social variables to support a multidimensional analysis. We leverage this method to analyze images generated by 3 popular TTI systems (Dall·E 2 , Stable Diffusion v 1.4 and 2) and find that all of their outputs show correlations with US labor demographics, they also consistently under-represent marginalized identities to different extents. We also release the datasets and low-code interactive bias exploration platforms developed forthis work, as well as the necessary tools to similarly evaluate additional TTI systems.
****
Locality Sensitive Hashing in Fourier Frequency Domain For Soft Set Containment Search
Indradyumna Roy, Rishi Agarwal, Soumen Chakrabarti, Anirban Dasgupta, Abir De
In many search applications related to passage retrieval, text entailment, and subgraph search, the query and each 'document' is a set of elements, with a document being relevant if it contains the query. These elements are not represented by atomic IDs, but by embedded representations, thereby extending set containment to soft set containment. Recent applications address soft set containment by encoding sets into fixed-size vectors and checking for elementwise vector dominance. This 0/1 property can be relaxed to an asymmetric hinge distance for scoring and ranking candidate documents. Here we focus on data-sensitive, trainable indices for fast retrieval of relevant documents. Existing LSH methods are designed for mostly symmetric or few simple asymmetric distance functions, which are not suitable for hinge distance. Instead, we transform hinge distance into a prop

osed dominance similarity measure, to which we then apply a Fourier transform, thereby expressing dominance similarity as an expectation of inner products of functions in the frequency domain. Next, we approximate the expectation with an importance-sampled estimate. The overall consequence is that now we can use a traditional LSH, but in the frequency domain. To ensure that the LSH uses hash bits efficiently, we learn hash functions that are sensitive to both corpus and query distributions, mapped to the frequency domain. Our experiments show that the proposed asymmetric dominance similarity is critical to the targeted applications, and that our LSH, which we call FourierHashNet, provides a better query time vs. retrieval quality trade-off, compared to several baselines. Both the Fourier transform and the trainable hash codes contribute to performance gains.
****
L-C2ST: Local Diagnostics for Posterior Approximations in Simulation-Based Inference
Julia Linhart, Alexandre Gramfort, Pedro Rodrigues
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Self-Supervised Reinforcement Learning that Transfers using Random Features
Boyuan Chen, Chuning Zhu, Pulkit Agrawal, Kaiqing Zhang, Abhishek Gupta
Model-free reinforcement learning algorithms have exhibited great potential in solving single-task sequential decision-making problems with high-dimensional observations and long horizons, but are known to be hard to generalize across tasks. Model-based RL, on the other hand, learns task-agnostic models of the world that naturally enables transfer across different reward functions, but struggles to scale to complex environments due to the compounding error. To get the best of both worlds, we propose a self-supervised reinforcement learning method that enables the transfer of behaviors across tasks with different rewards, while circumventing the challenges of model-based RL. In particular, we show self-supervised pre-training of model-free reinforcement learning with a number of random features as rewards allows implicit modeling of long-horizon environment dynamics. Then, planning techniques like model-predictive control using these implicit models enable fast adaptation to problems with new reward functions. Our method is self-supervised in that it can be trained on offline datasets without reward labels, but can then be quickly deployed on new tasks. We validate that our proposed method enables transfer across tasks on a variety of manipulation and locomotion domains in simulation, opening the door to generalist decision-making agents.
****
MGDD: A Meta Generator for Fast Dataset Distillation
Songhua Liu, Xinchao Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
New Complexity-Theoretic Frontiers of Tractability for Neural Network Training
Cornelius Brand, Robert Ganian, Mathis Rocton
In spite of the fundamental role of neural networks in contemporary machine learning research, our understanding of the computational complexity of optimally training neural networks remains limited even when dealing with the simplest kinds of activation functions. Indeed, while there has been a number of very recent results that establish ever-tighter lower bounds for the problem under linear and ReLU activation functions, little progress has been made towards the identification of novel polynomial-time tractable network architectures. In this article we obtain novel algorithmic upper bounds for training linear- and ReLU-activated neural networks to optimality which push the boundaries of tractability for these problems beyond the previous state of the art.
****

V-InFoR: A Robust Graph Neural Networks Explainer for Structurally Corrupted Graphs
Senzhang Wang, Jun Yin, Chaozhuo Li, Xing Xie, Jianxin Wang
GNN explanation method aims to identify an explanatory subgraph which contains the most informative components of the full graph. However, a major limitation of existing GNN explainers is that they are not robust to the structurally corrupted graphs, e.g., graphs with noisy or adversarial edges. On the one hand, existing GNN explainers mostly explore explanations based on either the raw graph features or the learned latent representations, both of which can be easily corrupted. On the other hand, the corruptions in graphs are irregular in terms of the structural properties, e.g., the size or connectivity of graphs, which makes the rigorous constraints used by previous GNN explainers unfeasible. To address these issues, we propose a robust GNN explainer called V-InfoR. Specifically, a robust graph representation extractor, which takes insights of variational inference, is proposed to infer the latent distribution of graph representations. Instead of directly using the corrupted raw features or representations of each single graph, we sample the graph representations from the inferred distribution for the downstream explanation generator, which can effectively eliminate the minor corruption. We next formulate the explanation exploration as a graph information bottleneck (GIB) optimization problem. As a more general method that does not need any rigorous structural constraints, our GIB-based method can adaptively capture both the regularity and irregularity of the severely corrupted graphs for explanation. Extensive evaluations on both synthetic and real-world datasets indicate that V-InfoR significantly improves the GNN explanation performance for the structurally corrupted graphs. Code and dataset are available at https://anonymous.4open.science/r/V-InfoR-EF88
****
Beyond Average Return in Markov Decision Processes
Alexandre Marthe, Aurélien Garivier, Claire Vernade
What are the functionals of the reward that can be computed and optimized exactly in Markov Decision Processes?In the finite-horizon, undiscounted setting, Dynamic Programming (DP) can only handle these operations efficiently for certain classes of statistics. We summarize the characterization of these classes for policy evaluation, and give a new answer for the planning problem. Interestingly, we prove that only generalized means can be optimized exactly, even in the more general framework of Distributional Reinforcement Learning (DistRL).DistRL permits, however, to evaluate other functionals approximately. We provide error bounds on the resulting estimators, and discuss the potential of this approach as well as its limitations.These results contribute to advancing the theory of Markov Decision Processes by examining overall characteristics of the return, and particularly risk-conscious strategies.
****
Latent exploration for Reinforcement Learning
Alberto Silvio Chiappa, Alessandro Marin Vargas, Ann Huang, Alexander Mathis
In Reinforcement Learning, agents learn policies by exploring and interacting with the environment. Due to the curse of dimensionality, learning policies that map high-dimensional sensory input to motor output is particularly challenging. During training, state of the art methods (SAC, PPO, etc.) explore the environment by perturbing the actuation with independent Gaussian noise. While this unstructured exploration has proven successful in numerous tasks, it can be suboptimal for overactuated systems. When multiple actuators, such as motors or muscles, drive behavior, uncorrelated perturbations risk diminishing each other's effect, or modifying the behavior in a task-irrelevant way. While solutions to introduce time correlation across action perturbations exist, introducing correlation across actuators has been largely ignored. Here, we propose LATent TIme-Correlated Exploration (Lattice), a method to inject temporally-correlated noise into the latent state of the policy network, which can be seamlessly integrated with on- and off-policy algorithms. We demonstrate that the noisy actions generated by perturbing the network's activations can be modeled as a multivariate Gaussian distribution with a full covariance matrix. In the PyBullet locomotion tasks, Lattic

e-SAC achieves state of the art results, and reaches 18\% higher reward than unstructured exploration in the Humanoid environment. In the musculoskeletal control environments of MyoSuite, Lattice-PPO achieves higher reward in most reaching and object manipulation tasks, while also finding more energy-efficient policies with reductions of 20-60\%. Overall, we demonstrate the effectiveness of structured action noise in time and actuator space for complex motor control tasks. The code is available at: https://github.com/amathislab/lattice.
****
Distributional Model Equivalence for Risk-Sensitive Reinforcement Learning
Tyler Kastner, Murat A. Erdogdu, Amir-massoud Farahmand
We consider the problem of learning models for risk-sensitive reinforcement learning. We theoretically demonstrate that proper value equivalence, a method of learning models which can be used to plan optimally in the risk-neutral setting, is not sufficient to plan optimally in the risk-sensitive setting. We leverage distributional reinforcement learning to introduce two new notions of model equivalence, one which is general and can be used to plan for any risk measure, but is intractable; and a practical variation which allows one to choose which risk measures they may plan optimally for. We demonstrate how our models can be used to augment any model-free risk-sensitive algorithm, and provide both tabular and large-scale experiments to demonstrate our method’s ability.
****
Group Robust Classification Without Any Group Information
Christos Tsirigotis, Joao Monteiro, Pau Rodriguez, David Vazquez, Aaron C. Courville
Empirical risk minimization (ERM) is sensitive to spurious correlations present in training data, which poses a significant risk when deploying systems trained under this paradigm in high-stake applications. While the existing literature focuses on maximizing group-balanced or worst-group accuracy, estimating these quantities is hindered by costly bias annotations. This study contends that current bias-unsupervised approaches to group robustness continue to rely on group information to achieve optimal performance. Firstly, these methods implicitly assume that all group combinations are represented during training. To illustrate this, we introduce a systematic generalization task on the MPI3D dataset and discover that current algorithms fail to improve the ERM baseline when combinations of observed attribute values are missing. Secondly, bias labels are still crucial for effective model selection, restricting the practicality of these methods in real-world scenarios. To address these limitations, we propose a revised methodology for training and validating debiased models in an entirely bias-unsupervised manner. We achieve this by employing pretrained self-supervised models to reliably extract bias information, which enables the integration of a logit adjustment training loss with our validation criterion. Our empirical analysis on synthetic and real-world tasks provides evidence that our approach overcomes the identified challenges and consistently enhances robust accuracy, attaining performance which is competitive with or outperforms that of state-of-the-art methods, which, conversely, rely on bias labels for validation.
****
Tackling Heavy-Tailed Rewards in Reinforcement Learning with Function Approximation: Minimax Optimal and Instance-Dependent Regret Bounds
Jiayi Huang, Han Zhong, Liwei Wang, Lin Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Dictionary for Visual Attention
Yingjie Liu, Xuan Liu, Hui Yu, XUAN TANG, Xian Wei
Recently, the attention mechanism has shown outstanding competence in capturing global structure information and long-range relationships within data, thus enhancing the performance of deep vision models on various computer vision tasks. In this work, we propose a novel dictionary learning-based attention (\textit{Dic-

Attn}) module, which models this issue as a decomposition and reconstruction problem with the sparsity prior, inspired by sparse coding in the human visual perception system. The proposed \textit{Dic-Attn} module decomposes the input into a dictionary and corresponding sparse representations, allowing for the disentanglement of underlying nonlinear structural information in visual data and the reconstruction of an attention embedding. By applying transformation operations in the spatial and channel domains, the module dynamically selects the dictionary's atoms and sparse representations. Finally, the updated dictionary and sparse representations capture the global contextual information and reconstruct the attention maps. The proposed \textit{Dic-Attn} module is designed with plug-and-play compatibility, allowing for integration into deep attention encoders. Our approach offers an intuitive and elegant means to exploit the discriminative information from data, promoting visual attention construction. Extensive experimental results on various computer vision tasks, e.g., image and point cloud classification, validate that our method achieves promising performance, and shows a strong competitive comparison with state-of-the-art attention methods.
****
A Bayesian Take on Gaussian Process Networks
Enrico Giudice, Jack Kuipers, Giusi Moffa
Gaussian Process Networks (GPNs) are a class of directed graphical models which employ Gaussian processes as priors for the conditional expectation of each variable given its parents in the network. The model allows the description of continuous joint distributions in a compact but flexible manner with minimal parametric assumptions on the dependencies between variables. Bayesian structure learning of GPNs requires computing the posterior over graphs of the network and is computationally infeasible even in low dimensions. This work implements Monte Carlo and Markov Chain Monte Carlo methods to sample from the posterior distribution of network structures. As such, the approach follows the Bayesian paradigm, comparing models via their marginal likelihood and computing the posterior probability of the GPN features. Simulation studies show that our method outperforms state-of-the-art algorithms in recovering the graphical structure of the network and provides an accurate approximation of its posterior distribution.
****
Exploring Question Decomposition for Zero-Shot VQA
Zaid Khan, Vijay Kumar B G, Samuel Schulter, Manmohan Chandraker, Yun Fu
Visual question answering (VQA) has traditionally been treated as a single-step task where each question receives the same amount of effort, unlike natural human question-answering strategies. We explore a question decomposition strategy for VQA to overcome this limitation. We probe the ability of recently developed large vision-language models to use human-written decompositions and produce their own decompositions of visual questions, finding they are capable of learning both tasks from demonstrations alone.However, we show that naive application of model-written decompositions can hurt performance.We introduce a model-driven selective decomposition approach for second-guessing predictions and correcting errors, and validate its effectiveness on eight VQA tasks across three domains, showing consistent improvements in accuracy, including improvements of >20% on medical VQA datasets and boosting the zero-shot performance of BLIP-2 above chance on a VQA reformulation of the challenging Winoground task. Project Site: https://zaidkhan.me/decomposition-0shot-vqa/
****
Sharp Recovery Thresholds of Tensor PCA Spectral Algorithms
Michael Feldman, David Donoho
Many applications seek to recover low-rank approximations of noisy tensor data. We consider several practical and effective matricization strategies which construct specific matrices from such tensors and then apply spectral methods; the strategies include tensor unfolding, partial tracing, power iteration, and recursive unfolding. We settle the behaviors of unfolding and partial tracing, identifying sharp thresholds in signal-to-noise ratio above which the signal is partially recovered. In particular, we extend previous results to a much larger class of tensor shapes where axis lengths may be different. For power iteration and recu

rsive unfolding, we prove that under conditions where previous algorithms partially recovery the signal, these methods achieve (asymptotically) exact recovery. Our analysis deploys random matrix theory to obtain sharp thresholds which elude perturbation and concentration bounds. Specifically, we rely upon recent disproportionate random matrix results, which describe sequences of matrices with diverging aspect ratio.
****
R-divergence for Estimating Model-oriented Distribution Discrepancy
Zhilin Zhao, Longbing Cao
Real-life data are often non-IID due to complex distributions and interactions, and the sensitivity to the distribution of samples can differ among learning models. Accordingly, a key question for any supervised or unsupervised model is whether the probability distributions of two given datasets can be considered identical. To address this question, we introduce R-divergence, designed to assess model-oriented distribution discrepancies. The core insight is that two distributions are likely identical if their optimal hypothesis yields the same expected risk for each distribution. To estimate the distribution discrepancy between two datasets, R-divergence learns a minimum hypothesis on the mixed data and then gauges the empirical risk difference between them. We evaluate the test power across various unsupervised and supervised tasks and find that R-divergence achieves state-of-the-art performance. To demonstrate the practicality of R-divergence, we employ R-divergence to train robust neural networks on samples with noisy labels.
****
On-the-Fly Adapting Code Summarization on Trainable Cost-Effective Language Models
Yufan Cai, Yun Lin, Chenyan Liu, Jinglian Wu, Yifan Zhang, Yiming Liu, Yeyun Gong, Jin Song Dong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Exploring and Interacting with the Set of Good Sparse Generalized Additive Models
Chudi Zhong, Zhi Chen, Jiachang Liu, Margo Seltzer, Cynthia Rudin
In real applications, interaction between machine learning models and domain experts is critical; however, the classical machine learning paradigm that usually produces only a single model does not facilitate such interaction. Approximating and exploring the Rashomon set, i.e., the set of all near-optimal models, addresses this practical challenge by providing the user with a searchable space containing a diverse set of models from which domain experts can choose. We present algorithms to efficiently and accurately approximate the Rashomon set of sparse, generalized additive models with ellipsoids for fixed support sets and use these ellipsoids to approximate Rashomon sets for many different support sets. The approximated Rashomon set serves as a cornerstone to solve practical challenges such as (1) studying the variable importance for the model class; (2) finding models under user-specified constraints (monotonicity, direct editing); and (3) investigating sudden changes in the shape functions. Experiments demonstrate the fidelity of the approximated Rashomon set and its effectiveness in solving practical challenges.
****
Convergence Analysis of Sequential Federated Learning on Heterogeneous Data
Yipeng Li, Xinchen Lyu
There are two categories of methods in Federated Learning (FL) for joint training across multiple clients: i) parallel FL (PFL), where clients train models in a parallel manner; and ii) sequential FL (SFL), where clients train models in a sequential manner. In contrast to that of PFL, the convergence theory of SFL on heterogeneous data is still lacking. In this paper, we establish the convergence guarantees of SFL for strongly/general/non-convex objectives on heterogeneous da

ta. The convergence guarantees of SFL are better than that of PFL on heterogeneous data with both full and partial client participation. Experimental results validate the counterintuitive analysis result that SFL outperforms PFL on extremely heterogeneous data in cross-device settings.
****
Disambiguated Attention Embedding for Multi-Instance Partial-Label Learning
Wei Tang, Weijia Zhang, Min-Ling Zhang
In many real-world tasks, the concerned objects can be represented as a multi-instance bag associated with a candidate label set, which consists of one ground-truth label and several false positive labels. Multi-instance partial-label learning (MIPL) is a learning paradigm to deal with such tasks and has achieved favorable performances. Existing MIPL approach follows the instance-space paradigm by assigning augmented candidate label sets of bags to each instance and aggregating bag-level labels from instance-level labels. However, this scheme may be suboptimal as global bag-level information is ignored and the predicted labels of bags are sensitive to predictions of negative instances. In this paper, we study an alternative scheme where a multi-instance bag is embedded into a single vector representation. Accordingly, an intuitive algorithm named DEMIPL, i.e., Disambiguated attention Embedding for Multi-Instance Partial-Label learning, is proposed. DEMIPL employs a disambiguation attention mechanism to aggregate a multi-instance bag into a single vector representation, followed by a momentum-based disambiguation strategy to identify the ground-truth label from the candidate label set. Furthermore, we introduce a real-world MIPL dataset for colorectal cancer classification. Experimental results on benchmark and real-world datasets validate the superiority of DEMIPL against the compared MIPL and partial-label learning approaches.
****
Efficient Batched Algorithm for Contextual Linear Bandits with Large Action Space via Soft Elimination
Osama Hanna, Lin Yang, Christina Fragouli
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Add and Thin: Diffusion for Temporal Point Processes
David Lüdke, Marin Biloš, Oleksandr Shchur, Marten Lienen, Stephan Günnemann
Autoregressive neural networks within the temporal point process (TPP) framework have become the standard for modeling continuous-time event data. Even though these models can expressively capture event sequences in a one-step-ahead fashion, they are inherently limited for long-term forecasting applications due to the accumulation of errors caused by their sequential nature. To overcome these limitations, we derive ADD-THIN, a principled probabilistic denoising diffusion model for TPPs that operates on entire event sequences. Unlike existing diffusion approaches, ADD-THIN naturally handles data with discrete and continuous components. In experiments on synthetic and real-world datasets, our model matches the state-of-the-art TPP models in density estimation and strongly outperforms them in forecasting.
****
Pitfall of Optimism: Distributional Reinforcement Learning by Randomizing Risk Criterion
Taehyun Cho, Seungyub Han, Heesoo Lee, Kyungjae Lee, Jungwoo Lee
Distributional reinforcement learning algorithms have attempted to utilize estimated uncertainty for exploration, such as optimism in the face of uncertainty. However, using the estimated variance for optimistic exploration may cause biased data collection and hinder convergence or performance. In this paper, we present a novel distributional reinforcement learning that selects actions by randomizing risk criterion without losing the risk-neutral objective. We provide a perturbed distributional Bellman optimality operator by distorting the risk measure. Also,we prove the convergence and optimality of the proposed method with the wea

ker contraction property. Our theoretical results support that the proposed meth
od does not fall into biased exploration and is guaranteed to converge to an opt
imal return. Finally, we empirically show that our method outperforms other exis
ting distribution-based algorithms in various environments including Atari 55 ga
mes.
****
Efficient Meta Neural Heuristic for Multi-Objective Combinatorial Optimization
Jinbiao Chen, Jiahai Wang, Zizhen Zhang, Zhiguang Cao, Te Ye, Siyuan Chen
Recently, neural heuristics based on deep reinforcement learning have exhibited
promise in solving multi-objective combinatorial optimization problems (MOCOPs).
 However, they are still struggling to achieve high learning efficiency and solu
tion quality. To tackle this issue, we propose an efficient meta neural heuristi
c (EMNH), in which a meta-model is first trained and then fine-tuned with a few
steps to solve corresponding single-objective subproblems. Specifically, for the
 training process, a (partial) architecture-shared multi-task model is leveraged
 to achieve parallel learning for the meta-model, so as to speed up the training
; meanwhile, a scaled symmetric sampling method with respect to the weight vecto
rs is designed to stabilize the training. For the fine-tuning process, an effici
ent hierarchical method is proposed to systematically tackle all the subproblems
. Experimental results on the multi-objective traveling salesman problem (MOTSP)
, multi-objective capacitated vehicle routing problem (MOCVRP), and multi-object
ive knapsack problem (MOKP) show that, EMNH is able to outperform the state-of-t
he-art neural heuristics in terms of solution quality and learning efficiency, a
nd yield competitive solutions to the strong traditional heuristics while consum
ing much shorter time.
****
QuantSR: Accurate Low-bit Quantization for Efficient Image Super-Resolution
Haotong Qin, Yulun Zhang, Yifu Ding, Yifan liu, Xianglong Liu, Martin Danelljan,
 Fisher Yu
Low-bit quantization in image super-resolution (SR) has attracted copious attent
ion in recent research due to its ability to reduce parameters and operations si
gnificantly. However, many quantized SR models suffer from accuracy degradation
compared to their full-precision counterparts, especially at ultra-low bit width
s (2-4 bits), limiting their practical applications. To address this issue, we p
ropose a novel quantized image SR network, called QuantSR, which achieves accura
te and efficient SR processing under low-bit quantization. To overcome the repre
sentation homogeneity caused by quantization in the network, we introduce the Re
distribution-driven Learnable Quantizer (RLQ). This is accomplished through an i
nference-agnostic efficient redistribution design, which adds additional informa
tion in both forward and backward passes to improve the representation ability o
f quantized networks. Furthermore, to achieve flexible inference and break the u
pper limit of accuracy, we propose the Depth-dynamic Quantized Architecture (DQA
). Our DQA allows for the trade-off between efficiency and accuracy during infer
ence through weight sharing. Our comprehensive experiments show that QuantSR out
performs existing state-of-the-art quantized SR networks in terms of accuracy wh
ile also providing more competitive computational efficiency. In addition, we de
monstrate the scheme's satisfactory architecture generality by providing QuantSR
-C and QuantSR-T for both convolution and Transformer versions, respectively. Ou
r code and models are released at https://github.com/htqin/QuantSR .
****
Streaming Factor Trajectory Learning for Temporal Tensor Decomposition
Shikai Fang, Xin Yu, Shibo Li, Zheng Wang, Mike Kirby, Shandian Zhe
Practical tensor data is often along with time information. Most existing tempor
al decomposition approaches estimate a set of fixed factors for the objects in e
ach tensor mode, and hence cannot capture the temporal evolution of the objects'
 representation. More important, we lack an effective approach to capture such e
volution from streaming data, which is common in real-world applications. To ad
dress these issues, we propose Streaming Factor Trajectory Learning (SFTL) for t
emporal tensor decomposition. We use Gaussian processes (GPs) to model the traje
ctory of factors so as to flexibly estimate their temporal evolution. To addres

s the computational challenges in handling streaming data, we convert the GPs in to a state-space prior by constructing an equivalent stochastic differential equation (SDE). We develop an efficient online filtering algorithm to estimate a decoupled running posterior of the involved factor states upon receiving new data. The decoupled estimation enables us to conduct standard Rauch-Tung-Striebel smoothing to compute the full posterior of all the trajectories in parallel, without the need for revisiting any previous data. We have shown the advantage of SFTL in both synthetic tasks and real-world applications.
****
DDF-HO: Hand-Held Object Reconstruction via Conditional Directed Distance Field
Chenyangguang Zhang, Yan Di, Ruida Zhang, Guangyao Zhai, Fabian Manhardt, Federico Tombari, Xiangyang Ji
Reconstructing hand-held objects from a single RGB image is an important and challenging problem. Existing works utilizing Signed Distance Fields (SDF) reveal limitations in comprehensively capturing the complex hand-object interactions, since SDF is only reliable within the proximity of the target, and hence, infeasible to simultaneously encode local hand and object cues. To address this issue, we propose DDF-HO, a novel approach leveraging Directed Distance Field (DDF) as the shape representation. Unlike SDF, DDF maps a ray in 3D space, consisting of an origin and a direction, to corresponding DDF values, including a binary visibility signal determining whether the ray intersects the objects and a distance value measuring the distance from origin to target in the given direction. We randomly sample multiple rays and collect local to global geometric features for them by introducing a novel 2D ray-based feature aggregation scheme and a 3D intersection-aware hand pose embedding, combining 2D-3D features to model hand-object interactions. Extensive experiments on synthetic and real-world datasets demonstrate that DDF-HO consistently outperforms all baseline methods by a large margin, especially under Chamfer Distance, with about 80% leap forward. Codes are available at https://github.com/ZhangCYG/DDFHO.
****
Effective Targeted Attacks for Adversarial Self-Supervised Learning
Minseon Kim, Hyeonjeong Ha, Sooel Son, Sung Ju Hwang
Recently, unsupervised adversarial training (AT) has been highlighted as a means of achieving robustness in models without any label information. Previous studies in unsupervised AT have mostly focused on implementing self-supervised learning (SSL) frameworks, which maximize the instance-wise classification loss to generate adversarial examples. However, we observe that simply maximizing the self-supervised training loss with an untargeted adversarial attack often results in generating ineffective adversaries that may not help improve the robustness of the trained model, especially for non-contrastive SSL frameworks without negative examples. To tackle this problem, we propose a novel positive mining for targeted adversarial attack to generate effective adversaries for adversarial SSL frameworks. Specifically, we introduce an algorithm that selects the most confusing yet similar target example for a given instance based on entropy and similarity, and subsequently perturbs the given instance towards the selected target. Our method demonstrates significant enhancements in robustness when applied to non-contrastive SSL frameworks, and less but consistent robustness improvements with contrastive SSL frameworks, on the benchmark datasets.
****
Statistical Guarantees for Variational Autoencoders using PAC-Bayesian Theory
Sokhna Diarra Mbacke, Florence Clerc, Pascal Germain
Since their inception, Variational Autoencoders (VAEs) have become central in machine learning. Despite their widespread use, numerous questions regarding their theoretical properties remain open. Using PAC-Bayesian theory, this work develops statistical guarantees for VAEs. First, we derive the first PAC-Bayesian bound for posterior distributions conditioned on individual samples from the data-generating distribution. Then, we utilize this result to develop generalization guarantees for the VAE's reconstruction loss, as well as upper bounds on the distance between the input and the regenerated distributions. More importantly, we provide upper bounds on the Wasserstein distance between the input distribution an

d the distribution defined by the VAE's generative model.
****
Multi-Head Adapter Routing for Cross-Task Generalization
Lucas Page-Caccia, Edoardo Maria Ponti, Zhan Su, Matheus Pereira, Nicolas Le Roux, Alessandro Sordoni
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
GenS: Generalizable Neural Surface Reconstruction from Multi-View Images
Rui Peng, Xiaodong Gu, Luyang Tang, Shihe Shen, Fanqi Yu, Ronggang Wang
Combining the signed distance function (SDF) and differentiable volume rendering has emerged as a powerful paradigm for surface reconstruction from multi-view images without 3D supervision. However, current methods are impeded by requiring long-time per-scene optimizations and cannot generalize to new scenes. In this paper, we present GenS, an end-to-end generalizable neural surface reconstruction model. Unlike coordinate-based methods that train a separate network for each scene, we construct a generalized multi-scale volume to directly encode all scenes. Compared with existing solutions, our representation is more powerful, which can recover high-frequency details while maintaining global smoothness. Meanwhile, we introduce a multi-scale feature-metric consistency to impose the multi-view consistency in a more discriminative multi-scale feature space, which is robust to the failures of the photometric consistency. And the learnable feature can be self-enhanced to continuously improve the matching accuracy and mitigate aggregation ambiguity. Furthermore, we design a view contrast loss to force the model to be robust to those regions covered by few viewpoints through distilling the geometric prior from dense input to sparse input. Extensive experiments on popular benchmarks show that our model can generalize well to new scenes and outperform existing state-of-the-art methods even those employing ground-truth depth supervision. Code will be available at https://github.com/prstrive/GenS.
****
Better with Less: A Data-Active Perspective on Pre-Training Graph Neural Networks
Jiarong Xu, Renhong Huang, XIN JIANG, Yuxuan Cao, Carl Yang, Chunping Wang, YANG YANG
Pre-training on graph neural networks (GNNs) aims to learn transferable knowledge for downstream tasks with unlabeled data, and it has recently become an active research area. The success of graph pre-training models is often attributed to the massive amount of input data. In this paper, however, we identify the curse of big data phenomenon in graph pre-training: more training data do not necessarily lead to better downstream performance. Motivated by this observation, we propose a better-with-less framework for graph pre-training: fewer, but carefully chosen data are fed into a GNN model to enhance pre-training. The proposed pre-training pipeline is called the data-active graph pre-training (APT) framework, and is composed of a graph selector and a pre-training model. The graph selector chooses the most representative and instructive data points based on the inherent properties of graphs as well as predictive uncertainty. The proposed predictive uncertainty, as feedback from the pre-training model, measures the confidence level of the model in the data. When fed with the chosen data, on the other hand, the pre-training model grasps an initial understanding of the new, unseen data, and at the same time attempts to remember the knowledge learned from previous data. Therefore, the integration and interaction between these two components form a unified framework (APT), in which graph pre-training is performed in a progressive and iterative way. Experiment results show that the proposed APT is able to obtain an efficient pre-training model with fewer training data and better downstream performance.
****
Brain-like Flexible Visual Inference by Harnessing Feedback Feedforward Alignment

Tahereh Toosi, Elias Issa
In natural vision, feedback connections support versatile visual inference capabilities such as making sense of the occluded or noisy bottom-up sensory information or mediating pure top-down processes such as imagination. However, the mechanisms by which the feedback pathway learns to give rise to these capabilities flexibly are not clear. We propose that top-down effects emerge through alignment between feedforward and feedback pathways, each optimizing its own objectives. To achieve this co-optimization, we introduce Feedback-Feedforward Alignment (FFA), a learning algorithm that leverages feedback and feedforward pathways as mutual credit assignment computational graphs, enabling alignment. In our study, we demonstrate the effectiveness of FFA in co-optimizing classification and reconstruction tasks on widely used MNIST and CIFAR10 datasets. Notably, the alignment mechanism in FFA endows feedback connections with emergent visual inference functions, including denoising, resolving occlusions, hallucination, and imagination. Moreover, FFA offers bio-plausibility compared to traditional backpropagation (BP) methods in implementation. By repurposing the computational graph of credit assignment into a goal-driven feedback pathway, FFA alleviates weight transport problems encountered in BP, enhancing the bio-plausibility of the learning algorithm. Our study presents FFA as a promising proof-of-concept for the mechanisms underlying how feedback connections in the visual cortex support flexible visual functions. This work also contributes to the broader field of visual inference underlying perceptual phenomena and has implications for developing more biologically inspired learning algorithms.
****
Latent Diffusion for Language Generation
Justin Lovelace, Varsha Kishore, Chao Wan, Eliot Shekhtman, Kilian Q. Weinberger
Diffusion models have achieved great success in modeling continuous data modalities such as images, audio, and video, but have seen limited use in discrete domains such as language. Recent attempts to adapt diffusion to language have presented diffusion as an alternative to existing pretrained language models. We view diffusion and existing language models as complementary. We demonstrate that encoder-decoder language models can be utilized to efficiently learn high-quality language autoencoders. We then demonstrate that continuous diffusion models can be learned in the latent space of the language autoencoder, enabling us to sample continuous latent representations that can be decoded into natural language with the pretrained decoder. We validate the effectiveness of our approach for unconditional, class-conditional, and sequence-to-sequence language generation. We demonstrate across multiple diverse data sets that our latent language diffusion models are significantly more effective than previous diffusion language models. Our code is available at \url{https://github.com/justinlovelace/latent-diffusion-for-language}.
****
The emergence of clusters in self-attention dynamics
Borjan Geshkovski, Cyril Letrouit, Yury Polyanskiy, Philippe Rigollet
Viewing Transformers as interacting particle systems, we describe the geometry of learned representations when the weights are not time-dependent. We show that particles, representing tokens, tend to cluster toward particular limiting objects as time tends to infinity. Using techniques from dynamical systems and partial differential equations, we show that type of limiting object that emerges depends on the spectrum of the value matrix. Additionally, in the one-dimensional case we prove that the self-attention matrix converges to a low-rank Boolean matrix. The combination of these results mathematically confirms the empirical observation made by Vaswani et al. [VSP`17] that leaders appear in a sequence of tokens when processed by Transformers.
****
Self-Consistent Velocity Matching of Probability Flows
Lingxiao Li, Samuel Hurault, Justin M. Solomon
We present a discretization-free scalable framework for solving a large class of mass-conserving partial differential equations (PDEs), including the time-dependent Fokker-Planck equation and the Wasserstein gradient flow. The main observat

ion is that the time-varying velocity field of the PDE solution needs to be self-consistent: it must satisfy a fixed-point equation involving the probability flow characterized by the same velocity field. Instead of directly minimizing the residual of the fixed-point equation with neural parameterization, we use an iterative formulation with a biased gradient estimator that bypasses significant computational obstacles with strong empirical performance. Compared to existing approaches, our method does not suffer from temporal or spatial discretization, covers a wider range of PDEs, and scales to high dimensions. Experimentally, our method recovers analytical solutions accurately when they are available and achieves superior performance in high dimensions with less training time compared to alternatives.
****
Deep Momentum Multi-Marginal Schrödinger Bridge
Tianrong Chen, Guan-Horng Liu, Molei Tao, Evangelos Theodorou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Semi-Supervised Contrastive Learning for Deep Regression with Ordinal Rankings from Spectral Seriation
Weihang Dai, Yao DU, Hanru Bai, Kwang-Ting Cheng, Xiaomeng Li
Contrastive learning methods can be applied to deep regression by enforcing label distance relationships in feature space. However, these methods are limited to labeled data only unlike for classification, where unlabeled data can be used for contrastive pretraining. In this work, we extend contrastive regression methods to allow unlabeled data to be used in a semi-supervised setting, thereby reducing the reliance on manual annotations. We observe that the feature similarity matrix between unlabeled samples still reflect inter-sample relationships, and that an accurate ordinal relationship can be recovered through spectral seriation algorithms if the level of error is within certain bounds. By using the recovered ordinal relationship for contrastive learning on unlabeled samples, we can allow more data to be used for feature representation learning, thereby achieve more robust results. The ordinal rankings can also be used to supervise predictions on unlabeled samples, which can serve as an additional training signal. We provide theoretical guarantees and empirical support through experiments on different datasets, demonstrating that our method can surpass existing state-of-the-art semi-supervised deep regression methods. To the best of our knowledge, this work is the first to explore using unlabeled data to perform contrastive learning for regression.
****
Domain Re-Modulation for Few-Shot Generative Domain Adaptation
Yi Wu, Ziqiang Li, Chaoyue Wang, Heliang Zheng, Shanshan Zhao, Bin Li, Dacheng Tao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Transformers as Statisticians: Provable In-Context Learning with In-Context Algorithm Selection
Yu Bai, Fan Chen, Huan Wang, Caiming Xiong, Song Mei
Neural sequence models based on the transformer architecture have demonstrated remarkable \emph{in-context learning} (ICL) abilities, where they can perform new tasks when prompted with training and test examples, without any parameter update to the model. This work first provides a comprehensive statistical theory for transformers to perform ICL. Concretely, we show that transformers can implement a broad class of standard machine learning algorithms in context, such as least squares, ridge regression, Lasso, learning generalized linear models, and gradient descent on two-layer neural networks, with near-optimal predictive power on

various in-context data distributions. Using an efficient implementation of in-context gradient descent as the underlying mechanism, our transformer constructions admit mild size bounds, and can be learned with polynomially many pretraining sequences.   Building on these ``base'' ICL algorithms, intriguingly, we show that transformers can implement more complex ICL procedures involving \emph{in-context algorithm selection}, akin to what a statistician can do in real life---A \emph{single} transformer can adaptively select different base ICL algorithms---or even perform qualitatively different tasks---on different input sequences, without any explicit prompting of the right algorithm or task. We both establish this in theory by explicit constructions, and also observe this phenomenon experimentally. In theory, we construct two general mechanisms for algorithm selection with concrete examples: pre-ICL testing, and post-ICL validation. As an example, we use the post-ICL validation mechanism to construct a transformer that can perform nearly Bayes-optimal ICL on a challenging task---noisy linear models with mixed noise levels. Experimentally, we demonstrate the strong in-context algorithm selection capabilities of standard transformer architectures.
****
Arbitrarily Scalable Environment Generators via Neural Cellular Automata
Yulun Zhang, Matthew Fontaine, Varun Bhatt, Stefanos Nikolaidis, Jiaoyang Li
We study the problem of generating arbitrarily large environments to improve the throughput of multi-robot systems. Prior work proposes Quality Diversity (QD) algorithms as an effective method for optimizing the environments of automated warehouses. However, these approaches optimize only relatively small environments, falling short when it comes to replicating real-world warehouse sizes. The challenge arises from the exponential increase in the search space as the environment size increases. Additionally, the previous methods have only been tested with up to 350 robots in simulations, while practical warehouses could host thousands of robots. In this paper, instead of optimizing environments, we propose to optimize Neural Cellular Automata (NCA) environment generators via QD algorithms. We train a collection of NCA generators with QD algorithms in small environments and then generate arbitrarily large environments from the generators at test time. We show that NCA environment generators maintain consistent, regularized patterns regardless of environment size, significantly enhancing the scalability of multi-robot systems in two different domains with up to 2,350 robots. Additionally, we demonstrate that our method scales a single-agent reinforcement learning policy to arbitrarily large environments with similar patterns. We include the source code at https://github.com/lunjohnzhang/warehouseenvgenncapublic.
****
FAMO: Fast Adaptive Multitask Optimization
Bo Liu, Yihao Feng, Peter Stone, Qiang Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Theory of Multimodal Learning
Zhou Lu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
IDEA: An Invariant Perspective for Efficient Domain Adaptive Image Retrieval
Haixin Wang, Hao Wu, Jinan Sun, Shikun Zhang, Chong Chen, Xian-Sheng Hua, Xiao Luo
In this paper, we investigate the problem of unsupervised domain adaptive hashing, which leverage knowledge from a label-rich source domain to expedite learning to hash on a label-scarce target domain. Although numerous existing approaches attempt to incorporate transfer learning techniques into deep hashing frameworks, they often neglect the essential invariance for adequate alignment between the

se two domains. Worse yet, these methods fail to distinguish between causal and non-causal effects embedded in images, rendering cross-domain retrieval ineffective. To address these challenges, we propose an Invariance-acquired Domain AdaptivE HAshing (IDEA) model. Our IDEA first decomposes each image into a causal feature representing label information, and a non-causal feature indicating domain information. Subsequently, we generate discriminative hash codes using causal features with consistency learning on both source and target domains. More importantly, we employ a generative model for synthetic samples to simulate the intervention of various non-causal effects, ultimately minimizing their impact on hash codes for domain invariance. Comprehensive experiments conducted on benchmark datasets validate the superior performance of our IDEA compared to a variety of competitive baselines.
****
Learning Provably Robust Estimators for Inverse Problems via Jittering
Anselm Krainovic, Mahdi Soltanolkotabi, Reinhard Heckel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Occlusions in Video Action Detection: Benchmark Datasets And Training Recipes
Rajat Modi, Vibhav Vineet, Yogesh Rawat
This paper explores the impact of occlusions in video action detection. We facilitatethis study by introducing five new benchmark datasets namely O-UCF and O-JHMDB consisting of synthetically controlled static/dynamic occlusions, OVIS-UCF and OVIS-JHMDB consisting of occlusions with realistic motions and Real-OUCF for occlusions in realistic-world scenarios. We formally confirm an intuitiveexpectation: existing models suffer a lot as occlusion severity is increased andexhibit different behaviours when occluders are static vs when they are moving.We discover several intriguing phenomenon emerging in neural nets: 1) transformerscan naturally outperform CNN models which might have even used occlusion as aform of data augmentation during training 2) incorporating symbolic-componentslike capsules to such backbones allows them to bind to occluders never even seenduring training and 3) Islands of agreement (similar to the ones hypothesized inHinton et Al’s GLOM) can emerge in realistic images/videos without instance-levelsupervision, distillation or contrastive-based objectives(eg. video-textual training).Such emergent properties allow us to derive simple yet effective training recipeswhich lead to robust occlusion models inductively satisfying the first two stages ofthe binding mechanism (grouping/segregation). Models leveraging these recipesoutperform existing video action-detectors under occlusion by 32.3% on O-UCF,32.7% on O-JHMDB & 2.6% on Real-OUCF in terms of the vMAP metric. The code for this work has been released at https: //github.com/rajatmodi62/OccludedActionBenchmark.
****
Black-box Backdoor Defense via Zero-shot Image Purification
Yucheng Shi, Mengnan Du, Xuansheng Wu, Zihan Guan, Jin Sun, Ninghao Liu
Backdoor attacks inject poisoned samples into the training data, resulting in the misclassification of the poisoned input during a model's deployment. Defending against such attacks is challenging, especially for real-world black-box models where only query access is permitted. In this paper, we propose a novel defense framework against backdoor attacks through Zero-shot Image Purification (ZIP). Our framework can be applied to poisoned models without requiring internal information about the model or any prior knowledge of the clean/poisoned samples. Our defense framework involves two steps. First, we apply a linear transformation (e.g., blurring) on the poisoned image to destroy the backdoor pattern. Then, we use a pre-trained diffusion model to recover the missing semantic information removed by the transformation. In particular, we design a new reverse process by using the transformed image to guide the generation of high-fidelity purified images, which works in zero-shot settings. We evaluate our ZIP framework on multiple datasets with different types of attacks. Experimental results demonstrate the

superiority of our ZIP framework compared to state-of-the-art backdoor defense baselines. We believe that our results will provide valuable insights for future defense methods for black-box models. Our code is available at https://github.com/sycny/ZIP.
****
Beyond NTK with Vanilla Gradient Descent: A Mean-Field Analysis of Neural Networks with Polynomial Width, Samples, and Time
Arvind Mahankali, Haochen Zhang, Kefan Dong, Margalit Glasgow, Tengyu Ma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Real-Time Motion Prediction via Heterogeneous Polyline Transformer with Relative Pose Encoding
Zhejun Zhang, Alexander Liniger, Christos Sakaridis, Fisher Yu, Luc V Gool
The real-world deployment of an autonomous driving system requires its components to run on-board and in real-time, including the motion prediction module that predicts the future trajectories of surrounding traffic participants. Existing agent-centric methods have demonstrated outstanding performance on public benchmarks. However, they suffer from high computational overhead and poor scalability as the number of agents to be predicted increases. To address this problem, we introduce the K-nearest neighbor attention with relative pose encoding (KNARPE), a novel attention mechanism allowing the pairwise-relative representation to be used by Transformers. Then, based on KNARPE we present the Heterogeneous Polyline Transformer with Relative pose encoding (HPTR), a hierarchical framework enabling asynchronous token update during the online inference. By sharing contexts among agents and reusing the unchanged contexts, our approach is as efficient as scene-centric methods, while performing on par with state-of-the-art agent-centric methods. Experiments on Waymo and Argoverse-2 datasets show that HPTR achieves superior performance among end-to-end methods that do not apply expensive post-processing or model ensembling. The code is available at https://github.com/zhejz/HPTR.
****
Customizable Image Synthesis with Multiple Subjects
Zhiheng Liu, Yifei Zhang, Yujun Shen, Kecheng Zheng, Kai Zhu, Ruili Feng, Yu Liu, Deli Zhao, Jingren Zhou, Yang Cao
Synthesizing images with user-specified subjects has received growing attention due to its practical applications. Despite the recent success in single subject customization, existing algorithms suffer from high training cost and low success rate along with increased number of subjects. Towards controllable image synthesis with multiple subjects as the constraints, this work studies how to efficiently represent a particular subject as well as how to appropriately compose different subjects. We find that the text embedding regarding the subject token already serves as a simple yet effective representation that supports arbitrary combinations without any model tuning. Through learning a residual on top of the base embedding, we manage to robustly shift the raw subject to the customized subject given various text conditions. We then propose to employ layout, a very abstract and easy-to-obtain prior, as the spatial guidance for subject arrangement. By rectifying the activations in the cross-attention map, the layout appoints and separates the location of different subjects in the image, significantly alleviating the interference across them. Using cross-attention map as the intermediary, we could strengthen the signal of target subjects and weaken the signal of irrelevant subjects within a certain region, significantly alleviating the interference across subjects. Both qualitative and quantitative experimental results demonstrate our superiority over state-of-the-art alternatives under a variety of settings for multi-subject customization.
****
How do Minimum-Norm Shallow Denoisers Look in Function Space?
Chen Zeno, Greg Ongie, Yaniv Blumenfeld, Nir Weinberger, Daniel Soudry

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Lookup Table meets Local Laplacian Filter: Pyramid Reconstruction Network for Tone Mapping
Feng Zhang, Ming Tian, Zhiqiang Li, Bin Xu, Qingbo Lu, Changxin Gao, Nong Sang
Tone mapping aims to convert high dynamic range (HDR) images to low dynamic range (LDR) representations, a critical task in the camera imaging pipeline. In recent years, 3-Dimensional LookUp Table (3D LUT) based methods have gained attention due to their ability to strike a favorable balance between enhancement performance and computational efficiency. However, these methods often fail to deliver satisfactory results in local areas since the look-up table is a global operator for tone mapping, which works based on pixel values and fails to incorporate crucial local information. To this end, this paper aims to address this issue by exploring a novel strategy that integrates global and local operators by utilizing closed-form Laplacian pyramid decomposition and reconstruction. Specifically, we employ image-adaptive 3D LUTs to manipulate the tone in the low-frequency image by leveraging the specific characteristics of the frequency information. Furthermore, we utilize local Laplacian filters to refine the edge details in the high-frequency components in an adaptive manner. Local Laplacian filters are widely used to preserve edge details in photographs, but their conventional usage involves manual tuning and fixed implementation within camera imaging pipelines or photo editing tools. We propose to learn parameter value maps progressively for local Laplacian filters from annotated data using a lightweight network. Our model achieves simultaneous global tone manipulation and local edge detail preservation in an end-to-end manner. Extensive experimental results on two benchmark datasets demonstrate that the proposed method performs favorably against state-of-the-art methods.
****
Masked Image Residual Learning for Scaling Deeper Vision Transformers
Guoxi Huang, Hongtao Fu, Adrian G. Bors
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Revisiting Area Convexity: Faster Box-Simplex Games and Spectrahedral Generalizations
Arun Jambulapati, Kevin Tian
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Adversarial Learning for Feature Shift Detection and Correction
Míriam Barrabés, Daniel Mas Montserrat, Margarita Geleta, Xavier Giró-i-Nieto, Alexander Ioannidis
Data shift is a phenomenon present in many real-world applications, and while there are multiple methods attempting to detect shifts, the task of localizing and correcting the features originating such shifts has not been studied in depth. Feature shifts can occur in many datasets, including in multi-sensor data, where some sensors are malfunctioning, or in tabular and structured data, including biomedical, financial, and survey data, where faulty standardization and data processing pipelines can lead to erroneous features. In this work, we explore using the principles of adversarial learning, where the information from several discriminators trained to distinguish between two distributions is used to both detect the corrupted features and fix them in order to remove the distribution shift between datasets. We show that mainstream supervised classifiers, such as rando

m forest or gradient boosting trees, combined with simple iterative heuristics, can localize and correct feature shifts, outperforming current statistical and neural network-based techniques. The code is available at https://github.com/AI-sandbox/DataFix.
****
General Munchausen Reinforcement Learning with Tsallis Kullback-Leibler Divergence
Lingwei Zhu, Zheng Chen, Matthew Schlegel, Martha White
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Residual Alignment: Uncovering the Mechanisms of Residual Networks
Jianing Li, Vardan Papyan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Globally injective and bijective neural operators
Takashi Furuya, Michael Puthawala, Matti Lassas, Maarten V. de Hoop
Recently there has been great interest in operator learning, where networks learn operators between function spaces from an essentially infinite-dimensional perspective. In this work we present results for when the operators learned by these networks are injective and surjective. As a warmup, we combine prior work in both the finite-dimensional ReLU and operator learning setting by giving sharp conditions under which ReLU layers with linear neural operators are injective. We then consider the case when the activation function is pointwise bijective and obtain sufficient conditions for the layer to be injective. We remark that this question, while trivial in the finite-rank setting, is subtler in the infinite-rank setting and is proven using tools from Fredholm theory. Next, we prove that our supplied injective neural operators are universal approximators and that their implementation, with finite-rank neural networks, are still injective. This ensures that injectivity is not 'lost' in the transcription from analytical operators to their finite-rank implementation with networks. Finally, we conclude with an increase in abstraction and consider general conditions when subnetworks, which may have many layers, are injective and surjective and provide an exact inversion from a 'linearization.’ This section uses general arguments from Fredholm theory and Leray-Schauder degree theory for non-linear integral equations to analyze the mapping properties of neural operators in function spaces. These results apply to subnetworks formed from the layers considered in this work, under natural conditions. We believe that our work has applications in Bayesian uncertainty quantification where injectivity enables likelihood estimation and in inverse problems where surjectivity and injectivity corresponds to existence and uniqueness of the solutions, respectively.
****
On the Convergence of CART under Sufficient Impurity Decrease Condition
Rahul Mazumder, Haoyue Wang
The decision tree is a flexible machine-learning model that finds its success in numerous applications. It is usually fitted in a recursively greedy manner using CART. In this paper, we study the convergence rate of CART under a regression setting. First, we prove an upper bound on the prediction error of CART under a sufficient impurity decrease (SID) condition \cite{chi2020asymptotic} -- our result is an improvement over the known result by \cite{chi2020asymptotic} under a similar assumption. We show via examples that this error bound cannot be further improved by more than a constant or a log factor. Second, we introduce a few easy-to-check sufficient conditions of the SID condition. In particular, we show that the SID condition can be satisfied by an additive model when the component functions satisfy a ``locally reverse Poincare inequality". We discuss a few fami

liar function classes in non-parametric estimation to demonstrate the usefulness
 of this conception.
****
PERFOGRAPH: A Numerical Aware Program Graph Representation for Performance Optim
ization and Program Analysis
Ali TehraniJamsaz, Quazi Ishtiaque Mahmud, Le Chen, Nesreen K. Ahmed, Ali Jannes
ari
The remarkable growth and significant success of machine learning have expanded
its applications into programming languages and program analysis. However, a key
 challenge in adopting the latest machine learning methods is the representation
 of programming languages which has a direct impact on the ability of machine le
arning methods to reason about programs. The absence of numerical awareness, agg
regate data structure information, and improper way of presenting variables in p
revious representation works have limited their performances. To overcome the l
imitations and challenges of current program representations, we propose a novel
 graph-based program representation called PERFOGRAPH. PERFOGRAPH can capture nu
merical information and the aggregate data structure by introducing new nodes an
d edges. Furthermore, we propose an adapted embedding method to incorporate nume
rical awareness.These enhancements make PERFOGRAPH a highly flexible and scalabl
e representation that can effectively capture programs' intricate dependencies a
nd semantics. Consequently, it serves as a powerful tool for various application
s such as program analysis, performance optimization, and parallelism discovery.
 Our experimental results demonstrate that PERFOGRAPH outperforms existing repre
sentations and sets new state-of-the-art results by reducing the error rate by 7
.4% (AMD dataset) and 10% (NVIDIA dataset) in the well-known Device Mapping chal
lenge. It also sets new state-of-the-art results in various performance optimiza
tion tasks like Parallelism Discovery and Numa and Prefetchers Configuration pre
diction.
****
Penalising the biases in norm regularisation enforces sparsity
Etienne Boursier, Nicolas Flammarion
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Distributional Learning of Variational AutoEncoder: Application to Synthetic Dat
a Generation
Seunghwan An, Jong-June Jeon
The Gaussianity assumption has been consistently criticized as a main limitation
 of the Variational Autoencoder (VAE) despite its efficiency in computational mo
deling. In this paper, we propose a new approach that expands the model capacity
 (i.e., expressive power of distributional family) without sacrificing the compu
tational advantages of the VAE framework. Our VAE model's decoder is composed of
 an infinite mixture of asymmetric Laplace distribution, which possesses general
 distribution fitting capabilities for continuous variables. Our model is repres
ented by a special form of a nonparametric M-estimator for estimating general qu
antile functions, and we theoretically establish the relevance between the propo
sed model and quantile estimation. We apply the proposed model to synthetic data
 generation, and particularly, our model demonstrates superiority in easily adju
sting the level of data privacy.
****
Optimistic Meta-Gradients
Sebastian Flennerhag, Tom Zahavy, Brendan O'Donoghue, Hado P. van Hasselt, Andrá
s György, Satinder Singh
We study the connection between gradient-based meta-learning and convex optimisa
tion. We observe that gradient descent with momentum is a special case of meta-g
radients, and building on recent results in optimisation, we prove convergence r
ates for meta learning in the single task setting. While a meta-learned update r
ule can yield faster convergence up to constant factor, it is not sufficient for

acceleration. Instead, some form of optimism is required. We show that optimism in meta-learning can be captured through the recently proposed Bootstrapped Meta-Gradient (Flennerhag et. al., 2022) method, providing deeper insight into its underlying mechanics.
****
Norm-guided latent space exploration for text-to-image generation
Dvir Samuel, Rami Ben-Ari, Nir Darshan, Haggai Maron, Gal Chechik
Text-to-image diffusion models show great potential in synthesizing a large variety of concepts in new compositions and scenarios. However, the latent space of initial seeds is still not well understood and its structure was shown to impact the generation of various concepts. Specifically, simple operations like interpolation and finding the centroid of a set of seeds perform poorly when using standard Euclidean or spherical metrics in the latent space. This paper makes the observation that, in current training procedures, diffusion models observed inputs with a narrow range of norm values. This has strong implications for methods that rely on seed manipulation for image generation, with applications to few-shot and long-tail learning tasks. To address this issue, we propose a novel method for interpolating between two seeds and demonstrate that it defines a new non-Euclidean metric that takes into account a norm-based prior on seeds. We describe a simple yet efficient algorithm for approximating this interpolation procedure and use it to further define centroids in the latent seed space. We show that our new interpolation and centroid techniques significantly enhance the generation of rare concept images. This further leads to state-of-the-art performance on few-shot and long-tail benchmarks, improving prior approaches in terms of generation speed, image quality, and semantic content.
****
Scale Alone Does not Improve Mechanistic Interpretability in Vision Models
Roland S. Zimmermann, Thomas Klein, Wieland Brendel
In light of the recent widespread adoption of AI systems, understanding the internal information processing of neural networks has become increasingly critical. Most recently, machine vision has seen remarkable progress by scaling neural networks to unprecedented levels in dataset and model size. We here ask whether this extraordinary increase in scale also positively impacts the field of mechanistic interpretability. In other words, has our understanding of the inner workings of scaled neural networks improved as well? We use a psychophysical paradigm to quantify one form of mechanistic interpretability for a diverse suite of nine models and find no scaling effect for interpretability - neither for model nor dataset size. Specifically, none of the investigated state-of-the-art models are easier to interpret than the GoogLeNet model from almost a decade ago. Latest-generation vision models appear even less interpretable than older architectures, hinting at a regression rather than improvement, with modern models sacrificing interpretability for accuracy. These results highlight the need for models explicitly designed to be mechanistically interpretable and the need for more helpful interpretability methods to increase our understanding of networks at an atomic level. We release a dataset containing more than 130'000 human responses from our psychophysical evaluation of 767 units across nine models. This dataset facilitates research on automated instead of human-based interpretability evaluations, which can ultimately be leveraged to directly optimize the mechanistic interpretability of models.
****
The Harvard USPTO Patent Dataset: A Large-Scale, Well-Structured, and Multi-Purpose Corpus of Patent Applications
Mirac Suzgun, Luke Melas-Kyriazi, Suproteem Sarkar, Scott D Kominers, Stuart Shieber
Innovation is a major driver of economic and social development, and information about many kinds of innovation is embedded in semi-structured data from patents and patent applications. Though the impact and novelty of innovations expressed in patent data are difficult to measure through traditional means, machine learning offers a promising set of techniques for evaluating novelty, summarizing contributions, and embedding semantics. In this paper, we introduce the Harvard US

PTO Patent Dataset (HUPD), a large-scale, well-structured, and multi-purpose corpus of English-language patent applications filed to the United States Patent and Trademark Office (USPTO) between 2004 and 2018. With more than 4.5 million patent documents, HUPD is two to three times larger than comparable corpora. Unlike other NLP patent datasets, HUPD contains the inventor-submitted versions of patent applications, not the final versions of granted patents, allowing us to study patentability at the time of filing using NLP methods for the first time. It is also novel in its inclusion of rich structured data alongside the text of patent filings: By providing each application's metadata along with all of its text fields, HUPD enables researchers to perform new sets of NLP tasks that leverage variation in structured covariates. As a case study on the types of research HUPD makes possible, we introduce a new task to the NLP community -- patent acceptance prediction. We additionally show the structured metadata provided in HUPD allows us to conduct explicit studies of concept shifts for this task. We find that performance on patent acceptance prediction decays when models trained in one context are evaluated on different innovation categories and over time. Finally, we demonstrate how HUPD can be used for three additional tasks: Multi-class classification of patent subject areas, language modeling, and abstractive summarization. Put together, our publicly-available dataset aims to advance research extending language and classification models to diverse and dynamic real-world data distributions.
****
MEMTO: Memory-guided Transformer for Multivariate Time Series Anomaly Detection
Junho Song, Keonwoo Kim, Jeonglyul Oh, Sungzoon Cho
Detecting anomalies in real-world multivariate time series data is challenging due to complex temporal dependencies and inter-variable correlations. Recently, reconstruction-based deep models have been widely used to solve the problem. However, these methods still suffer from an over-generalization issue and fail to deliver consistently high performance. To address this issue, we propose the MEMTO, a memory-guided Transformer using a reconstruction-based approach. It is designed to incorporate a novel memory module that can learn the degree to which each memory item should be updated in response to the input data. To stabilize the training procedure, we use a two-phase training paradigm which involves using K-means clustering for initializing memory items. Additionally, we introduce a bi-dimensional deviation-based detection criterion that calculates anomaly scores considering both input space and latent space. We evaluate our proposed method on five real-world datasets from diverse domains, and it achieves an average anomaly detection F1-score of 95.74%, significantly outperforming the previous state-of-the-art methods. We also conduct extensive experiments to empirically validate the effectiveness of our proposed model's key components.
****
Minimax Risks and Optimal Procedures for Estimation under Functional Local Differential Privacy
Bonwoo Lee, Jeongyoun Ahn, Cheolwoo Park
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Switching Autoregressive Low-rank Tensor Models
Hyun Dong Lee, Andrew Warrington, Joshua Glaser, Scott Linderman
An important problem in time-series analysis is modeling systems with time-varying dynamics. Probabilistic models with joint continuous and discrete latent states offer interpretable, efficient, and experimentally useful descriptions of such data. Commonly used models include autoregressive hidden Markov models (ARHMMs) and switching linear dynamical systems (SLDSs), each with its own advantages and disadvantages. ARHMMs permit exact inference and easy parameter estimation, but are parameter intensive when modeling long dependencies, and hence are prone to overfitting. In contrast, SLDSs can capture long-range dependencies in a parameter efficient way through Markovian latent dynamics, but present an intrac

table likelihood and a challenging parameter estimation task. In this paper, we propose switching autoregressive low-rank tensor SALT models, which retain the advantages of both approaches while ameliorating the weaknesses. SALT parameterizes the tensor of an ARHMM with a low-rank factorization to control the number of parameters and allow longer range dependencies without overfitting. We prove theoretical and discuss practical connections between SALT, linear dynamical systems, and SLDSs. We empirically demonstrate quantitative advantages of SALT models on a range of simulated and real prediction tasks, including behavioral and neural datasets. Furthermore, the learned low-rank tensor provides novel insights into temporal dependencies within each discrete state.
****
From Tempered to Benign Overfitting in ReLU Neural Networks
Guy Kornowski, Gilad Yehudai, Ohad Shamir
Overparameterized neural networks (NNs) are observed to generalize well even when trained to perfectly fit noisy data. This phenomenon motivated a large body of work on "benign overfitting", where interpolating predictors achieve near-optimal performance. Recently, it was conjectured and empirically observed that the behavior of NNs is often better described as "tempered overfitting", where the performance is non-optimal yet also non-trivial, and degrades as a function of the noise level. However, a theoretical justification of this claim for non-linear NNs has been lacking so far. In this work, we provide several results that aim at bridging these complementing views. We study a simple classification setting with 2-layer ReLU NNs, and prove that under various assumptions, the type of overfitting transitions from tempered in the extreme case of one-dimensional data, to benign in high dimensions. Thus, we show that the input dimension has a crucial role on the overfitting profile in this setting, which we also validate empirically for intermediate dimensions. Overall, our results shed light on the intricate connections between the dimension, sample size, architecture and training algorithm on the one hand, and the type of resulting overfitting on the other hand.
****
Tree-Rings Watermarks: Invisible Fingerprints for Diffusion Images
Yuxin Wen, John Kirchenbauer, Jonas Geiping, Tom Goldstein
Watermarking the outputs of generative models is a crucial technique for tracing copyright and preventing potential harm from AI-generated content. In this paper, we introduce a novel technique called Tree-Ring Watermarking that robustly fingerprints diffusion model outputs. Unlike existing methods that perform post-hoc modifications to images after sampling, Tree-Ring Watermarking subtly influences the entire sampling process, resulting in a model fingerprint that is invisible to humans. The watermark embeds a pattern into the initial noise vector used for sampling. These patterns are structured in Fourier space so that they are invariant to convolutions, crops, dilations, flips, and rotations. After image generation, the watermark signal is detected by inverting the diffusion process to retrieve the noise vector, which is then checked for the embedded signal. We demonstrate that this technique can be easily applied to arbitrary diffusion models, including text-conditioned Stable Diffusion, as a plug-in with negligible loss in FID. Our watermark is semantically hidden in the image space and is far more robust than watermarking alternatives that are currently deployed.
****
MVDoppler: Unleashing the Power of Multi-View Doppler for MicroMotion-based Gait Classification
Soheil Hor, Shubo Yang, Jaeho Choi, Amin Arbabian
Modern perception systems rely heavily on high-resolution cameras, LiDARs, and advanced deep neural networks, enabling exceptional performance across various applications. However, these optical systems predominantly depend on geometric features and shapes of objects, which can be challenging to capture in long-range perception applications. To overcome this limitation, alternative approaches such as Doppler-based perception using high-resolution radars have been proposed. Doppler-based systems are capable of measuring micro-motions of targets remotely and with very high precision. When compared to geometric features, the resolution

of micro-motion features exhibits significantly greater resilience to the influ
ence of distance. However, the true potential of Doppler-based perception has ye
t to be fully realized due to several factors. These include the unintuitive nat
ure of Doppler signals, the limited availability of public Doppler datasets, and
 the current datasets' inability to capture the specific co-factors that are uni
que to Doppler-based perception, such as the effect of the radar's observation a
ngle and the target's motion trajectory.This paper introduces a new large multi-
view Doppler dataset together with baseline perception models for micro-motion-b
ased gait analysis and classification. The dataset captures the impact of the su
bject's walking trajectory and radar's observation angle on the classification p
erformance. Additionally, baseline multi-view data fusion techniques are provide
d to mitigate these effects. This work demonstrates that sub-second micro-motion
 snapshots can be sufficient for reliable detection of hand movement patterns an
d even changes in a pedestrian's walking behavior when distracted by their phone
. Overall, this research not only showcases the potential of Doppler-based perce
ption, but also offers valuable solutions to tackle its fundamental challenges.
****
Understanding and Improving Ensemble Adversarial Defense
Yian Deng, Tingting Mu
The strategy of ensemble has become popular in adversarial defense, which trains
 multiple base classifiers to defend against adversarial attacks in a cooperativ
e manner. Despite the empirical success, theoretical explanations on why an ense
mble of adversarially trained classifiers is more robust than single ones remain
 unclear. To fill in this gap, we develop a new error theory dedicated to unders
tanding ensemble adversarial defense, demonstrating a provable 0-1 loss reducti
on on challenging sample sets in adversarial defense scenarios. Guided by this t
heory, we propose an effective approach to improve ensemble adversarial defense,
 named interactive global adversarial training (iGAT). The proposal includes (1)
 a probabilistic distributing rule that selectively allocates to different base
classifiers adversarial examples that are globally challenging to the ensemble,
and (2) a regularization term to rescue the severest weaknesses of the base clas
sifiers. Being tested over various existing ensemble adversarial defense techniq
ues, iGAT is capable of boosting their performance by up to 17\% evaluated usi
ng CIFAR10 and CIFAR100 datasets under both white-box and black-box attacks.
****
Adversarial Training for Graph Neural Networks: Pitfalls, Solutions, and New Dir
ections
Lukas Gosch, Simon Geisler, Daniel Sturm, Bertrand Charpentier, Daniel Zügner, S
tephan Günnemann
Despite its success in the image domain, adversarial training did not (yet) stan
d out as an effective defense for Graph Neural Networks (GNNs) against graph str
ucture perturbations. In the pursuit of fixing adversarial training (1) we show
 and overcome fundamental theoretical as well as practical limitations of the ad
opted graph learning setting in prior work; (2) we reveal that flexible GNNs bas
ed on learnable graph diffusion are able to adjust to adversarial perturbations,
 while the learned message passing scheme is naturally interpretable; (3) we int
roduce the first attack for structure perturbations that, while targeting multip
le nodes at once, is capable of handling global (graph-level) as well as local (
node-level) constraints. Including these contributions, we demonstrate that adve
rsarial training is a state-of-the-art defense against adversarial structure per
turbations.
****
A Massive Scale Semantic Similarity Dataset of Historical English
Emily Silcock, Abhishek Arora, Melissa Dell
A diversity of tasks use language models trained on semantic similarity data. Wh
ile there are a variety of datasets that capture semantic similarity, they are e
ither constructed from modern web data or are relatively small datasets created
in the past decade by human annotators. This study utilizes a novel source, newl
y digitized articles from off-copyright, local U.S. newspapers, to assemble a ma
ssive-scale semantic similarity dataset spanning 70 years from 1920 to 1989 and

containing nearly 400M positive semantic similarity pairs. Historically, around half of articles in U.S. local newspapers came from newswires like the Associated Press. While local papers reproduced articles from the newswire, they wrote their own headlines, which form abstractive summaries of the associated articles. We associate articles and their headlines by exploiting document layouts and language understanding. We then use deep neural methods to detect which articles are from the same underlying source, in the presence of substantial noise and abridgement. The headlines of reproduced articles form positive semantic similarity pairs. The resulting publicly available HEADLINES dataset is significantly larger than most existing semantic similarity datasets and covers a much longer span of time. It will facilitate the application of contrastively trained semantic similarity models to a variety of tasks, including the study of semantic change across space and time.
****
Joint Prompt Optimization of Stacked LLMs using Variational Inference
Alessandro Sordoni, Eric Yuan, Marc-Alexandre Côté, Matheus Pereira, Adam Trischler, Ziang Xiao, Arian Hosseini, Friederike Niedtner, Nicolas Le Roux
Large language models (LLMs) can be seen as atomic units of computation mapping sequences to a distribution over sequences. Thus, they can be seen as stochastic language layers in a language network, where the learnable parameters are the natural language prompts at each layer. By stacking two such layers and feeding the output of one layer to the next, we obtain a Deep Language Network (DLN). We first show how to effectively perform prompt optimization for a 1-Layer language network (DLN-1). Then, we present an extension that applies to 2-layer DLNs (DLN-2), where two prompts must be learned. The key idea is to consider the output of the first layer as a latent variable, which requires inference, and prompts to be learned as the parameters of the generative distribution. We first test the effectiveness of DLN-1 in multiple reasoning and natural language understanding tasks. Then, we show that DLN-2 can reach higher performance than a single layer, showing promise that we might reach comparable performance to GPT-4, even when each LLM in the network is smaller and less powerful.
****
On the Properties of Kullback-Leibler Divergence Between Multivariate Gaussian Distributions
Yufeng Zhang, Jialu Pan, Li Ken Li, Wanwei Liu, Zhenbang Chen, Xinwang Liu, J Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Implicit Bias of (Stochastic) Gradient Descent for Rank-1 Linear Neural Network
Bochen Lyu, Zhanxing Zhu
Studying the implicit bias of gradient descent (GD) and stochastic gradient descent (SGD) is critical to unveil the underlying mechanism of deep learning. Unfortunately, even for standard linear networks in regression setting, a comprehensive characterization of the implicit bias is still an open problem. This paper proposes to investigate a new proxy model of standard linear network, rank-1 linear network, where each weight matrix is parameterized as a rank-1 form. For over-parameterized regression problem, we precisely analyze the implicit bias of GD and SGD---by identifying a “potential” function such that GD converges to its minimizer constrained by zero training error (i.e., interpolation solution), and further characterizing the role of the noise introduced by SGD in perturbing the form of this potential. Our results explicitly connect the depth of the network and the initialization with the implicit bias of GD and SGD. Furthermore, we emphasize a new implicit bias of SGD jointly induced by stochasticity and over-parameterization, which can reduce the dependence of the SGD's solution on the initialization. Our findings regarding the implicit bias are different from that of a recently popular model, the diagonal linear network. We highlight that the induced bias of our rank-1 model is more consistent with standard linear network w

hile the diagonal one is not. This suggests that the proposed rank-1 linear network might be a plausible proxy for standard linear net.
****

AdaPlanner: Adaptive Planning from Feedback with Language Models
Haotian Sun, Yuchen Zhuang, Lingkai Kong, Bo Dai, Chao Zhang
Large language models (LLMs) have recently demonstrated the potential in acting as autonomous agents for sequential decision-making tasks. However, most existing methods either take actions greedily without planning or rely on static plans that are not adaptable to environmental feedback. Consequently, the sequential decision-making performance of LLM agents degenerates with problem complexity and plan horizons increase. We propose a closed-loop approach, AdaPlanner, which allows the LLM agent to refine its self-generated plan adaptively in response to environmental feedback. In AdaPlanner, the LLM agent adaptively refines its plan from feedback with both in-plan and out-of-plan refinement strategies. To mitigate hallucination, we develop a code-style LLM prompt structure that facilitates plan generation across a variety of tasks, environments, and agent capabilities. Furthermore, we propose a skill discovery mechanism that leverages successful plans as few-shot exemplars, enabling the agent to plan and refine with fewer task demonstrations. Our experiments in the ALFWorld and MiniWoB++ environments demonstrate that AdaPlanner outperforms state-of-the-art baselines by 3.73% and 4.11% while utilizing 2x and 600x fewer samples, respectively. The implementation of AdaPlanner is available at https://github.com/haotiansun14/AdaPlanner.
****

Fairness Aware Counterfactuals for Subgroups
Loukas Kavouras, Konstantinos Tsopelas, Giorgos Giannopoulos, Dimitris Sacharidis, Eleni Psaroudaki, Nikolaos Theologitis, Dimitrios Rontogiannis, Dimitris Fotakis, Ioannis Emiris
In this work, we present Fairness Aware Counterfactuals for Subgroups (FACTS), a framework for auditing subgroup fairness through counterfactual explanations. We start with revisiting (and generalizing) existing notions and introducing new, more refined notions of subgroup fairness. We aim to (a) formulate different aspects of the difficulty of individuals in certain subgroups to achieve recourse, i.e. receive the desired outcome, either at the micro level, considering members of the subgroup individually, or at the macro level, considering the subgroup as a whole, and (b) introduce notions of subgroup fairness that are robust, if not totally oblivious, to the cost of achieving recourse. We accompany these notions with an efficient, model-agnostic, highly parameterizable, and explainable framework for evaluating subgroup fairness. We demonstrate the advantages, the wide applicability, and the efficiency of our approach through a thorough experimental evaluation on different benchmark datasets.
****

ProteinShake: Building datasets and benchmarks for deep learning on protein structures
Tim Kucera, Carlos Oliver, Dexiong Chen, Karsten Borgwardt
We present ProteinShake, a Python software package that simplifies datasetcreation and model evaluation for deep learning on protein structures. Users cancreate custom datasets or load an extensive set of pre-processed datasets fromthe Protein Data Bank (PDB) and AlphaFoldDB. Each dataset is associated withprediction tasks and evaluation functions covering a broad array of biologicalchallenges. A benchmark on these tasks shows that pre-training almost alwaysimproves performance, the optimal data modality (graphs, voxel grids, or pointclouds) is task-dependent, and models struggle to generalize to new structures.ProteinShake makes protein structure data easily accessible and comparisonamong models straightforward, providing challenging benchmark settings withreal-world implications.ProteinShake is available at: https://proteinshake.ai
****

Lovász Principle for Unsupervised Graph Representation Learning
Ziheng Sun, Chris Ding, Jicong Fan
This paper focuses on graph-level representation learning that aims to represent graphs as vectors that can be directly utilized in downstream tasks such as gra

ph classification. We propose a novel graph-level representation learning principle called Lovász principle, which is motivated by the Lovász number in graph theory. The Lovász number of a graph is a real number that is an upper bound for graph Shannon capacity and is strongly connected with various global characteristics of the graph. Specifically, we show that the handle vector for computing the Lovász number is potentially a suitable choice for graph representation, as it captures a graph's global properties, though a direct application of the handle vector is difficult and problematic. We propose to use neural networks to address the problems and hence provide the Lovász principle. Moreover, we propose an enhanced Lovász principle that is able to exploit the subgraph Lovász numbers directly and efficiently. The experiments demonstrate that our Lovász principles achieve competitive performance compared to the baselines in unsupervised and semi-supervised graph-level representation learning tasks. The code of our Lovász principles is publicly available on GitHub.
****
ComSL: A Composite Speech-Language Model for End-to-End Speech-to-Text Translation
Chenyang Le, Yao Qian, Long Zhou, Shujie LIU, Yanmin Qian, Michael Zeng, Xuedong Huang
Joint speech-language training is challenging due to the large demand for training data and GPU consumption, as well as the modality gap between speech and language. We present ComSL, a speech-language model built atop a composite architecture of public pre-trained speech-only and language-only models and optimized data-efficiently for spoken language tasks. Particularly, we propose to incorporate cross-modality learning into transfer learning and conduct them simultaneously for downstream tasks in a multi-task learning manner. Our approach has demonstrated effectiveness in end-to-end speech-to-text translation tasks, achieving a new state-of-the-art average BLEU score of 31.5 on the multilingual speech to English text translation task for 21 languages, as measured on the public CoVoST2 evaluation set.
****
Reverse Engineering Self-Supervised Learning
Ido Ben-Shaul, Ravid Shwartz-Ziv, Tomer Galanti, Shai Dekel, Yann LeCun
Understanding the learned representation and underlying mechanisms of Self-Supervised Learning (SSL) often poses a challenge. In this paper, we ‘reverse engineer’ SSL, conducting an in-depth empirical analysis of its learned internal representations, encompassing diverse models, architectures, and hyperparameters. Our study reveals an intriguing process within the SSL training: an inherent facilitation of semantic label-based clustering, which is surprisingly driven by the regularization component of the SSL objective. This clustering not only enhances downstream classification, but also compresses the information. We further illustrate that the alignment of the SSL-trained representation is more pronounced with semantic classes rather than random functions. Remarkably, the learned representations align with semantic classes across various hierarchical levels, with this alignment intensifying when going deeper into the network. This ‘reverse engineering’ approach provides valuable insights into the inner mechanism of SSL and their influences on the performance across different class sets.
****
DinoSR: Self-Distillation and Online Clustering for Self-supervised Speech Representation Learning
Alexander H. Liu, Heng-Jui Chang, Michael Auli, Wei-Ning Hsu, Jim Glass
In this paper, we introduce self-distillation and online clustering for self-supervised speech representation learning (DinoSR) which combines masked language modeling, self-distillation, and online clustering. We show that these concepts complement each other and result in a strong representation learning model for speech. DinoSR first extracts contextualized embeddings from the input audio with a teacher network, then runs an online clustering system on the embeddings to yield a machine-discovered phone inventory, and finally uses the discretized tokens to guide a student network. We show that DinoSR surpasses previous state-of-the-art performance in several downstream tasks, and provide a detailed analysis o

f the model and the learned discrete units.
****
4M: Massively Multimodal Masked Modeling
David Mizrahi, Roman Bachmann, Oguzhan Kar, Teresa Yeo, Mingfei Gao, Afshin Dehghan, Amir Zamir
Current machine learning models for vision are often highly specialized and limited to a single modality and task. In contrast, recent large language models exhibit a wide range of capabilities, hinting at a possibility for similarly versatile models in computer vision.In this paper, we take a step in this direction and propose a multimodal training scheme called 4M. It consists of training a single unified Transformer encoder-decoder using a masked modeling objective across a wide range of input/output modalities – including text, images, geometric, and semantic modalities, as well as neural network feature maps. 4M achieves scalability by unifying the representation space of all modalities through mapping them into discrete tokens and performing multimodal masked modeling on a small randomized subset of tokens.4M leads to models that exhibit several key capabilities: (1) they can perform a diverse set of vision tasks out of the box, (2) they excel when fine-tuned for unseen downstream tasks or new input modalities, and (3) they can function as a generative model that can be conditioned on arbitrary modalities, enabling a wide variety of expressive multimodal editing capabilities with remarkable flexibility.Through experimental analyses, we demonstrate the potential of 4M for training versatile and scalable foundation models for vision tasks, setting the stage for further exploration in multimodal learning for vision and other domains.
****
Non-Rigid Shape Registration via Deep Functional Maps Prior
Puhua Jiang, Mingze Sun, Ruqi Huang
In this paper, we propose a learning-based framework for non-rigid shape registra- tion without correspondence supervision. Traditional shape registration techniques typically rely on correspondences induced by extrinsic proximity, therefore can fail in the presence of large intrinsic deformations. Spectral mapping methods overcome this challenge by embedding shapes into, geometric or learned, high- dimensional spaces, where shapes are easier to align. However, due to the dependency on abstract, non-linear embedding schemes, the latter can be vulnerable with respect to perturbed or alien input. In light of this, our framework takes the best of both worlds. Namely, we deform source mesh towards the target point cloud, guided by correspondences induced by high-dimensional embeddings learned from deep functional maps (DFM). In particular, the correspondences are dynamically updated according to the intermediate registrations and filtered by consistency prior, which prominently robustify the overall pipeline. Moreover, in order to alleviate the requirement of extrinsically aligned input, we train an orientation regressor on a set of aligned synthetic shapes independent of the training shapes for DFM. Empirical results show that, with as few as dozens of training shapes of limited variability, our pipeline achieves state-of-the-art results on several benchmarks of non-rigid point cloud matching, but also delivers high-quality correspondences between unseen challenging shape pairs that undergo both significant extrinsic and intrinsic defor- mations, in which case neither traditional registration methods nor intrinsic methods work. The code is available at https://github.com/rqhuang88/DFR.
****
Game Solving with Online Fine-Tuning
Ti-Rong Wu, Hung Guei, Ting Han Wei, Chung-Chin Shih, Jui-Te Chin, I-Chen Wu
Game solving is a similar, yet more difficult task than mastering a game. Solving a game typically means to find the game-theoretic value (outcome given optimal play), and optionally a full strategy to follow in order to achieve that outcome. The AlphaZero algorithm has demonstrated super-human level play, and its powerful policy and value predictions have also served as heuristics in game solving. However, to solve a game and obtain a full strategy, a winning response must be found for all possible moves by the losing player. This includes very poor lines of play from the losing side, for which the AlphaZero self-play process will

not encounter. AlphaZero-based heuristics can be highly inaccurate when evaluating these out-of-distribution positions, which occur throughout the entire search. To address this issue, this paper investigates applying online fine-tuning while searching and proposes two methods to learn tailor-designed heuristics for game solving. Our experiments show that using online fine-tuning can solve a series of challenging 7x7 Killall-Go problems, using only 23.54\% of computation time compared to the baseline without online fine-tuning. Results suggest that the savings scale with problem size. Our method can further be extended to any tree search algorithm for problem solving. Our code is available at https://rlg.iis.sinica.edu.tw/papers/neurips2023-online-fine-tuning-solver.
****
Beyond probability partitions: Calibrating neural networks with semantic aware grouping
Jia-Qi Yang, De-Chuan Zhan, Le Gan
Research has shown that deep networks tend to be overly optimistic about their predictions, leading to an underestimation of prediction errors. Due to the limited nature of data, existing studies have proposed various methods based on model prediction probabilities to bin the data and evaluate calibration error. We propose a more generalized definition of calibration error called Partitioned Calibration Error (PCE), revealing that the key difference among these calibration error metrics lies in how the data space is partitioned. We put forth an intuitive proposition that an accurate model should be calibrated across any partition, suggesting that the input space partitioning can extend beyond just the partitioning of prediction probabilities, and include partitions directly related to the input. Through semantic-related partitioning functions, we demonstrate that the relationship between model accuracy and calibration lies in the granularity of the partitioning function. This highlights the importance of partitioning criteria for training a calibrated and accurate model. To validate the aforementioned analysis, we propose a method that involves jointly learning a semantic aware grouping function based on deep model features and logits to partition the data space into subsets. Subsequently, a separate calibration function is learned for each subset. Experimental results demonstrate that our approach achieves significant performance improvements across multiple datasets and network architectures, thus highlighting the importance of the partitioning function for calibration.
****
Identifiable Contrastive Learning with Automatic Feature Importance Discovery
Qi Zhang, Yifei Wang, Yisen Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Revisiting Out-of-distribution Robustness in NLP: Benchmarks, Analysis, and LLMs Evaluations
Lifan Yuan, Yangyi Chen, Ganqu Cui, Hongcheng Gao, FangYuan Zou, Xingyi Cheng, Heng Ji, Zhiyuan Liu, Maosong Sun
This paper reexamines the research on out-of-distribution (OOD) robustness in the field of NLP. We find that the distribution shift settings in previous studies commonly lack adequate challenges, hindering the accurate evaluation of OOD robustness. To address these issues, we propose a benchmark construction protocol that ensures clear differentiation and challenging distribution shifts. Then we introduceBOSS, a Benchmark suite for Out-of-distribution robustneSS evaluation covering 5 tasks and 20 datasets. Based on BOSS, we conduct a series of experiments on pretrained language models for analysis and evaluation of OOD robustness. First, for vanilla fine-tuning, we examine the relationship between in-distribution (ID) and OOD performance. We identify three typical types that unveil the inner learningmechanism, which could potentially facilitate the forecasting of OOD robustness, correlating with the advancements on ID datasets. Then, we evaluate 5 classic methods on BOSS and find that, despite exhibiting some effectiveness in specific cases, they do not offer significant improvement compared to vanilla

fine-tuning. Further, we evaluate 5 LLMs with various adaptation paradigms and find that when sufficient ID data is available, fine-tuning domain-specific models outperform LLMs on ID examples significantly. However, in the case of OOD instances, prioritizing LLMs with in-context learning yields better results. We identify that both fine-tuned small models and LLMs face challenges in effectively addressing downstream tasks. The code is public at https://github.com/lifan-yuan/OOD_NLP.
****
Towards Consistent Video Editing with Text-to-Image Diffusion Models
Zicheng Zhang, Bonan Li, Xuecheng Nie, Congying Han, Tiande Guo, Luoqi Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Federated Spectral Clustering via Secure Similarity Reconstruction
Dong Qiao, Chris Ding, Jicong Fan
Federated learning has a significant advantage in protecting information privacy. Many scholars proposed various secure learning methods within the framework of federated learning but the study on secure federated unsupervised learning especially clustering is limited. We in this work propose a secure kernelized factorization method for federated spectral clustering on distributed dataset. The method is non-trivial because the kernel or similarity matrix for spectral clustering is computed by data pairs, which violates the principle of privacy protection. Our method implicitly constructs an approximation for the kernel matrix on distributed data such that we can perform spectral clustering under the constraint of privacy protection. We provide a convergence guarantee of the optimization algorithm, reconstruction error bounds of the Gaussian kernel matrix, and the sufficient condition of correct clustering of our method. We also present some results of differential privacy. Numerical results on synthetic and real datasets demonstrate that the proposed method is efficient and accurate in comparison to the baselines.
****
CS-Isolate: Extracting Hard Confident Examples by Content and Style Isolation
Yexiong Lin, Yu Yao, Xiaolong Shi, Mingming Gong, Xu Shen, Dong Xu, Tongliang Liu
Label noise widely exists in large-scale image datasets. To mitigate the side effects of label noise, state-of-the-art methods focus on selecting confident examples by leveraging semi-supervised learning. Existing research shows that the ability to extract hard confident examples, which are close to the decision boundary, significantly influences the generalization ability of the learned classifier.In this paper, we find that a key reason for some hard examples being close to the decision boundary is due to the entanglement of style factors with content factors. The hard examples become more discriminative when we focus solely on content factors, such as semantic information, while ignoring style factors. Nonetheless, given only noisy data, content factors are not directly observed and have to be inferred.To tackle the problem of inferring content factors for classification when learning with noisy labels, our objective is to ensure that the content factors of all examples in the same underlying clean class remain unchanged as their style information changes.To achieve this, we utilize different data augmentation techniques to alter the styles while regularizing content factors based on some confident examples. By training existing methods with our inferred content factors, CS-Isolate proves their effectiveness in learning hard examples on benchmark datasets. The implementation is available at https://github.com/tmllab/2023NeurIPSCS-isolate.
****
Conditional independence testing under misspecified inductive biases
Felipe Maia Polo, Yuekai Sun, Moulinath Banerjee
Conditional independence (CI) testing is a fundamental and challenging task in modern statistics and machine learning. Many modern methods for CI testing rely o

n powerful supervised learning methods to learn regression functions or Bayes pr edictors as an intermediate step; we refer to this class of tests as regression-based tests. Although these methods are guaranteed to control Type-I error when the supervised learning methods accurately estimate the regression functions or Bayes predictors of interest, their behavior is less understood when they fail d ue to misspecified inductive biases; in other words, when the employed models ar e not flexible enough or when the training algorithm does not induce the desired predictors. Then, we study the performance of regression-based CI tests under m isspecified inductive biases. Namely, we propose new approximations or upper bou nds for the testing errors of three regression-based tests that depend on misspe cification errors. Moreover, we introduce the Rao-Blackwellized Predictor Test (RBPT), a regression-based CI test robust against misspecified inductive biases. Finally, we conduct experiments with artificial and real data, showcasing the us efulness of our theory and methods.
****
Blurred-Dilated Method for Adversarial Attacks
Yang Deng, Weibin Wu, Jianping Zhang, Zibin Zheng
Deep neural networks (DNNs) are vulnerable to adversarial attacks, which lead to incorrect predictions. In black-box settings, transfer attacks can be convenien tly used to generate adversarial examples. However, such examples tend to overfi t the specific architecture and feature representations of the source model, res ulting in poor attack performance against other target models. To overcome this drawback, we propose a novel model modification-based transfer attack: Blurred-D ilated method (BD) in this paper. In summary, BD works by reducing downsampling while introducing BlurPool and dilated convolutions in the source model. Then BD employs the modified source model to generate adversarial samples. We think tha t BD can more comprehensively preserve the feature information than the original source model. It thus enables more thorough destruction of the image features, which can improve the transferability of the generated adversarial samples. Exte nsive experiments on the ImageNet dataset show that adversarial examples generat ed by BD achieve significantly higher transferability than the state-of-the-art baselines. Besides, BD can be conveniently combined with existing black-box atta ck techniques to further improve their performance.
****
Towards Distribution-Agnostic Generalized Category Discovery
Jianhong Bai, Zuozhu Liu, Hualiang Wang, Ruizhe Chen, Lianrui Mu, Xiaomeng Li, J oey Tianyi Zhou, YANG FENG, Jian Wu, Haoji Hu
Data imbalance and open-ended distribution are two intrinsic characteristics of the real visual world. Though encouraging progress has been made in tackling eac h challenge separately, few works dedicated to combining them towards real-world scenarios. While several previous works have focused on classifying close-set s amples and detecting open-set samples during testing, it's still essential to be able to classify unknown subjects as human beings. In this paper, we formally d efine a more realistic task as distribution-agnostic generalized category discov ery (DA-GCD): generating fine-grained predictions for both close- and open-set c lasses in a long-tailed open-world setting. To tackle the challenging problem, w e propose a Self-Balanced Co-Advice contrastive framework (BaCon), which consist s of a contrastive-learning branch and a pseudo-labeling branch, working collabo ratively to provide interactive supervision to resolve the DA-GCD task. In parti cular, the contrastive-learning branch provides reliable distribution estimation to regularize the predictions of the pseudo-labeling branch, which in turn guid es contrastive learning through self-balanced knowledge transfer and a proposed novel contrastive loss. We compare BaCon with state-of-the-art methods from two closely related fields: imbalanced semi-supervised learning and generalized cate gory discovery. The effectiveness of BaCon is demonstrated with superior perform ance over all baselines and comprehensive analysis across various datasets. Our code is publicly available.
****
Stable Diffusion is Unstable
Chengbin Du, Yanxi Li, Zhongwei Qiu, Chang Xu

Recently, text-to-image models have been thriving. Despite their powerful genera tive capacity, our research has uncovered a lack of robustness in this generatio n process. Specifically, the introduction of small perturbations to the text pro mpts can result in the blending of primary subjects with other categories or the ir complete disappearance in the generated images. In this paper, we propose Aut o-attack on Text-to-image Models (ATM), a gradient-based approach, to effectivel y and efficiently generate such perturbations. By learning a Gumbel Softmax dist ribution, we can make the discrete process of word replacement or extension cont inuous, thus ensuring the differentiability of the perturbation generation. Once the distribution is learned, ATM can sample multiple attack samples simultaneou sly. These attack samples can prevent the generative model from generating the d esired subjects without tampering with the category keywords in the prompt. ATM has achieved a 91.1\% success rate in short-text attacks and an 81.2\% success r ate in long-text attacks. Further empirical analysis revealed three attack patte rns based on: 1) variability in generation speed, 2) similarity of coarse-graine d characteristics, and 3) polysemy of words. The code is available at https://gi thub.com/duchengbin8/StableDiffusionis_Unstable
****
A Competitive Algorithm for Agnostic Active Learning
Yihan Zhou, Eric Price
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Efficient Hyper-parameter Optimization with Cubic Regularization
Zhenqian Shen, Hansi Yang, Yong Li, James Kwok, Quanming Yao
As hyper-parameters are ubiquitous and can significantly affect the model perfor mance, hyper-parameter optimization is extremely important in machine learning. In this paper, we consider a sub-class of hyper-parameter optimization problems, where the hyper-gradients are not available. Such problems frequently appear wh en the performance metric is non-differentiable or the hyper-parameter is not co ntinuous. However, existing algorithms, like Bayesian optimization and reinforce ment learning, often get trapped in local optimals with poor performance. To add ress the above limitations, we propose to use cubic regularization to accelerate convergence and avoid saddle points. First, we adopt stochastic relaxation, whi ch allows obtaining gradient and Hessian information without hyper-gradients. Th en, we exploit the rich curvature information by cubic regularization. Theoretic ally, we prove that the proposed method can converge to approximate second-order stationary points, and the convergence is also guaranteed when the lower-level problem is inexactly solved. Experiments on synthetic and real-world data demons trate the effectiveness of our proposed method.
****
Joint Attribute and Model Generalization Learning for Privacy-Preserving Action Recognition
Duo Peng, Li Xu, Qiuhong Ke, Ping Hu, Jun Liu
Privacy-Preserving Action Recognition (PPAR) aims to transform raw videos into a nonymous ones to prevent privacy leakage while maintaining action clues, which i s an increasingly important problem in intelligent vision applications. Despite recent efforts in this task, it is still challenging to deal with novel privacy attributes and novel privacy attack models that are unavailable during the train ing phase. In this paper, from the perspective of meta-learning (learning to lea rn), we propose a novel Meta Privacy-Preserving Action Recognition (MPPAR) frame work to improve both generalization abilities above (i.e., generalize to novel p rivacy attributes and novel privacy attack models) in a unified manner. Concrete ly, we simulate train/test task shifts by constructing disjoint support/query se ts w.r.t. privacy attributes or attack models. Then, a virtual training and test ing scheme is applied based on support/query sets to provide feedback to optimiz e the model's learning toward better generalization. Extensive experiments demon strate the effectiveness and generalization of the proposed framework compared t

o state-of-the-arts.
****
3D-Aware Visual Question Answering about Parts, Poses and Occlusions
Xingrui Wang, Wufei Ma, Zhuowan Li, Adam Kortylewski, Alan L. Yuille
Despite rapid progress in Visual question answering (\textit{VQA}), existing datasets and models mainly focus on testing reasoning in 2D. However, it is important that VQA models also understand the 3D structure of visual scenes, for example to support tasks like navigation or manipulation. This includes an understanding of the 3D object pose, their parts and occlusions. In this work, we introduce the task of 3D-aware VQA, which focuses on challenging questions that require a compositional reasoning over the 3D structure of visual scenes. We address 3D-aware VQA from both the dataset and the model perspective. First, we introduce Super-CLEVR-3D, a compositional reasoning dataset that contains questions about object parts, their 3D poses, and occlusions. Second, we propose PO3D-VQA, a 3D-aware VQA model that marries two powerful ideas: probabilistic neural symbolic program execution for reasoning and deep neural networks with 3D generative representations of objects for robust visual recognition. Our experimental results show our model PO3D-VQA outperforms existing methods significantly, but we still observe a significant performance gap compared to 2D VQA benchmarks, indicating that 3D-aware VQA remains an important open research area.
****
MixFormerV2: Efficient Fully Transformer Tracking
Yutao Cui, Tianhui Song, Gangshan Wu, Limin Wang
Transformer-based trackers have achieved strong accuracy on the standard benchmarks. However, their efficiency remains an obstacle to practical deployment on both GPU and CPU platforms. In this paper, to overcome this issue, we propose a fully transformer tracking framework, coined as \emph{MixFormerV2}, without any dense convolutional operation and complex score prediction module. Our key design is to introduce four special prediction tokens and concatenate them with the tokens from target template and search areas. Then, we apply the unified transformer backbone on these mixed token sequence. These prediction tokens are able to capture the complex correlation between target template and search area via mixed attentions. Based on them, we can easily predict the tracking box and estimate its confidence score through simple MLP heads. To further improve the efficiency of MixFormerV2, we present a new distillation-based model reduction paradigm, including dense-to-sparse distillation and deep-to-shallow distillation. The former one aims to transfer knowledge from the dense-head based MixViT to our fully transformer tracker, while the latter one is used to prune some layers of the backbone. We instantiate two types of MixForemrV2, where the MixFormerV2-B achieves an AUC of 70.6\% on LaSOT and AUC of 56.7\% on TNL2k with a high GPU speed of 165 FPS, and the MixFormerV2-S surpasses FEAR-L by 2.7\% AUC on LaSOT with a real-time CPU speed.
****
Generalized Information-theoretic Multi-view Clustering
Weitian Huang, Sirui Yang, Hongmin Cai
In an era of more diverse data modalities, multi-view clustering has become a fundamental tool for comprehensive data analysis and exploration. However, existing multi-view unsupervised learning methods often rely on strict assumptions on semantic consistency among samples. In this paper, we reformulate the multi-view clustering problem from an information-theoretic perspective and propose a general theoretical model. In particular, we define three desiderata under multi-view unsupervised learning in terms of mutual information, namely, comprehensiveness, concentration, and cross-diversity. The multi-view variational lower bound is then obtained by approximating the samples' high-dimensional mutual information. The Kullback–Leibler divergence is utilized to deduce sample assignments. Ultimately the information-based multi-view clustering model leverages deep neural networks and Stochastic Gradient Variational Bayes to achieve representation learning and clustering simultaneously. Extensive experiments on both synthetic and real datasets with wide types demonstrate that the proposed method exhibits a more stable and superior clustering performance than state-of-the-art algorithms.

****
Counterfactual Evaluation of Peer-Review Assignment Policies
Martin Saveski, Steven Jecmen, Nihar Shah, Johan Ugander
Peer review assignment algorithms aim to match research papers to suitable exper
t reviewers, working to maximize the quality of the resulting reviews. A key cha
llenge in designing effective assignment policies is evaluating how changes to t
he assignment algorithm map to changes in review quality. In this work, we lever
age recently proposed policies that introduce randomness in peer-review assignme
nt—in order to mitigate fraud—as a valuable opportunity to evaluate counterfactu
al assignment policies. Specifically, we exploit how such randomized assignments
 provide a positive probability of observing the reviews of many assignment poli
cies of interest. To address challenges in applying standard off-policy evaluati
on methods, such as violations of positivity, we introduce novel methods for par
tial identification based on monotonicity and Lipschitz smoothness assumptions f
or the mapping between reviewer-paper covariates and outcomes. We apply our meth
ods to peer-review data from two computer science venues: the TPDP'21 workshop (
95 papers and 35 reviewers) and the AAAI'22 conference (8,450 papers and 3,145 r
eviewers). We consider estimates of (i) the effect on review quality when changi
ng weights in the assignment algorithm, e.g., weighting reviewers' bids vs. text
ual similarity (between the review's past papers and the submission), and (ii) t
he "cost of randomization", capturing the difference in expected quality between
 the perturbed and unperturbed optimal match. We find that placing higher weight
 on text similarity results in higher review quality and that introducing random
ization in the reviewer-paper assignment only marginally reduces the review qual
ity. Our methods for partial identification may be of independent interest, whil
e our off-policy approach can likely find use in evaluating a broad class of alg
orithmic matching systems.
****
Temporal Causal Mediation through a Point Process: Direct and Indirect Effects o
f Healthcare Interventions
Ça■lar H■zl■, ST John, Anne Juuti, Tuure Saarinen, Kirsi Pietiläinen, Pekka Mart
tinen
Deciding on an appropriate intervention requires a causal model of a treatment,
the outcome, and potential mediators. Causal mediation analysis lets us distingu
ish between direct and indirect effects of the intervention, but has mostly been
 studied in a static setting. In healthcare, data come in the form of complex, i
rregularly sampled time-series, with dynamic interdependencies between a treatme
nt, outcomes, and mediators across time. Existing approaches to dynamic causal m
ediation analysis are limited to regular measurement intervals, simple parametri
c models, and disregard long-range mediator--outcome interactions. To address th
ese limitations, we propose a non-parametric mediator--outcome model where the m
ediator is assumed to be a temporal point process that interacts with the outcom
e process. With this model, we estimate the direct and indirect effects of an ex
ternal intervention on the outcome, showing how each of these affects the whole
future trajectory. We demonstrate on semi-synthetic data that our method can acc
urately estimate direct and indirect effects. On real-world healthcare data, our
 model infers clinically meaningful direct and indirect effect trajectories for
 blood glucose after a surgery.
****
Randomized and Deterministic Maximin-share Approximations for Fractionally Subad
ditive Valuations
Hannaneh Akrami, Kurt Mehlhorn, Masoud Seddighin, Golnoosh Shahkarami
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Causal normalizing flows: from theory to practice
Adrián Javaloy, Pablo Sanchez-Martin, Isabel Valera
In this work, we deepen on the use of normalizing flows for causal reasoning. Sp

ecifically, we first leverage recent results on non-linear ICA to show that causal models are identifiable from observational data given a causal ordering, and thus can be recovered using autoregressive normalizing flows (NFs). Second, we analyze different design and learning choices for causal normalizing flows to capture the underlying causal data-generating process. Third, we describe how to implement the do-operator in causal NFs, and thus, how to answer interventional and counterfactual questions. Finally, in our experiments, we validate our design and training choices through a comprehensive ablation study; compare causal NFs to other approaches for approximating causal models; and empirically demonstrate that causal NFs can be used to address real-world problems—where the presence of mixed discrete-continuous data and partial knowledge on the causal graph is the norm. The code for this work can be found at https://github.com/psanch21/causal-flows.
****
Maximum Average Randomly Sampled: A Scale Free and Non-parametric Algorithm for Stochastic Bandits
Masoud Moravej Khorasani, Erik Weyer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Cappy: Outperforming and Boosting Large Multi-Task LMs with a Small Scorer
Bowen Tan, Yun Zhu, Lijuan Liu, Eric Xing, Zhiting Hu, Jindong Chen
Large language models (LLMs) such as T0, FLAN, and OPT-IML excel in multi-tasking under a unified instruction-following paradigm, where they also exhibit remarkable generalization abilities to unseen tasks. Despite their impressive performance, these LLMs, with sizes ranging from several billion to hundreds of billions of parameters, demand substantial computational resources, making their training and inference expensive and inefficient. Furthermore, adapting these models to downstream applications, particularly complex tasks, is often unfeasible due to the extensive hardware requirements for finetuning, even when utilizing parameter-efficient approaches such as prompt tuning. Additionally, the most powerful multi-task LLMs, such as OPT-IML-175B and FLAN-PaLM-540B, are not publicly accessible, severely limiting their customization potential. To address these challenges, we introduce a pretrained small scorer, \textit{Cappy}, designed to enhance the performance and efficiency of multi-task LLMs. With merely 360 million parameters, Cappy functions either independently on classification tasks or serve as an auxiliary component for LLMs, boosting their performance. Moreover, Cappy enables efficiently integrating downstream supervision without requiring LLM finetuning nor the access to their parameters. Our experiments demonstrate that, when working independently on 11 language understanding tasks from PromptSource, Cappy outperforms LLMs that are several orders of magnitude larger. Besides, on 45 complex tasks from BIG-Bench, Cappy boosts the performance of the advanced multi-task LLM, FLAN-T5, by a large margin. Furthermore, Cappy is flexible to cooperate with other LLM adaptations, including finetuning and in-context learning, offering additional performance enhancement.
****
Enhancing Adaptive History Reserving by Spiking Convolutional Block Attention Module in Recurrent Neural Networks
Qi Xu, Yuyuan Gao, Jiangrong Shen, Yaxin Li, Xuming Ran, Huajin Tang, Gang Pan
Spiking neural networks (SNNs) serve as one type of efficient model to process spatio-temporal patterns in time series, such as the Address-Event Representation data collected from Dynamic Vision Sensor (DVS). Although convolutional SNNs have achieved remarkable performance on these AER datasets, benefiting from the predominant spatial feature extraction ability of convolutional structure, they ignore temporal features related to sequential time points. In this paper, we develop a recurrent spiking neural network (RSNN) model embedded with an advanced spiking convolutional block attention module (SCBAM) component to combine both spatial and temporal features of spatio-temporal patterns. It invokes the history i

nformation in spatial and temporal channels adaptively through SCBAM, which brin
gs the advantages of efficient memory calling and history redundancy elimination
. The performance of our model was evaluated in DVS128-Gesture dataset and other
 time-series datasets. The experimental results show that the proposed SRNN-SCBA
M model makes better use of the history information in spatial and temporal dime
nsions with less memory space, and achieves higher accuracy compared to other mo
dels.
****
Reducing Shape-Radiance Ambiguity in Radiance Fields with a Closed-Form Color Es
timation Method
Qihang Fang, Yafei Song, Keqiang Li, Liefeng Bo
A neural radiance field (NeRF) enables the synthesis of cutting-edge realistic n
ovel view images of a 3D scene. It includes density and color fields to model th
e shape and radiance of a scene, respectively. Supervised by the photometric los
s in an end-to-end training manner, NeRF inherently suffers from the shape-radia
nce ambiguity problem, i.e., it can perfectly fit training views but does not gu
arantee decoupling the two fields correctly. To deal with this issue, existing w
orks have incorporated prior knowledge to provide an independent supervision sig
nal for the density field, including total variation loss, sparsity loss, distor
tion loss, etc. These losses are based on general assumptions about the density
field, e.g., it should be smooth, sparse, or compact, which are not adaptive to
a specific scene. In this paper, we propose a more adaptive method to reduce the
 shape-radiance ambiguity. The key is a rendering method that is only based on t
he density field. Specifically, we first estimate the color field based on the d
ensity field and posed images in a closed form. Then NeRF's rendering process ca
n proceed. We address the problems in estimating the color field, including occl
usion and non-uniformly distributed views. Afterwards, it is applied to regulari
ze NeRF's density field. As our regularization is guided by photometric loss, it
 is more adaptive compared to existing ones. Experimental results show that our
method improves the density field of NeRF both qualitatively and quantitatively.
 Our code is available at https://github.com/qihangGH/Closed-form-color-field.
****
Text-to-Image Diffusion Models are Zero Shot Classifiers
Kevin Clark, Priyank Jaini
The excellent generative capabilities of text-to-image diffusion models suggest
they learn informative representations of image-text data.However, what knowledg
e their representations capture is not fully understood, and they have not been
thoroughly explored on downstream tasks.We investigate diffusion models by propo
sing a method for evaluating them as zero-shot classifiers.The key idea is using
 a diffusion model's ability to denoise a noised image given a text description
of a label as a proxy for that label's likelihood.We apply our method to Stable
Diffusion and Imagen, using it to probe fine-grained aspects of the models' know
ledge and comparing them with CLIP's zero-shot abilities. They perform competiti
vely with CLIP on a wide range of zero-shot image classification datasets. Addit
ionally, they achieve state-of-the-art results on shape/texture bias tests and c
an successfully perform attribute binding while CLIP cannot.Although generative
pre-training is prevalent in NLP, visual foundation models often use other metho
ds such as contrastive learning. Based on our findings, we argue that generative
 pre-training should be explored as a compelling alternative for vision and visi
on-language problems.
****
MADG: Margin-based Adversarial Learning for Domain Generalization
Aveen Dayal, Vimal K B, Linga Reddy Cenkeramaddi, C Mohan, Abhinav Kumar, Vineet
h N Balasubramanian
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Bridging RL Theory and Practice with the Effective Horizon

Cassidy Laidlaw, Stuart J Russell, Anca Dragan
Deep reinforcement learning (RL) works impressively in some environments and fails catastrophically in others. Ideally, RL theory should be able to provide an understanding of why this is, i.e. bounds predictive of practical performance. Unfortunately, current theory does not quite have this ability. We compare standard deep RL algorithms to prior sample complexity bounds by introducing a new dataset, BRIDGE. It consists of 155 MDPs from common deep RL benchmarks, along with their corresponding tabular representations, which enables us to exactly compute instance-dependent bounds. We find that prior bounds do not correlate well with when deep RL succeeds vs. fails, but discover a surprising property that does. When actions with the highest Q-values under the random policy also have the highest Q-values under the optimal policy—i.e., when it is optimal to act greedily with respect to the random's policy Q function—deep RL tends to succeed; when they don't, deep RL tends to fail. We generalize this property into a new complexity measure of an MDP that we call the effective horizon, which roughly corresponds to how many steps of lookahead search would be needed in that MDP in order to identify the next optimal action, when leaf nodes are evaluated with random rollouts. Using BRIDGE, we show that the effective horizon-based bounds are more closely reflective of the empirical performance of PPO and DQN than prior sample complexity bounds across four metrics. We also show that, unlike existing bounds, the effective horizon can predict the effects of using reward shaping or a pre-trained exploration policy. Our code and data are available at https://github.com/cassidylaidlaw/effective-horizon.
****
Fine-Grained Human Feedback Gives Better Rewards for Language Model Training
Zeqiu Wu, Yushi Hu, Weijia Shi, Nouha Dziri, Alane Suhr, Prithviraj Ammanabrolu, Noah A. Smith, Mari Ostendorf, Hannaneh Hajishirzi
Language models (LMs) often exhibit undesirable text generation behaviors, including generating false, toxic, or irrelevant outputs. Reinforcement learning from human feedback (RLHF)---where human preference judgments on LM outputs are transformed into a learning signal---has recently shown promise in addressing these issues. However, such holistic feedback conveys limited information on long text outputs; it does not indicate which aspects of the outputs influenced user preference; e.g., which parts contain what type(s) of errors. In this paper, we use fine-grained human feedback (e.g., which sentence is false, which sub-sentence is irrelevant) as an explicit training signal. We introduce Fine-Grained RLHF, a framework that enables training and learning from reward functions that are fine-grained in two respects: (1) density, providing a reward after every segment (e.g., a sentence) is generated; and (2) incorporating multiple reward models associated with different feedback types (e.g., factual incorrectness, irrelevance, and information incompleteness). We conduct experiments on detoxification and long-form question answering to illustrate how learning with this reward function leads to improved performance, supported by both automatic and human evaluation. Additionally, we show that LM behaviors can be customized using different combinations of fine-grained reward models. We release all data, collected human feedback, and codes at https://FineGrainedRLHF.github.io.
****
Towards Optimal Effective Resistance Estimation
Rajat Vadiraj Dwaraknath, Ishani Karmarkar, Aaron Sidford
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
TradeMaster: A Holistic Quantitative Trading Platform Empowered by Reinforcement Learning
Shuo Sun, Molei Qin, Wentao Zhang, Haochong Xia, Chuqiao Zong, Jie Ying, Yonggang Xie, Lingxuan Zhao, Xinrun Wang, Bo An
The financial markets, which involve over \$90 trillion market capitals, attract the attention of innumerable profit-seeking investors globally. Recent explosio

n of reinforcement learning in financial trading (RLFT) research has shown stell
ar performance on many quantitative trading tasks. However, it is still challeng
ing to deploy reinforcement learning (RL) methods into real-world financial mark
ets due to the highly composite nature of this domain, which entails design choi
ces and interactions between components that collect financial data, conduct fea
ture engineering, build market environments, make investment decisions, evaluate
 model behaviors and offers user interfaces. Despite the availability of abundan
t financial data and advanced RL techniques, a remarkable gap still exists betwe
en the potential and realized utilization of RL in financial trading. In particu
lar, orchestrating an RLFT project lifecycle poses challenges in engineering (i.
e. hard to build), benchmarking (i.e. hard to compare) and usability (i.e. hard
to optimize, maintain and use). To overcome these challenges, we introduce Trade
Master, a holistic open-source RLFT platform that serves as a i) software toolki
t, ii) empirical benchmark, and iii) user interface. Our ultimate goal is to pro
vide infrastructures for transparent and reproducible RLFT research and facilita
te their real-world deployment with industry impact. TradeMaster will be updated
 continuously and welcomes contributions from both RL and finance communities.
****
Towards Optimal Caching and Model Selection for Large Model Inference
Banghua Zhu, Ying Sheng, Lianmin Zheng, Clark Barrett, Michael Jordan, Jiantao J
iao
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Deep Non-line-of-sight Imaging from Under-scanning Measurements
Yue Li, Yueyi Zhang, Juntian Ye, Feihu Xu, Zhiwei Xiong
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Learning Adversarial Low-rank Markov Decision Processes with Unknown Transition
and Full-information Feedback
Canzhe Zhao, Ruofeng Yang, Baoxiang Wang, Xuezhou Zhang, Shuai Li
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
UE4-NeRF:Neural Radiance Field for Real-Time Rendering of Large-Scale Scene
Jiaming Gu, Minchao Jiang, Hongsheng Li, Xiaoyuan Lu, Guangming Zhu, Syed Afaq A
li Shah, Liang Zhang, Mohammed Bennamoun
Neural Radiance Fields (NeRF) is a novel implicit 3D reconstruction method that
shows immense potential and has been gaining increasing attention. It enables th
e reconstruction of 3D scenes solely from a set of photographs. However, its rea
l-time rendering capability, especially for interactive real-time rendering of l
arge-scale scenes, still has significant limitations. To address these challenge
s, in this paper, we propose a novel neural rendering system called UE4-NeRF, sp
ecifically designed for real-time rendering of large-scale scenes. We partitione
d each large scene into different sub-NeRFs. In order to represent the partition
ed independent scene, we initialize polygonal meshes by constructing multiple re
gular octahedra within the scene and the vertices of the polygonal faces are co
ntinuously optimized during the training process. Drawing inspiration from Level
 of Detail (LOD) techniques, we trained meshes of varying levels of detail for d
ifferent observation levels. Our approach combines with the rasterization pipeli
ne in Unreal Engine 4 (UE4), achieving real-time rendering of large-scale scenes
 at 4K resolution with a frame rate of up to 43 FPS. Rendering within UE4 also f
acilitates scene editing in subsequent stages. Furthermore, through experiments

, we have demonstrated that our method achieves rendering quality comparable to state-of-the-art approaches. Project page: https://jamchaos.github.io/UE4-NeRF/.
****
H2RBox-v2: Incorporating Symmetry for Boosting Horizontal Box Supervised Oriented Object Detection
Yi Yu, Xue Yang, Qingyun Li, Yue Zhou, Feipeng Da, Junchi Yan
With the rapidly increasing demand for oriented object detection, e.g. in autonomous driving and remote sensing, the recently proposed paradigm involving weakly-supervised detector H2RBox for learning rotated box (RBox) from the more readily-available horizontal box (HBox) has shown promise. This paper presents H2RBox-v2, to further bridge the gap between HBox-supervised and RBox-supervised oriented object detection. Specifically, we propose to leverage the reflection symmetry via flip and rotate consistencies, using a weakly-supervised network branch similar to H2RBox, together with a novel self-supervised branch that learns orientations from the symmetry inherent in visual objects. The detector is further stabilized and enhanced by practical techniques to cope with peripheral issues e.g. angular periodicity. To our best knowledge, H2RBox-v2 is the first symmetry-aware self-supervised paradigm for oriented object detection. In particular, our method shows less susceptibility to low-quality annotation and insufficient training data compared to H2RBox. Specifically, H2RBox-v2 achieves very close performance to a rotation annotation trained counterpart -- Rotated FCOS: 1) DOTA-v1.0/1.5/2.0: 72.31%/64.76%/50.33% vs. 72.44%/64.53%/51.77%; 2) HRSC: 89.66% vs. 88.99%; 3) FAIR1M: 42.27% vs. 41.25%.
****
The Waymo Open Sim Agents Challenge
Nico Montali, John Lambert, Paul Mougin, Alex Kuefler, Nicholas Rhinehart, Michelle Li, Cole Gulino, Tristan Emrich, Zoey Yang, Shimon Whiteson, Brandyn White, Dragomir Anguelov
Simulation with realistic, interactive agents represents a key task for autonomous vehicle software development. In this work, we introduce the Waymo Open Sim Agents Challenge (WOSAC). WOSAC is the first public challenge to tackle this task and propose corresponding metrics. The goal of the challenge is to stimulate the design of realistic simulators that can be used to evaluate and train a behavior model for autonomous driving. We outline our evaluation methodology, present results for a number of different baseline simulation agent methods, and analyze several submissions to the 2023 competition which ran from March 16, 2023 to May 23, 2023. The WOSAC evaluation server remains open for submissions and we discuss open problems for the task.
****
Online RL in Linearly q^π-Realizable MDPs Is as Easy as in Linear MDPs If You Learn What to Ignore
Gellert Weisz, András György, Csaba Szepesvari
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On Transfer of Adversarial Robustness from Pretraining to Downstream Tasks
Laura F. Nern, Harsh Raj, Maurice André Georgi, Yash Sharma
As large-scale training regimes have gained popularity, the use of pretrained models for downstream tasks has become common practice in machine learning. While pretraining has been shown to enhance the performance of models in practice, the transfer of robustness properties from pretraining to downstream tasks remains poorly understood. In this study, we demonstrate that the robustness of a linear predictor on downstream tasks can be constrained by the robustness of its underlying representation, regardless of the protocol used for pretraining. We prove (i) a bound on the loss that holds independent of any downstream task, as well as (ii) a criterion for robust classification in particular. We validate our theoretical results in practical applications, show how our results can be used for calibrating expectations of downstream robustness, and when our results are usef

ul for optimal transfer learning. Taken together, our results offer an initial step towards characterizing the requirements of the representation function for reliable post-adaptation performance.
****
Entropy-dissipation Informed Neural Network for McKean-Vlasov Type PDEs
Zebang Shen, Zhenfu Wang
The McKean-Vlasov equation (MVE) describes the collective behavior of particles subject to drift, diffusion, and mean-field interaction. In physical systems, the interaction term can be singular, i.e. it diverges when two particles collide. Notable examples of such interactions include the Coulomb interaction, fundamental in plasma physics, and the Biot-Savart interaction, present in the vorticity formulation of the 2D Navier-Stokes equation (NSE) in fluid dynamics. Solving MVEs that involve singular interaction kernels presents a significant challenge, especially when aiming to provide rigorous theoretical guarantees. In this work, we propose a novel approach based on the concept of entropy dissipation in the underlying system. We derive a potential function that effectively controls the KL divergence between a hypothesis solution and the ground truth. Building upon this theoretical foundation, we introduce the Entropy-dissipation Informed Neural Network (EINN) framework for solving MVEs. In EINN, we utilize neural networks (NN) to approximate the underlying velocity field and minimize the proposed potential function. By leveraging the expressive power of NNs, our approach offers a promising avenue for tackling the complexities associated with singular interactions. To assess the empirical performance of our method, we compare EINN with SOTA NN-based MVE solvers. The results demonstrate the effectiveness of our approach in solving MVEs across various example problems.
****
Revisiting Visual Model Robustness: A Frequency Long-Tailed Distribution View
Zhiyu Lin, Yifei Gao, Yunfan Yang, Jitao Sang
A widely discussed hypothesis regarding the cause of visual models' lack of robustness is that they can exploit human-imperceptible high-frequency components (HFC) in images, which in turn leads to model vulnerabilities, such as the adversarial examples. However, (1) inconsistent findings regarding the validation of this hypothesis reflect in a limited understanding of HFC, and (2) solutions inspired by the hypothesis tend to involve a robustness-accuracy trade-off and leaning towards suppressing the model's learning on HFC. In this paper, inspired by the long-tailed characteristic observed in frequency spectrum, we first formally define the HFC from long-tailed perspective and then revisit the relationship between HFC and model robustness. In the frequency long-tailed scenario, experimental results on common datasets and various network structures consistently indicate that models in standard training exhibit high sensitivity to HFC. We investigate the reason of the sensitivity, which reflects in model's under-fitting behavior on HFC. Furthermore, the cause of the model's under-fitting behavior is attributed to the limited information content in HFC. Based on these findings, we propose a Balance Spectrum Sampling (BaSS) strategy, which effectively counteracts the long-tailed effect and enhances the model's learning on HFC. Extensive experimental results demonstrate that our method achieves a substantially better robustness-accuracy trade-off when combined with existing defense methods, while also indicating the potential of encouraging HFC learning in improving model performance.
****
How to Fine-tune the Model: Unified Model Shift and Model Bias Policy Optimization
Hai Zhang, Hang Yu, Junqiao Zhao, Di Zhang, xiao zhang, Hongtu Zhou, Chang Huang, Chen Ye
Designing and deriving effective model-based reinforcement learning (MBRL) algorithms with a performance improvement guarantee is challenging, mainly attributed to the high coupling between model learning and policy optimization. Many prior methods that rely on return discrepancy to guide model learning ignore the impacts of model shift, which can lead to performance deterioration due to excessive model updates. Other methods use performance difference bound to explicitly con

sider model shift. However, these methods rely on a fixed threshold to constrain model shift, resulting in a heavy dependence on the threshold and a lack of adaptability during the training process. In this paper, we theoretically derive an optimization objective that can unify model shift and model bias and then formulate a fine-tuning process. This process adaptively adjusts the model updates to get a performance improvement guarantee while avoiding model overfitting. Based on these, we develop a straightforward algorithm USB-PO (Unified model Shift and model Bias Policy Optimization). Empirical results show that USB-PO achieves state-of-the-art performance on several challenging benchmark tasks.
****

Dynamic Pricing and Learning with Bayesian Persuasion
Shipra Agrawal, Yiding Feng, Wei Tang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

RH-BrainFS: Regional Heterogeneous Multimodal Brain Networks Fusion Strategy
Hongting Ye, Yalu Zheng, Yueying Li, Ke Zhang, Youyong Kong, Yonggui Yuan
Multimodal fusion has become an important research technique in neuroscience that completes downstream tasks by extracting complementary information from multiple modalities. Existing multimodal research on brain networks mainly focuses on two modalities, structural connectivity (SC) and functional connectivity (FC). Recently, extensive literature has shown that the relationship between SC and FC is complex and not a simple one-to-one mapping. The coupling of structure and function at the regional level is heterogeneous. However, all previous studies have neglected the modal regional heterogeneity between SC and FC and fused their representations via "simple patterns", which are inefficient ways of multimodal fusion and affect the overall performance of the model. In this paper, to alleviate the issue of regional heterogeneity of multimodal brain networks, we propose a novel Regional Heterogeneous multimodal Brain networks Fusion Strategy (RH-BrainFS). Briefly, we introduce a brain subgraph networks module to extract regional characteristics of brain networks, and further use a new transformer-based fusion bottleneck module to alleviate the issue of regional heterogeneity between SC and FC. To the best of our knowledge, this is the first paper to explicitly state the issue of structural-functional modal regional heterogeneity and to propose asolution. Extensive experiments demonstrate that the proposed method outperforms several state-of-the-art methods in a variety of neuroscience tasks.
****

To Repeat or Not To Repeat: Insights from Scaling LLM under Token-Crisis
Fuzhao Xue, Yao Fu, Wangchunshu Zhou, Zangwei Zheng, Yang You
Recent research has highlighted the importance of dataset size in scaling language models. However, large language models (LLMs) are notoriously token-hungry during pre-training, and high-quality text data on the web is likely to be approaching its scaling limit for LLMs. To further enhance LLMs, a straightforward approach is to repeat the pre-training data for additional epochs. In this study, we empirically investigate three key aspects under this approach. First, we explore the consequences of repeating pre-training data, revealing that the model is susceptible to overfitting, leading to multi-epoch degradation. Second, we examine the key factors contributing to multi-epoch degradation, finding that significant factors include dataset size, model parameters, and training objectives, while less influential factors consist of dataset quality and model FLOPs. Finally, we explore whether widely used regularization can alleviate multi-epoch degradation. Most regularization techniques do not yield significant improvements, except for dropout, which demonstrates remarkable effectiveness but requires careful tuning when scaling up the model size. Additionally, we discover that leveraging mixture-of-experts (MoE) enables cost-effective and efficient hyper-parameter tuning for computationally intensive dense LLMs with comparable trainable parameters, potentially impacting efficient LLM development on a broader scale.
****

A*Net: A Scalable Path-based Reasoning Approach for Knowledge Graphs
Zhaocheng Zhu, Xinyu Yuan, Michael Galkin, Louis-Pascal Xhonneux, Ming Zhang, Maxime Gazeau, Jian Tang
Reasoning on large-scale knowledge graphs has been long dominated by embedding methods. While path-based methods possess the inductive capacity that embeddings lack, their scalability is limited by the exponential number of paths. Here we present A*Net, a scalable path-based method for knowledge graph reasoning. Inspired by the A* algorithm for shortest path problems, our A*Net learns a priority function to select important nodes and edges at each iteration, to reduce time and memory footprint for both training and inference. The ratio of selected nodes and edges can be specified to trade off between performance and efficiency. Experiments on both transductive and inductive knowledge graph reasoning benchmarks show that A*Net achieves competitive performance with existing state-of-the-art path-based methods, while merely visiting 10% nodes and 10% edges at each iteration. On a million-scale dataset ogbl-wikikg2, A*Net not only achieves a new state-of-the-art result, but also converges faster than embedding methods. A*Net is the first path-based method for knowledge graph reasoning at such scale.
****
HASSOD: Hierarchical Adaptive Self-Supervised Object Detection
Shengcao Cao, Dhiraj Joshi, Liangyan Gui, Yu-Xiong Wang
The human visual perception system demonstrates exceptional capabilities in learning without explicit supervision and understanding the part-to-whole composition of objects. Drawing inspiration from these two abilities, we propose Hierarchical Adaptive Self-Supervised Object Detection (HASSOD), a novel approach that learns to detect objects and understand their compositions without human supervision. HASSOD employs a hierarchical adaptive clustering strategy to group regions into object masks based on self-supervised visual representations, adaptively determining the number of objects per image. Furthermore, HASSOD identifies the hierarchical levels of objects in terms of composition, by analyzing coverage relations between masks and constructing tree structures. This additional self-supervised learning task leads to improved detection performance and enhanced interpretability. Lastly, we abandon the inefficient multi-round self-training process utilized in prior methods and instead adapt the Mean Teacher framework from semi-supervised learning, which leads to a smoother and more efficient training process. Through extensive experiments on prevalent image datasets, we demonstrate the superiority of HASSOD over existing methods, thereby advancing the state of the art in self-supervised object detection. Notably, we improve Mask AR from 20.2 to 22.5 on LVIS, and from 17.0 to 26.0 on SA-1B. Project page: https://HASSOD-NeurIPS23.github.io.
****
Addressing the speed-accuracy simulation trade-off for adaptive spiking neurons
Luke Taylor, Andrew King, Nicol S Harper
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Re-Think and Re-Design Graph Neural Networks in Spaces of Continuous Graph Diffusion Functionals
Tingting Dan, Jiaqi Ding, Ziquan Wei, Shahar Kovalsky, Minjeong Kim, Won Hwa Kim, Guorong Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Adapting Fairness Interventions to Missing Values
Raymond Feng, Flavio Calmon, Hao Wang
Missing values in real-world data pose a significant and unique challenge to algorithmic fairness. Different demographic groups may be unequally affected by mis

sing data, and the standard procedure for handling missing values where first da
ta is imputed, then the imputed data is used for classification—a procedure refe
rred to as "impute-then-classify"—can exacerbate discrimination. In this paper,
we analyze how missing values affect algorithmic fairness. We first prove that t
raining a classifier from imputed data can significantly worsen the achievable v
alues of group fairness and average accuracy. This is because imputing data resu
lts in the loss of the missing pattern of the data, which often conveys informat
ion about the predictive label. We present scalable and adaptive algorithms for
fair classification with missing values. These algorithms can be combined with a
ny preexisting fairness-intervention algorithm to handle all possible missing pa
tterns while preserving information encoded within the missing patterns. Numeric
al experiments with state-of-the-art fairness interventions demonstrate that our
 adaptive algorithms consistently achieve higher fairness and accuracy than impu
te-then-classify across different datasets.
****
Probabilistic Weight Fixing: Large-scale training of neural network weight uncer
tainties for quantisation.
Chris Subia-Waud, Srinandan Dasmahapatra
Weight-sharing quantization has emerged as a technique to reduce energy expendit
ure during inference in large neural networks by constraining their weights to a
 limited set of values. However, existing methods often assume weights are treat
ed solely based on value, neglecting the unique role of weight position. This pa
per proposes a probabilistic framework based on Bayesian neural networks (BNNs)
and a variational relaxation to identify which weights can be moved to which clu
ster center and to what degree based on their individual position-specific learn
ed uncertainty distributions. We introduce a new initialization setting and a re
gularization term, enabling the training of BNNs with complex dataset-model comb
inations. Leveraging the flexibility of weight values from probability distribut
ions, we enhance noise resilience and compressibility. Our iterative clustering
procedure demonstrates superior compressibility and higher accuracy compared to
state-of-the-art methods on both ResNet models and the more complex transformer-
based architectures. In particular, our method outperforms the state-of-the-art
quantization method top-1 accuracy by 1.6\% on ImageNet using DeiT-Tiny, with it
s 5 million+ weights now represented by only 296 unique values. Code available
at https://github.com/subiawaud/PWFN.
****
PID-Inspired Inductive Biases for Deep Reinforcement Learning in Partially Obser
vable Control Tasks
Ian Char, Jeff Schneider
Deep reinforcement learning (RL) has shown immense potential for learning to con
trol systems through data alone. However, one challenge deep RL faces is that th
e full state of the system is often not observable. When this is the case, the p
olicy needs to leverage the history of observations to infer the current state.
At the same time, differences between the training and testing environments make
s it critical for the policy not to overfit to the sequence of observations it s
ees at training time. As such, there is an important balancing act between havin
g the history encoder be flexible enough to extract relevant information, yet be
 robust to changes in the environment. To strike this balance, we look to the PI
D controller for inspiration. We assert the PID controller's success shows that
only summing and differencing are needed to accumulate information over time for
 many control tasks. Following this principle, we propose two architectures for
encoding history: one that directly uses PID features and another that extends t
hese core ideas and can be used in arbitrary control tasks. When compared with p
rior approaches, our encoders produce policies that are often more robust and ac
hieve better performance on a variety of tracking tasks. Going beyond tracking t
asks, our policies achieve 1.7x better performance on average over previous stat
e-of-the-art methods on a suite of locomotion control tasks.
****
SustainGym: Reinforcement Learning Environments for Sustainable Energy Systems
Christopher Yeh, Victor Li, Rajeev Datta, Julio Arroyo, Nicolas Christianson, Ch

i Zhang, Yize Chen, Mohammad Mehdi Hosseini, Azarang Golmohammadi, Yuanyuan Shi, Yisong Yue, Adam Wierman
The lack of standardized benchmarks for reinforcement learning (RL) in sustainability applications has made it difficult to both track progress on specific domains and identify bottlenecks for researchers to focus their efforts. In this paper, we present SustainGym, a suite of five environments designed to test the performance of RL algorithms on realistic sustainable energy system tasks, ranging from electric vehicle charging to carbon-aware data center job scheduling. The environments test RL algorithms under realistic distribution shifts as well as in multi-agent settings. We show that standard off-the-shelf RL algorithms leave significant room for improving performance and highlight the challenges ahead for introducing RL to real-world sustainability tasks.
****
Policy Gradient for Rectangular Robust Markov Decision Processes
Navdeep Kumar, Esther Derman, Matthieu Geist, Kfir Y. Levy, Shie Mannor
Policy gradient methods have become a standard for training reinforcement learning agents in a scalable and efficient manner. However, they do not account for transition uncertainty, whereas learning robust policies can be computationally expensive. In this paper, we introduce robust policy gradient (RPG), a policy-based method that efficiently solves rectangular robust Markov decision processes (MDPs). We provide a closed-form expression for the worst occupation measure. Incidentally, we find that the worst kernel is a rank-one perturbation of the nominal. Combining the worst occupation measure with a robust Q-value estimation yields an explicit form of the robust gradient. Our resulting RPG can be estimated from data with the same time complexity as its non-robust equivalent. Hence, it relieves the computational burden of convex optimization problems required for training robust policies by current policy gradient approaches.
****
MultiFusion: Fusing Pre-Trained Models for Multi-Lingual, Multi-Modal Image Generation
Marco Bellagente, Manuel Brack, Hannah Teufel, Felix Friedrich, Björn Deiseroth, Constantin Eichenberg, Andrew M. Dai, Robert Baldock, Souradeep Nanda, Koen Oostermeijer, Andres Felipe Cruz-Salinas, Patrick Schramowski, Kristian Kersting, Samuel Weinbach
The recent popularity of text-to-image diffusion models (DM) can largely be attributed to the intuitive interface they provide to users. The intended generation can be expressed in natural language, with the model producing faithful interpretations of text prompts. However, expressing complex or nuanced ideas in text alone can be difficult. To ease image generation, we propose MultiFusion that allows one to express complex and nuanced concepts with arbitrarily interleaved inputs of multiple modalities and languages. MultiFusion leverages pre-trained models and aligns them for integration into a cohesive system, thereby avoiding the need for extensive training from scratch. Our experimental results demonstrate the efficient transfer of capabilities from individual modules to the downstream model. Specifically, the fusion of all independent components allows the image generation module to utilize multilingual, interleaved multimodal inputs despite being trained solely on monomodal data in a single language.
****
What Planning Problems Can A Relational Neural Network Solve?
Jiayuan Mao, Tomás Lozano-Pérez, Josh Tenenbaum, Leslie Kaelbling
Goal-conditioned policies are generally understood to be "feed-forward" circuits, in the form of neural networks that map from the current state and the goal specification to the next action to take. However, under what circumstances such a policy can be learned and how efficient the policy will be are not well understood. In this paper, we present a circuit complexity analysis for relational neural networks (such as graph neural networks and transformers) representing policies for planning problems, by drawing connections with serialized goal regression search (S-GRS). We show that there are three general classes of planning problems, in terms of the growth of circuit width and depth as a function of the number of objects and planning horizon, providing constructive proofs. We also illust

rate the utility of this analysis for designing neural networks for policy learn ing.
****

Selectively Sharing Experiences Improves Multi-Agent Reinforcement Learning
Matthias Gerstgrasser, Tom Danino, Sarah Keren
We present a novel multi-agent RL approach, Selective Multi-Agent Prioritized Ex perience Relay, in which agents share with other agents a limited number of tran sitions they observe during training. The intuition behind this is that even a s mall number of relevant experiences from other agents could help each agent lear n. Unlike many other multi-agent RL algorithms, this approach allows for largely decentralized training, requiring only a limited communication channel between agents. We show that our approach outperforms baseline no-sharing decentralized training and state-of-the art multi-agent RL algorithms. Further, sharing only a small number of highly relevant experiences outperforms sharing all experiences between agents, and the performance uplift from selective experience sharing is robust across a range of hyperparameters and DQN variants.
****

Learning From Biased Soft Labels
Hua Yuan, Yu Shi, Ning Xu, Xu Yang, Xin Geng, Yong Rui
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Learning from Visual Observation via Offline Pretrained State-to-Go Transformer
Bohan Zhou, Ke Li, Jiechuan Jiang, Zongqing Lu
Learning from visual observation (LfVO), aiming at recovering policies from only visual observation data, is promising yet a challenging problem. Existing LfVO approaches either only adopt inefficient online learning schemes or require addi tional task-specific information like goal states, making them not suited for op en-ended tasks. To address these issues, we propose a two-stage framework for le arning from visual observation. In the first stage, we introduce and pretrain St ate-to-Go (STG) Transformer offline to predict and differentiate latent transiti ons of demonstrations. Subsequently, in the second stage, the STG Transformer pr ovides intrinsic rewards for downstream reinforcement learning tasks where an ag ent learns merely from intrinsic rewards. Empirical results on Atari and Minecra ft show that our proposed method outperforms baselines and in some tasks even ac hieves performance comparable to the policy learned from environmental rewards. These results shed light on the potential of utilizing video-only data to solve difficult visual reinforcement learning tasks rather than relying on complete of fline datasets containing states, actions, and rewards. The project’s website an d code can befound at https://sites.google.com/view/stgtransformer.
****

Rotating Features for Object Discovery
Sindy Löwe, Phillip Lippe, Francesco Locatello, Max Welling
The binding problem in human cognition, concerning how the brain represents and connects objects within a fixed network of neural connections, remains a subject of intense debate. Most machine learning efforts addressing this issue in an un supervised setting have focused on slot-based methods, which may be limiting due to their discrete nature and difficulty to express uncertainty. Recently, the C omplex AutoEncoder was proposed as an alternative that learns continuous and dis tributed object-centric representations. However, it is only applicable to simpl e toy data. In this paper, we present Rotating Features, a generalization of com plex-valued features to higher dimensions, and a new evaluation procedure for ex tracting objects from distributed representations. Additionally, we show the app licability of our approach to pre-trained features. Together, these advancements enable us to scale distributed object-centric representations from simple toy t o real-world data. We believe this work advances a new paradigm for addressing t he binding problem in machine learning and has the potential to inspire further innovation in the field.

****
Grounded Decoding: Guiding Text Generation with Grounded Models for Embodied Age
nts
Wenlong Huang, Fei Xia, Dhruv Shah, Danny Driess, Andy Zeng, Yao Lu, Pete Floren
ce, Igor Mordatch, Sergey Levine, Karol Hausman, brian ichter
Recent progress in large language models (LLMs) has demonstrated the ability to
learn and leverage Internet-scale knowledge through pre-training with autoregres
sive models. Unfortunately, applying such models to settings with embodied agent
s, such as robots, is challenging due to their lack of experience with the physi
cal world, inability to parse non-language observations, and ignorance of reward
s or safety constraints that robots may require. On the other hand, language-con
ditioned robotic policies that learn from interaction data can provide the neces
sary grounding that allows the agent to be correctly situated in the real world,
 but such policies are limited by the lack of high-level semantic understanding
due to the limited breadth of the interaction data available for training them.
Thus, if we want to make use of the semantic knowledge in a language model while
 still situating it in an embodied setting, we must construct an action sequence
 that is both likely according to the language model and also realizable accordi
ng to grounded models of the environment. We frame this as a problem similar to
probabilistic filtering: decode a sequence that both has high probability under
the language model and high probability under a set of grounded model objectives
. We demonstrate how such grounded models can be obtained across three simulatio
n and real-world domains, and that the proposed decoding strategy is able to sol
ve complex, long-horizon embodiment tasks in a robotic setting by leveraging the
 knowledge of both models.
****
What can Large Language Models do in chemistry? A comprehensive benchmark on eig
ht tasks
Taicheng Guo, kehan Guo, Bozhao Nan, Zhenwen Liang, Zhichun Guo, Nitesh Chawla,
Olaf Wiest, Xiangliang Zhang
Large Language Models (LLMs) with strong abilities in natural language processin
g tasks have emerged and have been applied in various kinds of areas such as sci
ence, finance and software engineering. However, the capability of LLMs to advan
ce the field of chemistry remains unclear. In this paper, rather than pursuing s
tate-of-the-art performance, we aim to evaluate capabilities of LLMs in a wide r
ange of tasks across the chemistry domain. We identify three key chemistry-relat
ed capabilities including understanding, reasoning and explaining to explore in
LLMs and establish a benchmark containing eight chemistry tasks. Our analysis dr
aws on widely recognized datasets facilitating a broad exploration of the capaci
ties of LLMs within the context of practical chemistry. Five LLMs (GPT-4,GPT-3.5
, Davinci-003, Llama and Galactica) are evaluated for each chemistry task in zer
o-shot and few-shot in-context learning settings with carefully selected demonst
ration examples and specially crafted prompts. Our investigation found that GPT-
4 outperformed other models and LLMs exhibit different competitive levels in eig
ht chemistry tasks. In addition to the key findings from the comprehensive bench
mark analysis, our work provides insights into the limitation of current LLMs an
d the impact of in-context learning settings on LLMs’ performance across various
 chemistry tasks. The code and datasets used in this study are available at http
s://github.com/ChemFoundationModels/ChemLLMBench.
****
Focus Your Attention when Few-Shot Classification
Haoqing Wang, Shibo Jie, Zhihong Deng
Since many pre-trained vision transformers emerge and provide strong representat
ion for various downstream tasks, we aim to adapt them to few-shot image classif
ication tasks in this work. The input images typically contain multiple entities
. The model may not focus on the class-related entities for the current few-shot
 task, even with fine-tuning on support samples, and the noise information from
the class-independent ones harms performance. To this end, we first propose a me
thod that uses the attention and gradient information to automatically locate th
e positions of key entities, denoted as position prompts, in the support images.

Then we employ the cross-entropy loss between their many-hot presentation and the attention logits to optimize the model to focus its attention on the key entities during fine-tuning. This ability then can generalize to the query samples. Our method is applicable to different vision transformers (e.g., columnar or pyramidal ones), and also to different pre-training ways (e.g., single-modal or vision-language pre-training). Extensive experiments show that our method can improve the performance of full or parameter-efficient fine-tuning methods on few-shot tasks. Code is available at https://github.com/Haoqing-Wang/FORT.
****
AndroidInTheWild: A Large-Scale Dataset For Android Device Control
Christopher Rawles, Alice Li, Daniel Rodriguez, Oriana Riva, Timothy Lillicrap
There is a growing interest in device-control systems that can interpret human natural language instructions and execute them on a digital device by directly controlling its user interface. We present a dataset for device-control research, Android in the Wild (AitW), which is orders of magnitude larger than current datasets. The dataset contains human demonstrations of device interactions, including the screens and actions, and corresponding natural language instructions. It consists of 715k episodes spanning 30k unique instructions, four versions of Android (v10–13), and eight device types (Pixel 2 XL to Pixel 6) with varying screen resolutions. It contains multi-step tasks that require semantic understanding of language and visual context. This dataset poses a new challenge: actions available through the user interface must be inferred from their visual appearance, and, instead of simple UI element-based actions, the action space consists of precise gestures (e.g., horizontal scrolls to operate carousel widgets). We organize our dataset to encourage robustness analysis of device-control systems, i.e., how well a system performs in the presence of new task descriptions, new applications, or new platform versions. We develop two agents and report performance across the dataset. The dataset is available at https://github.com/google-research/google-research/tree/master/androidinthe_wild.
****
Unifying GANs and Score-Based Diffusion as Generative Particle Models
Jean-Yves Franceschi, Mike Gartrell, Ludovic Dos Santos, Thibaut Issenhuth, Emmanuel de Bézenac, Mickael Chen, Alain Rakotomamonjy
Particle-based deep generative models, such as gradient flows and score-based diffusion models, have recently gained traction thanks to their striking performance. Their principle of displacing particle distributions using differential equations is conventionally seen as opposed to the previously widespread generative adversarial networks (GANs), which involve training a pushforward generator network. In this paper we challenge this interpretation, and propose a novel framework that unifies particle and adversarial generative models by framing generator training as a generalization of particle models. This suggests that a generator is an optional addition to any such generative model. Consequently, integrating a generator into a score-based diffusion model and training a GAN without a generator naturally emerge from our framework. We empirically test the viability of these original models as proofs of concepts of potential applications of our framework.
****
Path Regularization: A Convexity and Sparsity Inducing Regularization for Parallel ReLU Networks
Tolga Ergen, Mert Pilanci
Understanding the fundamental principles behind the success of deep neural networks is one of the most important open questions in the current literature. To this end, we study the training problem of deep neural networks and introduce an analytic approach to unveil hidden convexity in the optimization landscape. We consider a deep parallel ReLU network architecture, which also includes standard deep networks and ResNets as its special cases. We then show that pathwise regularized training problems can be represented as an exact convex optimization problem. We further prove that the equivalent convex problem is regularized via a group sparsity inducing norm. Thus, a path regularized parallel ReLU network can be viewed as a parsimonious convex model in high dimensions. More importantly, sin

ce the original training problem may not be trainable in polynomial-time, we pro pose an approximate algorithm with a fully polynomial-time complexity in all dat a dimensions. Then, we prove strong global optimality guarantees for this algori thm. We also provide experiments corroborating our theory.
****
SSL4EO-L: Datasets and Foundation Models for Landsat Imagery
Adam Stewart, Nils Lehmann, Isaac Corley, Yi Wang, Yi-Chia Chang, Nassim Ait Ait Ali Braham, Shradha Sehgal, Caleb Robinson, Arindam Banerjee
The Landsat program is the longest-running Earth observation program in history, with 50+ years of data acquisition by 8 satellites. The multispectral imagery c aptured by sensors onboard these satellites is critical for a wide range of scie ntific fields. Despite the increasing popularity of deep learning and remote sen sing, the majority of researchers still use decision trees and random forests fo r Landsat image analysis due to the prevalence of small labeled datasets and lac k of foundation models. In this paper, we introduce SSL4EO-L, the first ever dat aset designed for Self-Supervised Learning for Earth Observation for the Landsat family of satellites (including 3 sensors and 2 product levels) and the largest Landsat dataset in history (5M image patches). Additionally, we modernize and r e-release the L7 Irish and L8 Biome cloud detection datasets, and introduce the first ML benchmark datasets for Landsats 4–5 TM and Landsat 7 ETM+ SR. Finally, we pre-train the first foundation models for Landsat imagery using SSL4EO-L and evaluate their performance on multiple semantic segmentation tasks. All datasets and model weights are available via the TorchGeo library, making reproducibilit y and experimentation easy, and enabling scientific advancements in the burgeoni ng field of remote sensing for a multitude of downstream applications.
****
Fast Attention Over Long Sequences With Dynamic Sparse Flash Attention
Matteo Pagliardini, Daniele Paliotta, Martin Jaggi, François Fleuret
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Learning Better with Less: Effective Augmentation for Sample-Efficient Visual Re inforcement Learning
Guozheng Ma, Linrui Zhang, Haoyu Wang, Lu Li, Zilin Wang, Zhen Wang, Li Shen, Xu eqian Wang, Dacheng Tao
Data augmentation (DA) is a crucial technique for enhancing the sample efficienc y of visual reinforcement learning (RL) algorithms.Notably, employing simple obs ervation transformations alone can yield outstanding performance without extra a uxiliary representation tasks or pre-trained encoders. However, it remains uncle ar which attributes of DA account for its effectiveness in achieving sample-effi cient visual RL. To investigate this issue and further explore the potential of DA, this work conducts comprehensive experiments to assess the impact of DA's at tributes on its efficacy and provides the following insights and improvements: (1) For individual DA operations, we reveal that both ample spatial diversity and slight hardness are indispensable. Building on this finding, we introduce Rando m PadResize (Rand PR), a new DA operation that offers abundant spatial diversity with minimal hardness. (2) For multi-type DA fusion schemes, the increased DA h ardness and unstable data distribution result in the current fusion schemes bein g unable to achieve higher sample efficiency than their corresponding individual operations. Taking the non-stationary nature of RL into account, we propose a R L-tailored multi-type DA fusion scheme called Cycling Augmentation (CycAug), whi ch performs periodic cycles of different DA operations to increase type diversit y while maintaining data distribution consistency. Extensive evaluations on the DeepMind Control suite and CARLA driving simulator demonstrate that our methods achieve superior sample efficiency compared with the prior state-of-the-art meth ods.
****
Pairwise GUI Dataset Construction Between Android Phones and Tablets

han hu, Haolan Zhan, Yujin Huang, Di Liu
In the current landscape of pervasive smartphones and tablets, apps frequently exist across both platforms.Although apps share most graphic user interfaces (GUIs) and functionalities across phones and tablets, developers often rebuild from scratch for tablet versions, escalating costs and squandering existing design resources.Researchers are attempting to collect data and employ deep learning in automated GUIs development to enhance developers' productivity.There are currently several publicly accessible GUI page datasets for phones, but none for pairwise GUIs between phones and tablets.This poses a significant barrier to the employment of deep learning in automated GUI development.In this paper, we introduce the Papt dataset, a pioneering pairwise GUI dataset tailored for Android phones and tablets, encompassing 10,035 phone-tablet GUI page pairs sourced from 5,593 unique app pairs.We propose novel pairwise GUI collection approaches for constructing this dataset and delineate its advantages over currently prevailing datasets in the field.Through preliminary experiments on this dataset, we analyze the present challenges of utilizing deep learning in automated GUI development.
****
Improved Communication Efficiency in Federated Natural Policy Gradient via ADMM-based Gradient Updates
Guangchen Lan, Han Wang, James Anderson, Christopher Brinton, Vaneet Aggarwal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Equivariant flow matching
Leon Klein, Andreas Krämer, Frank Noe
Normalizing flows are a class of deep generative models that are especially interesting for modeling probability distributions in physics, where the exact likelihood of flows allows reweighting to known target energy functions and computing unbiased observables. For instance, Boltzmann generators tackle the long-standing sampling problem in statistical physics by training flows to produce equilibrium samples of many-body systems such as small molecules and proteins. To build effective models for such systems, it is crucial to incorporate the symmetries of the target energy into the model, which can be achieved by equivariant continuous normalizing flows (CNFs). However, CNFs can be computationally expensive to train and generate samples from, which has hampered their scalability and practical application.In this paper, we introduce equivariant flow matching, a new training objective for equivariant CNFs that is based on the recently proposed optimal transport flow matching. Equivariant flow matching exploits the physical symmetries of the target energy for efficient, simulation-free training of equivariant CNFs.We demonstrate the effectiveness of flow matching on rotation and permutation invariant many-particle systems and a small molecule, alanine dipeptide, where for the first time we obtain a Boltzmann generator with significant sampling efficiency without relying on tailored internal coordinate featurization. Our results show that the equivariant flow matching objective yields flows with shorter integration paths, improved sampling efficiency, and higher scalability compared to existing methods.
****
InsActor: Instruction-driven Physics-based Characters
Jiawei Ren, Mingyuan Zhang, Cunjun Yu, Xiao Ma, Liang Pan, Ziwei Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
An Efficient Doubly-Robust Test for the Kernel Treatment Effect
Diego Martinez Taboada, Aaditya Ramdas, Edward Kennedy
The average treatment effect, which is the difference in expectation of the counterfactuals, is probably the most popular target effect in causal inference with

binary treatments. However, treatments may have effects beyond the mean, for instance decreasing or increasing the variance. We propose a new kernel-based test for distributional effects of the treatment. It is, to the best of our knowledge, the first kernel-based, doubly-robust test with provably valid type-I error. Furthermore, our proposed algorithm is computationally efficient, avoiding the use of permutations.
****
Policy Finetuning in Reinforcement Learning via Design of Experiments using Offline Data
Ruiqi Zhang, Andrea Zanette
In some applications of reinforcement learning, a dataset of pre-collected experience is already availablebut it is also possible to acquire some additional online data to help improve the quality of the policy.However, it may be preferable to gather additional data with a single, non-reactive exploration policyand avoid the engineering costs associated with switching policies. In this paper we propose an algorithm with provable guarantees that can leverage an offline dataset to design a single non-reactive policy for exploration. We theoretically analyze the algorithm and measure the quality of the final policy as a function of the local coverage of the original dataset and the amount of additional data collected.
****
Zero-sum Polymatrix Markov Games: Equilibrium Collapse and Efficient Computation of Nash Equilibria
Fivos Kalogiannis, Ioannis Panageas
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning to Configure Separators in Branch-and-Cut
Sirui Li, Wenbin Ouyang, Max Paulus, Cathy Wu
Cutting planes are crucial in solving mixed integer linear programs (MILP) as they facilitate bound improvements on the optimal solution. Modern MILP solvers rely on a variety of separators to generate a diverse set of cutting planes by invoking the separators frequently during the solving process. This work identifies that MILP solvers can be drastically accelerated by appropriately selecting separators to activate. As the combinatorial separator selection space imposes challenges for machine learning, we learn to separate by proposing a novel data-driven strategy to restrict the selection space and a learning-guided algorithm on the restricted space. Our method predicts instance-aware separator configurations which can dynamically adapt during the solve, effectively accelerating the open source MILP solver SCIP by improving the relative solve time up to 72% and 37% on synthetic and real-world MILP benchmarks. Our work complements recent work on learning to select cutting planes and highlights the importance of separator management.
****
Multi-Object Representation Learning via Feature Connectivity and Object-Centric Regularization
Alex Foo, Wynne Hsu, Mong Li Lee
Discovering object-centric representations from images has the potential to greatly improve the robustness, sample efficiency and interpretability of machine learning algorithms. Current works on multi-object images typically follow a generative approach that optimizes for input reconstruction and fail to scale to real-world datasets despite significant increases in model capacity. We address this limitation by proposing a novel method that leverages feature connectivity to cluster neighboring pixels likely to belong to the same object. We further design two object-centric regularization terms to refine object representations in the latent space, enabling our approach to scale to complex real-world images. Experimental results on simulated, real-world, complex texture and common object images demonstrate a substantial improvement in the quality of discovered objects c

ompared to state-of-the-art methods, as well as the sample efficiency and generalizability of our approach. We also show that the discovered object-centric representations can accurately predict key object properties in downstream tasks, highlighting the potential of our method to advance the field of multi-object representation learning.
****
Ess-InfoGAIL: Semi-supervised Imitation Learning from Imbalanced Demonstrations
Huiqiao Fu, Kaiqiang Tang, Yuanyang Lu, Yiming Qi, Guizhou Deng, Flood Sung, Chunlin Chen
Imitation learning aims to reproduce expert behaviors without relying on an explicit reward signal. However, real-world demonstrations often present challenges, such as multi-modal, data imbalance, and expensive labeling processes. In this work, we propose a novel semi-supervised imitation learning architecture that learns disentangled behavior representations from imbalanced demonstrations using limited labeled data. Specifically, our method consists of three key components. First, we adapt the concept of semi-supervised generative adversarial networks to the imitation learning context. Second, we employ a learnable latent distribution to align the generated and expert data distributions. Finally, we utilize a regularized information maximization approach in conjunction with an approximate label prior to further improve the semi-supervised learning performance. Experimental results demonstrate the efficiency of our method in learning multi-modal behaviors from imbalanced demonstrations compared to baseline methods.
****
Revisiting Implicit Differentiation for Learning Problems in Optimal Control
Ming Xu, Timothy L. Molloy, Stephen Gould
This paper proposes a new method for differentiating through optimal trajectories arising from non-convex, constrained discrete-time optimal control (COC) problems using the implicit function theorem (IFT). Previous works solve a differential Karush-Kuhn-Tucker (KKT) system for the trajectory derivative, and achieve this efficiently by solving an auxiliary Linear Quadratic Regulator (LQR) problem. In contrast, we directly evaluate the matrix equations which arise from applying variable elimination on the Lagrange multiplier terms in the (differential) KKT system. By appropriately accounting for the structure of the terms within the resulting equations, we show that the trajectory derivatives scale linearly with the number of timesteps. Furthermore, our approach allows for easy parallelization, significantly improved scalability with model size, direct computation of vector-Jacobian products and improved numerical stability compared to prior works. As an additional contribution, we unify prior works, addressing claims that computing trajectory derivatives using IFT scales quadratically with the number of timesteps. We evaluate our method on a both synthetic benchmark and four challenging, learning from demonstration benchmarks including a 6-DoF maneuvering quadrotor and 6-DoF rocket powered landing.
****
p-Poisson surface reconstruction in curl-free flow from point clouds
Yesom Park, Taekyung Lee, Jooyoung Hahn, Myungjoo Kang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Binarized Neural Machine Translation
Yichi Zhang, Ankush Garg, Yuan Cao, Lukasz Lew, Behrooz Ghorbani, Zhiru Zhang, Orhan Firat
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Greedy Pruning with Group Lasso Provably Generalizes for Matrix Sensing
Nived Rajaraman, Fnu Devvrit, Aryan Mokhtari, Kannan Ramchandran

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
EgoEnv: Human-centric environment representations from egocentric video
Tushar Nagarajan, Santhosh Kumar Ramakrishnan, Ruta Desai, James Hillis, Kristen Grauman
First-person video highlights a camera-wearer's activities in the context of their persistent environment. However, current video understanding approaches reason over visual features from short video clips that are detached from the underlying physical space and capture only what is immediately visible. To facilitate human-centric environment understanding, we present an approach that links egocentric video and the environment by learning representations that are predictive of the camera-wearer's (potentially unseen) local surroundings. We train such models using videos from agents in simulated 3D environments where the environment is fully observable, and test them on human-captured real-world videos from unseen environments. On two human-centric video tasks, we show that models equipped with our environment-aware features consistently outperform their counterparts with traditional clip features. Moreover, despite being trained exclusively on simulated videos, our approach successfully handles real-world videos from HouseTours and Ego4D, and achieves state-of-the-art results on the Ego4D NLQ challenge.
****
Revisiting Evaluation Metrics for Semantic Segmentation: Optimization and Evaluation of Fine-grained Intersection over Union
Zifu Wang, Maxim Berman, Amal Rannen-Triki, Philip Torr, Devis Tuia, Tinne Tuytelaars, Luc V Gool, Jiaqian Yu, Matthew Blaschko
Semantic segmentation datasets often exhibit two types of imbalance: \textit{class imbalance}, where some classes appear more frequently than others and \textit{size imbalance}, where some objects occupy more pixels than others. This causes traditional evaluation metrics to be biased towards \textit{majority classes} (e.g. overall pixel-wise accuracy) and \textit{large objects} (e.g. mean pixel-wise accuracy and per-dataset mean intersection over union). To address these shortcomings, we propose the use of fine-grained mIoUs along with corresponding worst-case metrics, thereby offering a more holistic evaluation of segmentation techniques. These fine-grained metrics offer less bias towards large objects, richer statistical information, and valuable insights into model and dataset auditing. Furthermore, we undertake an extensive benchmark study, where we train and evaluate 15 modern neural networks with the proposed metrics on 12 diverse natural and aerial segmentation datasets. Our benchmark study highlights the necessity of not basing evaluations on a single metric and confirms that fine-grained mIoUs reduce the bias towards large objects. Moreover, we identify the crucial role played by architecture designs and loss functions, which lead to best practices in optimizing fine-grained metrics. The code is available at \href{https://github.com/zifuwanggg/JDTLosses}{https://github.com/zifuwanggg/JDTLosses}.
****
Optimal Unbiased Randomizers for Regression with Label Differential Privacy
Ashwinkumar Badanidiyuru Varadaraja, Badih Ghazi, Pritish Kamath, Ravi Kumar, Ethan Leeman, Pasin Manurangsi, Avinash V Varadarajan, Chiyuan Zhang
We propose a new family of label randomizers for training regression models under the constraint of label differential privacy (DP). In particular, we leverage the trade-offs between bias and variance to construct better label randomizers depending on a privately estimated prior distribution over the labels. We demonstrate that these randomizers achieve state-of-the-art privacy-utility trade-offs on several datasets, highlighting the importance of reducing bias when training neural networks with label DP. We also provide theoretical results shedding light on the structural properties of the optimal unbiased randomizers.
****
Understanding, Predicting and Better Resolving Q-Value Divergence in Offline-RL

Yang Yue, Rui Lu, Bingyi Kang, Shiji Song, Gao Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Online List Labeling with Predictions
Samuel McCauley, Ben Moseley, Aidin Niaparast, Shikha Singh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
How a Student becomes a Teacher: learning and forgetting through Spectral methods
Lorenzo Giambagli, Lorenzo Buffoni, Lorenzo Chicchi, Duccio Fanelli
In theoretical Machine Learning, the teacher-student paradigm is often employed as an effective metaphor for real-life tuition. A student network is trained on data generated by a fixed teacher network until it matches the instructor's ability to cope with the assigned task. The above scheme proves particularly relevant when the student network is overparameterized (namely, when larger layer sizes are employed) as compared to the underlying teacher network. Under these operating conditions, it is tempting to speculate that the student ability to handle the given task could be eventually stored in a sub-portion of the whole network. This latter should be to some extent reminiscent of the frozen teacher structure, according to suitable metrics, while being approximately invariant across different architectures of the student candidate network. Unfortunately, state-of-the-art conventional learning techniques could not help in identifying the existence of such an invariant subnetwork, due to the inherent degree of non-convexity that characterizes the examined problem. In this work, we take a decisive leap forward by proposing a radically different optimization scheme which builds on a spectral representation of the linear transfer of information between layers. The gradient is hence calculated with respect to both eigenvalues and eigenvectors with negligible increase in terms of computational and complexity load, as compared to standard training algorithms. Working in this framework, we could isolate a stable student substructure, that mirrors the true complexity of the teacher in terms of computing neurons, path distribution and topological attributes. When pruning unimportant nodes of the trained student, as follows a ranking that reflects the optimized eigenvalues, no degradation in the recorded performance is seen above a threshold that corresponds to the effective teacher size. The observed behavior can be pictured as a genuine second-order phase transition that bears universality traits.
****
PAPR: Proximity Attention Point Rendering
Yanshu Zhang, Shichong Peng, Alireza Moazeni, Ke Li
Learning accurate and parsimonious point cloud representations of scene surfaces from scratch remains a challenge in 3D representation learning. Existing point-based methods often suffer from the vanishing gradient problem or require a large number of points to accurately model scene geometry and texture. To address these limitations, we propose Proximity Attention Point Rendering (PAPR), a novel method that consists of a point-based scene representation and a differentiable renderer. Our scene representation uses a point cloud where each point is characterized by its spatial position, influence score, and view-independent feature vector. The renderer selects the relevant points for each ray and produces accurate colours using their associated features. PAPR effectively learns point cloud positions to represent the correct scene geometry, even when the initialization drastically differs from the target geometry. Notably, our method captures fine texture details while using only a parsimonious set of points. We also demonstrate four practical applications of our method: zero-shot geometry editing, object manipulation, texture transfer, and exposure control. More results and code ar

e available on our project website at https://zvict.github.io/papr/.
****

Enhancing Minority Classes by Mixing: An Adaptative Optimal Transport Approach for Long-tailed Classification

Jintong Gao, He Zhao, Zhuo Li, Dandan Guo

Real-world data usually confronts severe class-imbalance problems, where several majority classes have a significantly larger presence in the training set than minority classes. One effective solution is using mixup-based methods to generate synthetic samples to enhance the presence of minority classes. Previous approaches mix the background images from the majority classes and foreground images from theminority classes in a random manner, which ignores the sample-level semantic similarity, possibly resulting in less reasonable or less useful images. In this work, we propose an adaptive image-mixing method based on optimal transport (OT) to incorporate both class-level and sample-level information, which is able to generate semantically reasonable and meaningful mixed images for minority classes. Due toits flexibility, our method can be combined with existing long-tailed classification methods to enhance their performance and it can also serve as a general data augmentation method for balanced datasets. Extensive experiments indicate that our method achieves effective performance for long-tailed classification tasks. The code is available at https://github.com/JintongGao/Enhancing-Minority-Classes-by-Mixing.
****

Robust Data Valuation with Weighted Banzhaf Values

Weida Li, Yaoliang Yu

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Multi-Modal Inverse Constrained Reinforcement Learning from a Mixture of Demonstrations

Guanren Qiao, Guiliang Liu, Pascal Poupart, Zhiqiang Xu

Inverse Constraint Reinforcement Learning (ICRL) aims to recover the underlying constraints respected by expert agents in a data-driven manner. Existing ICRL algorithms typically assume that the demonstration data is generated by a single type of expert. However, in practice, demonstrations often comprise a mixture of trajectories collected from various expert agents respecting different constraints, making it challenging to explain expert behaviors with a unified constraint function. To tackle this issue, we propose a Multi-Modal Inverse Constrained Reinforcement Learning (MMICRL) algorithm for simultaneously estimating multiple constraints corresponding to different types of experts. MMICRL constructs a flow-based density estimator that enables unsupervised expert identification from demonstrations, so as to infer the agent-specific constraints. Following these constraints, MMICRL imitates expert policies with a novel multi-modal constrained policy optimization objective that minimizes the agent-conditioned policy entropy and maximizes the unconditioned one. To enhance robustness, we incorporate this objective into the contrastive learning framework. This approach enables imitation policies to capture the diversity of behaviors among expert agents. Extensive experiments in both discrete and continuous environments show that MMICRL outperforms other baselines in terms of constraint recovery and control performance.
****

FedFed: Feature Distillation against Data Heterogeneity in Federated Learning

Zhiqin Yang, Yonggang Zhang, Yu Zheng, Xinmei Tian, Hao Peng, Tongliang Liu, Bo Han

Federated learning (FL) typically faces data heterogeneity, i.e., distribution shifting among clients. Sharing clients' information has shown great potentiality in mitigating data heterogeneity, yet incurs a dilemma in preserving privacy and promoting model performance. To alleviate the dilemma, we raise a fundamental question: Is it possible to share partial features in the data to tackle data heterogeneity?In this work, we give an affirmative answer to this question by prop

osing a novel approach called Federated Feature distillation (FedFed).Specifical ly, FedFed partitions data into performance-sensitive features (i.e., greatly co ntributing to model performance) and performance-robust features (i.e., limitedl y contributing to model performance).The performance-sensitive features are glob ally shared to mitigate data heterogeneity, while the performance-robust feature s are kept locally.FedFed enables clients to train models over local and shared data. Comprehensive experiments demonstrate the efficacy of FedFed in promoting model performance.
****

A Privacy-Friendly Approach to Data Valuation
Jiachen (Tianhao) Wang, Yuqing Zhu, Yu-Xiang Wang, Ruoxi Jia, Prateek Mittal
Data valuation, a growing field that aims at quantifying the usefulness of indiv idual data sources for training machine learning (ML) models, faces notable yet often overlooked privacy challenges. This paper studies these challenges with a focus on KNN-Shapley, one of the most practical data valuation methods nowadays. We first emphasize the inherent privacy risks of KNN-Shapley, and demonstrate t he significant technical challenges in adapting KNN-Shapley to accommodate diffe rential privacy (DP). To overcome these challenges, we introduce TKNN-Shapley, a refined variant of KNN-Shapley that is privacy-friendly, allowing for straightf orward modifications to incorporate DP guarantee (DP-TKNN-Shapley). We show that DP-TKNN-Shapley has several advantages and offers a superior privacy-utility tr adeoff compared to naively privatized KNN-Shapley. Moreover, even non-private TK NN-Shapley matches KNN-Shapley's performance in discerning data quality. Overall , our findings suggest that TKNN-Shapley is a promising alternative to KNN-Shapl ey, particularly for real-world applications involving sensitive data.
****

Learning Nonparametric Latent Causal Graphs with Unknown Interventions
Yibo Jiang, Bryon Aragam
We establish conditions under which latent causal graphs are nonparametrically i dentifiable and can be reconstructed from unknown interventions in the latent sp ace. Our primary focus is the identification of the latent structure in measurem ent models without parametric assumptions such as linearity or Gaussianity. More over, we do not assume the number of hidden variables is known, and we show that at most one unknown intervention per hidden variable is needed. This extends a recent line of work on learning causal representations from observations and int erventions. The proofs are constructive and introduce two new graphical concepts ---imaginary subsets and isolated edges---that may be useful in their own right. As a matter of independent interest, the proofs also involve a novel characteri zation of the limits of edge orientations within the equivalence class of DAGs i nduced by unknown interventions. These are the first results to characterize the conditions under which causal representations are identifiable without making a ny parametric assumptions in a general setting with unknown interventions and wi thout faithfulness.
****

Kullback-Leibler Maillard Sampling for Multi-armed Bandits with Bounded Rewards
Hao Qin, Kwang-Sung Jun, Chicheng Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

The Pursuit of Human Labeling: A New Perspective on Unsupervised Learning
Artyom Gadetsky, Maria Brbic
We present HUME, a simple model-agnostic framework for inferring human labeling of a given dataset without any external supervision. The key insight behind our approach is that classes defined by many human labelings are linearly separable regardless of the representation space used to represent a dataset. HUME utilize s this insight to guide the search over all possible labelings of a dataset to d iscover an underlying human labeling. We show that the proposed optimization obj ective is strikingly well-correlated with the ground truth labeling of the datas

et. In effect, we only train linear classifiers on top of pretrained representat ions that remain fixed during training, making our framework compatible with any large pretrained and self-supervised model. Despite its simplicity, HUME outper forms a supervised linear classifier on top of self-supervised representations o n the STL-10 dataset by a large margin and achieves comparable performance on th e CIFAR-10 dataset. Compared to the existing unsupervised baselines, HUME achiev es state-of-the-art performance on four benchmark image classification datasets including the large-scale ImageNet-1000 dataset. Altogether, our work provides a fundamentally new view to tackle unsupervised learning by searching for consist ent labelings between different representation spaces.
****
Improving Compositional Generalization using Iterated Learning and Simplicial Em beddings
Yi Ren, Samuel Lavoie, Michael Galkin, Danica J. Sutherland, Aaron C. Courville
Compositional generalization, the ability of an agent to generalize to unseen co mbinations of latent factors, is easy for humans but hard for deep neural networ ks. A line of research in cognitive science has hypothesized a process, "iterate d learning," to help explain how human language developed this ability; the theo ry rests on simultaneous pressures towards compressibility (when an ignorant age nt learns from an informed one) and expressivity (when it uses the representatio n for downstream tasks). Inspired by this process, we propose to improve the com positional generalization of deep networks by using iterated learning on models with simplicial embeddings, which can approximately discretize representations. This approach is further motivated by an analysis of compositionality based on K olmogorov complexity. We show that this combination of changes improves composit ional generalization over other approaches, demonstrating these improvements bot h on vision tasks with well-understood latent factors and on real molecular grap h prediction tasks where the latent structure is unknown.
****
Geometric Analysis of Matrix Sensing over Graphs
Haixiang Zhang, Ying Chen, Javad Lavaei
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Towards Combinatorial Generalization for Catalysts: A Kohn-Sham Charge-Density A pproach
Phillip Pope, David Jacobs
The Kohn-Sham equations underlie many important applications such as the discove ry of new catalysts. Recent machine learning work on catalyst modeling has focus ed on prediction of the energy, but has so far not yet demonstrated significant out-of-distribution generalization. Here we investigate another approach based o n the pointwise learning of the Kohn-Sham charge-density. On a new dataset of bu lk catalysts with charge densities, we show density models can generalize to new structures with combinations of elements not seen at train time, a form of comb inatorial generalization. We show that over 80% of binary and ternary test cases achieve faster convergence than standard baselines in Density Functional Theory , amounting to an average reduction of 13% in the number of iterations required to reach convergence, which may be of independent interest. Our results suggest that density learning is a viable alternative, trading greater inference costs f or a step towards combinatorial generalization, a key property for applications.
****
Reward-Directed Conditional Diffusion: Provable Distribution Estimation and Rewa rd Improvement
Hui Yuan, Kaixuan Huang, Chengzhuo Ni, Minshuo Chen, Mengdi Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.

****

Unifying Predictions of Deterministic and Stochastic Physics in Mesh-reduced Space with Sequential Flow Generative Model

Luning Sun, Xu Han, Han Gao, Jian-Xun Wang, Liping Liu

Accurate prediction of dynamical systems in unstructured meshes has recently shown successes in scientific simulations. Many dynamical systems have a nonnegligible level of stochasticity introduced by various factors (e.g. chaoticity), so there is a need for a unified framework that captures both deterministic and stochastic components in the rollouts of these systems. Inspired by regeneration learning, we propose a new model that combines generative and sequential networks to model dynamical systems. Specifically, we use an autoencoder to learn compact representations of full-space physical variables in a low-dimensional space. We then integrate a transformer with a conditional normalizing flow model to model the temporal sequence of latent representations. We evaluate the new model in both deterministic and stochastic systems. The model outperforms several competitive baseline models and makes more accurate predictions of deterministic systems. Its own prediction error is also reflected in its uncertainty estimations. When predicting stochastic systems, the proposed model generates high-quality rollout samples. The mean and variance of these samples well match the statistics of samples computed from expensive numerical simulations.

****

Transitivity Recovering Decompositions: Interpretable and Robust Fine-Grained Relationships

ABHRA CHAUDHURI, Massimiliano Mancini, Zeynep Akata, Anjan Dutta

Recent advances in fine-grained representation learning leverage local-to-global (emergent) relationships for achieving state-of-the-art results. The relational representations relied upon by such methods, however, are abstract. We aim to deconstruct this abstraction by expressing them as interpretable graphs over image views. We begin by theoretically showing that abstract relational representations are nothing but a way of recovering transitive relationships among local views. Based on this, we design Transitivity Recovering Decompositions (TRD), a graph-space search algorithm that identifies interpretable equivalents of abstract emergent relationships at both instance and class levels, and with no post-hoc computations. We additionally show that TRD is provably robust to noisy views, with empirical evidence also supporting this finding. The latter allows TRD to perform at par or even better than the state-of-the-art, while being fully interpretable. Implementation is available at https://github.com/abhrac/trd.

****

Dynamic Regret of Adversarial Linear Mixture MDPs

Long-Fei Li, Peng Zhao, Zhi-Hua Zhou

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

Neural Harmonics: Bridging Spectral Embedding and Matrix Completion in Self-Supervised Learning

Marina Munkhoeva, Ivan Oseledets

Self-supervised methods received tremendous attention thanks to their seemingly heuristic approach to learning representations that respect the semantics of the data without any apparent supervision in the form of labels. A growing body of literature is already being published in an attempt to build a coherent and theoretically grounded understanding of the workings of a zoo of losses used in modern self-supervised representation learning methods. In this paper, we attempt to provide an understanding from the perspective of a Laplace operator and connect the inductive bias stemming from the augmentation process to a low-rank matrix completion problem.To this end, we leverage the results from low-rank matrix completion to provide theoretical analysis on the convergence of modern SSL methods and a key property that affects their downstream performance.

****

Keypoint-Augmented Self-Supervised Learning for Medical Image Segmentation with Limited Annotation
Zhangsihao Yang, Mengwei Ren, Kaize Ding, Guido Gerig, Yalin Wang
Pretraining CNN models (i.e., UNet) through self-supervision has become a powerful approach to facilitate medical image segmentation under low annotation regimes. Recent contrastive learning methods encourage similar global representations when the same image undergoes different transformations, or enforce invariance across different image/patch features that are intrinsically correlated. However, CNN-extracted global and local features are limited in capturing long-range spatial dependencies that are essential in biological anatomy. To this end, we present a keypoint-augmented fusion layer that extracts representations preserving both short- and long-range self-attention. In particular, we augment the CNN feature map at multiple scales by incorporating an additional input that learns long-range spatial self-attention among localized keypoint features. Further, we introduce both global and local self-supervised pretraining for the framework. At the global scale, we obtain global representations from both the bottleneck of the UNet, and by aggregating multiscale keypoint features. These global features are subsequently regularized through image-level contrastive objectives. At the local scale, we define a distance-based criterion to first establish correspondences among keypoints and encourage similarity between their features. Through extensive experiments on both MRI and CT segmentation tasks, we demonstrate the architectural advantages of our proposed method in comparison to both CNN and Transformer-based UNets, when all architectures are trained with randomly initialized weights. With our proposed pretraining strategy, our method further outperforms existing SSL methods by producing more robust self-attention and achieving state-of-the-art segmentation results. The code is available at https://github.com/zshyang/kaf.git.
****
Aiming towards the minimizers: fast convergence of SGD for overparametrized problems
Chaoyue Liu, Dmitriy Drusvyatskiy, Misha Belkin, Damek Davis, Yian Ma
Modern machine learning paradigms, such as deep learning, occur in or close to the interpolation regime, wherein the number of model parameters is much larger than the number of data samples. In this work, we propose a regularity condition within the interpolation regime which endows the stochastic gradient method with the same worst-case iteration complexity as the deterministic gradient method, while using only a single sampled gradient (or a minibatch) in each iteration. In contrast, all existing guarantees require the stochastic gradient method to take small steps, thereby resulting in a much slower linear rate of convergence. Finally, we demonstrate that our condition holds when training sufficiently wide feedforward neural networks with a linear output layer.
****
ResMem: Learn what you can and memorize the rest
Zitong Yang, MICHAL LUKASIK, Vaishnavh Nagarajan, Zonglin Li, Ankit Rawat, Manzil Zaheer, Aditya K. Menon, Sanjiv Kumar
The impressive generalization performance of modern neural networks is attributed in part to their ability to implicitly memorize complex training patterns.Inspired by this, we explore a novel mechanism to improve model generalization via explicit memorization.Specifically, we propose the residual-memorization (ResMem) algorithm, a new method that augments an existing prediction model (e.g., a neural network) by fitting the model's residuals with a nearest-neighbor based regressor.The final prediction is then the sum of the original model and the fitted residual regressor.By construction, ResMem can explicitly memorize the training labels.We start by formulating a stylized linear regression problem and rigorously show that ResMem results in a more favorable test risk over a base linear neural network.Then, we empirically show that ResMem consistently improves the test set generalization of the original prediction model across standard vision and natural language processing benchmarks.
****
Generalized Semi-Supervised Learning via Self-Supervised Feature Adaptation

Jiachen Liang, RuiBing Hou, Hong Chang, Bingpeng MA, Shiguang Shan, Xilin Chen
Traditional semi-supervised learning (SSL) assumes that the feature distributions of labeled and unlabeled data are consistent which rarely holds in realistic scenarios. In this paper, we propose a novel SSL setting, where unlabeled samples are drawn from a mixed distribution that deviates from the feature distribution of labeled samples.Under this setting, previous SSL methods tend to predict wrong pseudo-labels with the model fitted on labeled data, resulting in noise accumulation. To tackle this issue, we propose \emph{Self-Supervised Feature Adaptation} (SSFA), a generic framework for improving SSL performance when labeled and unlabeled data come from different distributions. SSFA decouples the prediction of pseudo-labels from the current model to improve the quality of pseudo-labels. Particularly, SSFA incorporates a self-supervised task into the SSL framework and uses it to adapt the feature extractor of the model to the unlabeled data. In this way, the extracted features better fit the distribution of unlabeled data, thereby generating high-quality pseudo-labels. Extensive experiments show that our proposed SSFA is applicable to various pseudo-label-based SSL learners and significantly improves performance in labeled, unlabeled, and even unseen distributions.
****
Soft-Unification in Deep Probabilistic Logic
Jaron Maene, Luc De Raedt
A fundamental challenge in neuro-symbolic AI is to devise primitives that fuse the logical and neural concepts. The Neural Theorem Prover has proposed the notion of soft-unification to turn the symbolic comparison between terms (i.e. unification) into a comparison in embedding space. It has been shown that soft-unification is a powerful mechanism that can be used to learn logic rules in an end-to-end differentiable manner. We study soft-unification from a conceptual point and outline several desirable properties of this operation. These include non-redundancy in the proof, well-defined proof scores, and non-sparse gradients. Unfortunately, these properties are not satisfied by previous systems such as the Neural Theorem Prover. Therefore, we introduce a more principled framework called DeepSoftLog based on probabilistic rather than fuzzy semantics. Our experiments demonstrate that DeepSoftLog can outperform the state-of-the-art on neuro-symbolic benchmarks, highlighting the benefits of these properties.
****
Scaling MLPs: A Tale of Inductive Bias
Gregor Bachmann, Sotiris Anagnostidis, Thomas Hofmann
In this work we revisit the most fundamental building block in deep learning, the multi-layer perceptron (MLP), and study the limits of its performance on vision tasks. Empirical insights into MLPs are important for multiple reasons. (1) Given the recent narrative "less inductive bias is better", popularized due to transformers eclipsing convolutional models, it is natural to explore the limits of this hypothesis. To that end, MLPs offer an ideal test bed, as they lack any vision-specific inductive bias. (2) MLPs have almost exclusively been the main protagonist in the deep learning theory literature due to their mathematical simplicity, serving as a proxy to explain empirical phenomena observed for more complex architectures. Surprisingly, experimental datapoints for MLPs are very difficult to find in the literature, especially when coupled with large pre-training protocols. This discrepancy between practice and theory is worrying: \textit{Do MLPs reflect the empirical advances exhibited by practical models?} Or do theorists need to rethink the role of MLPs as a proxy? We provide insights into both these aspects.We show that the performance of MLPs drastically improves with scale (95% on CIFAR10, 82% on CIFAR100, 58% on ImageNet ReaL), highlighting that lack of inductive bias can indeed be compensated. We observe that MLPs mimic the behaviour of their modern counterparts faithfully, with some components in the learning setting however exhibiting stronger or unexpected behaviours. Due to their inherent computational efficiency, large pre-training experiments become more accessible for academic researchers. All of our experiments were run on a single GPU.
****

Local Convergence of Gradient Methods for Min-Max Games: Partial Curvature Generically Suffices
Guillaume Wang, Lénaïc Chizat
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Going Beyond Linear Mode Connectivity: The Layerwise Linear Feature Connectivity
Zhanpeng Zhou, Yongyi Yang, Xiaojiang Yang, Junchi Yan, Wei Hu
Recent work has revealed many intriguing empirical phenomena in neural network training, despite the poorly understood and highly complex loss landscapes and training dynamics. One of these phenomena, Linear Mode Connectivity (LMC), has gained considerable attention due to the intriguing observation that different solutions can be connected by a linear path in the parameter space while maintaining near-constant training and test losses. In this work, we introduce a stronger notion of linear connectivity, Layerwise Linear Feature Connectivity (LLFC), which says that the feature maps of every layer in different trained networks are also linearly connected. We provide comprehensive empirical evidence for LLFC across a wide range of settings, demonstrating that whenever two trained networks satisfy LMC (via either spawning or permutation methods), they also satisfy LLFC in nearly all the layers. Furthermore, we delve deeper into the underlying factors contributing to LLFC, which reveal new insights into the permutation approaches. The study of LLFC transcends and advances our understanding of LMC by adopting a feature-learning perspective.
****
Dream the Impossible: Outlier Imagination with Diffusion Models
Xuefeng Du, Yiyou Sun, Jerry Zhu, Yixuan Li
Utilizing auxiliary outlier datasets to regularize the machine learning model has demonstrated promise for out-of-distribution (OOD) detection and safe prediction. Due to the labor intensity in data collection and cleaning, automating outlier data generation has been a long-desired alternative. Despite the appeal, generating photo-realistic outliers in the high dimensional pixel space has been an open challenge for the field. To tackle the problem, this paper proposes a new framework Dream-OOD, which enables imagining photo-realistic outliers by way of diffusion models, provided with only the in-distribution (ID) data and classes. Specifically, Dream-OOD learns a text-conditioned latent space based on ID data, and then samples outliers in the low-likelihood region via the latent, which can be decoded into images by the diffusion model. Different from prior works [16, 95], Dream-OOD enables visualizing and understanding the imagined outliers, directly in the pixel space. We conduct comprehensive quantitative and qualitative studies to understand the efficacy of Dream-OOD, and show that training with the samples generated by Dream-OOD can significantly benefit OOD detection performance.
****
RETVec: Resilient and Efficient Text Vectorizer
Elie Bursztein, Marina Zhang, Owen Vallis, XINYU JIA, Alexey Kurakin
This paper describes RETVec, an efficient, resilient, and multilingual text vectorizer designed for neural-based text processing. RETVec combines a novel character encoding with an optional small embedding model to embed words into a 256-dimensional vector space. The RETVec embedding model is pre-trained using pair-wise metric learning to be robust against typos and character-level adversarial attacks. In this paper, we evaluate and compare RETVec to state-of-the-art vectorizers and word embeddings on popular model architectures and datasets. These comparisons demonstrate that RETVec leads to competitive, multilingual models that are significantly more resilient to typos and adversarial text attacks. RETVec is available under the Apache 2 license at https://github.com/google-research/retvec.
****
An Alternative to Variance: Gini Deviation for Risk-averse Policy Gradient

Yudong Luo, Guiliang Liu, Pascal Poupart, Yangchen Pan
Restricting the variance of a policy’s return is a popular choice in risk-averse
 Reinforcement Learning (RL) due to its clear mathematical definition and easy i
nterpretability. Traditional methods directly restrict the total return variance
. Recent methods restrict the per-step reward variance as a proxy. We thoroughly
 examine the limitations of these variance-based methods, such as sensitivity to
 numerical scale and hindering of policy learning, and propose to use an alterna
tive risk measure, Gini deviation, as a substitute. We study various properties
of this new risk measure and derive a policy gradient algorithm to minimize it.
Empirical evaluation in domains where risk-aversion can be clearly defined, show
s that our algorithm can mitigate the limitations of variance-based risk measure
s and achieves high return with low risk in terms of variance and Gini deviation
 when others fail to learn a reasonable policy.
****
Do Not Marginalize Mechanisms, Rather Consolidate!
Moritz Willig, Matej Ze■evi■, Devendra Dhami, Kristian Kersting
Structural causal models (SCMs) are a powerful tool for understanding the comple
x causal relationships that underlie many real-world systems. As these systems g
row in size, the number of variables and complexity of interactions between them
 does, too. Thus, becoming convoluted and difficult to analyze. This is particul
arly true in the context of machine learning and artificial intelligence, where
an ever increasing amount of data demands for new methods to simplify and compre
ss large scale SCM. While methods for marginalizing and abstracting SCM already
exist today, they may destroy the causality of the marginalized model. To allevi
ate this, we introduce the concept of consolidating causal mechanisms to transfo
rm large-scale SCM while preserving consistent interventional behaviour. We show
 consolidation is a powerful method for simplifying SCM, discuss reduction of co
mputational complexity and give a perspective on generalizing abilities of conso
lidated SCM.
****
Learning Environment-Aware Affordance for 3D Articulated Object Manipulation und
er Occlusions
Ruihai Wu, Kai Cheng, Yan Zhao, Chuanruo Ning, Guanqi Zhan, Hao Dong
Perceiving and manipulating 3D articulated objects in diverse environments is es
sential for home-assistant robots. Recent studies have shown that point-level af
fordance provides actionable priors for downstream manipulation tasks. However,
existing works primarily focus on single-object scenarios with homogeneous agent
s, overlooking the realistic constraints imposed by the environment and the agen
t's morphology, e.g., occlusions and physical limitations. In this paper, we pro
pose an environment-aware affordance framework that incorporates both object-lev
el actionable priors and environment constraints. Unlike object-centric affordan
ce approaches, learning environment-aware affordance faces the challenge of comb
inatorial explosion due to the complexity of various occlusions, characterized b
y their quantities, geometries, positions and poses. To address this and enhance
 data efficiency, we introduce a novel contrastive affordance learning framework
 capable of training on scenes containing a single occluder and generalizing to
scenes with complex occluder combinations. Experiments demonstrate the effective
ness of our proposed approach in learning affordance considering environment con
straints.
****
Enhancing CLIP with CLIP: Exploring Pseudolabeling for Limited-Label Prompt Tuni
ng
Cristina Menghini, Andrew Delworth, Stephen Bach
Fine-tuning vision-language models (VLMs) like CLIP to downstream tasks is often
 necessary to optimize their performance. However, a major obstacle is the limit
ed availability of labeled data. We study the use of pseudolabels, i.e., heurist
ic labels for unlabeled data, to enhance CLIP via prompt tuning. Conventional ps
eudolabeling trains a model on labeled data and then generates labels for unlabe
led data. VLMs' zero-shot capabilities enable a ``second generation'' of pseudol
abeling approaches that do not require task-specific training on labeled data. B

y using zero-shot pseudolabels as a source of supervision, we observe that learning paradigms such as semi-supervised, transductive zero-shot, and unsupervised learning can all be seen as optimizing the same loss function. This unified view enables the development of versatile training strategies that are applicable across learning paradigms. We investigate them on image classification tasks where CLIP exhibits limitations, by varying prompt modalities, e.g., textual or visual prompts, and learning paradigms. We find that(1) unexplored prompt tuning strategies that iteratively refine pseudolabels consistently improve CLIP accuracy, by 19.5 points in semi-supervised learning, by 28.4 points in transductive zero-shot learning, and by 15.2 points in unsupervised learning, and (2) unlike conventional semi-supervised pseudolabeling, which exacerbates model biases toward classes with higher-quality pseudolabels, prompt tuning leads to a more equitable distribution of per-class accuracy. The code to reproduce the experiments is at https://github.com/BatsResearch/menghini-neurips23-code.
****

Accelerating Monte Carlo Tree Search with Probability Tree State Abstraction
Yangqing Fu, Ming Sun, Buqing Nie, Yue Gao
Monte Carlo Tree Search (MCTS) algorithms such as AlphaGo and MuZero have achieved superhuman performance in many challenging tasks. However, the computational complexity of MCTS-based algorithms is influenced by the size of the search space. To address this issue, we propose a novel probability tree state abstraction (PTSA) algorithm to improve the search efficiency of MCTS. A general tree state abstraction with path transitivity is defined. In addition, the probability tree state abstraction is proposed for fewer mistakes during the aggregation step. Furthermore, the theoretical guarantees of the transitivity and aggregation error bound are justified. To evaluate the effectiveness of the PTSA algorithm, we integrate it with state-of-the-art MCTS-based algorithms, such as Sampled MuZero and Gumbel MuZero. Experimental results on different tasks demonstrate that our method can accelerate the training process of state-of-the-art algorithms with 10%-45% search space reduction.
****

MeCo: Zero-Shot NAS with One Data and Single Forward Pass via Minimum Eigenvalue of Correlation
Tangyu Jiang, Haodi Wang, Rongfang Bie
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

On the Constrained Time-Series Generation Problem
Andrea Coletta, Sriram Gopalakrishnan, Daniel Borrajo, Svitlana Vyetrenko
Synthetic time series are often used in practical applications to augment the historical time series dataset, amplify the occurrence of rare events and also create counterfactual scenarios.Distributional-similarity (which we refer to as realism) as well as the satisfaction of certain numerical constraints are common requirements for counterfactual time series generation. For instance, the US Federal Reserve publishes synthetic market stress scenarios given by the constrained time series for financial institutions to assess their performance in hypothetical recessions.Existing approaches for generating constrained time series usually penalize training loss to enforce constraints, and reject non-conforming samples. However, these approaches would require re-training if we change constraints, and rejection sampling can be computationally expensive, or impractical for complex constraints.In this paper, we propose a novel set of methods to tackle the constrained time series generation problem and provide efficient sampling while ensuring the realism of generated time series. In particular, we frame the problem using a constrained optimization framework and then we propose a set of generative methods including 'GuidedDiffTime', a guided diffusion model. We empirically evaluate our work on several datasets for financial and energy data, where incorporating constraints is critical. We show that our approaches outperform existing work both qualitatively and quantitatively, and that 'GuidedDiffTime' does

not require re-training for new constraints, resulting in a significant carbon footprint reduction, up to 92% w.r.t. existing deep learning methods.
****
InfoPrompt: Information-Theoretic Soft Prompt Tuning for Natural Language Understanding
Junda Wu, Tong Yu, Rui Wang, Zhao Song, Ruiyi Zhang, Handong Zhao, Chaochao Lu, Shuai Li, Ricardo Henao
Soft prompt tuning achieves superior performances across a wide range of few-shot tasks. However, the performances of prompt tuning can be highly sensitive to the initialization of the prompts. We have also empirically observed that conventional prompt tuning methods cannot encode and learn sufficient task-relevant information from prompt tokens. In this work, we develop an information-theoretic framework that formulates soft prompt tuning as maximizing the mutual information between prompts and other model parameters (or encoded representations). This novel view helps us to develop a more efficient, accurate and robust soft prompt tuning method, InfoPrompt. With this framework, we develop two novel mutual information based loss functions, to (i) explore proper prompt initialization for the downstream tasks and learn sufficient task-relevant information from prompt tokens and (ii) encourage the output representation from the pretrained language model to be more aware of the task-relevant information captured in the learnt prompts. Extensive experiments validate that InfoPrompt can significantly accelerate the convergence of the prompt tuning and outperform traditional prompt tuning methods. Finally, we provide a formal theoretical result to show that a gradient descent type algorithm can be used to train our mutual information loss.
****
On the Size and Approximation Error of Distilled Datasets
Alaa Maalouf, Murad Tukan, Noel Loo, Ramin Hasani, Mathias Lechner, Daniela Rus
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Unified Approach for Maximizing Continuous DR-submodular Functions
Mohammad Pedramfar, Christopher Quinn, Vaneet Aggarwal
This paper presents a unified approach for maximizing continuous DR-submodular functions that encompasses a range of settings and oracle access types. Our approach includes a Frank-Wolfe type offline algorithm for both monotone and non-monotone functions, with different restrictions on the general convex set. We consider settings where the oracle provides access to either the gradient of the function or only the function value, and where the oracle access is either deterministic or stochastic. We determine the number of required oracle accesses in all cases. Our approach gives new/improved results for nine out of the sixteen considered cases, avoids computationally expensive projections in three cases, with the proposed framework matching performance of state-of-the-art approaches in the remaining four cases. Notably, our approach for the stochastic function value-based oracle enables the first regret bounds with bandit feedback for stochastic DR-submodular functions.
****
On Sample-Efficient Offline Reinforcement Learning: Data Diversity, Posterior Sampling and Beyond
Thanh Nguyen-Tang, Raman Arora
We seek to understand what facilitates sample-efficient learning from historical datasets for sequential decision-making, a problem that is popularly known as offline reinforcement learning (RL). Further, we are interested in algorithms that enjoy sample efficiency while leveraging (value) function approximation. In this paper, we address these fundamental questions by (i) proposing a notion of data diversity that subsumes the previous notions of coverage measures in offline RL and (ii) using this notion to \emph{unify} three distinct classes of offline RL algorithms based on version spaces (VS), regularized optimization (RO), and posterior sampling (PS). We establish that VS-based, RO-based, and PS-based algor

ithms, under standard assumptions, achieve \emph{comparable} sample efficiency, which recovers the state-of-the-art sub-optimality bounds for finite and linear model classes with the standard assumptions. This result is surprising, given that the prior work suggested an unfavorable sample complexity of the RO-based algorithm compared to the VS-based algorithm, whereas posterior sampling is rarely considered in offline RL due to its explorative nature. Notably, our proposed model-free PS-based algorithm for offline RL is \emph{novel}, with sub-optimality bounds that are \emph{frequentist} (i.e., worst-case) in nature.
****
GRAND-SLAMIN' Interpretable Additive Modeling with Structural Constraints
Shibal Ibrahim, Gabriel Afriat, Kayhan Behdin, Rahul Mazumder
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
S-CLIP: Semi-supervised Vision-Language Learning using Few Specialist Captions
Sangwoo Mo, Minkyu Kim, Kyungmin Lee, Jinwoo Shin
Vision-language models, such as contrastive language-image pre-training (CLIP), have demonstrated impressive results in natural image domains. However, these models often struggle when applied to specialized domains like remote sensing, and adapting to such domains is challenging due to the limited number of image-text pairs available for training. To address this, we propose S-CLIP, a semi-supervised learning method for training CLIP that utilizes additional unpaired images. S-CLIP employs two pseudo-labeling strategies specifically designed for contrastive learning and the language modality. The caption-level pseudo-label is given by a combination of captions of paired images, obtained by solving an optimal transport problem between unpaired and paired images. The keyword-level pseudo-label is given by a keyword in the caption of the nearest paired image, trained through partial label learning that assumes a candidate set of labels for supervision instead of the exact one. By combining these objectives, S-CLIP significantly enhances the training of CLIP using only a few image-text pairs, as demonstrated in various specialist domains, including remote sensing, fashion, scientific figures, and comics. For instance, S-CLIP improves CLIP by 10% for zero-shot classification and 4% for image-text retrieval on the remote sensing benchmark, matching the performance of supervised CLIP while using three times fewer image-text pairs.
****
A3FL: Adversarially Adaptive Backdoor Attacks to Federated Learning
Hangfan Zhang, Jinyuan Jia, Jinghui Chen, Lu Lin, Dinghao Wu
Federated Learning (FL) is a distributed machine learning paradigm that allows multiple clients to train a global model collaboratively without sharing their local training data. Due to its distributed nature, many studies have shown that it is vulnerable to backdoor attacks. However, existing studies usually used a predetermined, fixed backdoor trigger or optimized it based solely on the local data and model without considering the global training dynamics. This leads to sub-optimal and less durable attack effectiveness, i.e., their attack success rate is low when the attack budget is limited and decreases quickly if the attacker can no longer perform attacks anymore. To address these limitations, we propose A3FL, a new backdoor attack which adversarially adapts the backdoor trigger to make it less likely to be removed by the global training dynamics. Our key intuition is that the difference between the global model and the local model in FL makes the local-optimized trigger much less effective when transferred to the global model. We solve this by optimizing the trigger to even survive the worst-case scenario where the global model was trained to directly unlearn the trigger. Extensive experiments on benchmark datasets are conducted for twelve existing defenses to comprehensively evaluate the effectiveness of our A3FL. Our code is available at https://github.com/hfzhang31/A3FL.
****
Towards Understanding the Dynamics of Gaussian-Stein Variational Gradient Descen

t
Tianle Liu, Promit Ghosal, Krishnakumar Balasubramanian, Natesh Pillai
Stein Variational Gradient Descent (SVGD) is a nonparametric particle-based deterministic sampling algorithm. Despite its wide usage, understanding the theoretical properties of SVGD has remained a challenging problem. For sampling from a Gaussian target, the SVGD dynamics with a bilinear kernel will remain Gaussian as long as the initializer is Gaussian. Inspired by this fact, we undertake a detailed theoretical study of the Gaussian-SVGD, i.e., SVGD projected to the family of Gaussian distributions via the bilinear kernel, or equivalently Gaussian variational inference (GVI) with SVGD. We present a complete picture by considering both the mean-field PDE and discrete particle systems. When the target is strongly log-concave, the mean-field Gaussian-SVGD dynamics is proven to converge linearly to the Gaussian distribution closest to the target in KL divergence. In the finite-particle setting, there is both uniform in time convergence to the mean-field limit and linear convergence in time to the equilibrium if the target is Gaussian. In the general case, we propose a density-based and a particle-based implementation of the Gaussian-SVGD, and show that several recent algorithms for GVI, proposed from different perspectives, emerge as special cases of our unified framework. Interestingly, one of the new particle-based instance from this framework empirically outperforms existing approaches. Our results make concrete contributions towards obtaining a deeper understanding of both SVGD and GVI.
****
Validated Image Caption Rating Dataset
Lothar D Narins, Andrew Scott, Aakash Gautam, Anagha Kulkarni, Mar Castanon, Benjamin Kao, Shasta Ihorn, Yue-Ting Siu, James M. Mason, Alexander Blum, Ilmi Yoon
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Provable benefits of score matching
Chirag Pabbaraju, Dhruv Rohatgi, Anish Prasad Sevekari, Holden Lee, Ankur Moitra, Andrej Risteski
Score matching is an alternative to maximum likelihood (ML) for estimating a probability distribution parametrized up to a constant of proportionality. By fitting the ''score'' of the distribution, it sidesteps the need to compute this constant of proportionality (which is often intractable).While score matching and variants thereof are popular in practice, precise theoretical understanding of the benefits and tradeoffs with maximum likelihood---both computational and statistical---are not well understood. In this work, we give the first example of a natural exponential family of distributions such that the score matching loss is computationally efficient to optimize, and has a comparable statistical efficiency to ML, while the ML loss is intractable to optimize using a gradient-based method. The family consists of exponentials of polynomials of fixed degree, and our result can be viewed as a continuous analogue of recent developments in the discrete setting. Precisely, we show: (1) Designing a zeroth-order or first-order oracle for optimizing the maximum likelihood loss is NP-hard. (2) Maximum likelihood has a statistical efficiency polynomial in the ambient dimension and the radius of the parameters of the family. (3) Minimizing the score matching loss is both computationally and statistically efficient, with complexity polynomial in the ambient dimension.
****
Oracle Complexity of Single-Loop Switching Subgradient Methods for Non-Smooth Weakly Convex Functional Constrained Optimization
Yankun Huang, Qihang Lin
We consider a non-convex constrained optimization problem, where the objective function is weakly convex and the constraint function is either convex or weakly convex. To solve this problem, we consider the classical switching subgradient method, which is an intuitive and easily implementable first-order method whose oracle complexity was only known for convex problems. This paper provides the fir

st analysis on the oracle complexity of the switching subgradient method for finding a nearly stationary point of non-convex problems. Our results are derived separately for convex and weakly convex constraints. Compared to existing approaches, especially the double-loop methods, the switching gradient method can be applied to non-smooth problems and achieves the same complexity using only a single loop, which saves the effort on tuning the number of inner iterations.
****

Performance Scaling via Optimal Transport: Enabling Data Selection from Partially Revealed Sources

Feiyang Kang, Hoang Anh Just, Anit Kumar Sahu, Ruoxi Jia

Traditionally, data selection has been studied in settings where all samples from prospective sources are fully revealed to a machine learning developer. However, in practical data exchange scenarios, data providers often reveal only a limited subset of samples before an acquisition decision is made. Recently, there have been efforts to fit scaling functions that predict model performance at any size and data source composition using the limited available samples. However, these scaling functions are usually black-box, computationally expensive to fit, highly susceptible to overfitting, or/and difficult to optimize for data selection. This paper proposes a framework called , which predicts model performance and supports data selection decisions based on partial samples of prospective data sources. Our approach distinguishes itself from existing work by introducing a novel two-stage performance inference process. In the first stage, we leverage the Optimal Transport distance to predict the model's performance for any data mixture ratio within the range of disclosed data sizes. In the second stage, we extrapolate the performance to larger undisclosed data sizes based on a novel parameter-free mapping technique inspired by neural scaling laws. We further derive an efficient gradient-based method to select data sources based on the projected model performance. Evaluation over a diverse range of applications (e.g., vision, text, fine-tuning, noisy data sources, etc.) demonstrates that significantly improves existing performance scaling approaches in terms of both the accuracy of performance inference and the computation costs associated with constructing the performance predictor. Also, outperforms by a wide margin in data selection effectiveness compared to a range of other off-the-shelf solutions. We provide an open-source toolkit.
****

The Rank-Reduced Kalman Filter: Approximate Dynamical-Low-Rank Filtering In High Dimensions

Jonathan Schmidt, Philipp Hennig, Jörg Nick, Filip Tronarp

Inference and simulation in the context of high-dimensional dynamical systems remain computationally challenging problems.Some form of dimensionality reduction is required to make the problem tractable in general.In this paper, we propose a novel approximate Gaussian filtering and smoothing methodwhich propagates low-rank approximations of the covariance matrices.This is accomplished by projecting the Lyapunov equations associated with the prediction step to a manifold of low-rank matrices,which are then solved by a recently developed, numerically stable, dynamical low-rank integrator.Meanwhile, the update steps are made tractable by noting that the covariance update only transforms the column space of the covariance matrix, which is low-rank by construction.The algorithm differentiates itself from existing ensemble-based approaches in thatthe low-rank approximations of the covariance matrices are deterministic, rather than stochastic.Crucially, this enables the method to reproduce the exact Kalman filter as the low-rank dimension approaches the true dimensionality of the problem.Our method reduces computational complexity from cubic (for the Kalman filter) to quadratic in the state-space size in the worst-case, and can achieve linear complexity if the state-space model satisfies certain criteria.Through a set of experiments in classical data-assimilation and spatio-temporal regression, we show that the proposed method consistently outperforms the ensemble-based methods in terms of error in the mean and covariance with respect to the exact Kalman filter. This comes at no additional cost in terms of asymptotic computational complexity.
****

Cognitive Model Discovery via Disentangled RNNs
Kevin Miller, Maria Eckstein, Matt Botvinick, Zeb Kurth-Nelson
Computational cognitive models are a fundamental tool in behavioral neuroscience. They embody in software precise hypotheses about the cognitive mechanisms underlying a particular behavior. Constructing these models is typically a difficult iterative process that requires both inspiration from the literature and the creativity of an individual researcher. Here, we adopt an alternative approach to learn parsimonious cognitive models directly from data. We fit behavior data using a recurrent neural network that is penalized for carrying excess information between timesteps, leading to sparse, interpretable representations and dynamics. When fitting synthetic behavioral data from known cognitive models, our method recovers the underlying form of those models. When fit to choice data from rats performing a bandit task, our method recovers simple and interpretable models that make testable predictions about neural mechanisms.
****
Offline Reinforcement Learning with Differential Privacy
Dan Qiao, Yu-Xiang Wang
The offline reinforcement learning (RL) problem is often motivated by the need to learn data-driven decision policies in financial, legal and healthcare applications. However, the learned policy could retain sensitive information of individuals in the training data (e.g., treatment and outcome of patients), thus susceptible to various privacy risks. We design offline RL algorithms with differential privacy guarantees which provably prevent such risks. These algorithms also enjoy strong instance-dependent learning bounds under both tabular and linear Markov Decision Process (MDP) settings. Our theory and simulation suggest that the privacy guarantee comes at (almost) no drop in utility comparing to the non-private counterpart for a medium-size dataset.
****
Chatting Makes Perfect: Chat-based Image Retrieval
Matan Levy, Rami Ben-Ari, Nir Darshan, Dani Lischinski
Chats emerge as an effective user-friendly approach for information retrieval, and are successfully employed in many domains, such as customer service, healthcare, and finance. However, existing image retrieval approaches typically address the case of a single query-to-image round, and the use of chats for image retrieval has been mostly overlooked. In this work, we introduce ChatIR: a chat-based image retrieval system that engages in a conversation with the user to elicit information, in addition to an initial query, in order to clarify the user's search intent. Motivated by the capabilities of today's foundation models, we leverage Large Language Models to generate follow-up questions to an initial image description. These questions form a dialog with the user in order to retrieve the desired image from a large corpus. In this study, we explore the capabilities of such a system tested on a large dataset and reveal that engaging in a dialog yields significant gains in image retrieval. We start by building an evaluation pipeline from an existing manually generated dataset and explore different modules and training strategies for ChatIR. Our comparison includes strong baselines derived from related applications trained with Reinforcement Learning. Our system is capable of retrieving the target image from a pool of 50K images with over 78% success rate after 5 dialogue rounds, compared to 75% when questions are asked by humans, and 64% for a single shot text-to-image retrieval. Extensive evaluations reveal the strong capabilities and examine the limitations of CharIR under different settings. Project repository is available at https://github.com/levymsn/ChatIR.
****
SHOT: Suppressing the Hessian along the Optimization Trajectory for Gradient-Based Meta-Learning
JunHoo Lee, Jayeon Yoo, Nojun Kwak
In this paper, we hypothesize that gradient-based meta-learning (GBML) implicitly suppresses the Hessian along the optimization trajectory in the inner loop. Based on this hypothesis, we introduce an algorithm called SHOT (Suppressing the Hessian along the Optimization Trajectory) that minimizes the distance between

the parameters of the target and reference models to suppress the Hessian in the inner loop. Despite dealing with high-order terms, SHOT does not increase the computational complexity of the baseline model much. It is agnostic to both the algorithm and architecture used in GBML, making it highly versatile and applicable to any GBML baseline. To validate the effectiveness of SHOT, we conduct empirical tests on standard few-shot learning tasks and qualitatively analyze its dynamics. We confirm our hypothesis empirically and demonstrate that SHOT outperforms the corresponding baseline.
****
ViCA-NeRF: View-Consistency-Aware 3D Editing of Neural Radiance Fields
Jiahua Dong, Yu-Xiong Wang
We introduce ViCA-NeRF, the first view-consistency-aware method for 3D editing with text instructions. In addition to the implicit neural radiance field (NeRF) modeling, our key insight is to exploit two sources of regularization that explicitly propagate the editing information across different views, thus ensuring multi-view consistency. For geometric regularization, we leverage the depth information derived from NeRF to establish image correspondences between different views. For learned regularization, we align the latent codes in the 2D diffusion model between edited and unedited images, enabling us to edit key views and propagate the update throughout the entire scene. Incorporating these two strategies, our ViCA-NeRF operates in two stages. In the initial stage, we blend edits from different views to create a preliminary 3D edit. This is followed by a second stage of NeRF training, dedicated to further refining the scene's appearance. Experimental results demonstrate that ViCA-NeRF provides more flexible, efficient (3 times faster) editing with higher levels of consistency and details, compared with the state of the art. Our code is available at: https://github.com/Dongjiahua/VICA-NeRF
****
Are aligned neural networks adversarially aligned?
Nicholas Carlini, Milad Nasr, Christopher A. Choquette-Choo, Matthew Jagielski, Irena Gao, Pang Wei W. Koh, Daphne Ippolito, Florian Tramer, Ludwig Schmidt
Large language models are now tuned to align with the goals of their creators, namely to be "helpful and harmless." These models should respond helpfully to user questions, but refuse to answer requests that could cause harm. However, adversarial users can construct inputs which circumvent attempts at alignment. In this work, we study adversarial alignment, and ask to what extent these models remain aligned when interacting with an adversarial user who constructs worst-case inputs (adversarial examples). These inputs are designed to cause the model to emit harmful content that would otherwise be prohibited.We show that existing NLP-based optimization attacks are insufficiently powerful to reliably attack aligned text models: even when current NLP-based attacks fail, we can find adversarial inputs with brute force. As a result, the failure of current attacks should not be seen as proof that aligned text models remain aligned under adversarial inputs. However the recent trend in large-scale ML models is multimodal models that allow users to provide images that influence the text that is generated. We show these models can be easily attacked, i.e., induced to perform arbitrary un-aligned behavior through adversarial perturbation of the input image. We conjecture that improved NLP attacks may demonstrate this same level of adversarial control over text-only models.
****
VisionLLM: Large Language Model is also an Open-Ended Decoder for Vision-Centric Tasks
Wenhai Wang, Zhe Chen, Xiaokang Chen, Jiannan Wu, Xizhou Zhu, Gang Zeng, Ping Luo, Tong Lu, Jie Zhou, Yu Qiao, Jifeng Dai
Large language models (LLMs) have notably accelerated progress towards artificial general intelligence (AGI), with their impressive zero-shot capacity for user-tailored tasks, endowing them with immense potential across a range of applications. However, in the field of computer vision, despite the availability of numerous powerful vision foundation models (VFMs), they are still restricted to tasks in a pre-defined form, struggling to match the open-ended task capabilities of

LLMs. In this work, we present an LLM-based framework for vision-centric tasks, termed VisionLLM. This framework provides a unified perspective for vision and language tasks by treating images as a foreign language and aligning vision-centric tasks with language tasks that can be flexibly defined and managed using language instructions. An LLM-based decoder can then make appropriate predictions based on these instructions for open-ended tasks. Extensive experiments show that the proposed VisionLLM can achieve different levels of task customization through language instructions, from fine-grained object-level to coarse-grained task-level customization, all with good results. It's noteworthy that, with a generalist LLM-based framework, our model can achieve over 60% mAP on COCO, on par with detection-specific models. We hope this model can set a new baseline for generalist vision and language models. The code shall be released.
****
Object-Centric Learning for Real-World Videos by Predicting Temporal Feature Similarities
Andrii Zadaianchuk, Maximilian Seitzer, Georg Martius
Unsupervised video-based object-centric learning is a promising avenue to learn structured representations from large, unlabeled video collections, but previous approaches have only managed to scale to real-world datasets in restricted domains.Recently, it was shown that the reconstruction of pre-trained self-supervised features leads to object-centric representations on unconstrained real-world image datasets.Building on this approach, we propose a novel way to use such pre-trained features in the form of a temporal feature similarity loss.This loss encodes semantic and temporal correlations between image patches and is a natural way to introduce a motion bias for object discovery.We demonstrate that this loss leads to state-of-the-art performance on the challenging synthetic MOVi datasets.When used in combination with the feature reconstruction loss, our model is the first object-centric video model that scales to unconstrained video datasets such as YouTube-VIS.https://martius-lab.github.io/videosaur/
****
Regret Matching+: (In)Stability and Fast Convergence in Games
Gabriele Farina, Julien Grand-Clément, Christian Kroer, Chung-Wei Lee, Haipeng Luo
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
For SALE: State-Action Representation Learning for Deep Reinforcement Learning
Scott Fujimoto, Wei-Di Chang, Edward Smith, Shixiang (Shane) Gu, Doina Precup, David Meger
In reinforcement learning (RL), representation learning is a proven tool for complex image-based tasks, but is often overlooked for environments with low-level states, such as physical control problems. This paper introduces SALE, a novel approach for learning embeddings that model the nuanced interaction between state and action, enabling effective representation learning from low-level states. We extensively study the design space of these embeddings and highlight important design considerations. We integrate SALE and an adaptation of checkpoints for RL into TD3 to form the TD7 algorithm, which significantly outperforms existing continuous control algorithms. On OpenAI gym benchmark tasks, TD7 has an average performance gain of 276.7% and 50.7% over TD3 at 300k and 5M time steps, respectively, and works in both the online and offline settings.
****
First- and Second-Order Bounds for Adversarial Linear Contextual Bandits
Julia Olkhovskaya, Jack Mayo, Tim van Erven, Gergely Neu, Chen-Yu Wei
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Calibrated Stackelberg Games: Learning Optimal Commitments Against Calibrated Agents
Nika Haghtalab, Chara Podimata, Kunhe Yang
In this paper, we introduce a generalization of the standard Stackelberg Games (SGs) framework: Calibrated Stackelberg Games. In CSGs, a principal repeatedly interacts with an agent who (contrary to standard SGs) does not have direct access to the principal's action but instead best responds to calibrated forecasts about it. CSG is a powerful modeling tool that goes beyond assuming that agents use ad hoc and highly specified algorithms for interacting in strategic settings to infer the principal's actions and thus more robustly addresses real-life applications that SGs were originally intended to capture. Along with CSGs, we also introduce a stronger notion of calibration, termed adaptive calibration, that provides fine-grained any-time calibration guarantees against adversarial sequences. We give a general approach for obtaining adaptive calibration algorithms and specialize them for finite CSGs. In our main technical result, we show that in CSGs, the principal can achieve utility that converges to the optimum Stackelberg value of the game both in finite and continuous settings and that no higher utility is achievable. Two prominent and immediate applications of our results are the settings of learning in Stackelberg Security Games and strategic classification, both against calibrated agents.
****
Density of States Prediction of Crystalline Materials via Prompt-guided Multi-Modal Transformer
Namkyeong Lee, Heewoong Noh, Sungwon Kim, Dongmin Hyun, Gyoung S. Na, Chanyoung Park
The density of states (DOS) is a spectral property of crystalline materials, which provides fundamental insights into various characteristics of the materials.While previous works mainly focus on obtaining high-quality representations of crystalline materials for DOS prediction, we focus on predicting the DOS from the obtained representations by reflecting the nature of DOS: DOS determines the general distribution of states as a function of energy.That is, DOS is not solely determined by the crystalline material but also by the energy levels, which has been neglected in previous works.In this paper, we propose to integrate heterogeneous information obtained from the crystalline materials and the energies via a multi-modal transformer, thereby modeling the complex relationships between the atoms in the crystalline materials and various energy levels for DOS prediction. Moreover, we propose to utilize prompts to guide the model to learn the crystal structural system-specific interactions between crystalline materials and energies.Extensive experiments on two types of DOS, i.e., Phonon DOS and Electron DOS, with various real-world scenarios demonstrate the superiority of DOSTransformer.The source code for DOSTransformer is available at https://github.com/HeewoongNoh/DOSTransformer.
****
A Single-Loop Accelerated Extra-Gradient Difference Algorithm with Improved Complexity Bounds for Constrained Minimax Optimization
Yuanyuan Liu, Fanhua Shang, Weixin An, Junhao Liu, Hongying Liu, Zhouchen Lin
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Unleashing the Full Potential of Product Quantization for Large-Scale Image Retrieval
Yu Liang, Shiliang Zhang, Li Ken Li, Xiaoyu Wang
Due to its promising performance, deep hashing has become a prevalent method for approximate nearest neighbors search (ANNs). However, most of current deep hashing methods are validated on relatively small-scale datasets, leaving potential threats when are applied to large-scale real-world scenarios. Specifically, they can be constrained either by the computational cost due to the large number of training categories and samples, or unsatisfactory accuracy. To tackle those iss

ues, we propose a novel deep hashing framework based on product quantization (PQ). It uses a softmax-based differentiable PQ branch to learn a set of predefined PQ codes of the classes. Our method is easy to implement, does not involve large-scale matrix operations, and learns highly discriminate compact codes. We validate our method on multiple large-scaled datasets, including ImageNet100, ImageNet1K, and Glint360K, where the category size scales from 100 to 360K and sample number scales from 10K to 17 million, respectively. Extensive experiments demonstrate the superiority of our method. Code is available at https://github.com/yuleung/FPPQ.
****
The Bayesian Stability Zoo
Shay Moran, Hilla Schefler, Jonathan Shafer
We show that many definitions of stability found in the learning theory literature are equivalent to one another. We distinguish between two families of definitions of stability: distribution-dependent and distribution-independent Bayesian stability. Within each family, we establish equivalences between various definitions, encompassing approximate differential privacy, pure differential privacy, replicability, global stability, perfect generalization, TV stability, mutual information stability, KL-divergence stability, and Rényi-divergence stability. Along the way, we prove boosting results that enable the amplification of the stability of a learning rule. This work is a step towards a more systematic taxonomy of stability notions in learning theory, which can promote clarity and an improved understanding of an array of stability concepts that have emerged in recent years.
****
Improving the Knowledge Gradient Algorithm
Le Yang, Siyang Gao, Chin Pang Ho
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Hierarchical Multi-Agent Skill Discovery
Mingyu Yang, Yaodong Yang, Zhenbo Lu, Wengang Zhou, Houqiang Li
Skill discovery has shown significant progress in unsupervised reinforcement learning. This approach enables the discovery of a wide range of skills without any extrinsic reward, which can be effectively combined to tackle complex tasks. However, such unsupervised skill learning has not been well applied to multi-agent reinforcement learning (MARL) due to two primary challenges. One is how to learn skills not only for the individual agents but also for the entire team, and the other is how to coordinate the skills of different agents to accomplish multi-agent tasks. To address these challenges, we present Hierarchical Multi-Agent Skill Discovery (HMASD), a two-level hierarchical algorithm for discovering both team and individual skills in MARL. The high-level policy employs a transformer structure to realize sequential skill assignment, while the low-level policy learns to discover valuable team and individual skills. We evaluate HMASD on sparse reward multi-agent benchmarks, and the results show that HMASD achieves significant performance improvements compared to strong MARL baselines.
****
Deep Optimal Transport: A Practical Algorithm for Photo-realistic Image Restoration
Theo Adrai, Guy Ohayon, Michael Elad, Tomer Michaeli
We propose an image restoration algorithm that can control the perceptual quality and/or the mean square error (MSE) of any pre-trained model, trading one over the other at test time. Our algorithm is few-shot: Given about a dozen images restored by the model, it can significantly improve the perceptual quality and/or the MSE of the model for newly restored images without further training. Our approach is motivated by a recent theoretical result that links between the minimum MSE (MMSE) predictor and the predictor that minimizes the MSE under a perfect perceptual quality constraint. Specifically, it has been shown that the latter ca

n be obtained by optimally transporting the output of the former, such that its distribution matches that of the source data. Thus, to improve the perceptual qu ality of a predictor that was originally trained to minimize MSE, we approximate the optimal transport by a linear transformation in the latent space of a varia tional auto-encoder, which we compute in closed-form using empirical means and c ovariances. Going beyond the theory, we find that applying the same procedure on models that were initially trained to achieve high perceptual quality, typicall y improves their perceptual quality even further. And by interpolating the resul ts with the original output of the model, we can improve their MSE on the expens e of perceptual quality. We illustrate our method on a variety of degradations a pplied to general content images with arbitrary dimensions.
****
DAW: Exploring the Better Weighting Function for Semi-supervised Semantic Segmen tation
Rui Sun, Huayu Mai, Tianzhu Zhang, Feng Wu
The critical challenge of semi-supervised semantic segmentation lies in how to f ully exploit a large volume of unlabeled data to improve the model’s generalizat ion performance for robust segmentation. Existing methods tend to employ certai n criteria (weighting function) to select pixel-level pseudo labels. However, th e trade-off exists between inaccurate yet utilized pseudo-labels, and correct ye t discarded pseudo-labels in these methods when handling pseudo-labels without t houghtful consideration of the weighting function, hindering the generalization ability of the model. In this paper, we systematically analyze the trade-off in previous methods when dealing with pseudo-labels. We formally define the trade-o ff between inaccurate yet utilized pseudo-labels, and correct yet discarded pseu do-labels by explicitly modeling the confidence distribution of correct and inac curate pseudo-labels, equipped with a unified weighting function. To this end, w e propose Distribution-Aware Weighting (DAW) to strive to minimize the negative equivalence impact raised by the trade-off. We find an interesting fact that the optimal solution for the weighting function is a hard step function, with the j ump point located at the intersection of the two confidence distributions. Besid es, we devise distribution alignment to mitigate the issue of the discrepancy be tween the prediction distributions of labeled and unlabeled data. Extensive expe rimental results on multiple benchmarks including mitochondria segmentation demo nstrate that DAW performs favorably against state-of-the-art methods.
****
Feature Dropout: Revisiting the Role of Augmentations in Contrastive Learning
Alex Tamkin, Margalit Glasgow, Xiluo He, Noah Goodman
What role do augmentations play in contrastive learning? Recent work suggests th at good augmentations are label-preserving with respect to a specific downstream task. We complicate this picture by showing that label-destroying augmentations can be useful in the foundation model setting, where the goal is to learn diver se, general-purpose representations for multiple downstream tasks. We perform co ntrastive learning experiments on a range of image and audio datasets with multi ple downstream tasks (e.g. for digits superimposed on photographs, predicting th e class of one vs. the other). We find that Viewmaker Networks, a recently propo sed model for learning augmentations for contrastive learning, produce label-des troying augmentations that stochastically destroy features needed for different downstream tasks. These augmentations are interpretable (e.g. altering shapes, d igits, or letters added to images) and surprisingly often result in better perfo rmance compared to expert-designed augmentations, despite not preserving label i nformation. To support our empirical results, we theoretically analyze a simple contrastive learning setting with a linear model. In this setting, label-destroy ing augmentations are crucial for preventing one set of features from suppressin g the learning of features useful for another downstream task. Our results highl ight the need for analyzing the interaction between multiple downstream tasks wh en trying to explain the success of foundation models.
****
On the Exploitability of Instruction Tuning
Manli Shu, Jiongxiao Wang, Chen Zhu, Jonas Geiping, Chaowei Xiao, Tom Goldstein

Instruction tuning is an effective technique to align large language models (LLMs) with human intent. In this work, we investigate how an adversary can exploit instruction tuning by injecting specific instruction-following examples into the training data that intentionally changes the model's behavior. For example, an adversary can achieve content injection by injecting training examples that mention target content and eliciting such behavior from downstream models. To achieve this goal, we propose \textit{AutoPoison}, an automated data poisoning pipeline. It naturally and coherently incorporates versatile attack goals into poisoned data with the help of an oracle LLM. We showcase two example attacks: content injection and over-refusal attacks, each aiming to induce a specific exploitable behavior. We quantify and benchmark the strength and the stealthiness of our data poisoning scheme. Our results show that AutoPoison allows an adversary to change a model's behavior by poisoning only a small fraction of data while maintaining a high level of stealthiness in the poisoned examples. We hope our work sheds light on how data quality affects the behavior of instruction-tuned models and raises awareness of the importance of data quality for responsible deployments of LLMs.
****
Residual Q-Learning: Offline and Online Policy Customization without Value
Chenran Li, Chen Tang, Haruki Nishimura, Jean Mercat, Masayoshi TOMIZUKA, Wei Zhan
Imitation Learning (IL) is a widely used framework for learning imitative behavior from demonstrations. It is especially appealing for solving complex real-world tasks where handcrafting reward function is difficult, or when the goal is to mimic human expert behavior. However, the learned imitative policy can only follow the behavior in the demonstration. When applying the imitative policy, we may need to customize the policy behavior to meet different requirements coming from diverse downstream tasks. Meanwhile, we still want the customized policy to maintain its imitative nature. To this end, we formulate a new problem setting called policy customization. It defines the learning task as training a policy that inherits the characteristics of the prior policy while satisfying some additional requirements imposed by a target downstream task. We propose a novel and principled approach to interpret and determine the trade-off between the two task objectives. Specifically, we formulate the customization problem as a Markov Decision Process (MDP) with a reward function that combines 1) the inherent reward of the demonstration; and 2) the add-on reward specified by the downstream task. We propose a novel framework, Residual Q-learning, which can solve the formulated MDP by leveraging the prior policy without knowing the inherent reward or value function of the prior policy. We derive a family of residual Q-learning algorithms that can realize offline and online policy customization, and show that the proposed algorithms can effectively accomplish policy customization tasks in various environments. Demo videos and code are available on our website: https://sites.google.com/view/residualq-learning.
****
LICO: Explainable Models with Language-Image COnsistency
Yiming Lei, Zilong Li, Yangyang Li, Junping Zhang, Hongming Shan
Interpreting the decisions of deep learning models has been actively studied since the explosion of deep neural networks. One of the most convincing interpretation approaches is salience-based visual interpretation, such as Grad-CAM, where the generation of attention maps depends merely on categorical labels. Although existing interpretation methods can provide explainable decision clues, they often yield partial correspondence between image and saliency maps due to the limited discriminative information from one-hot labels. This paper develops a Language-Image COnsistency model for explainable image classification, termed LICO, by correlating learnable linguistic prompts with corresponding visual features in a coarse-to-fine manner. Specifically, we first establish a coarse global manifold structure alignment by minimizing the distance between the distributions of image and language features. We then achieve fine-grained saliency maps by applying optimal transport (OT) theory to assign local feature maps with class-specific prompts. Extensive experimental results on eight benchmark datasets demonstrat

e that the proposed LICO achieves a significant improvement in generating more explainable attention maps in conjunction with existing interpretation methods such as Grad-CAM. Remarkably, LICO improves the classification performance of existing models without introducing any computational overhead during inference.
****
Solving Inverse Physics Problems with Score Matching
Benjamin Holzschuh, Simona Vegetti, Nils Thuerey
We propose to solve inverse problems involving the temporal evolution of physics systems by leveraging recent advances from diffusion models. Our method moves the system's current state backward in time step by step by combining an approximate inverse physics simulator and a learned correction function. A central insight of our work is that training the learned correction with a single-step loss is equivalent to a score matching objective, while recursively predicting longer parts of the trajectory during training relates to maximum likelihood training of a corresponding probability flow.We highlight the advantages of our algorithm compared to standard denoising score matching and implicit score matching, as well as fully learned baselines for a wide range of inverse physics problems. The resulting inverse solver has excellent accuracy and temporal stability and, in contrast to other learned inverse solvers, allows for sampling the posterior of the solutions. Code and experiments are available at https://github.com/tum-pbs/SMDP.
****
Embedding Space Interpolation Beyond Mini-Batch, Beyond Pairs and Beyond Examples
Shashanka Venkataramanan, Ewa Kijak, laurent amsaleg, Yannis Avrithis
Mixup refers to interpolation-based data augmentation, originally motivated as a way to go beyond empirical risk minimization (ERM). Its extensions mostly focus on the definition of interpolation and the space (input or feature) where it takes place, while the augmentation process itself is less studied. In most methods, the number of generated examples is limited to the mini-batch size and the number of examples being interpolated is limited to two (pairs), in the input space.We make progress in this direction by introducing MultiMix, which generates an arbitrarily large number of interpolated examples beyond the mini-batch size and interpolates the entire mini-batch in the embedding space. Effectively, we sample on the entire convex hull of the mini-batch rather than along linear segments between pairs of examples.On sequence data, we further extend to Dense MultiMix. We densely interpolate features and target labels at each spatial location and also apply the loss densely. To mitigate the lack of dense labels, we inherit labels from examples and weight interpolation factors by attention as a measure of confidence.Overall, we increase the number of loss terms per mini-batch by orders of magnitude at little additional cost. This is only possible because of interpolating in the embedding space. We empirically show that our solutions yield significant improvement over state-of-the-art mixup methods on four different benchmarks, despite interpolation being only linear. By analyzing the embedding space, we show that the classes are more tightly clustered and uniformly spread over the embedding space, thereby explaining the improved behavior.
****
Approximation-Generalization Trade-offs under (Approximate) Group Equivariance
Mircea Petrache, Shubhendu Trivedi
The explicit incorporation of task-specific inductive biases through symmetry has emerged as a general design precept in the development of high-performance machine learning models. For example, group equivariant neural networks have demonstrated impressive performance across various domains and applications such as protein and drug design. A prevalent intuition about such models is that the integration of relevant symmetry results in enhanced generalization. Moreover, it is posited that when the data and/or the model exhibits only approximate or partial symmetry, the optimal or best-performing model is one where the model symmetry aligns with the data symmetry. In this paper, we conduct a formal unified investigation of these intuitions. To begin, we present quantitative bounds that demonstrate how models capturing task-specific symmetries lead to improved generaliza

tion. Utilizing this quantification, we examine the more general question of dea
ling with approximate/partial symmetries. We establish, for a given symmetry gro
up, a quantitative comparison between the approximate equivariance of the model
and that of the data distribution, precisely connecting model equivariance error
 and data equivariance error. Our result delineates the conditions under which t
he model equivariance error is optimal, thereby yielding the best-performing mod
el for the given task and data.
****
Equivariant Neural Operator Learning with Graphon Convolution
Chaoran Cheng, Jian Peng
We propose a general architecture that combines the coefficient learning scheme
with a residual operator layer for learning mappings between continuous function
s in the 3D Euclidean space. Our proposed model is guaranteed to achieve SE(3)-e
quivariance by design. From the graph spectrum view, our method can be interpret
ed as convolution on graphons (dense graphs with infinitely many nodes), which w
e term InfGCN. By leveraging both the continuous graphon structure and the discr
ete graph structure of the input data, our model can effectively capture the geo
metric information while preserving equivariance. Through extensive experiments
on large-scale electron density datasets, we observed that our model significant
ly outperformed the current state-of-the-art architectures. Multiple ablation st
udies were also carried out to demonstrate the effectiveness of the proposed arc
hitecture.
****
Reinforcement Learning with Simple Sequence Priors
Tankred Saanum, Noémi Éltet■, Peter Dayan, Marcel Binz, Eric Schulz
In reinforcement learning (RL), simplicity is typically quantified on an action-
by-action basis -- but this timescale ignores temporal regularities, like repeti
tions, often present in sequential strategies. We therefore propose an RL algori
thm that learns to solve tasks with sequences of actions that are compressible.
We explore two possible sources of simple action sequences: Sequences that can b
e learned by autoregressive models, and sequences that are compressible with off
-the-shelf data compression algorithms. Distilling these preferences into sequen
ce priors, we derive a novel information-theoretic objective that incentivizes a
gents to learn policies that maximize rewards while conforming to these priors.
We show that the resulting RL algorithm leads to faster learning, and attains hi
gher returns than state-of-the-art model-free approaches in a series of continuo
us control tasks from the DeepMind Control Suite. These priors also produce a po
werful information-regularized agent that is robust to noisy observations and ca
n perform open-loop control.
****
Fed-FA: Theoretically Modeling Client Data Divergence for Federated Language Bac
kdoor Defense
Zhiyuan Zhang, Deli Chen, Hao Zhou, Fandong Meng, Jie Zhou, Xu Sun
Federated learning algorithms enable neural network models to be trained across
multiple decentralized edge devices without sharing private data. However, they
are susceptible to backdoor attacks launched by malicious clients. Existing robu
st federated aggregation algorithms heuristically detect and exclude suspicious
clients based on their parameter distances, but they are ineffective on Natural
Language Processing (NLP) tasks. The main reason is that, although text backdoor
 patterns are obvious at the underlying dataset level, they are usually hidden a
t the parameter level, since injecting backdoors into texts with discrete featur
e space has less impact on the statistics of the model parameters. To settle thi
s issue, we propose to identify backdoor clients by explicitly modeling the data
 divergence among clients in federated NLP systems. Through theoretical analysis
, we derive the f-divergence indicator to estimate the client data divergence wi
th aggregation updates and Hessians. Furthermore, we devise a dataset synthesiza
tion method with a Hessian reassignment mechanism guided by the diffusion theory
 to address the key challenge of inaccessible datasets in calculating clients' d
ata Hessians.We then present the novel Federated F-Divergence-Based Aggregation~
(\textbf{Fed-FA}) algorithm, which leverages the f-divergence indicator to detec

t and discard suspicious clients. Extensive empirical results show that Fed-FA o
utperforms all the parameter distance-based methods in defending against backdoo
r attacks among various natural language backdoor attack scenarios.
****
One Less Reason for Filter Pruning: Gaining Free Adversarial Robustness with Str
uctured Grouped Kernel Pruning
Shaochen (Henry) Zhong, Zaichuan You, Jiamu Zhang, Sebastian Zhao, Zachary LeCla
ire, Zirui Liu, Daochen Zha, Vipin Chaudhary, Shuai Xu, Xia Hu
Densely structured pruning methods utilizing simple pruning heuristics can deliv
er immediate compression and acceleration benefits with acceptable benign perfor
mances. However, empirical findings indicate such naively pruned networks are ex
tremely fragile under simple adversarial attacks. Naturally, we would be interes
ted in knowing if such a phenomenon also holds for carefully designed modern str
uctured pruning methods. If so, then to what extent is the severity? And what ki
nd of remedies are available? Unfortunately, both the questions and the solution
 remain largely unaddressed: no prior art is able to provide a thorough investig
ation on the adversarial performance of modern structured pruning methods (spoil
er: it is not good), yet the few works that attempt to provide mitigation often
do so at various extra costs with only to-be-desired performance.In this work, w
e answer both questions by fairly and comprehensively investigating the adversar
ial performance of 10+ popular structured pruning methods. Solution-wise, we tak
e advantage of Grouped Kernel Pruning (GKP)'s recent success in pushing densely
structured pruning freedom to a more fine-grained level. By mixing up kernel smo
othness — a classic robustness-related kernel-level metric — into a modified GKP
 procedure, we present a one-shot-post-train-weight-dependent GKP method capable
 of advancing SOTA performance on both the benign and adversarial scale, while r
equiring no extra (in fact, often less) cost than a standard pruning procedure.
Please refer to our GitHub repository for code implementation, tool sharing, and
 model checkpoints.
****
Survival Instinct in Offline Reinforcement Learning
Anqi Li, Dipendra Misra, Andrey Kolobov, Ching-An Cheng
We present a novel observation about the behavior of offline reinforcement learn
ing (RL) algorithms: on many benchmark datasets, offline RL can produce well-per
forming and safe policies even when trained with "wrong" reward labels, such as
those that are zero everywhere or are negatives of the true rewards. This phenom
enon cannot be easily explained by offline RL's return maximization objective. M
oreover, it gives offline RL a degree of robustness that is uncharacteristic of
its online RL counterparts, which are known to be sensitive to reward design. We
 demonstrate that this surprising robustness property is attributable to an inte
rplay between the notion of pessimism in offline RL algorithms and certain impli
cit biases in common data collection practices. As we prove in this work, pessim
ism endows the agent with a survival instinct, i.e., an incentive to stay within
 the data support in the long term, while the limited and biased data coverage f
urther constrains the set of survival policies. Formally, given a reward class -
- which may not even contain the true reward -- we identify conditions on the tr
aining data distribution that enable offline RL to learn a near-optimal and safe
 policy from any reward within the class. We argue that the survival instinct sh
ould be taken into account when interpreting results from existing offline RL be
nchmarks and when creating future ones. Our empirical and theoretical results su
ggest a new paradigm for offline RL, whereby an agent is "nudged" to learn a des
irable behavior with imperfect reward but purposely biased data coverage. Please
 visit our website https://survival-instinct.github.io for accompanied code and
videos.
****
Recurrent Hypernetworks are Surprisingly Strong in Meta-RL
Jacob Beck, Risto Vuorio, Zheng Xiong, Shimon Whiteson
Deep reinforcement learning (RL) is notoriously impractical to deploy due to sam
ple inefficiency. Meta-RL directly addresses this sample inefficiency by learnin
g to perform few-shot learning when a distribution of related tasks is available

for meta-training. While many specialized meta-RL methods have been proposed, recent work suggests that end-to-end learning in conjunction with an off-the-shelf sequential model, such as a recurrent network, is a surprisingly strong baseline. However, such claims have been controversial due to limited supporting evidence, particularly in the face of prior work establishing precisely the opposite. In this paper, we conduct an empirical investigation. While we likewise find that a recurrent network can achieve strong performance, we demonstrate that the use of hypernetworks is crucial to maximizing their potential. Surprisingly, when combined with hypernetworks, the recurrent baselines that are far simpler than existing specialized methods actually achieve the strongest performance of all methods evaluated. We provide code at https://github.com/jacooba/hyper.
****
The Target-Charging Technique for Privacy Analysis across Interactive Computations
Edith Cohen, Xin Lyu
We propose the \emph{Target Charging Technique} (TCT), a unified privacy analysis framework for interactive settings where a sensitive dataset is accessed multiple times using differentially private algorithms. Unlike traditional composition, where privacy guarantees deteriorate quickly with the number of accesses, TCT allows computations that don't hit a specified \emph{target}, often the vast majority, to be essentially free (while incurring instead a small overhead on those that do hit their targets). TCT generalizes tools such as the sparse vector technique and top-k selection from private candidates and extends their remarkable privacy enhancement benefits from noisy Lipschitz functions to general private algorithms.
****
EV-Eye: Rethinking High-frequency Eye Tracking through the Lenses of Event Cameras
Guangrong Zhao, Yurun Yang, Jingwei Liu, Ning Chen, Yiran Shen, Hongkai Wen, Guohao Lan
In this paper, we present EV-Eye, a first-of-its-kind large scale multimodal eye tracking dataset aimed at inspiring research on high-frequency eye/gaze tracking. EV-Eye utilizes an emerging bio-inspired event camera to capture independent pixel-level intensity changes induced by eye movements, achieving sub-microsecond latency. Our dataset was curated over a two-week period and collected from 48 participants encompassing diverse genders and age groups. It comprises over 1.5 million near-eye grayscale images and 2.7 billion event samples generated by two DAVIS346 event cameras. Additionally, the dataset contains 675 thousands scene images and 2.7 million gaze references captured by Tobii Pro Glasses 3 eye tracker for cross-modality validation. Compared with existing event-based high-frequency eye tracking datasets, our dataset is significantly larger in size, and the gaze references involve more natural eye movement patterns, i.e., fixation, saccade and smooth pursuit. Alongside the event data, we also present a hybrid eye tracking method as benchmark, which leverages both the near-eye grayscale images and event data for robust and high-frequency eye tracking. We show that our method achieves higher accuracy for both pupil and gaze estimation tasks compared to the existing solution.
****
Diffusion Schrödinger Bridge Matching
Yuyang Shi, Valentin De Bortoli, Andrew Campbell, Arnaud Doucet
Solving transport problems, i.e. finding a map transporting one given distribution to another, has numerous applications in machine learning. Novel mass transport methods motivated by generative modeling have recently been proposed, e.g. Denoising Diffusion Models (DDMs) and Flow Matching Models (FMMs) implement such a transport through a Stochastic Differential Equation (SDE) or an Ordinary Differential Equation (ODE). However, while it is desirable in many applications to approximate the deterministic dynamic Optimal Transport (OT) map which admits attractive properties, DDMs and FMMs are not guaranteed to provide transports close to the OT map. In contrast, Schrödinger bridges (SBs) compute stochastic dynamic mappings which recover entropy-regularized versions of OT. Unfortunately, exis

ting numerical methods approximating SBs either scale poorly with dimension or accumulate errors across iterations. In this work, we introduce Iterative Markovian Fitting (IMF), a new methodology for solving SB problems, and Diffusion Schrödinger Bridge Matching (DSBM), a novel numerical algorithm for computing IMF iterates. DSBM significantly improves over previous SB numerics and recovers as special/limiting cases various recent transport methods. We demonstrate the performance of DSBM on a variety of problems.
****
Interpreting Unsupervised Anomaly Detection in Security via Rule Extraction
Ruoyu Li, Qing Li, Yu Zhang, Dan Zhao, Yong Jiang, Yong Yang
Many security applications require unsupervised anomaly detection, as malicious data are extremely rare and often only unlabeled normal data are available for training (i.e., zero-positive). However, security operators are concerned about the high stakes of trusting black-box models due to their lack of interpretability. In this paper, we propose a post-hoc method to globally explain a black-box unsupervised anomaly detection model via rule extraction.First, we propose the concept of distribution decomposition rules that decompose the complex distribution of normal data into multiple compositional distributions. To find such rules, we design an unsupervised Interior Clustering Tree that incorporates the model prediction into the splitting criteria. Then, we propose the Compositional Boundary Exploration (CBE) algorithm to obtain the boundary inference rules that estimate the decision boundary of the original model on each compositional distribution. By merging these two types of rules into a rule set, we can present the inferential process of the unsupervised black-box model in a human-understandable way, and build a surrogate rule-based model for online deployment at the same time. We conduct comprehensive experiments on the explanation of four distinct unsupervised anomaly detection models on various real-world datasets. The evaluation shows that our method outperforms existing methods in terms of diverse metrics including fidelity, correctness and robustness.
****
Cal-QL: Calibrated Offline RL Pre-Training for Efficient Online Fine-Tuning
Mitsuhiko Nakamoto, Simon Zhai, Anikait Singh, Max Sobol Mark, Yi Ma, Chelsea Finn, Aviral Kumar, Sergey Levine
A compelling use case of offline reinforcement learning (RL) is to obtain a policy initialization from existing datasets followed by fast online fine-tuning with limited interaction. However, existing offline RL methods tend to behave poorly during fine-tuning. In this paper, we devise an approach for learning an effective initialization from offline data that also enables fast online fine-tuning capabilities. Our approach, calibrated Q-learning (Cal-QL), accomplishes this by learning a conservative value function initialization that underestimates the value of the learned policy from offline data, while also being calibrated, in the sense that the learned Q-values are at a reasonable scale. We refer to this property as calibration, and define it formally as providing a lower bound on the true value function of the learned policy and an upper bound on the value of some other (suboptimal) reference policy, which may simply be the behavior policy. We show that offline RL algorithms that learn such calibrated value functions lead to effective online fine-tuning, enabling us to take the benefits of offline initializations in online fine-tuning. In practice, Cal-QL can be implemented on top of the conservative Q learning (CQL) for offline RL within a one-line code change. Empirically, Cal-QL outperforms state-of-the-art methods on 9/11 fine-tuning benchmark tasks that we study in this paper. Code and video are available at https://nakamotoo.github.io/Cal-QL
****
PLASTIC: Improving Input and Label Plasticity for Sample Efficient Reinforcement Learning
Hojoon Lee, Hanseul Cho, HYUNSEUNG KIM, DAEHOON GWAK, Joonkee Kim, Jaegul Choo, Se-Young Yun, Chulhee Yun
In Reinforcement Learning (RL), enhancing sample efficiency is crucial, particularly in scenarios when data acquisition is costly and risky. In principle, off-policy RL algorithms can improve sample efficiency by allowing multiple updates p

er environment interaction. However, these multiple updates often lead the model to overfit to earlier interactions, which is referred to as the loss of plasticity. Our study investigates the underlying causes of this phenomenon by dividing plasticity into two aspects. Input plasticity, which denotes the model's adaptability to changing input data, and label plasticity, which denotes the model's adaptability to evolving input-output relationships. Synthetic experiments on the CIFAR-10 dataset reveal that finding smoother minima of loss landscape enhances input plasticity, whereas refined gradient propagation improves label plasticity. Leveraging these findings, we introduce the PLASTIC algorithm, which harmoniously combines techniques to address both concerns. With minimal architectural modifications, PLASTIC achieves competitive performance on benchmarks including Atari-100k and Deepmind Control Suite. This result emphasizes the importance of preserving the model's plasticity to elevate the sample efficiency in RL. The code is available at https://github.com/dojeon-ai/plastic.
****

Metropolis Sampling for Constrained Diffusion Models
Nic Fishman, Leo Klarner, Emile Mathieu, Michael Hutchinson, Valentin De Bortoli
Denoising diffusion models have recently emerged as the predominant paradigm for generative modelling on image domains. In addition, their extension to Riemannian manifolds has facilitated a range of applications across the natural sciences. While many of these problems stand to benefit from the ability to specify arbitrary, domain-informed constraints, this setting is not covered by the existing (Riemannian) diffusion model methodology. Recent work has attempted to address this issue by constructing novel noising processes based on the reflected Brownian motion and logarithmic barrier methods. However, the associated samplers are either computationally burdensome or only apply to convex subsets of Euclidean space. In this paper, we introduce an alternative, simple noising scheme based on Metropolis sampling that affords substantial gains in computational efficiency and empirical performance compared to the earlier samplers. Of independent interest, we prove that this new process corresponds to a valid discretisation of the reflected Brownian motion. We demonstrate the scalability and flexibility of our approach on a range of problem settings with convex and non-convex constraints, including applications from geospatial modelling, robotics and protein design.
****

ReTR: Modeling Rendering Via Transformer for Generalizable Neural Surface Reconstruction
Yixun Liang, Hao He, Yingcong Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

FETV: A Benchmark for Fine-Grained Evaluation of Open-Domain Text-to-Video Generation
Yuanxin Liu, Lei Li, Shuhuai Ren, Rundong Gao, Shicheng Li, Sishuo Chen, Xu Sun, Lu Hou
Recently, open-domain text-to-video (T2V) generation models have made remarkable progress. However, the promising results are mainly shown by the qualitative cases of generated videos, while the quantitative evaluation of T2V models still faces two critical problems. Firstly, existing studies lack fine-grained evaluation of T2V models on different categories of text prompts. Although some benchmarks have categorized the prompts, their categorization either only focuses on a single aspect or fails to consider the temporal information in video generation. Secondly, it is unclear whether the automatic evaluation metrics are consistent with human standards. To address these problems, we propose FETV, a benchmark for Fine-grained Evaluation of Text-to-Video generation. FETV is multi-aspect, categorizing the prompts based on three orthogonal aspects: the major content, the attributes to control and the prompt complexity. FETV is also temporal-aware, which introduces several temporal categories tailored for video generation. Based on FETV, we conduct comprehensive manual evaluations of four representative T2V

models, revealing their pros and cons on different categories of prompts from different aspects. We also extend FETV as a testbed to evaluate the reliability of automatic T2V metrics. The multi-aspect categorization of FETV enables fine-grained analysis of the metrics' reliability in different scenarios. We find that existing automatic metrics (e.g., CLIPScore and FVD) correlate poorly with human evaluation. To address this problem, we explore several solutions to improve CLIPScore and FVD, and develop two automatic metrics that exhibit significant higher correlation with humans than existing metrics. Benchmark page: https://github.com/llyx97/FETV.
****
Stability Guarantees for Feature Attributions with Multiplicative Smoothing
Anton Xue, Rajeev Alur, Eric Wong
Explanation methods for machine learning models tend not to provide any formal guarantees and may not reflect the underlying decision-making process.In this work, we analyze stability as a property for reliable feature attribution methods. We prove that relaxed variants of stability are guaranteed if the model is sufficiently Lipschitz with respect to the masking of features. We develop a smoothing method called Multiplicative Smoothing (MuS) to achieve such a model.We show that MuS overcomes the theoretical limitations of standard smoothing techniques and can be integrated with any classifier and feature attribution method.We evaluate MuS on vision and language models with various feature attribution methods, such as LIME and SHAP, and demonstrate that MuS endows feature attributions with non-trivial stability guarantees.
****
Pruning vs Quantization: Which is Better?
Andrey Kuzmin, Markus Nagel, Mart van Baalen, Arash Behboodi, Tijmen Blankevoort
Neural network pruning and quantization techniques are almost as old as neural networks themselves. However, to date, only ad-hoc comparisons between the two have been published. In this paper, we set out to answer the question of which is better: neural network quantization or pruning? By answering this question, we hope to inform design decisions made on neural network hardware going forward. We provide an extensive comparison between the two techniques for compressing deep neural networks. First, we give an analytical comparison of expected quantization and pruning error for general data distributions.Then, we provide lower and upper bounds for the per-layer pruning and quantization error in trained networks and compare these to empirical error after optimization.Finally, we provide an extensive experimental comparison for training 8 large-scale models trained on 3 tasks and provide insights into the representations learned during fine-tuning with quantization and pruning in the loop.Our results show that in most cases quantization outperforms pruning. Only in some scenarios with a very high compression ratio, compression might be beneficial from an accuracy standpoint.
****
EvoFed: Leveraging Evolutionary Strategies for Communication-Efficient Federated Learning
Mohammad Mahdi Rahimi, Hasnain Irshad Bhatti, Younghyun Park, Humaira Kousar, Do-Yeon Kim, Jaekyun Moon
Federated Learning (FL) is a decentralized machine learning paradigm that enables collaborative model training across dispersed nodes without having to force individual nodes to share data.However, its broad adoption is hindered by the high communication costs of transmitting a large number of model parameters. This paper presents EvoFed, a novel approach that integrates Evolutionary Strategies (ES) with FL to address these challenges.EvoFed employs a concept of `fitness-based information sharing’, deviating significantly from the conventional model-based FL. Rather than exchanging the actual updated model parameters, each node transmits a distance-based similarity measure between the locally updated model and each member of the noise-perturbed model population. Each node, as well as the server, generates an identical population set of perturbed models in a completely synchronized fashion using the same random seeds. With properly chosen noise variance and population size, perturbed models can be combined to closely reflect the actual model updated using the local dataset, allowing the transmitted simil

arity measures (or fitness values) to carry nearly the complete information about the model parameters.As the population size is typically much smaller than the number of model parameters, the savings in communication load is large. The server aggregates these fitness values and is able to update the global model. This global fitness vector is then disseminated back to the nodes, each of which applies the same update to be synchronized to the global model. Our analysis shows that EvoFed converges, and our experimental results validate that at the cost of increased local processing loads, EvoFed achieves performance comparable to FedAvg while reducing overall communication requirements drastically in various practical settings.
****
UUKG: Unified Urban Knowledge Graph Dataset for Urban Spatiotemporal Prediction
Yansong Ning, Hao Liu, Hao Wang, Zhenyu Zeng, Hui Xiong
Accurate Urban SpatioTemporal Prediction (USTP) is of great importance to the development and operation of the smart city. As an emerging building block, multi-sourced urban data are usually integrated as urban knowledge graphs (UrbanKGs) to provide critical knowledge for urban spatiotemporal prediction models. However, existing UrbanKGs are often tailored for specific downstream prediction tasks and are not publicly available, which limits the potential advancement. This paper presents UUKG, the unified urban knowledge graph dataset for knowledge-enhanced urban spatiotemporal predictions. Specifically, we first construct UrbanKGs consisting of millions of triplets for two metropolises by connecting heterogeneous urban entities such as administrative boroughs, POIs, and road segments. Moreover, we conduct qualitative and quantitative analysis on constructed UrbanKGs and uncover diverse high-order structural patterns, such as hierarchies and cycles, that can be leveraged to benefit downstream USTP tasks. To validate and facilitate the use of UrbanKGs, we implement and evaluate 15 KG embedding methods on the KG completion task and integrate the learned KG embeddings into 9 spatiotemporal models for five different USTP tasks. The extensive experimental results not only provide benchmarks of knowledge-enhanced USTP models under different task settings but also highlight the potential of state-of-the-art high-order structure-aware UrbanKG embedding methods. We hope the proposed UUKG fosters research on urban knowledge graphs and broad smart city applications. The dataset and source code are available at https://github.com/usail-hkust/UUKG/.
****
StateMask: Explaining Deep Reinforcement Learning through State Mask
Zelei Cheng, Xian Wu, Jiahao Yu, Wenhai Sun, Wenbo Guo, Xinyu Xing
Despite the promising performance of deep reinforcement learning (DRL) agents in many challenging scenarios, the black-box nature of these agents greatly limits their applications in critical domains. Prior research has proposed several explanation techniques to understand the deep learning-based policies in RL. Most existing methods explain why an agent takes individual actions rather than pinpointing the critical steps to its final reward. To fill this gap, we propose StateMask, a novel method to identify the states most critical to the agent's final reward. The high-level idea of StateMask is to learn a mask net that blinds a target agent and forces it to take random actions at some steps without compromising the agent's performance. Through careful design, we can theoretically ensure that the masked agent performs similarly to the original agent. We evaluate StateMask in various popular RL environments and show its superiority over existing explainers in explanation fidelity. We also show that StateMask has better utilities, such as launching adversarial attacks and patching policy errors.
****
Faster Margin Maximization Rates for Generic Optimization Methods
Guanghui Wang, Zihao Hu, Vidya Muthukumar, Jacob D. Abernethy
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Managing Temporal Resolution in Continuous Value Estimation: A Fundamental Trade

-off
Zichen (Vincent) Zhang, Johannes Kirschner, Junxi Zhang, Francesco Zanini, Alex Ayoub, Masood Dehghan, Dale Schuurmans
A default assumption in reinforcement learning (RL) and optimal control is that observations arrive at discrete time points on a fixed clock cycle. Yet, many ap plications involve continuous-time systems where the time discretization, in pri nciple, can be managed. The impact of time discretization on RL methods has not been fully characterized in existing theory, but a more detailed analysis of its effect could reveal opportunities for improving data-efficiency. We address thi s gap by analyzing Monte-Carlo policy evaluation for LQR systems and uncover a f undamental trade-off between approximation and statistical error in value estima tion. Importantly, these two errors behave differently to time discretization, l eading to an optimal choice of temporal resolution for a given data budget. Thes e findings show that managing the temporal resolution can provably improve polic y evaluation efficiency in LQR systems with finite data. Empirically, we demonst rate the trade-off in numerical simulations of LQR instances and standard RL ben chmarks for non-linear continuous control.
****
Federated Linear Bandits with Finite Adversarial Actions
Li Fan, Ruida Zhou, Chao Tian, Cong Shen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
BERT Lost Patience Won't Be Robust to Adversarial Slowdown
Zachary Coalson, Gabriel Ritter, Rakesh Bobba, Sanghyun Hong
In this paper, we systematically evaluate the robustness of multi-exit language models against adversarial slowdown. To audit their robustness, we design a slow down attack that generates natural adversarial text bypassing early-exit points. We use the resulting WAFFLE attack as a vehicle to conduct a comprehensive eval uation of three multi-exit mechanisms with the GLUE benchmark against adversaria l slowdown. We then show our attack significantly reduces the computational savi ngs provided by the three methods in both white-box and black-box settings. The more complex a mechanism is, the more vulnerable it is to adversarial slowdown. We also perform a linguistic analysis of the perturbed text inputs, identifying common perturbation patterns that our attack generates, and comparing them with standard adversarial text attacks. Moreover, we show that adversarial training i s ineffective in defeating our slowdown attack, but input sanitization with a co nversational model, e.g., ChatGPT, can remove perturbations effectively. This re sult suggests that future work is needed for developing efficient yet robust mul ti-exit models. Our code is available at: https://github.com/ztcoalson/WAFFLE
****
RECKONING: Reasoning through Dynamic Knowledge Encoding
Zeming Chen, Gail Weiss, Eric Mitchell, Asli Celikyilmaz, Antoine Bosselut
Recent studies on transformer-based language models show that they can answer qu estions by reasoning over knowledge provided as part of the context (i.e., in-co ntext reasoning). However, since the available knowledge is often not filtered f or a particular question, in-context reasoning can be sensitive to distractor fa cts, additional content that is irrelevant to a question but that may be relevan t for a different question (i.e., not necessarily random noise). In these situat ions, the model fails todistinguish the necessary knowledge to answer the questi on, leading to spurious reasoning and degraded performance. This reasoning failu re contrasts with the model’s apparent ability to distinguish its contextual kno wledge from all the knowledge it has memorized during pre-training. Following th is observation, we propose teaching the model to reason more robustly by folding the provided contextual knowledge into the model’s parameters before presenting it with a question. Our method, RECKONING, is a bi-level learning algorithm tha t teaches language models to reason by updating their parametric knowledge throu gh back-propagation, allowing them to answer questions using the updated paramet

ers. During training, the inner loop rapidly adapts a copy of the model weights to encode contextual knowledge into its parameters. In the outer loop, the model learns to use the updated weights to reproduce and answer reasoning questions about the memorized knowledge. Our experiments on three diverse multi-hop reasoning datasets show that RECKONING's performance improves over the in-context reasoning baseline (by up to 4.5%). We also find that compared to in-context reasoning, RECKONING generalizes better to longer reasoning chains unseen during training, is more robust to distractors in the context, and is computationally more efficient when multiple questions are asked about the same knowledge.
****

Regularity as Intrinsic Reward for Free Play
Cansu Sancaktar, Justus Piater, Georg Martius
We propose regularity as a novel reward signal for intrinsically-motivated reinforcement learning. Taking inspiration from child development, we postulate that striving for structure and order helps guide exploration towards a subspace of tasks that are not favored by naive uncertainty-based intrinsic rewards. Our generalized formulation of Regularity as Intrinsic Reward (RaIR) allows us to operationalize it within model-based reinforcement learning. In a synthetic environment, we showcase the plethora of structured patterns that can emerge from pursuing this regularity objective. We also demonstrate the strength of our method in a multi-object robotic manipulation environment. We incorporate RaIR into free play and use it to complement the model's epistemic uncertainty as an intrinsic reward. Doing so, we witness the autonomous construction of towers and other regular structures during free play, which leads to a substantial improvement in zero-shot downstream task performance on assembly tasks.
****

Guiding Large Language Models via Directional Stimulus Prompting
Zekun Li, Baolin Peng, Pengcheng He, Michel Galley, Jianfeng Gao, Xifeng Yan
We introduce Directional Stimulus Prompting, a novel framework for guiding black-box large language models (LLMs) towards specific desired outputs. Instead of directly adjusting LLMs, our method employs a small tunable policy model (e.g., T5) to generate an auxiliary directional stimulus prompt for each input instance. These directional stimulus prompts act as nuanced, instance-specific hints and clues to guide LLMs in generating desired outcomes, such as including specific keywords in the generated summary. Our approach sidesteps the challenges of direct LLM tuning by optimizing the policy model to explore directional stimulus prompts that align LLMs with desired behaviors. The policy model can be optimized through 1) supervised fine-tuning using labeled data and 2) reinforcement learning from offline or online rewards based on the LLM's output. We evaluate our method across various tasks, including summarization, dialogue response generation, and chain-of-thought reasoning. Our experiments indicate a consistent improvement in the performance of LLMs such as ChatGPT, Codex, and InstructGPT on these supervised tasks with minimal labeled data. Remarkably, by utilizing merely 80 dialogues from the MultiWOZ dataset, our approach boosts ChatGPT's performance by a relative 41.4%, achieving or exceeding the performance of some fully supervised state-of-the-art models. Moreover, the instance-specific chain-of-thought prompt generated through our method enhances InstructGPT's reasoning accuracy, outperforming both generalized human-crafted prompts and those generated through automatic prompt engineering. The code and data are publicly available at https://github.com/Leezekun/Directional-Stimulus-Prompting.
****

Distributionally Robust Ensemble of Lottery Tickets Towards Calibrated Sparse Network Training
Hitesh Sapkota, Dingrong Wang, Zhiqiang Tao, Qi Yu
The recently developed sparse network training methods, such as Lottery Ticket Hypothesis (LTH) and its variants, have shown impressive learning capacity by finding sparse sub-networks from a dense one. While these methods could largely sparsify deep networks, they generally focus more on realizing comparable accuracy to dense counterparts yet neglect network calibration. However, how to achieve calibrated network predictions lies at the core of improving model reliability, e

specially when it comes to addressing the overconfident issue and out-of-distribution cases. In this study, we propose a novel Distributionally Robust Optimization (DRO) framework to achieve an ensemble of lottery tickets towards calibrated network sparsification. Specifically, the proposed DRO ensemble aims to learn multiple diverse and complementary sparse sub-networks (tickets) with the guidance of uncertainty sets, which encourage tickets to gradually capture different data distributions from easy to hard and naturally complement each other. We theoretically justify the strong calibration performance by showing how the proposed robust training process guarantees to lower the confidence of incorrect predictions. Extensive experimental results on several benchmarks show that our proposed lottery ticket ensemble leads to a clear calibration improvement without sacrificing accuracy and burdening inference costs. Furthermore, experiments on OOD datasets demonstrate the robustness of our approach in the open-set environment.
****

A Recurrent Neural Circuit Mechanism of Temporal-scaling Equivariant Representation

Junfeng Zuo, Xiao Liu, Ying Nian Wu, Si Wu, Wenhao Zhang

Time perception is critical in our daily life. An important feature of time perception is temporal scaling (TS): the ability to generate temporal sequences (e.g., motor actions) at different speeds. However, it is largely unknown about the math principle underlying temporal scaling in recurrent circuits in the brain. To shed insight, the present study investigates the temporal scaling from the Lie group point of view. We propose a canonical nonlinear recurrent circuit dynamics, modeled as a continuous attractor network, whose neuronal population responses embed a temporal sequence that is TS equivariant. Furthermore, we found the TS group operators can be explicitly represented by a control input fed into the recurrent circuit, where the input gain determines the temporal scaling factor (group parameter), and the spatial offset between the control input and network state emerges the generator. The neuronal responses in the recurrent circuit are also consistent with experimental findings. We illustrated that the recurrent circuit can drive a feedforward circuit to generate complex temporal sequences with different time scales, even in the case of negative time scaling (''time reversal''). Our work for the first time analytically links the abstract temporal scaling group and concrete neural circuit dynamics.
****

Sample Complexity of Goal-Conditioned Hierarchical Reinforcement Learning

Arnaud Robert, Ciara Pike-Burke, Aldo A. Faisal

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

MiliPoint: A Point Cloud Dataset for mmWave Radar

Han Cui, Shu Zhong, Jiacheng Wu, Zichao Shen, Naim Dahnoun, Yiren Zhao

Millimetre-wave (mmWave) radar has emerged as an attractive and cost-effective alternative for human activity sensing compared to traditional camera-based systems. mmWave radars are also non-intrusive, providing better protection for user privacy. However, as a Radio Frequency based technology, mmWave radars rely on capturing reflected signals from objects, making them more prone to noise compared to cameras. This raises an intriguing question for the deep learning community: Can we develop more effective point set-based deep learning methods for such attractive sensors?   To answer this question, our work, termed MiliPoint, delves into this idea by providing a large-scale, open dataset for the community to explore how mmWave radars can be utilised for human activity recognition. Moreover, MiliPoint stands out as it is larger in size than existing datasets, has more diverse human actions represented, and encompasses all three key tasks in human activity recognition. We have also established a range of point-based deep neural networks such as DGCNN, PointNet++ and PointTransformer, on MiliPoint, which can serve to set the ground baseline for further development.
****

NAR-Former V2: Rethinking Transformer for Universal Neural Network Representation Learning
Yun Yi, Haokui Zhang, Rong Xiao, Nannan Wang, Xiaoyu Wang
As more deep learning models are being applied in real-world applications, there is a growing need for modeling and learning the representations of neural networks themselves. An effective representation can be used to predict target attributes of networks without the need for actual training and deployment procedures, facilitating efficient network design and deployment. Recently, inspired by the success of Transformer, some Transformer-based representation learning frameworks have been proposed and achieved promising performance in handling cell-structured models. However, graph neural network (GNN) based approaches still dominate the field of learning representation for the entire network. In this paper, we revisit the Transformer and compare it with GNN to analyze their different architectural characteristics. We then propose a modified Transformer-based universal neural network representation learning model NAR-Former V2. It can learn efficient representations from both cell-structured networks and entire networks. Specifically, we first take the network as a graph and design a straightforward tokenizer to encode the network into a sequence. Then, we incorporate the inductive representation learning capability of GNN into Transformer, enabling Transformer to generalize better when encountering unseen architecture. Additionally, we introduce a series of simple yet effective modifications to enhance the ability of the Transformer in learning representation from graph structures. In encoding entire networks and then predicting the latency, our proposed method surpasses the GNN-based method NNLP by a significant margin on the NNLQP dataset. Furthermore, regarding accuracy prediction on the cell-structured NASBench101 and NASBench201 datasets, our method achieves highly comparable performance to other state-of-the-art methods. The code is available at https://github.com/yuny220/NAR-Former-V2.
****
Sensitivity in Translation Averaging
Lalit Manam, Venu Madhav Govindu
In 3D computer vision, translation averaging solves for absolute translations given a set of pairwise relative translation directions. While there has been much work on robustness to outliers and studies on the uniqueness of the solution, this paper deals with a distinctly different problem of sensitivity in translation averaging under uncertainty. We first analyze sensitivity in estimating scales corresponding to relative directions under small perturbations of the relative directions. Then, we formally define the conditioning of the translation averaging problem, which assesses the reliability of estimated translations based solely on the input directions. We give a sufficient criterion to ensure that the problem is well-conditioned. Subsequently, we provide an efficient algorithm to identify and remove combinations of directions which make the problem ill-conditioned while ensuring uniqueness of the solution. We demonstrate the utility of such analysis in global structure-from-motion pipelines for obtaining 3D reconstructions, which reveals the benefits of filtering the ill-conditioned set of directions in translation averaging in terms of reduced translation errors, a higher number of 3D points triangulated and faster convergence of bundle adjustment.
****
Data-Dependent Bounds for Online Portfolio Selection Without Lipschitzness and Smoothness
Chung-En Tsai, Ying-Ting Lin, Yen-Huan Li
This work introduces the first small-loss and gradual-variation regret bounds for online portfolio selection, marking the first instances of data-dependent bounds for online convex optimization with non-Lipschitz, non-smooth losses. The algorithms we propose exhibit sublinear regret rates in the worst cases and achieve logarithmic regrets when the data is "easy," with per-round time almost linear in the number of investment alternatives. The regret bounds are derived using novel smoothness characterizations of the logarithmic loss, a local norm-based analysis of following the regularized leader (FTRL) with self-concordant regularizers, which are not necessarily barriers, and an implicit variant of optimistic FT

RL with the log-barrier.
****
Outlier-Robust Wasserstein DRO
Sloan Nietert, Ziv Goldfeld, Soroosh Shafiee
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Certified Minimax Unlearning with Generalization Rates and Deletion Capacity
Jiaqi Liu, Jian Lou, Zhan Qin, Kui Ren
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
An Empirical Study Towards Prompt-Tuning for Graph Contrastive Pre-Training in Recommendations
Haoran Yang, Xiangyu Zhao, Yicong Li, Hongxu Chen, Guandong Xu
Graph contrastive learning (GCL) has emerged as a potent technology for numerous graph learning tasks. It has been successfully applied to real-world recommender systems, where the contrastive loss and the downstream recommendation objectives are always combined to form the overall objective function. Such a strategy is inconsistent with the original GCL paradigm, where graph embeddings are pre-trained without involving downstream training objectives. In this paper, we innovatively propose a prompt-enhanced framework for GCL-based recommender systems, namely CPTPP, which can fully leverage the advantages of the original GCL protocol through prompt tuning. Specifically, we first summarise user profiles in graph recommender systems to automatically generate personalized user prompts. These prompts will then be combined with pre-trained user embeddings to conduct prompt-tuning in downstream tasks, thereby narrowing the distinct targets between pre-training and downstream tasks. Extensive experiments on three benchmark datasets validate the effectiveness of CPTPP against state-of-the-art baselines. A further visualization experiment demonstrates that user embeddings generated by CPTPP have a more uniform distribution, indicating a better capacity to model the diversity of user preferences.The implementation code is available online to ease reproducibility: https://anonymous.4open.science/r/CPTPP-F8F4
****
Can Language Models Teach? Teacher Explanations Improve Student Performance via Personalization
Swarnadeep Saha, Peter Hase, Mohit Bansal
A hallmark property of explainable AI models is the ability to teach other agents, communicating knowledge of how to perform a task. While Large Language Models (LLMs) perform complex reasoning by generating explanations for their predictions, it is unclear whether they also make good teachers for weaker agents. To address this, we consider a student-teacher framework between two LLM agents and study if, when, and how the teacher should intervene with natural language explanations to improve the student’s performance. Since communication is expensive, we define a budget such that the teacher only communicates explanations for a fraction of the data, after which the student should perform well on its own. We decompose the teaching problem along four axes: (1) if teacher’s test time in- tervention improve student predictions, (2) when it is worth explaining a data point, (3) how the teacher should personalize explanations to better teach the student, and (4) if teacher explanations also improve student performance on future unexplained data. We first show that teacher LLMs can indeed intervene on student reasoning to improve their performance. Next, inspired by the Theory of Mind abilities of effective teachers, we propose building two few-shot mental models of the student. The first model defines an Intervention Function that simulates the utility of an intervention, allowing the teacher to intervene when this utility is the highest and improving student performance at lower budgets. The second m

odel enables the teacher to personalize explanations for a particular student and outperform unpersonalized teachers. We also demonstrate that in multi-turn interactions, teacher explanations generalize and learning from explained data improves student performance on future unexplained data. Finally, we also verify that misaligned teachers can lower student performance to random chance by intentionally misleading them.
****
Finding Local Minima Efficiently in Decentralized Optimization
Wenhan Xian, Heng Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Clifford Group Equivariant Neural Networks
David Ruhe, Johannes Brandstetter, Patrick Forré
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
C-Eval: A Multi-Level Multi-Discipline Chinese Evaluation Suite for Foundation Models
Yuzhen Huang, Yuzhuo Bai, Zhihao Zhu, Junlei Zhang, Jinghan Zhang, Tangjun Su, Junteng Liu, Chuancheng Lv, Yikai Zhang, jiayi lei, Yao Fu, Maosong Sun, Junxian He
New NLP benchmarks are urgently needed to align with the rapid development of large language models (LLMs). We present C-Eval, the first comprehensive Chinese evaluation suite designed to assess advanced knowledge and reasoning abilities of foundation models in a Chinese context. C-Eval comprises multiple-choice questions across four difficulty levels: middle school, high school, college, and professional. The questions span 52 diverse disciplines, ranging from humanities to science and engineering. C-Eval is accompanied by C-Eval Hard, a subset of very challenging subjects in C-Eval that requires advanced reasoning abilities to solve. We conduct a comprehensive evaluation of the most advanced LLMs on C-Eval, including both English- and Chinese-oriented models. Results indicate that only GPT-4 could achieve an average accuracy of over 60%, suggesting that there is still significant room for improvement for current LLMs. We anticipate C-Eval will help analyze important strengths and shortcomings of foundation models, and foster their development and growth for Chinese users.
****
NU-MCC: Multiview Compressive Coding with Neighborhood Decoder and Repulsive UDF
Stefan Lionar, Xiangyu Xu, Min Lin, Gim Hee Lee
Remarkable progress has been made in 3D reconstruction from single-view RGB-D inputs. MCC is the current state-of-the-art method in this field, which achieves unprecedented success by combining vision Transformers with large-scale training. However, we identified two key limitations of MCC: 1) The Transformer decoder is inefficient in handling large number of query points; 2) The 3D representation struggles to recover high-fidelity details. In this paper, we propose a new approach called NU-MCC that addresses these limitations. NU-MCC includes two key innovations: a Neighborhood decoder and a Repulsive Unsigned Distance Function (Repulsive UDF). First, our Neighborhood decoder introduces center points as an efficient proxy of input visual features, allowing each query point to only attend to a small neighborhood. This design not only results in much faster inference speed but also enables the exploitation of finer-scale visual features for improved recovery of 3D textures. Second, our Repulsive UDF is a novel alternative to the occupancy field used in MCC, significantly improving the quality of 3D object reconstruction. Compared to standard UDFs that suffer from holes in results, our proposed Repulsive UDF can achieve more complete surface reconstruction. Experimental results demonstrate that NU-MCC is able to learn a strong 3D representa

tion, significantly advancing the state of the art in single-view 3D reconstruct
ion. Particularly, it outperforms MCC by 9.7% in terms of the F1-score on the CO
3D-v2 dataset with more than 5x faster running speed.
****
Convergence analysis of ODE models for accelerated first-order methods via posit
ive semidefinite kernels
Jungbin Kim, Insoon Yang
We propose a novel methodology that systematically analyzes ordinary differentia
l equation (ODE) models for first-order optimization methods by converting the t
ask of proving convergence rates into verifying the positive semidefiniteness of
 specific Hilbert-Schmidt integral operators. Our approach is based on the perfo
rmance estimation problems (PEP) introduced by Drori and Teboulle. Unlike previo
us works on PEP, which rely on finite-dimensional linear algebra, we use tools f
rom functional analysis. Using the proposed method, we establish convergence rat
es of various accelerated gradient flow models, some of which are new. As an imm
ediate consequence of our framework, we show a correspondence between minimizing
 function values and minimizing gradient norms.
****
Curvature Filtrations for Graph Generative Model Evaluation
Joshua Southern, Jeremy Wayland, Michael Bronstein, Bastian Rieck
Graph generative model evaluation necessitates understanding differences between
 graphs on the distributional level. This entails being able to harness salient
attributes of graphs in an efficient manner. Curvature constitutes one such prop
erty of graphs, and has recently started to prove useful in characterising graph
s. Its expressive properties, stability, and practical utility in model evaluati
on remain largely unexplored, however. We combine graph curvature descriptors wi
th emerging methods from topological data analysis to obtain robust, expressive
descriptors for evaluating graph generative models.
****
DiffUTE: Universal Text Editing Diffusion Model
Haoxing Chen, Zhuoer Xu, Zhangxuan Gu, jun lan, ■ ■, Yaohui Li, Changhua Meng, H
uijia Zhu, Weiqiang Wang
Diffusion model based language-guided image editing has achieved great success r
ecently. However, existing state-of-the-art diffusion models struggle with rende
ring correct text and text style during generation. To tackle this problem, we p
ropose a universal self-supervised text editing diffusion model (DiffUTE), which
 aims to replace or modify words in the source image with another one while main
taining its realistic appearance. Specifically, we build our model on a diffusio
n model and carefully modify the network structure to enable the model for drawi
ng multilingual characters with the help of glyph and position information. More
over, we design a self-supervised learning framework to leverage large amounts o
f web data to improve the representation ability of the model. Experimental resu
lts show that our method achieves an impressive performance and enables controll
able editing on in-the-wild images with high fidelity. Our code will be avaliabl
e in \url{https://github.com/chenhaoxing/DiffUTE}.
****
Sampling weights of deep neural networks
Erik L Bolager, Iryna Burak, Chinmay Datar, Qing Sun, Felix Dietrich
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Fast Attention Requires Bounded Entries
Josh Alman, Zhao Song
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****

Open Compound Domain Adaptation with Object Style Compensation for Semantic Segmentation
Tingliang Feng, Hao Shi, Xueyang Liu, Wei Feng, Liang Wan, Yanlin Zhou, Di Lin
Many methods of semantic image segmentation have borrowed the success of open compound domain adaptation. They minimize the style gap between the images of source and target domains, more easily predicting the accurate pseudo annotations for target domain's images that train segmentation network. The existing methods globally adapt the scene style of the images, whereas the object styles of different categories or instances are adapted improperly. This paper proposes the Object Style Compensation, where we construct the Object-Level Discrepancy Memory with multiple sets of discrepancy features. The discrepancy features in a set capture the style changes of the same category's object instances adapted from target to source domains. We learn the discrepancy features from the images of source and target domains, storing the discrepancy features in memory. With this memory, we select appropriate discrepancy features for compensating the style information of the object instances of various categories, adapting the object styles to a unified style of source domain. Our method enables a more accurate computation of the pseudo annotations for target domain's images, thus yielding state-of-the-art results on different datasets.
****
Going beyond persistent homology using persistent homology
Johanna Immonen, Amauri Souza, Vikas Garg
Representational limits of message-passing graph neural networks (MP-GNNs), e.g., in terms of the Weisfeiler-Leman (WL) test for isomorphism, are well understood. Augmenting these graph models with topological features via persistent homology (PH) has gained prominence, but identifying the class of attributed graphs that PH can recognize remains open. We introduce a novel concept of color-separating sets to provide a complete resolution to this important problem. Specifically, we establish the necessary and sufficient conditions for distinguishing graphs based on the persistence of their connected components, obtained from filter functions on vertex and edge colors. Our constructions expose the limits of vertex- and edge-level PH, proving that neither category subsumes the other. Leveraging these theoretical insights, we propose RePHINE for learning topological features on graphs. RePHINE efficiently combines vertex- and edge-level PH, achieving a scheme that is provably more powerful than both. Integrating RePHINE into MP-GNNs boosts their expressive power, resulting in gains over standard PH on several benchmarks for graph classification.
****
Explore to Generalize in Zero-Shot RL
Ev Zisselman, Itai Lavie, Daniel Soudry, Aviv Tamar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CoLLAT: On Adding Fine-grained Audio Understanding to Language Models using Token-Level Locked-Language Tuning
Dadallage A R Silva, Spencer Whitehead, Christopher Lengerich, Hugh Leather
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Abide by the law and follow the flow: conservation laws for gradient flows
Sibylle Marcotte, Remi Gribonval, Gabriel Peyré
Understanding the geometric properties of gradient descent dynamics is a key ingredient in deciphering the recent success of very large machine learning models. A striking observation is that trained over-parameterized models retain some properties of the optimization initialization. This "implicit bias" is believed to be responsible for some favorable properties of the trained models and could ex

plain their good generalization properties. The purpose of this article is three fold. First, we rigorously expose the definition and basic properties of "conser vation laws", that define quantities conserved during gradient flows of a given model (e.g. of a ReLU network with a given architecture) with any training data and any loss. Then we explain how to find the maximal number of independent cons ervation lawsby performing finite-dimensional algebraic manipulations on the Lie algebra generated by the Jacobian of the model. Finally, we provide algorithms to: a) compute a family of polynomial laws; b) compute the maximal number of (no t necessarily polynomial) independent conservation laws. We provide showcase exa mples that we fully work out theoretically. Besides, applying the two algorithms confirms for a number of ReLU network architectures that all known laws are rec overed by the algorithm, and that there are no other independent laws. Such comp utational tools pave the way to understanding desirable properties of optimizati on initialization in large machine learning models.
****
Breadcrumbs to the Goal: Goal-Conditioned Exploration from Human-in-the-Loop Fee dback
Marcel Torne Villasevil, Max Balsells I Pamies, Zihan Wang, Samedh Desai, Tao Ch en, Pulkit Agrawal, Abhishek Gupta
Exploration and reward specification are fundamental and intertwined challenges for reinforcement learning. Solving sequential decision making tasks with a non-trivial element of exploration requires either specifying carefully designed rew ard functions or relying on indiscriminate, novelty seeking exploration bonuses. Human supervisors can provide effective guidance in the loop to direct the expl oration process, but prior methods to leverage this guidance require constant sy nchronous high-quality human feedback, which is expensive and impractical to obt ain. In this work, we propose a technique - Human Guided Exploration (HUGE), tha t is able to leverage low-quality feedback from non-expert users, which is infre quent, asynchronous and noisy, to guide exploration for reinforcement learning, without requiring careful reward specification. The key idea is to separate the challenges of directed exploration and policy learning - human feedback is used to direct exploration, while self-supervised policy learning is used to independ ently learn unbiased behaviors from the collected data. We show that this proced ure can leverage noisy, asynchronous human feedback to learn tasks with no hand-crafted reward design or exploration bonuses. We show that HUGE is able to learn a variety of challenging multi-stage robotic navigation and manipulation tasks in simulation using crowdsourced feedback from non-expert users. Moreover, this paradigm can be scaled to learning directly on real-world robots.
****
Embroid: Unsupervised Prediction Smoothing Can Improve Few-Shot Classification
Neel Guha, Mayee Chen, Kush Bhatia, Azalia Mirhoseini, Frederic Sala, Christophe r Ré
Recent work has shown that language models' (LMs) prompt-based learning capabili ties make them well suited for automating data labeling in domains where manual annotation is expensive. The challenge is that while writing an initial prompt i s cheap, improving a prompt is costly---practitioners often require significant labeled data in order to evaluate the impact of prompt modifications. Our work a sks whether it is possible to improve prompt-based learning without additional l abeled data. We approach this problem by attempting to modify the predictions of a prompt, rather than the prompt itself. Our intuition is that accurate predict ions should also be consistent: samples which are similar under some feature rep resentation should receive the same prompt prediction. We propose Embroid, a met hod which computes multiple representations of a dataset under different embeddi ng functions, and uses the consistency between the LM predictions for neighborin g samples to identify mispredictions. Embroid then uses these neighborhoods to c reate additional predictions for each sample, and combines these predictions wit h a simple latent variable graphical model in order to generate a final correcte d prediction. In addition to providing a theoretical analysis of Embroid, we con duct a rigorous empirical evaluation across six different LMs and up to 95 diffe rent tasks. We find that (1) Embroid substantially improves performance over ori

ginal prompts (e.g., by an average of 7.3 points on GPT-JT), (2) also realizes improvements for more sophisticated prompting strategies (e.g., chain-of-thought), and (3) can be specialized to domains like law through the embedding functions.

****

Perceptual Kalman Filters: Online State Estimation under a Perfect Perceptual-Quality Constraint

Dror Freirich, Tomer Michaeli, Ron Meir

Many practical settings call for the reconstruction of temporal signals from corrupted or missing data. Classic examples include decoding, tracking, signal enhancement and denoising. Since the reconstructed signals are ultimately viewed by humans, it is desirable to achieve reconstructions that are pleasing to human perception.Mathematically, perfect perceptual-quality is achieved when the distribution of restored signals is the same as that of natural signals, a requirement which has been heavily researched in static estimation settings (i.e. when a whole signal is processed at once). Here, we study the problem of optimal causal filtering under a perfect perceptual-quality constraint, which is a task of fundamentally different nature. Specifically, we analyze a Gaussian Markov signal observed through a linear noisy transformation. In the absence of perceptual constraints, the Kalman filter is known to be optimal in the MSE sense for this setting. Here, we show that adding the perfect perceptual quality constraint (i.e. the requirement of temporal consistency), introduces a fundamental dilemma whereby the filter may have to ``knowingly'' ignore new information revealed by the observations in order to conform to its past decisions. This often comes at the cost of a significant increase in the MSE (beyond that encountered in static settings). Our analysis goes beyond the classic innovation process of the Kalman filter, and introduces the novel concept of an unutilized information process. Using this tool, we present a recursive formula for perceptual filters, and demonstrate the qualitative effects of perfect perceptual-quality estimation on a video reconstruction problem.

****

SEENN: Towards Temporal Spiking Early Exit Neural Networks

Yuhang Li, Tamar Geller, Youngeun Kim, Priyadarshini Panda

Spiking Neural Networks (SNNs) have recently become more popular as a biologically plausible substitute for traditional Artificial Neural Networks (ANNs). SNNs are cost-efficient and deployment-friendly because they process input in both spatial and temporal manner using binary spikes. However, we observe that the information capacity in SNNs is affected by the number of timesteps, leading to an accuracy-efficiency tradeoff. In this work, we study a fine-grained adjustment of the number of timesteps in SNNs. Specifically, we treat the number of timesteps as a variable conditioned on different input samples to reduce redundant timesteps for certain data. We call our method Spiking Early-Exit Neural Networks (SEENNs). To determine the appropriate number of timesteps, we propose SEENN-I which uses a confidence score thresholding to filter out the uncertain predictions, and SEENN-II which determines the number of timesteps by reinforcement learning. Moreover, we demonstrate that SEENN is compatible with both the directly trained SNN and the ANN-SNN conversion. By dynamically adjusting the number of timesteps, our SEENN achieves a remarkable reduction in the average number of timesteps during inference. For example, our SEENN-II ResNet-19 can achieve 96.1\% accuracy with an average of 1.08 timesteps on the CIFAR-10 test dataset. Code is shared at https://github.com/Intelligent-Computing-Lab-Yale/SEENN.

****

Distributionally Robust Skeleton Learning of Discrete Bayesian Networks

Yeshu Li, Brian Ziebart

We consider the problem of learning the exact skeleton of general discrete Bayesian networks from potentially corrupted data. Building on distributionally robust optimization and a regression approach, we propose to optimize the most adverse risk over a family of distributions within bounded Wasserstein distance or KL divergence to the empirical distribution. The worst-case risk accounts for the effect of outliers. The proposed approach applies for general categorical random

variables without assuming faithfulness, an ordinal relationship or a specific form of conditional distribution. We present efficient algorithms and show the proposed methods are closely related to the standard regularized regression approach. Under mild assumptions, we derive non-asymptotic guarantees for successful structure learning with logarithmic sample complexities for bounded-degree graphs. Numerical study on synthetic and real datasets validates the effectiveness of our method.
****
Test-Time Amendment with a Coarse Classifier for Fine-Grained Classification
Kanishk Jain, Shyamgopal Karthik, Vineet Gandhi
We investigate the problem of reducing mistake severity for fine-grained classification. Fine-grained classification can be challenging, mainly due to the requirement of knowledge or domain expertise for accurate annotation. However, humans are particularly adept at performing coarse classification as it requires relatively low levels of expertise. To this end, we present a novel approach for Post-Hoc Correction called Hierarchical Ensembles (HiE) that utilizes label hierarchy to improve the performance of fine-grained classification at test-time using the coarse-grained predictions. By only requiring the parents of leaf nodes, our method significantly reduces avg. mistake severity while improving top-1 accuracy on the iNaturalist-19 and tieredImageNet-H datasets, achieving a new state-of-the-art on both benchmarks. We also investigate the efficacy of our approach in the semi-supervised setting. Our approach brings notable gains in top-1 accuracy while significantly decreasing the severity of mistakes as training data decreases for the fine-grained classes. The simplicity and post-hoc nature of HiE renders it practical to be used with any off-the-shelf trained model to improve its predictions further.
****
Robust Matrix Sensing in the Semi-Random Model
Xing Gao, Yu Cheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Implicit variance regularization in non-contrastive SSL
Manu Srinath Halvagal, Axel Laborieux, Friedemann Zenke
Non-contrastive SSL methods like BYOL and SimSiam rely on asymmetric predictor networks to avoid representational collapse without negative samples. Yet, how predictor networks facilitate stable learning is not fully understood. While previous theoretical analyses assumed Euclidean losses, most practical implementations rely on cosine similarity. To gain further theoretical insight into non-contrastive SSL, we analytically study learning dynamics in conjunction with Euclidean and cosine similarity in the eigenspace of closed-form linear predictor networks. We show that both avoid collapse through implicit variance regularization albeit through different dynamical mechanisms. Moreover, we find that the eigenvalues act as effective learning rate multipliers and propose a family of isotropic loss functions (IsoLoss) that equalize convergence rates across eigenmodes. Empirically, IsoLoss speeds up the initial learning dynamics and increases robustness, thereby allowing us to dispense with the EMA target network typically used with non-contrastive methods. Our analysis sheds light on the variance regularization mechanisms of non-contrastive SSL and lays the theoretical grounds for crafting novel loss functions that shape the learning dynamics of the predictor's spectrum.
****
ATMAN: Understanding Transformer Predictions Through Memory Efficient Attention Manipulation
Björn Deiseroth, Mayukh Deb, Samuel Weinbach, Manuel Brack, Patrick Schramowski, Kristian Kersting
Generative transformer models have become increasingly complex, with large numbers of parameters and the ability to process multiple input modalities. Current m

ethods for explaining their predictions are resource-intensive. Most crucially, they require prohibitively large amounts of additional memory, since they rely on backpropagation which allocates almost twice as much GPU memory as the forward pass. This makes it difficult, if not impossible, to use explanations in production. We present AtMan that provides explanations of generative transformer models at almost no extra cost. Specifically, AtMan is a modality-agnostic perturbation method that manipulates the attention mechanisms of transformers to produce relevance maps for the input with respect to the output prediction. Instead of using backpropagation, AtMan applies a parallelizable token-based search method relying on cosine similarity neighborhood in the embedding space. Our exhaustive experiments on text and image-text benchmarks demonstrate that AtMan outperforms current state-of-the-art gradient-based methods on several metrics while being computationally efficient. As such, AtMan is suitable for use in large model inference deployments.
****
Wasserstein Quantum Monte Carlo: A Novel Approach for Solving the Quantum Many-Body Schrödinger Equation
Kirill Neklyudov, Jannes Nys, Luca Thiede, Juan Carrasquilla, Qiang Liu, Max Welling, Alireza Makhzani
Solving the quantum many-body Schrödinger equation is a fundamental and challenging problem in the fields of quantum physics, quantum chemistry, and material sciences. One of the common computational approaches to this problem is Quantum Variational Monte Carlo (QVMC), in which ground-state solutions are obtained by minimizing the energy of the system within a restricted family of parameterized wave functions. Deep learning methods partially address the limitations of traditional QVMC by representing a rich family of wave functions in terms of neural networks. However, the optimization objective in QVMC remains notoriously hard to minimize and requires second-order optimization methods such as natural gradient. In this paper, we first reformulate energy functional minimization in the space of Born distributions corresponding to particle-permutation (anti-)symmetric wave functions, rather than the space of wave functions. We then interpret QVMC as the Fisher--Rao gradient flow in this distributional space, followed by a projection step onto the variational manifold. This perspective provides us with a principled framework to derive new QMC algorithms, by endowing the distributional space with better metrics, and following the projected gradient flow induced by those metrics. More specifically, we propose "Wasserstein Quantum Monte Carlo" (WQMC), which uses the gradient flow induced by the Wasserstein metric, rather than the Fisher--Rao metric, and corresponds to transporting the probability mass, rather than teleporting it. We demonstrate empirically that the dynamics of WQMC results in faster convergence to the ground state of molecular systems.
****
Textually Pretrained Speech Language Models
Michael Hassid, Tal Remez, Tu Anh Nguyen, Itai Gat, Alexis CONNEAU, Felix Kreuk, Jade Copet, Alexandre Defossez, Gabriel Synnaeve, Emmanuel Dupoux, Roy Schwartz, Yossi Adi
Speech language models (SpeechLMs) process and generate acoustic data only, without textual supervision. In this work, we propose TWIST, a method for training SpeechLMs using a warm-start from a pretrained textual language models. We show using both automatic and human evaluations that TWIST outperforms a cold-start SpeechLM across the board. We empirically analyze the effect of different model design choices such as the speech tokenizer, the pretrained textual model, and the dataset size. We find that model and dataset scale both play an important role in constructing better-performing SpeechLMs. Based on our observations, we present the largest (to the best of our knowledge) SpeechLM both in terms of number of parameters and training data. We additionally introduce two spoken versions of the StoryCloze textual benchmark to further improve model evaluation and advance future research in the field. We make speech samples, code and models publicly available.
****
Riemannian Residual Neural Networks

Isay Katsman, Eric Chen, Sidhanth Holalkere, Anna Asch, Aaron Lou, Ser Nam Lim, Christopher M. De Sa
Recent methods in geometric deep learning have introduced various neural networks to operate over data that lie on Riemannian manifolds. Such networks are often necessary to learn well over graphs with a hierarchical structure or to learn over manifold-valued data encountered in the natural sciences. These networks are often inspired by and directly generalize standard Euclidean neural networks. However, extending Euclidean networks is difficult and has only been done for a select few manifolds. In this work, we examine the residual neural network (ResNet) and show how to extend this construction to general Riemannian manifolds in a geometrically principled manner. Originally introduced to help solve the vanishing gradient problem, ResNets have become ubiquitous in machine learning due to their beneficial learning properties, excellent empirical results, and easy-to-incorporate nature when building varied neural networks. We find that our Riemannian ResNets mirror these desirable properties: when compared to existing manifold neural networks designed to learn over hyperbolic space and the manifold of symmetric positive definite matrices, we outperform both kinds of networks in terms of relevant testing metrics and training dynamics.
****
Aligning Gradient and Hessian for Neural Signed Distance Function
Ruian Wang, Zixiong Wang, Yunxiao Zhang, Shuangmin Chen, Shiqing Xin, Changhe Tu, Wenping Wang
The Signed Distance Function (SDF), as an implicit surface representation, provides a crucial method for reconstructing a watertight surface from unorganized point clouds. The SDF has a fundamental relationship with the principles of surface vector calculus. Given a smooth surface, there exists a thin-shell space in which the SDF is differentiable everywhere such that the gradient of the SDF is an eigenvector of its Hessian matrix, with a corresponding eigenvalue of zero. In this paper, we introduce a method to directly learn the SDF from point clouds in the absence of normals. Our motivation is grounded in a fundamental observation: aligning the gradient and the Hessian of the SDF provides a more efficient mechanism to govern gradient directions. This, in turn, ensures that gradient changes more accurately reflect the true underlying variations in shape. Extensive experimental results demonstrate its ability to accurately recover the underlying shape while effectively suppressing the presence of ghost geometry.
****
Achieving Cross Modal Generalization with Multimodal Unified Representation
Yan Xia, Hai Huang, Jieming Zhu, Zhou Zhao
This paper introduces a novel task called Cross Modal Generalization (CMG), which addresses the challenge of learning a unified discrete representation from paired multimodal data during pre-training. Then in downstream tasks, the model can achieve zero-shot generalization ability in other modalities when only one modal is labeled. Existing approaches in multimodal representation learning focus more on coarse-grained alignment or rely on the assumption that information from different modalities is completely aligned, which is impractical in real-world scenarios. To overcome this limitation, we propose \textbf{Uni-Code}, which contains two key contributions: the Dual Cross-modal Information Disentangling (DCID) module and the Multi-Modal Exponential Moving Average (MM-EMA). These methods facilitate bidirectional supervision between modalities and align semantically equivalent information in a shared discrete latent space, enabling fine-grained unified representation of multimodal sequences. During pre-training, we investigate various modality combinations, including audio-visual, audio-text, and the tri-modal combination of audio-visual-text. Extensive experiments on various downstream tasks, i.e., cross-modal event classification, localization, cross-modal retrieval, query-based video segmentation, and cross-dataset event localization, demonstrate the effectiveness of our proposed methods. The code is available at https://github.com/haihuangcode/CMG.
****
Temperature Balancing, Layer-wise Weight Analysis, and Neural Network Training
Yefan Zhou, TIANYU PANG, Keqin Liu, charles martin, Michael W. Mahoney, Yaoqing

Yang
Regularization in modern machine learning is crucial, and it can take various forms in algorithmic design: training set, model family, error function, regularization terms, and optimizations. In particular, the learning rate, which can be interpreted as a temperature-like parameter within the statistical mechanics of learning, plays a crucial role in neural network training. Indeed, many widely adopted training strategies basically just define the decay of the learning rate over time. This process can be interpreted as decreasing a temperature, using either a global learning rate (for the entire model) or a learning rate that varies for each parameter. This paper proposes TempBalance, a straightforward yet effective layer-wise learning rate method. TempBalance is based on Heavy-Tailed Self-Regularization (HT-SR) Theory, an approach which characterizes the implicit self-regularization of different layers in trained models. We demonstrate the efficacy of using HT-SR-motivated metrics to guide the scheduling and balancing of temperature across all network layers during model training, resulting in improved performance during testing. We implement TempBalance on CIFAR10, CIFAR100, SVHN, and TinyImageNet datasets using ResNets, VGGs and WideResNets with various depths and widths. Our results show that TempBalance significantly outperforms ordinary SGD and carefully-tuned spectral norm regularization. We also show that TempBalance outperforms a number of state-of-the-art optimizers and learning rate schedulers.
****
Gaussian Process Probes (GPP) for Uncertainty-Aware Probing
Zi Wang, Alexander Ku, Jason Baldridge, Tom Griffiths, Been Kim
Understanding which concepts models can and cannot represent has been fundamental to many tasks: from effective and responsible use of models to detecting out of distribution data. We introduce Gaussian process probes (GPP), a unified and simple framework for probing and measuring uncertainty about concepts represented by models. As a Bayesian extension of linear probing methods, GPP asks what kind of distribution over classifiers (of concepts) is induced by the model. This distribution can be used to measure both what the model represents and how confident the probe is about what the model represents. GPP can be applied to any pre-trained model with vector representations of inputs (e.g., activations). It does not require access to training data, gradients, or the architecture. We validate GPP on datasets containing both synthetic and real images. Our experiments show it can (1) probe a model's representations of concepts even with a very small number of examples, (2) accurately measure both epistemic uncertainty (how confident the probe is) and aleatory uncertainty (how fuzzy the concepts are to the model), and (3) detect out of distribution data using those uncertainty measures as well as classic methods do. By using Gaussian processes to expand what probing can offer, GPP provides a data-efficient, versatile and uncertainty-aware tool for understanding and evaluating the capabilities of machine learning models.
****
Inferring Hybrid Neural Fluid Fields from Videos
Hong-Xing Yu, Yang Zheng, Yuan Gao, Yitong Deng, Bo Zhu, Jiajun Wu
We study recovering fluid density and velocity from sparse multiview videos. Existing neural dynamic reconstruction methods predominantly rely on optical flows; therefore, they cannot accurately estimate the density and uncover the underlying velocity due to the inherent visual ambiguities of fluid velocity, as fluids are often shapeless and lack stable visual features. The challenge is further pronounced by the turbulent nature of fluid flows, which calls for properly designed fluid velocity representations. To address these challenges, we propose hybrid neural fluid fields (HyFluid), a neural approach to jointly infer fluid density and velocity fields. Specifically, to deal with visual ambiguities of fluid velocity, we introduce a set of physics-based losses that enforce inferring a physically plausible velocity field, which is divergence-free and drives the transport of density. To deal with the turbulent nature of fluid velocity, we design a hybrid neural velocity representation that includes a base neural velocity field that captures most irrotational energy and a vortex particle-based velocity that models residual turbulent velocity. We show that our method enables recovering

vortical flow details. Our approach opens up possibilities for various learning and reconstruction applications centered around 3D incompressible flow, including fluid re-simulation and editing, future prediction, and neural dynamic scene composition. Project website: https://kovenyu.com/HyFluid/
****
MeGraph: Capturing Long-Range Interactions by Alternating Local and Hierarchical Aggregation on Multi-Scaled Graph Hierarchy
Honghua Dong, Jiawei Xu, Yu Yang, Rui Zhao, Shiwen Wu, Chun Yuan, Xiu Li, Chris J. Maddison, Lei Han
Graph neural networks, which typically exchange information between local neighbors, often struggle to capture long-range interactions (LRIs) within the graph. Building a graph hierarchy via graph pooling methods is a promising approach to address this challenge; however, hierarchical information propagation cannot entirely take over the role of local information aggregation. To balance locality and hierarchy, we integrate the local and hierarchical structures, represented by intra- and inter-graphs respectively, of a multi-scale graph hierarchy into a single mega graph. Our proposed MeGraph model consists of multiple layers alternating between local and hierarchical information aggregation on the mega graph. Each layer first performs local-aware message-passing on graphs of varied scales via the intra-graph edges, then fuses information across the entire hierarchy along the bidirectional pathways formed by inter-graph edges. By repeating this fusion process, local and hierarchical information could intertwine and complement each other. To evaluate our model, we establish a new Graph Theory Benchmark designed to assess LRI capture ability, in which MeGraph demonstrates dominant performance. Furthermore, MeGraph exhibits superior or equivalent performance to state-of-the-art models on the Long Range Graph Benchmark. The experimental results on commonly adopted real-world datasets further demonstrate the broad applicability of MeGraph.
****
Double and Single Descent in Causal Inference with an Application to High-Dimensional Synthetic Control
Jann Spiess, guido imbens, Amar Venugopal
Motivated by a recent literature on the double-descent phenomenon in machine learning, we consider highly over-parameterized models in causal inference, including synthetic control with many control units. In such models, there may be so many free parameters that the model fits the training data perfectly. We first investigate high-dimensional linear regression for imputing wage data and estimating average treatment effects, where we find that models with many more covariates than sample size can outperform simple ones. We then document the performance of high-dimensional synthetic control estimators with many control units. We find that adding control units can help improve imputation performance even beyond the point where the pre-treatment fit is perfect. We provide a unified theoretical perspective on the performance of these high-dimensional models. Specifically, we show that more complex models can be interpreted as model-averaging estimators over simpler ones, which we link to an improvement in average performance. This perspective yields concrete insights into the use of synthetic control when control units are many relative to the number of pre-treatment periods.
****
IPMix: Label-Preserving Data Augmentation Method for Training Robust Classifiers
Zhenglin Huang, Xiaoan Bao, Na Zhang, Qingqi Zhang, Xiao Tu, Biao Wu, Xi Yang
Data augmentation has been proven effective for training high-accuracy convolutional neural network classifiers by preventing overfitting. However, building deep neural networks in real-world scenarios requires not only high accuracy on clean data but also robustness when data distributions shift. While prior methods have proposed that there is a trade-off between accuracy and robustness, we propose IPMix, a simple data augmentation approach to improve robustness without hurting clean accuracy. IPMix integrates three levels of data augmentation (image-level, patch-level, and pixel-level) into a coherent and label-preserving technique to increase the diversity of training data with limited computational overhead. To further improve the robustness, IPMix introduces structural complexity at d

ifferent levels to generate more diverse images and adopts the random mixing method for multi-scale information fusion. Experiments demonstrate that IPMix outperforms state-of-the-art corruption robustness on CIFAR-C and ImageNet-C. In addition, we show that IPMix also significantly improves the other safety measures, including robustness to adversarial perturbations, calibration, prediction consistency, and anomaly detection, achieving state-of-the-art or comparable results on several benchmarks, including ImageNet-R, ImageNet-A, and ImageNet-O.
****

Discovering Hierarchical Achievements in Reinforcement Learning via Contrastive Learning

Seungyong Moon, Junyoung Yeom, Bumsoo Park, Hyun Oh Song

Discovering achievements with a hierarchical structure in procedurally generated environments presents a significant challenge.This requires an agent to possess a broad range of abilities, including generalization and long-term reasoning. Many prior methods have been built upon model-based or hierarchical approaches, with the belief that an explicit module for long-term planning would be advantageous for learning hierarchical dependencies. However, these methods demand an excessive number of environment interactions or large model sizes, limiting their practicality. In this work, we demonstrate that proximal policy optimization (PPO), a simple yet versatile model-free algorithm, outperforms previous methods when optimized with recent implementation practices. Moreover, we find that the PPO agent can predict the next achievement to be unlocked to some extent, albeit with limited confidence. Based on this observation, we introduce a novel contrastive learning method, called achievement distillation, which strengthens the agent's ability to predict the next achievement. Our method exhibits a strong capacity for discovering hierarchical achievements and shows state-of-the-art performance on the challenging Crafter environment in a sample-efficient manner while utilizing fewer model parameters.
****

VisoGender: A dataset for benchmarking gender bias in image-text pronoun resolution

Siobhan Mackenzie Hall, Fernanda Gonçalves Abrantes, Hanwen Zhu, Grace Sodunke, Aleksandar Shtedritski, Hannah Rose Kirk

We introduce VisoGender, a novel dataset for benchmarking gender bias in vision-language models. We focus on occupation-related biases within a hegemonic system of binary gender, inspired by Winograd and Winogender schemas, where each image is associated with a caption containing a pronoun relationship of subjects and objects in the scene. VisoGender is balanced by gender representation in professional roles, supporting bias evaluation in two ways: i) resolution bias, where we evaluate the difference between pronoun resolution accuracies for image subjects with gender presentations perceived as masculine versus feminine by human annotators and ii) retrieval bias, where we compare ratios of professionals perceived to have masculine and feminine gender presentations retrieved for a gender-neutral search query. We benchmark several state-of-the-art vision-language models and find that they demonstrate bias in resolving binary gender in complex scenes. While the direction and magnitude of gender bias depends on the task and the model being evaluated, captioning models are generally less biased than Vision-Language Encoders.
****

Concept Distillation: Leveraging Human-Centered Explanations for Model Improvement

Avani Gupta, Saurabh Saini, P J Narayanan

Humans use abstract concepts for understanding instead of hard features. Recent interpretability research has focused on human-centered concept explanations of neural networks. Concept Activation Vectors (CAVs) estimate a model's sensitivity and possible biases to a given concept. We extend CAVs from post-hoc analysis to ante-hoc training to reduce model bias through fine-tuning using an additional Concept Loss. Concepts are defined on the final layer of the network in the past. We generalize it to intermediate layers, including the last convolution layer. We also introduce Concept Distillation, a method to define rich and effective

concepts using a pre-trained knowledgeable model as the teacher. Our method can sensitize or desensitize a model towards concepts. We show applications of concept-sensitive training to debias several classification problems. We also show a way to induce prior knowledge into a reconstruction problem. We show that concept-sensitive training can improve model interpretability, reduce biases, and induce prior knowledge.
****
Mitigating the Effect of Incidental Correlations on Part-based Learning
Gaurav Bhatt, Deepayan Das, Leonid Sigal, Vineeth N Balasubramanian
Intelligent systems possess a crucial characteristic of breaking complicated problems into smaller reusable components or parts and adjusting to new tasks using these part representations. However, current part-learners encounter difficulties in dealing with incidental correlations resulting from the limited observations of objects that may appear only in specific arrangements or with specific backgrounds. These incidental correlations may have a detrimental impact on the generalization and interpretability of learned part representations. This study asserts that part-based representations could be more interpretable and generalize better with limited data, employing two innovative regularization methods. The first regularization separates foreground and background information's generative process via a unique mixture-of-parts formulation. Structural constraints are imposed on the parts using a weakly-supervised loss, guaranteeing that the mixture-of-parts for foreground and background entails soft, object-agnostic masks. The second regularization assumes the form of a distillation loss, ensuring the invariance of the learned parts to the incidental background correlations. Furthermore, we incorporate sparse and orthogonal constraints to facilitate learning high-quality part representations.By reducing the impact of incidental background correlations on the learned parts, we exhibit state-of-the-art (SoTA) performance on few-shot learning tasks on benchmark datasets, including MiniImagenet, TieredImageNet, and FC100. We also demonstrate that the part-based representations acquired through our approach generalize better than existing techniques, even under domain shifts of the background and common data corruption on the ImageNet-9 dataset.
****
Towards In-context Scene Understanding
Ivana Balazevic, David Steiner, Nikhil Parthasarathy, Relja Arandjelovi■, Olivier Henaff
In-context learning––the ability to configure a model's behavior with different prompts––has revolutionized the field of natural language processing, alleviating the need for task-specific models and paving the way for generalist models capable of assisting with any query. Computer vision, in contrast, has largely stayed in the former regime: specialized decoders and finetuning protocols are generally required to perform dense tasks such as semantic segmentation and depth estimation. In this work we explore a simple mechanism for in-context learning of such scene understanding tasks: nearest neighbor retrieval from a prompt of annotated features. We propose a new pretraining protocol––leveraging attention within and across images––which yields representations particularly useful in this regime. The resulting Hummingbird model, suitably prompted, performs various scene understanding tasks without modification while approaching the performance of specialists that have been finetuned for each task. Moreover, Hummingbird can be configured to perform new tasks much more efficiently than finetuned models, raising the possibility of scene understanding in the interactive assistant regime.
****
Prediction and Control in Continual Reinforcement Learning
Nishanth Anand, Doina Precup
Temporal difference (TD) learning is often used to update the estimate of the value function which is used by RL agents to extract useful policies. In this paper, we focus on value function estimation in continual reinforcement learning. We propose to decompose the value function into two components which update at different timescales: a permanent value function, which holds general knowledge that persists over time, and a transient value function, which allows quick adaptat

ion to new situations. We establish theoretical results showing that our approach is well suited for continual learning and draw connections to the complementary learning systems (CLS) theory from neuroscience. Empirically, this approach improves performance significantly on both prediction and control problems.
****
EDGI: Equivariant Diffusion for Planning with Embodied Agents
Johann Brehmer, Joey Bose, Pim de Haan, Taco S. Cohen
Embodied agents operate in a structured world, often solving tasks with spatial, temporal, and permutation symmetries. Most algorithms for planning and model-based reinforcement learning (MBRL) do not take this rich geometric structure into account, leading to sample inefficiency and poor generalization. We introduce the Equivariant Diffuser for Generating Interactions (EDGI), an algorithm for MBRL and planning that is equivariant with respect to the product of the spatial symmetry group SE(3), the discrete-time translation group ■, and the object permutation group S■. EDGI follows the Diffuser framework by Janner et al. (2022) in treating both learning a world model and planning in it as a conditional generative modeling problem, training a diffusion model on an offline trajectory dataset. We introduce a new SE(3) × ■ × S■-equivariant diffusion model that supports multiple representations. We integrate this model in a planning loop, where conditioning and classifier guidance let us softly break the symmetry for specific tasks as needed. On object manipulation and navigation tasks, EDGI is substantially more sample efficient and generalizes better across the symmetry group than non-equivariant models.
****
Topological RANSAC for instance verification and retrieval without fine-tuning
Guoyuan An, Ju-hyeong Seon, Inkyu An, Yuchi Huo, Sung-eui Yoon
This paper presents an innovative approach to enhancing explainable image retrieval, particularly in situations where a fine-tuning set is unavailable. The widely-used SPatial verification (SP) method, despite its efficacy, relies on a spatial model and the hypothesis-testing strategy for instance recognition, leading to inherent limitations, including the assumption of planar structures and neglect of topological relations among features. To address these shortcomings, we introduce a pioneering technique that replaces the spatial model with a topological one within the RANSAC process. We propose bio-inspired saccade and fovea functions to verify the topological consistency among features, effectively circumventing the issues associated with SP's spatial model. Our experimental results demonstrate that our method significantly outperforms SP, achieving state-of-the-art performance in non-fine-tuning retrieval. Furthermore, our approach can enhance performance when used in conjunction with fine-tuned features. Importantly, our method retains high explainability and is lightweight, offering a practical and adaptable solution for a variety of real-world applications.
****
An Alternating Optimization Method for Bilevel Problems under the Polyak-■ojasiewicz Condition
Quan Xiao, Songtao Lu, Tianyi Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Sub-optimality of the Naive Mean Field approximation for proportional high-dimensional Linear Regression
Jiaze Qiu
The Naïve Mean Field (NMF) approximation is widely employed in modern Machine Learning due to the huge computational gains it bestows on the statistician. Despite its popularity in practice, theoretical guarantees for high-dimensional problems are only available under strong structural assumptions (e.g. sparsity). Moreover, existing theory often does not explain empirical observations noted in the existing literature. In this paper, we take a step towards addressing these problems by deriving sharp asymptotic characterizations for the NMF approximation

in high-dimensional linear regression. Our results apply to a wide class of natural priors and allow for model mismatch (i.e. the underlying statistical model can be different from the fitted model). We work under an iid Gaussian design and the proportional asymptotic regime, where the number of features and number of observations grow at a proportional rate. As a consequence of our asymptotic characterization, we establish two concrete corollaries: (a) we establish the inaccuracy of the NMF approximation for the log-normalizing constant in this regime, and (b) we provide theoretical results backing the empirical observation that the NMF approximation can be overconfident in terms of uncertainty quantification.Our results utilize recent advances in the theory of Gaussian comparison inequalities. To the best of our knowledge, this is the first application of these ideas to the analysis of Bayesian variational inference problems. Our theoretical results are corroborated by numerical experiments. Lastly, we believe our results can be generalized to non-Gaussian designs and provide empirical evidence to support it.
****
The Gain from Ordering in Online Learning
Vasilis Kontonis, Mingchen Ma, Christos Tzamos
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Variational Annealing on Graphs for Combinatorial Optimization
Sebastian Sanokowski, Wilhelm Berghammer, Sepp Hochreiter, Sebastian Lehner
Several recent unsupervised learning methods use probabilistic approaches to solve combinatorial optimization (CO) problems based on the assumption of statistically independent solution variables. We demonstrate that this assumption imposes performance limitations in particular on difficult problem instances. Our results corroborate that an autoregressive approach which captures statistical dependencies among solution variables yields superior performance on many popular CO problems. We introduce Subgraph Tokenization in which the configuration of a set of solution variables is represented by a single token. This tokenization technique alleviates the drawback of the long sequential sampling procedure which is inherent to autoregressive methods without sacrificing expressivity. Importantly, we theoretically motivate an annealed entropy regularization and show empirically that it is essential for efficient and stable learning.
****
Chasing Fairness Under Distribution Shift: A Model Weight Perturbation Approach
Zhimeng (Stephen) Jiang, Xiaotian Han, Hongye Jin, Guanchu Wang, Rui Chen, Na Zou, Xia Hu
Fairness in machine learning has attracted increasing attention in recent years. The fairness methods improving algorithmic fairness for in-distribution data may not perform well under distribution shifts. In this paper, we first theoretically demonstrate the inherent connection between distribution shift, data perturbation, and model weight perturbation.Subsequently, we analyze the sufficient conditions to guarantee fairness (i.e., low demographic parity) for the target dataset, including fairness for the source dataset, and low prediction difference between the source and target datasets for each sensitive attribute group. Motivated by these sufficient conditions, we propose robust fairness regularization (RFR) by considering the worst case within the model weight perturbation ball for each sensitive attribute group. We evaluate the effectiveness of our proposed RFR algorithm on synthetic and real distribution shifts across various datasets. Experimental results demonstrate that RFR achieves better fairness-accuracy trade-off performance compared with several baselines. The source code is available at \url{https://github.com/zhimengj0326/RFR_NeurIPS23}.
****
Fast Scalable and Accurate Discovery of DAGs Using the Best Order Score Search and Grow Shrink Trees
Bryan Andrews, Joseph Ramsey, Ruben Sanchez Romero, Jazmin Camchong, Erich Kumme

rfeld
Learning graphical conditional independence structures is an important machine learning problem and a cornerstone of causal discovery. However, the accuracy and execution time of learning algorithms generally struggle to scale to problems with hundreds of highly connected variables---for instance, recovering brain networks from fMRI data. We introduce the best order score search (BOSS) and grow-shrink trees (GSTs) for learning directed acyclic graphs (DAGs) in this paradigm. BOSS greedily searches over permutations of variables, using GSTs to construct and score DAGs from permutations. GSTs efficiently cache scores to eliminate redundant calculations. BOSS achieves state-of-the-art performance in accuracy and execution time, comparing favorably to a variety of combinatorial and gradient-based learning algorithms under a broad range of conditions. To demonstrate its practicality, we apply BOSS to two sets of resting-state fMRI data: simulated data with pseudo-empirical noise distributions derived from randomized empirical fMRI cortical signals and clinical data from 3T fMRI scans processed into cortical parcels. BOSS is available for use within the TETRAD project which includes Python and R wrappers.
****
Bayesian Active Causal Discovery with Multi-Fidelity Experiments
Zeyu Zhang, Chaozhuo Li, Xu Chen, Xing Xie
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Meta-Learning with Neural Bandit Scheduler
Yunzhe Qi, Yikun Ban, Tianxin Wei, Jiaru Zou, Huaxiu Yao, Jingrui He
Meta-learning has been proven an effective learning paradigm for training machine learning models with good generalization ability. Apart from the common practice of uniformly sampling the meta-training tasks, existing methods working on task scheduling strategies are mainly based on pre-defined sampling protocols or the assumed task-model correlations, and greedily make scheduling decisions, which can lead to sub-optimal performance bottlenecks of the meta-model. In this paper, we propose a novel task scheduling framework under Contextual Bandits settings, named BASS, which directly optimizes the task scheduling strategy based on the status of the meta-model. By balancing the exploitation and exploration in meta-learning task scheduling, BASS can help tackle the challenge of limited knowledge about the task distribution during the early stage of meta-training, while simultaneously exploring potential benefits for forthcoming meta-training iterations through an adaptive exploration strategy. Theoretical analysis and extensive experiments are presented to show the effectiveness of our proposed framework.
****
ClusterFomer: Clustering As A Universal Visual Learner
James Liang, Yiming Cui, Qifan Wang, Tong Geng, Wenguan Wang, Dongfang Liu
This paper presents ClusterFormer, a universal vision model that is based on the Clustering paradigm with TransFormer. It comprises two novel designs: 1) recurrent cross-attention clustering, which reformulates the cross-attention mechanism in Transformer and enables recursive updates of cluster centers to facilitate strong representation learning; and 2) feature dispatching, which uses the updated cluster centers to redistribute image features through similarity-based metrics, resulting in a transparent pipeline. This elegant design streamlines an explainable and transferable workflow, capable of tackling heterogeneous vision tasks (i.e., image classification, object detection, and image segmentation) with varying levels of clustering granularity (i.e., image-, box-, and pixel-level). Empirical results demonstrate that ClusterFormer outperforms various well-known specialized architectures, achieving 83.41% top-1 acc. over ImageNet-1K for image classification, 54.2% and 47.0% mAP over MS COCO for object detection and instance segmentation, 52.4% mIoU over ADE20K for semantic segmentation, and 55.8% PQ over COCO Panoptic for panoptic segmentation. This work aims to initiate a paradigm shift in universal visual understanding and to benefit the broader field.

****

Spike-driven Transformer
Man Yao, JiaKui Hu, Zhaokun Zhou, Li Yuan, Yonghong Tian, Bo Xu, Guoqi Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Dis-inhibitory neuronal circuits can control the sign of synaptic plasticity
Julian Rossbroich, Friedemann Zenke
How neuronal circuits achieve credit assignment remains a central unsolved question in systems neuroscience. Various studies have suggested plausible solutions for back-propagating error signals through multi-layer networks. These purely functionally motivated models assume distinct neuronal compartments to represent local error signals that determine the sign of synaptic plasticity. However, this explicit error modulation is inconsistent with phenomenological plasticity models in which the sign depends primarily on postsynaptic activity. Here we show how a plausible microcircuit model and Hebbian learning rule derived within an adaptive control theory framework can resolve this discrepancy. Assuming errors are encoded in top-down dis-inhibitory synaptic afferents, we show that error-modulated learning emerges naturally at the circuit level when recurrent inhibition explicitly influences Hebbian plasticity. The same learning rule accounts for experimentally observed plasticity in the absence of inhibition and performs comparably to back-propagation of error (BP) on several non-linearly separable benchmarks. Our findings bridge the gap between functional and experimentally observed plasticity rules and make concrete predictions on inhibitory modulation of excitatory plasticity.
****

Accelerated Zeroth-order Method for Non-Smooth Stochastic Convex Optimization Problem with Infinite Variance
Nikita Kornilov, Ohad Shamir, Aleksandr Lobanov, Darina Dvinskikh, Alexander Gasnikov, Innokentiy Shibaev, Eduard Gorbunov, Samuel Horváth
In this paper, we consider non-smooth stochastic convex optimization with two function evaluations per round under infinite noise variance. In the classical setting when noise has finite variance, an optimal algorithm, built upon the batched accelerated gradient method, was proposed in (Gasnikov et. al., 2022). This optimality is defined in terms of iteration and oracle complexity, as well as the maximal admissible level of adversarial noise. However, the assumption of finite variance is burdensome and it might not hold in many practical scenarios. To address this, we demonstrate how to adapt a refined clipped version of the accelerated gradient (Stochastic Similar Triangles) method from (Sadiev et al., 2023) for a two-point zero-order oracle. This adaptation entails extending the batching technique to accommodate infinite variance — a non-trivial task that stands as a distinct contribution of this paper.
****

Generator Identification for Linear SDEs with Additive and Multiplicative Noise
Yuanyuan Wang, Xi Geng, Wei Huang, Biwei Huang, Mingming Gong
In this paper, we present conditions for identifying the generator of a linear stochastic differential equation (SDE) from the distribution of its solution process with a given fixed initial state. These identifiability conditions are crucial in causal inference using linear SDEs as they enable the identification of the post-intervention distributions from its observational distribution. Specifically, we derive a sufficient and necessary condition for identifying the generator of linear SDEs with additive noise, as well as a sufficient condition for identifying the generator of linear SDEs with multiplicative noise. We show that the conditions derived for both types of SDEs are generic. Moreover, we offer geometric interpretations of the derived identifiability conditions to enhance their understanding. To validate our theoretical results, we perform a series of simulations, which support and substantiate the established findings.
****

Post-processing Private Synthetic Data for Improving Utility on Selected Measures
Hao Wang, Shivchander Sudalairaj, John Henning, Kristjan Greenewald, Akash Srivastava
Existing private synthetic data generation algorithms are agnostic to downstream tasks. However, end users may have specific requirements that the synthetic data must satisfy. Failure to meet these requirements could significantly reduce the utility of the data for downstream use. We introduce a post-processing technique that improves the utility of the synthetic data with respect to measures selected by the end user, while preserving strong privacy guarantees and dataset quality. Our technique involves resampling from the synthetic data to filter out samples that do not meet the selected utility measures, using an efficient stochastic first-order algorithm to find optimal resampling weights. Through comprehensive numerical experiments, we demonstrate that our approach consistently improves the utility of synthetic data across multiple benchmark datasets and state-of-the-art synthetic data generation algorithms.
****
A Bayesian Approach To Analysing Training Data Attribution In Deep Learning
Elisa Nguyen, Minjoon Seo, Seong Joon Oh
Training data attribution (TDA) techniques find influential training data for the model's prediction on the test data of interest. They approximate the impact of down- or up-weighting a particular training sample. While conceptually useful, they are hardly applicable to deep models in practice, particularly because of their sensitivity to different model initialisation. In this paper, we introduce a Bayesian perspective on the TDA task, where the learned model is treated as a Bayesian posterior and the TDA estimates as random variables. From this novel viewpoint, we observe that the influence of an individual training sample is often overshadowed by the noise stemming from model initialisation and SGD batch composition. Based on this observation, we argue that TDA can only be reliably used for explaining deep model predictions that are consistently influenced by certain training data, independent of other noise factors. Our experiments demonstrate the rarity of such noise-independent training-test data pairs but confirm their existence. We recommend that future researchers and practitioners trust TDA estimates only in such cases. Further, we find a disagreement between ground truth and estimated TDA distributions and encourage future work to study this gap. Code is provided at https://github.com/ElisaNguyen/bayesian-tda.
****
GPEX, A Framework For Interpreting Artificial Neural Networks
Amir Hossein Hosseini Akbarnejad, Gilbert Bigras, Nilanjan Ray
The analogy between Gaussian processes (GPs) and deep artificial neural networks (ANNs) has received a lot of interest, and has shown promise to unbox the blackbox of deep ANNs. Existing theoretical works put strict assumptions on the ANN (e.g. requiring all intermediate layers to be wide, or using specific activation functions). Accommodating those theoretical assumptions is hard in recent deep architectures, and those theoretical conditions need refinement as new deep architectures emerge. In this paper we derive an evidence lower-bound that encourages the GP's posterior to match the ANN's output without any requirement on the ANN. Using our method we find out that on 5 datasets, only a subset of those theoretical assumptions are sufficient. Indeed, in our experiments we used a normal ResNet-18 or feed-forward backbone with a single wide layer in the end. One limitation of training GPs is the lack of scalability with respect to the number of inducing points. We use novel computational techniques that allow us to train GPs with hundreds of thousands of inducing points and with GPU acceleration. As shown in our experiments, doing so has been essential to get a close match between the GPs and the ANNs on 5 datasets. We implement our method as a publicly available tool called GPEX: https://github.com/amirakbarnejad/gpex. On 5 datasets (4 image datasets, and 1 biological dataset) and ANNs with 2 types of functionality (classifier or attention-mechanism) we were able to find GPs whose outputs closely match those of the corresponding ANNs. After matching the GPs to the ANNs, we used the GPs' kernel functions to explain the ANNs' decisions. We provide more

than 200 explanations (around 30 in the paper and the rest in the supplementary) which are highly interpretable by humans and show the ability of the obtained GPs to unbox the ANNs' decisions.
****
On the Trade-off of Intra-/Inter-class Diversity for Supervised Pre-training
Jieyu Zhang, Bohan Wang, Zhengyu Hu, Pang Wei W. Koh, Alexander J. Ratner
Pre-training datasets are critical for building state-of-the-art machine learning models, motivating rigorous study on their impact on downstream tasks. In this work, we study the impact of the trade-off between the intra-class diversity (the number of samples per class) and the inter-class diversity (the number of classes) of a supervised pre-training dataset. Empirically, we found that with the size of the pre-training dataset fixed, the best downstream performance comes with a balance on the intra-/inter-class diversity. To understand the underlying mechanism, we show theoretically that the downstream performance depends monotonically on both types of diversity. Notably, our theory reveals that the optimal class-to-sample ratio (#classes / #samples per class) is invariant to the size of the pre-training dataset, which motivates an application of predicting the optimal number of pre-training classes. We demonstrate the effectiveness of this application by an improvement of around 2 points on the downstream tasks when using ImageNet as the pre-training dataset.
****
An information-theoretic quantification of the content of communication between brain regions
Marco Celotto, Jan Bím, Alejandro Tlaie, Vito De Feo, Alessandro Toso, Stefan Lemke, Daniel Chicharro, Hamed Nili, Malte Bieler, Ileana Hanganu-Opatz, Tobias Donner, Andrea Brovelli, Stefano Panzeri
Quantifying the amount, content and direction of communication between brain regions is key to understanding brain function. Traditional methods to analyze brain activity based on the Wiener-Granger causality principle quantify the overall information propagated by neural activity between simultaneously recorded brain regions, but do not reveal the information flow about specific features of interest (such as sensory stimuli). Here, we develop a new information theoretic measure termed Feature-specific Information Transfer (FIT), quantifying how much information about a specific feature flows between two regions. FIT merges the Wiener-Granger causality principle with information-content specificity. We first derive FIT and prove analytically its key properties. We then illustrate and test them with simulations of neural activity, demonstrating that FIT identifies, within the total information propagated between regions, the information that is transmitted about specific features. We then analyze three neural datasets obtained with different recording methods, magneto- and electro-encephalography, and spiking activity, to demonstrate the ability of FIT to uncover the content and direction of information flow between brain regions beyond what can be discerned with traditional analytical methods. FIT can improve our understanding of how brain regions communicate by uncovering previously unaddressed feature-specific information flow.
****
Efficient Sampling of Stochastic Differential Equations with Positive Semi-Definite Models
Anant Raj, Umut Simsekli, Alessandro Rudi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
TopoSRL: Topology preserving self-supervised Simplicial Representation Learning
Hiren Madhu, Sundeep Prabhakar Chepuri
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****
Occ3D: A Large-Scale 3D Occupancy Prediction Benchmark for Autonomous Driving
Xiaoyu Tian, Tao Jiang, Longfei Yun, Yucheng Mao, Huitong Yang, Yue Wang, Yilun Wang, Hang Zhao
Robotic perception requires the modeling of both 3D geometry and semantics. Existing methods typically focus on estimating 3D bounding boxes, neglecting finer geometric details and struggling to handle general, out-of-vocabulary objects. 3D occupancy prediction, which estimates the detailed occupancy states and semantics of a scene, is an emerging task to overcome these limitations.To support 3D occupancy prediction, we develop a label generation pipeline that produces dense, visibility-aware labels for any given scene. This pipeline comprises three stages: voxel densification, occlusion reasoning, and image-guided voxel refinement. We establish two benchmarks, derived from the Waymo Open Dataset and the nuScenes Dataset, namely Occ3D-Waymo and Occ3D-nuScenes benchmarks. Furthermore, we provide an extensive analysis of the proposed dataset with various baseline models. Lastly, we propose a new model, dubbed Coarse-to-Fine Occupancy (CTF-Occ) network, which demonstrates superior performance on the Occ3D benchmarks.The code, data, and benchmarks are released at \url{https://tsinghua-mars-lab.github.io/Occ3D/}.
****
ProteinGym: Large-Scale Benchmarks for Protein Fitness Prediction and Design
Pascal Notin, Aaron Kollasch, Daniel Ritter, Lood van Niekerk, Steffanie Paul, Han Spinner, Nathan Rollins, Ada Shaw, Rose Orenbuch, Ruben Weitzman, Jonathan Frazer, Mafalda Dias, Dinko Franceschi, Yarin Gal, Debora Marks
Predicting the effects of mutations in proteins is critical to many applications, from understanding genetic disease to designing novel proteins to address our most pressing challenges in climate, agriculture and healthcare. Despite an increase in machine learning-based protein modeling methods, assessing their effectiveness is problematic due to the use of distinct, often contrived, experimental datasets and variable performance across different protein families. Addressing these challenges requires scale. To that end we introduce ProteinGym v1.0, a large-scale and holistic set of benchmarks specifically designed for protein fitness prediction and design. It encompasses both a broad collection of over 250 standardized deep mutational scanning assays, spanning millions of mutated sequences, as well as curated clinical datasets providing high-quality expert annotations about mutation effects. We devise a robust evaluation framework that combines metrics for both fitness prediction and design, factors in known limitations of the underlying experimental methods, and covers both zero-shot and supervised settings. We report the performance of a diverse set of over 40 high-performing models from various subfields (eg., mutation effects, inverse folding) into a unified benchmark. We open source the corresponding codebase, datasets, MSAs, structures, predictions and develop a user-friendly website that facilitates comparisons across all settings.
****
On the spectral bias of two-layer linear networks
Aditya Vardhan Varre, Maria-Luiza Vladarean, Loucas PILLAUD-VIVIEN, Nicolas Flammarion
This paper studies the behaviour of two-layer fully connected networks with linear activations trained with gradient flow on the square loss. We show how the optimization process carries an implicit bias on the parameters that depends on the scale of its initialization. The main result of the paper is a variational characterization of the loss minimizers retrieved by the gradient flow for a specific initialization shape. This characterization reveals that, in the small scale initialization regime, the linear neural network's hidden layer is biased toward having a low-rank structure. To complement our results, we showcase a hidden mirror flow that tracks the dynamics of the singular values of the weights matrices and describe their time evolution. We support our findings with numerical experiments illustrating the phenomena.
****
GMSF: Global Matching Scene Flow

Yushan Zhang, Johan Edstedt, Bastian Wandt, Per-Erik Forssen, Maria Magnusson, Michael Felsberg
We tackle the task of scene flow estimation from point clouds. Given a source and a target point cloud, the objective is to estimate a translation from each point in the source point cloud to the target, resulting in a 3D motion vector field. Previous dominant scene flow estimation methods require complicated coarse-to-fine or recurrent architectures as a multi-stage refinement. In contrast, we propose a significantly simpler single-scale one-shot global matching to address the problem. Our key finding is that reliable feature similarity between point pairs is essential and sufficient to estimate accurate scene flow. We thus propose to decompose the feature extraction step via a hybrid local-global-cross transformer architecture which is crucial to accurate and robust feature representations. Extensive experiments show that the proposed Global Matching Scene Flow (GMSF) sets a new state-of-the-art on multiple scene flow estimation benchmarks. On FlyingThings3D, with the presence of occlusion points, GMSF reduces the outlier percentage from the previous best performance of 27.4% to 5.6%. On KITTI Scene Flow, without any fine-tuning, our proposed method shows state-of-the-art performance. On the Waymo-Open dataset, the proposed method outperforms previous methods by a large margin. The code is available at https://github.com/ZhangYushan3/GMSF.
****
Efficient Uncertainty Quantification and Reduction for Over-Parameterized Neural Networks
Ziyi Huang, Henry Lam, Haofeng Zhang
Uncertainty quantification (UQ) is important for reliability assessment and enhancement of machine learning models. In deep learning, uncertainties arise not only from data, but also from the training procedure that often injects substantial noises and biases. These hinder the attainment of statistical guarantees and, moreover, impose computational challenges on UQ due to the need for repeated network retraining. Building upon the recent neural tangent kernel theory, we create statistically guaranteed schemes to principally \emph{characterize}, and \emph{remove}, the uncertainty of over-parameterized neural networks with very low computation effort. In particular, our approach, based on what we call a procedural-noise-correcting (PNC) predictor, removes the procedural uncertainty by using only \emph{one} auxiliary network that is trained on a suitably labeled dataset, instead of many retrained networks employed in deep ensembles. Moreover, by combining our PNC predictor with suitable light-computation resampling methods, we build several approaches to construct asymptotically exact-coverage confidence intervals using as low as four trained networks without additional overheads.
****
LuminAIRe: Illumination-Aware Conditional Image Repainting for Lighting-Realistic Generation
Jiajun Tang, Haofeng Zhong, Shuchen Weng, Boxin Shi
We present the ilLumination-Aware conditional Image Repainting (LuminAIRe) task to address the unrealistic lighting effects in recent conditional image repainting (CIR) methods. The environment lighting and 3D geometry conditions are explicitly estimated from given background images and parsing masks using a parametric lighting representation and learning-based priors. These 3D conditions are then converted into illumination images through the proposed physically-based illumination rendering and illumination attention module. With the injection of illumination images, physically-correct lighting information is fed into the lighting-realistic generation process and repainted images with harmonized lighting effects in both foreground and background regions can be acquired, whose superiority over the results of state-of-the-art methods is confirmed through extensive experiments. For facilitating and validating the LuminAIRe task, a new dataset Car-LuminAIRe with lighting annotations and rich appearance variants is collected.
****
A graphon-signal analysis of graph neural networks
Ron Levie
We present an approach for analyzing message passing graph neural networks (MPNN

s) based on an extension of graphon analysis to a so called graphon-signal analy sis. A MPNN is a function that takes a graph and a signal on the graph (a graph-signal) and returns some value. Since the input space of MPNNs is non-Euclidean, i.e., graphs can be of any size and topology, properties such as generalization are less well understood for MPNNs than for Euclidean neural networks. We claim that one important missing ingredient in past work is a meaningful notion of gr aph-signal similarity measure, that endows the space of inputs to MPNNs with a r egular structure. We present such a similarity measure, called the graphon-signa l cut distance, which makes the space of all graph-signals a dense subset of a c ompact metric space -- the graphon-signal space. Informally, two deterministic g raph-signals are close in cut-distance if they ``look like'' they were sampled f rom the same random graph-signal model. Hence, our cut distance is a natural not ion of graph-signal similarity, which allows comparing any pair of graph-signals of any size and topology. We prove that MPNNs are Lipschitz continuous function s over the graphon-signal metric space. We then give two applications of this re sult: 1) a generalization bound for MPNNs, and, 2) the stability of MPNNs to sub sampling of graph-signals. Our results apply to any regular enough MPNN on any d istribution of graph-signals, making the analysis rather universal.
****
Lo-Hi: Practical ML Drug Discovery Benchmark
Simon Steshin
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Class-Conditional Conformal Prediction with Many Classes
Tiffany Ding, Anastasios Angelopoulos, Stephen Bates, Michael Jordan, Ryan J. Ti bshirani
Standard conformal prediction methods provide a marginal coverage guarantee,whic h means that for a random test point, the conformal prediction set contains the true label with a user-specified probability. In many classificationproblems, we would like to obtain a stronger guarantee--that for test pointsof a specific cl ass, the prediction set contains the true label with thesame user-chosen probabi lity. For the latter goal, existing conformal predictionmethods do not work well when there is a limited amount of labeled data perclass, as is often the case i n real applications where the number of classes islarge. We propose a method cal led clustered conformal prediction thatclusters together classes having "similar " conformal scores and performs conformal prediction at the cluster level. Based on empirical evaluation acrossfour image data sets with many (up to 1000) class es, we find that clusteredconformal typically outperforms existing methods in te rms of class-conditionalcoverage and set size metrics.
****
Reward Imputation with Sketching for Contextual Batched Bandits
Xiao Zhang, Ninglu Shao, Zihua Si, Jun Xu, Wenhan Wang, Hanjing Su, Ji-Rong Wen
Contextual batched bandit (CBB) is a setting where a batch of rewards is observe d from the environment at the end of each episode, but the rewards of the non-ex ecuted actions are unobserved, resulting in partial-information feedback. Existi ng approaches for CBB often ignore the rewards of the non-executed actions, lead ing to underutilization of feedback information. In this paper, we propose an ef ficient approach called Sketched Policy Updating with Imputed Rewards (SPUIR) th at completes the unobserved rewards using sketching, which approximates the full -information feedbacks. We formulate reward imputation as an imputation regulari zed ridge regression problem that captures the feedback mechanisms of both execu ted and non-executed actions. To reduce time complexity, we solve the regression problem using randomized sketching. We prove that our approach achieves an inst antaneous regret with controllable bias and smaller variance than approaches wit hout reward imputation. Furthermore, our approach enjoys a sublinear regret boun d against the optimal policy. We also present two extensions, a rate-scheduled v ersion and a version for nonlinear rewards, making our approach more practical.

Experimental results show that SPUIR outperforms state-of-the-art baselines on synthetic, public benchmark, and real-world datasets.
****
A Unified Model and Dimension for Interactive Estimation
Nataly Brukhim, Miro Dudik, Aldo Pacchiano, Robert E. Schapire
We study an abstract framework for interactive learning called interactive estimation in which the goal is to estimate a target from its ``similarity'' to points queried by the learner.We introduce a combinatorial measure called Dissimilarity dimension which largely captures learnability in our model.We present a simple, general, and broadly-applicable algorithm, for which we obtain both regret and PAC generalization bounds that are polynomial in the new dimension. We show that our framework subsumes and thereby unifies two classic learning models:statistical-query learning and structured bandits. We also delineate how the Dissimilarity dimension is related to well-known parameters for both frameworks, in some cases yielding significantly improved analyses.
****
Simple, Scalable and Effective Clustering via One-Dimensional Projections
Moses Charikar, Monika Henzinger, Lunjia Hu, Maximilian Vötsch, Erik Waingarten
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Streaming PCA for Markovian Data
Syamantak Kumar, Purnamrita Sarkar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Generalized Logit Adjustment: Calibrating Fine-tuned Models by Removing Label Bias in Foundation Models
Beier Zhu, Kaihua Tang, QIANRU SUN, Hanwang Zhang
Foundation models like CLIP allow zero-shot transfer on various tasks without additional training data. Yet, the zero-shot performance is less competitive than a fully supervised one. Thus, to enhance the performance, fine-tuning and ensembling are also commonly adopted to better fit the downstream tasks. However, we argue that such prior work has overlooked the inherent biases in foundation models. Due to the highly imbalanced Web-scale training set, these foundation models are inevitably skewed toward frequent semantics, and thus the subsequent fine-tuning or ensembling is still biased. In this study, we systematically examine the biases in foundation models and demonstrate the efficacy of our proposed Generalized Logit Adjustment (GLA) method. Note that bias estimation in foundation models is challenging, as most pre-train data cannot be explicitly assessed like in traditional long-tailed classification tasks.To this end, GLA has an optimization-based bias estimation approach for debiasing foundation models. As our work resolves a fundamental flaw in the pre-training, the proposed GLA demonstrates significant improvements across a diverse range of tasks: it achieves 1.5 pp accuracy gains on ImageNet, an large average improvement (1.4-4.6 pp) on 11 few-shot datasets, 2.4 pp gains on long-tailed classification. Codes are in https://github.com/BeierZhu/GLA.
****
Learning to Parameterize Visual Attributes for Open-set Fine-grained Retrieval
Shijie Wang, Jianlong Chang, Haojie Li, Zhihui Wang, Wanli Ouyang, Qi Tian
Open-set fine-grained retrieval is an emerging challenging task that allows to retrieve unknown categories beyond the training set. The best solution for handling unknown categories is to represent them using a set of visual attributes learnt from known categories, as widely used in zero-shot learning. Though important, attribute modeling usually requires significant manual annotations and thus is labor-intensive. Therefore, it is worth to investigate how to transform retriev

al models trained by image-level supervision from category semantic extraction to attribute modeling. To this end, we propose a novel Visual Attribute Parameterization Network (VAPNet) to learn visual attributes from known categories and parameterize them into the retrieval model, without the involvement of any attribute annotations.In this way, VAPNet could utilize its parameters to parse a set of visual attributes from unknown categories and precisely represent them.Technically, VAPNet explicitly attains some semantics with rich details via making use of local image patches and distills the visual attributes from these discovered semantics. Additionally, it integrates the online refinement of these visual attributes into the training process to iteratively enhance their quality. Simultaneously, VAPNet treats these attributes as supervisory signals to tune the retrieval models, thereby achieving attribute parameterization. Extensive experiments on open-set fine-grained retrieval datasets validate the superior performance of our VAPNet over existing solutions.
****
Learning List-Level Domain-Invariant Representations for Ranking
Ruicheng Xian, Honglei Zhuang, Zhen Qin, Hamed Zamani, Jing Lu, Ji Ma, Kai Hui, Han Zhao, Xuanhui Wang, Michael Bendersky
Domain adaptation aims to transfer the knowledge learned on (data-rich) source domains to (low-resource) target domains, and a popular method is invariant representation learning, which matches and aligns the data distributions on the feature space. Although this method is studied extensively and applied on classification and regression problems, its adoption on ranking problems is sporadic, and the few existing implementations lack theoretical justifications. This paper revisits invariant representation learning for ranking. Upon reviewing prior work, we found that they implement what we call item-level alignment, which aligns the distributions of the items being ranked from all lists in aggregate but ignores their list structure. However, the list structure should be leveraged, because it is intrinsic to ranking problems where the data and the metrics are defined and computed on lists, not the items by themselves. To close this discrepancy, we propose list-level alignment—learning domain-invariant representations at the higher level of lists. The benefits are twofold: it leads to the first domain adaptation generalization bound for ranking, in turn providing theoretical support for the proposed method, and it achieves better empirical transfer performance for unsupervised domain adaptation on ranking tasks, including passage reranking.
****
Homotopy-based training of NeuralODEs for accurate dynamics discovery
Joon-Hyuk Ko, Hankyul Koh, Nojun Park, Wonho Jhe
Neural Ordinary Differential Equations (NeuralODEs) present an attractive way to extract dynamical laws from time series data, as they bridge neural networks with the differential equation-based modeling paradigm of the physical sciences. However, these models often display long training times and suboptimal results, especially for longer duration data. While a common strategy in the literature imposes strong constraints to the NeuralODE architecture to inherently promote stable model dynamics, such methods are ill-suited for dynamics discovery as the unknown governing equation is not guaranteed to satisfy the assumed constraints. In this paper, we develop a new training method for NeuralODEs, based on synchronization and homotopy optimization, that does not require changes to the model architecture. We show that synchronizing the model dynamics and the training data tames the originally irregular loss landscape, which homotopy optimization can then leverage to enhance training. Through benchmark experiments, we demonstrate our method achieves competitive or better training loss while often requiring less than half the number of training epochs compared to other model-agnostic techniques. Furthermore, models trained with our method display better extrapolation capabilities, highlighting the effectiveness of our method.
****
SGFormer: Simplifying and Empowering Transformers for Large-Graph Representations
Qitian Wu, Wentao Zhao, Chenxiao Yang, Hengrui Zhang, Fan Nie, Haitian Jiang, Yatao Bian, Junchi Yan

Learning representations on large-sized graphs is a long-standing challenge due to the inter-dependence nature involved in massive data points. Transformers, as an emerging class of foundation encoders for graph-structured data, have shown promising performance on small graphs due to its global attention capable of capturing all-pair influence beyond neighboring nodes. Even so, existing approaches tend to inherit the spirit of Transformers in language and vision tasks, and embrace complicated models by stacking deep multi-head attentions. In this paper, we critically demonstrate that even using a one-layer attention can bring up surprisingly competitive performance across node property prediction benchmarks where node numbers range from thousand-level to billion-level. This encourages us to rethink the design philosophy for Transformers on large graphs, where the global attention is a computation overhead hindering the scalability. We frame the proposed scheme as Simplified Graph Transformers (SGFormer), which is empowered by a simple attention model that can efficiently propagate information among arbitrary nodes in one layer. SGFormer requires none of positional encodings, feature/graph pre-processing or augmented loss. Empirically, SGFormer successfully scales to the web-scale graph ogbn-papers100M and yields up to 141x inference acceleration over SOTA Transformers on medium-sized graphs. Beyond current results, we believe the proposed methodology alone enlightens a new technical path of independent interest for building Transformers on large graphs.
****
Understanding the Limitations of Deep Models for Molecular property prediction: Insights and Solutions
Jun Xia, Lecheng Zhang, Xiao Zhu, Yue Liu, Zhangyang Gao, Bozhen Hu, Cheng Tan, Jiangbin Zheng, Siyuan Li, Stan Z. Li
Molecular Property Prediction (MPP) is a crucial task in the AI-driven Drug Discovery (AIDD) pipeline, which has recently gained considerable attention thanks to advancements in deep learning. However, recent research has revealed that deep models struggle to beat traditional non-deep ones on MPP. In this study, we benchmark 12 representative models (3 non-deep models and 9 deep models) on 15 molecule datasets. Through the most comprehensive study to date, we make the following key observations: \textbf{(\romannumeral 1)} Deep models are generally unable to outperform non-deep ones; \textbf{(\romannumeral 2)} The failure of deep models on MPP cannot be solely attributed to the small size of molecular datasets; \textbf{(\romannumeral 3)} In particular, some traditional models including XGB and RF that use molecular fingerprints as inputs tend to perform better than other competitors. Furthermore, we conduct extensive empirical investigations into the unique patterns of molecule data and inductive biases of various models underlying these phenomena. These findings stimulate us to develop a simple-yet-effective feature mapping method for molecule data prior to feeding them into deep models. Empirically, deep models equipped with this mapping method can beat non-deep ones in most MoleculeNet datasets. Notably, the effectiveness is further corroborated by extensive experiments on cutting-edge dataset related to COVID-19 and activity cliff datasets.
****
Neural Circuits for Fast Poisson Compressed Sensing in the Olfactory Bulb
Jacob Zavatone-Veth, Paul Masset, William Tong, Joseph D. Zak, Venkatesh Murthy, Cengiz Pehlevan
Within a single sniff, the mammalian olfactory system can decode the identity and concentration of odorants wafted on turbulent plumes of air. Yet, it must do so given access only to the noisy, dimensionally-reduced representation of the odor world provided by olfactory receptor neurons. As a result, the olfactory system must solve a compressed sensing problem, relying on the fact that only a handful of the millions of possible odorants are present in a given scene. Inspired by this principle, past works have proposed normative compressed sensing models for olfactory decoding. However, these models have not captured the unique anatomy and physiology of the olfactory bulb, nor have they shown that sensing can be achieved within the 100-millisecond timescale of a single sniff. Here, we propose a rate-based Poisson compressed sensing circuit model for the olfactory bulb. This model maps onto the neuron classes of the olfactory bulb, and recapitulate

s salient features of their connectivity and physiology. For circuit sizes compa rable to the human olfactory bulb, we show that this model can accurately detect tens of odors within the timescale of a single sniff. We also show that this mo del can perform Bayesian posterior sampling for accurate uncertainty estimation. Fast inference is possible only if the geometry of the neural code is chosen to match receptor properties, yielding a distributed neural code that is not axis-aligned to individual odor identities. Our results illustrate how normative mode ling can help us map function onto specific neural circuits to generate new hypo theses.
****
SUPA: A Lightweight Diagnostic Simulator for Machine Learning in Particle Physic s
Atul Kumar Sinha, Daniele Paliotta, Bálint Máté, John Raine, Tobias Golling, Fra nçois Fleuret
Deep learning methods have gained popularity in high energy physics for fast mod eling of particle showers in detectors. Detailed simulation frameworks such as t he gold standard \textsc{Geant4} are computationally intensive, and current deep generative architectures work on discretized, lower resolution versions of the detailed simulation. The development of models that work at higher spatial resol utions is currently hindered by the complexity of the full simulation data, and by the lack of simpler, more interpretable benchmarks. Our contribution is \text sc{SUPA}, the SUrrogate PArticle propagation simulator, an algorithm and softwar e package for generating data by simulating simplified particle propagation, sca ttering and shower development in matter. The generation is extremely fast and e asy to use compared to \textsc{Geant4}, but still exhibits the key characteristi cs and challenges of the detailed simulation. The proposed simulator generates t housands of particle showers per second on a desktop machine, a speed up of up t o 6 orders of magnitudes over \textsc{Geant4}, and stores detailed geometric inf ormation about the shower propagation. \textsc{\textsc{SUPA}} provides much grea ter flexibility for setting initial conditions and defining multiple benchmarks for the development of models. Moreover, interpreting particle showers as point clouds creates a connection to geometric machine learning and provides challengi ng and fundamentally new datasets for the field.
****
LagrangeBench: A Lagrangian Fluid Mechanics Benchmarking Suite
Artur Toshev, Gianluca Galletti, Fabian Fritz, Stefan Adami, Nikolaus Adams
Machine learning has been successfully applied to grid-based PDE modeling in var ious scientific applications. However, learned PDE solvers based on Lagrangian p article discretizations, which are the preferred approach to problems with free surfaces or complex physics, remain largely unexplored. We present LagrangeBench , the first benchmarking suite for Lagrangian particle problems, focusing on tem poral coarse-graining. In particular, our contribution is: (a) seven new fluid m echanics datasets (four in 2D and three in 3D) generated with the Smoothed Parti cle Hydrodynamics (SPH) method including the Taylor-Green vortex, lid-driven cav ity, reverse Poiseuille flow, and dam break, each of which includes different ph ysics like solid wall interactions or free surface, (b) efficient JAX-based API with various recent training strategies and three neighbor search routines, and (c) JAX implementation of established Graph Neural Networks (GNNs) like GNS and SEGNN with baseline results. Finally, to measure the performance of learned surr ogates we go beyond established position errors and introduce physical metrics l ike kinetic energy MSE and Sinkhorn distance for the particle distribution. Our codebase is available under the URL: https://github.com/tumaer/lagrangebench.
****
Group Fairness in Peer Review
Haris Aziz, Evi Micha, Nisarg Shah
Large conferences such as NeurIPS and AAAI serve as crossroads of various AI fi elds, since they attract submissions from a vast number of communities. However, in some cases, this has resulted in a poor reviewing experience for some commun ities, whose submissions get assigned to less qualified reviewers outside of the ir communities. An often-advocated solution is to break up any such large confer

ence into smaller conferences, but this can lead to isolation of communities and harm interdisciplinary research. We tackle this challenge by introducing a notion of group fairness, called the core, which requires that every possible community (subset of researchers) to be treated in a way that prevents them from unilaterally benefiting by withdrawing from a large conference. We study a simple peer review model, prove that it always admits a reviewing assignment in the core, and design an efficient algorithm to find one such assignment. We use real data from CVPR and ICLR conferences to compare our algorithm to existing reviewing assignment algorithms on a number of metrics.
****
Diffusion Model is an Effective Planner and Data Synthesizer for Multi-Task Reinforcement Learning
Haoran He, Chenjia Bai, Kang Xu, Zhuoran Yang, Weinan Zhang, Dong Wang, Bin Zhao, Xuelong Li
Diffusion models have demonstrated highly-expressive generative capabilities in vision and NLP. Recent studies in reinforcement learning (RL) have shown that diffusion models are also powerful in modeling complex policies or trajectories in offline datasets. However, these works have been limited to single-task settings where a generalist agent capable of addressing multi-task predicaments is absent. In this paper, we aim to investigate the effectiveness of a single diffusion model in modeling large-scale multi-task offline data, which can be challenging due to diverse and multimodal data distribution. Specifically, we propose Multi-Task Diffusion Model (\textsc{MTDiff}), a diffusion-based method that incorporates Transformer backbones and prompt learning for generative planning and data synthesis in multi-task offline settings. \textsc{MTDiff} leverages vast amounts of knowledge available in multi-task data and performs implicit knowledge sharing among tasks. For generative planning, we find \textsc{MTDiff} outperforms state-of-the-art algorithms across 50 tasks on Meta-World and 8 maps on Maze2D. For data synthesis, \textsc{MTDiff} generates high-quality data for testing tasks given a single demonstration as a prompt, which enhances the low-quality datasets for even unseen tasks.
****
Single-Call Stochastic Extragradient Methods for Structured Non-monotone Variational Inequalities: Improved Analysis under Weaker Conditions
Sayantan Choudhury, Eduard Gorbunov, Nicolas Loizou
Single-call stochastic extragradient methods, like stochastic past extragradient (SPEG) and stochastic optimistic gradient (SOG), have gained a lot of interest in recent years and are one of the most efficient algorithms for solving large-scale min-max optimization and variational inequalities problems (VIP) appearing in various machine learning tasks. However, despite their undoubted popularity, current convergence analyses of SPEG and SOG require strong assumptions like bounded variance or growth conditions. In addition, several important questions regarding the convergence properties of these methods are still open, including mini-batching, efficient step-size selection, and convergence guarantees under different sampling strategies. In this work, we address these questions and provide convergence guarantees for two large classes of structured non-monotone VIPs: (i) quasi-strongly monotone problems (a generalization of strongly monotone problems) and (ii) weak Minty variational inequalities (a generalization of monotone and Minty VIPs). We introduce the expected residual condition, explain its benefits, and show how it allows us to obtain a strictly weaker bound than previously used growth conditions, expected co-coercivity, or bounded variance assumptions. Finally, our convergence analysis holds under the arbitrary sampling paradigm, which includes importance sampling and various mini-batching strategies as special cases.
****
Assessor360: Multi-sequence Network for Blind Omnidirectional Image Quality Assessment
Tianhe Wu, Shuwei Shi, Haoming Cai, Mingdeng Cao, Jing Xiao, Yinqiang Zheng, Yujiu Yang
Blind Omnidirectional Image Quality Assessment (BOIQA) aims to objectively asses

s the human perceptual quality of omnidirectional images (ODIs) without relying on pristine-quality image information. It is becoming more significant with the increasing advancement of virtual reality (VR) technology. However, the quality assessment of ODIs is severely hampered by the fact that the existing BOIQA pipeline lacks the modeling of the observer's browsing process. To tackle this issue, we propose a novel multi-sequence network for BOIQA called Assessor360, which is derived from the realistic multi-assessor ODI quality assessment procedure. Specifically, we propose a generalized Recursive Probability Sampling (RPS) method for the BOIQA task, combining content and details information to generate multiple pseudo viewport sequences from a given starting point. Additionally, we design a Multi-scale Feature Aggregation (MFA) module with a Distortion-aware Block (DAB) to fuse distorted and semantic features of each viewport. We also devise Temporal Modeling Module (TMM) to learn the viewport transition in the temporal domain. Extensive experimental results demonstrate that Assessor360 outperforms state-of-the-art methods on multiple OIQA datasets. The code and models are available at https://github.com/TianheWu/Assessor360.
****
Lossy Image Compression with Conditional Diffusion Models
Ruihan Yang, Stephan Mandt
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Leveraging the two-timescale regime to demonstrate convergence of neural networks
Pierre Marion, Raphaël Berthier
We study the training dynamics of shallow neural networks, in a two-timescale regime in which the stepsizes for the inner layer are much smaller than those for the outer layer. In this regime, we prove convergence of the gradient flow to a global optimum of the non-convex optimization problem in a simple univariate setting. The number of neurons need not be asymptotically large for our result to hold, distinguishing our result from popular recent approaches such as the neural tangent kernel or mean-field regimes. Experimental illustration is provided, showing that the stochastic gradient descent behaves according to our description of the gradient flow and thus converges to a global optimum in the two-timescale regime, but can fail outside of this regime.
****
Grammar Prompting for Domain-Specific Language Generation with Large Language Models
Bailin Wang, Zi Wang, Xuezhi Wang, Yuan Cao, Rif A. Saurous, Yoon Kim
Large language models (LLMs) can learn to perform a wide range of natural language tasks from just a handful of in-context examples. However, for generating strings from highly structured languages (e.g., semantic parsing to complex domain-specific languages), it is challenging for the LLM to generalize from just a few exemplars. We propose \emph{grammar prompting}, a simple approach to enable LLMs to use external knowledge and domain-specific constraints, expressed through a grammar in Backus--Naur Form (BNF), during in-context learning. Grammar prompting augments each demonstration example with a specialized grammar that is minimally sufficient for generating the particular output example, where the specialized grammar is a subset of the full DSL grammar. For inference, the LLM first predicts a BNF grammar given a test input, and then generates the output according to the rules of the grammar. Experiments demonstrate that grammar prompting can enable LLMs to perform competitively on a diverse set of DSL generation tasks, including semantic parsing (SMCalFlow, Overnight, GeoQuery), PDDL planning, and SMILES-based molecule generation.
****
Don’t just prune by magnitude! Your mask topology is a secret weapon
Duc Hoang, Souvik Kundu, Shiwei Liu, Zhangyang "Atlas" Wang
Recent years have witnessed significant progress in understanding the relationsh

ip between the connectivity of a deep network's architecture as a graph, and the network's performance. A few prior arts connected deep architectures to expander graphs or Ramanujan graphs, and particularly,[7] demonstrated the use of such graph connectivity measures with ranking and relative performance of various obtained sparse sub-networks (i.e. models with prune masks) without the need for training. However, no prior work explicitly explores the role of parameters in the graph's connectivity, making the graph-based understanding of prune masks and the magnitude/gradient-based pruning practice isolated from one another. This paper strives to fill in this gap, by analyzing the Weighted Spectral Gap of Ramanujan structures in sparse neural networks and investigates its correlation with final performance. We specifically examine the evolution of sparse structures under a popular dynamic sparse-to-sparse network training scheme, and intriguingly find that the generated random topologies inherently maximize Ramanujan graphs. We also identify a strong correlation between masks, performance, and the weighted spectral gap. Leveraging this observation, we propose to construct a new "full-spectrum coordinate'' aiming to comprehensively characterize a sparse neural network's promise. Concretely, it consists of the classical Ramanujan's gap (structure), our proposed weighted spectral gap (parameters), and the constituent nested regular graphs within. In this new coordinate system, a sparse subnetwork's L2-distance from its original initialization is found to have nearly linear correlated with its performance. Eventually, we apply this unified perspective to develop a new actionable pruning method, by sampling sparse masks to maximize the L2-coordinate distance. Our method can be augmented with the "pruning at initialization" (PaI) method, and significantly outperforms existing PaI methods. With only a few iterations of training (e.g 500 iterations), we can get LTH-comparable performance as that yielded via "pruning after training", significantly saving pre-training costs. Codes can be found at: https://github.com/VITA-Group/FullSpectrum-PAI.
****
Learning Human Action Recognition Representations Without Real Humans
Howard Zhong, Samarth Mishra, Donghyun Kim, SouYoung Jin, Rameswar Panda, Hilde Kuehne, Leonid Karlinsky, Venkatesh Saligrama, Aude Oliva, Rogerio Feris
Pre-training on massive video datasets has become essential to achieve high action recognition performance on smaller downstream datasets. However, most large-scale video datasets contain images of people and hence are accompanied with issues related to privacy, ethics, and data protection, often preventing them from being publicly shared for reproducible research. Existing work has attempted to alleviate these problems by blurring faces, downsampling videos, or training on synthetic data. On the other hand, analysis on the {\em transferability} of privacy-preserving pre-trained models to downstream tasks has been limited. In this work, we study this problem by first asking the question: can we pre-train models for human action recognition with data that does not include real humans? To this end, we present, for the first time, a benchmark that leverages real-world videos with {\em humans removed} and synthetic data containing virtual humans to pre-train a model. We then evaluate the transferability of the representation learned on this data to a diverse set of downstream action recognition benchmarks. Furthermore, we propose a novel pre-training strategy, called Privacy-Preserving MAE-Align, to effectively combine synthetic data and human-removed real data. Our approach outperforms previous baselines by up to 5\% and closes the performance gap between human and no-human action recognition representations on downstream tasks, for both linear probing and fine-tuning. Our benchmark, code, and models are available at https://github.com/howardzh01/PPMA.
****
Primal-Attention: Self-attention through Asymmetric Kernel SVD in Primal Representation
Yingyi Chen, Qinghua Tao, Francesco Tonin, Johan Suykens
Recently, a new line of works has emerged to understand and improve self-attention in Transformers by treating it as a kernel machine. However, existing works apply the methods for symmetric kernels to the asymmetric self-attention, resulting in a nontrivial gap between the analytical understanding and numerical implem

entation. In this paper, we provide a new perspective to represent and optimize self-attention through asymmetric Kernel Singular Value Decomposition (KSVD), which is also motivated by the low-rank property of self-attention normally observed in deep layers. Through asymmetric KSVD, i) a primal-dual representation of self-attention is formulated, where the optimization objective is cast to maximize the projection variances in the attention outputs; ii) a novel attention mechanism, i.e., Primal-Attention, is proposed via the primal representation of KSVD, avoiding explicit computation of the kernel matrix in the dual; iii) with KKT conditions, we prove that the stationary solution to the KSVD optimization in Primal-Attention yields a zero-value objective. In this manner, KSVD optimization can be implemented by simply minimizing a regularization loss, so that low-rank property is promoted without extra decomposition. Numerical experiments show state-of-the-art performance of our Primal-Attention with improved efficiency. Moreover, we demonstrate that the deployed KSVD optimization regularizes Primal-Attention with a sharper singular value decay than that of the canonical self-attention, further verifying the great potential of our method. To the best of our knowledge, this is the first work that provides a primal-dual representation for the asymmetric kernel in self-attention and successfully applies it to modelling and optimization.
****
End-To-End Latent Variational Diffusion Models for Inverse Problems in High Energy Physics
Alexander Shmakov, Kevin Greif, Michael Fenton, Aishik Ghosh, Pierre Baldi, Daniel Whiteson
High-energy collisions at the Large Hadron Collider (LHC) provide valuable insights into open questions in particle physics. However, detector effects must be corrected before measurements can be compared to certain theoretical predictions or measurements from other detectors. Methods to solve this inverse problem of mapping detector observations to theoretical quantities of the underlying collision are essential parts of many physics analyses at the LHC. We investigate and compare various generative deep learning methods to approximate this inverse mapping. We introduce a novel unified architecture, termed latent variational diffusion models, which combines the latent learning of cutting-edge generative art approaches with an end-to-end variational framework. We demonstrate the effectiveness of this approach for reconstructing global distributions of theoretical kinematic quantities, as well as for ensuring the adherence of the learned posterior distributions to known physics constraints. Our unified approach achieves a distribution-free distance to the truth of over 20 times smaller than non-latent state-of-the-art baseline and 3 times smaller than traditional latent diffusion models.
****
AVeriTeC: A Dataset for Real-world Claim Verification with Evidence from the Web
Michael Schlichtkrull, Zhijiang Guo, Andreas Vlachos
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DiffTraj: Generating GPS Trajectory with Diffusion Probabilistic Model
Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, James Yu
Pervasive integration of GPS-enabled devices and data acquisition technologies has led to an exponential increase in GPS trajectory data, fostering advancements in spatial-temporal data mining research. Nonetheless, GPS trajectories contain personal geolocation information, rendering serious privacy concerns when working with raw data. A promising approach to address this issue is trajectory generation, which involves replacing original data with generated, privacy-free alternatives. Despite the potential of trajectory generation, the complex nature of human behavior and its inherent stochastic characteristics pose challenges in generating high-quality trajectories. In this work, we propose a spatial-temporal diffusion probabilistic model for trajectory generation (DiffTraj). This model ef

fectively combines the generative abilities of diffusion models with the spatial-temporal features derived from real trajectories. The core idea is to reconstruct and synthesize geographic trajectories from white noise through a reverse trajectory denoising process. Furthermore, we propose a Trajectory UNet (Traj-UNet) deep neural network to embed conditional information and accurately estimate noise levels during the reverse process. Experiments on two real-world datasets show that DiffTraj can be intuitively applied to generate high-fidelity trajectories while retaining the original distributions. Moreover, the generated results can support downstream trajectory analysis tasks and significantly outperform other methods in terms of geo-distribution evaluations.
****
Meta-in-context learning in large language models
Julian Coda-Forno, Marcel Binz, Zeynep Akata, Matt Botvinick, Jane Wang, Eric Schulz
Large language models have shown tremendous performance in a variety of tasks. In-context learning -- the ability to improve at a task after being provided with a number of demonstrations -- is seen as one of the main contributors to their success. In the present paper, we demonstrate that the in-context learning abilities of large language models can be recursively improved via in-context learning itself. We coin this phenomenon meta-in-context learning. Looking at two idealized domains, a one-dimensional regression task and a two-armed bandit task, we show that meta-in-context learning adaptively reshapes a large language model's priors over expected tasks. Furthermore, we find that meta-in-context learning modifies the in-context learning strategies of such models. Finally, we broaden the scope of our investigation to encompass two diverse benchmarks: one focusing on real-world regression problems and the other encompassing multiple NLP tasks. In both cases, we observe competitive performance comparable to that of traditional learning algorithms. Taken together, our work improves our understanding of in-context learning and paves the way toward adapting large language models to the environment they are applied purely through meta-in-context learning rather than traditional finetuning.
****
Dynamic Context Pruning for Efficient and Interpretable Autoregressive Transformers
Sotiris Anagnostidis, Dario Pavllo, Luca Biggio, Lorenzo Noci, Aurelien Lucchi, Thomas Hofmann
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SPQR: Controlling Q-ensemble Independence with Spiked Random Model for Reinforcement Learning
Dohyeok Lee, Seungyub Han, Taehyun Cho, Jungwoo Lee
Alleviating overestimation bias is a critical challenge for deep reinforcement learning to achieve successful performance on more complex tasks or offline datasets containing out-of-distribution data. In order to overcome overestimation bias, ensemble methods for Q-learning have been investigated to exploit the diversity of multiple Q-functions. Since network initialization has been the predominant approach to promote diversity in Q-functions, heuristically designed diversity injection methods have been studied in the literature. However, previous studies have not attempted to approach guaranteed independence over an ensemble from a theoretical perspective. By introducing a novel regularization loss for Q-ensemble independence based on random matrix theory, we propose spiked Wishart Q-ensemble independence regularization (SPQR) for reinforcement learning. Specifically, we modify the intractable hypothesis testing criterion for the Q-ensemble independence into a tractable KL divergence between the spectral distribution of the Q-ensemble and the target Wigner's semicircle distribution. We implement SPQR in several online and offline ensemble Q-learning algorithms. In the experiments, SPQR outperforms the baseline algorithms in both online and offline RL benchmar

ks.
****
SwapPrompt: Test-Time Prompt Adaptation for Vision-Language Models
XIAOSONG MA, Jie ZHANG, Song Guo, Wenchao Xu
Test-time adaptation (TTA) is a special and practical setting in unsupervised domain adaptation, which allows a pre-trained model in a source domain to adapt to unlabeled test data in another target domain. To avoid the computation-intensive backbone fine-tuning process, the zero-shot generalization potentials of the emerging pre-trained vision-language models (e.g., CLIP, CoOp) are leveraged to only tune the run-time prompt for unseen test domains. However, existing solutions have yet to fully exploit the representation capabilities of pre-trained models as they only focus on the entropy-based optimization and the performance is far below the supervised prompt adaptation methods, e.g., CoOp. In this paper, we propose SwapPrompt, a novel framework that can effectively leverage the self-supervised contrastive learning to facilitate the test-time prompt adaptation. SwapPrompt employs a dual prompts paradigm, i.e., an online prompt and a target prompt that averaged from the online prompt to retain historical information. In addition, SwapPrompt applies a swapped prediction mechanism, which takes advantage of the representation capabilities of pre-trained models to enhance the online prompt via contrastive learning. Specifically, we use the online prompt together with an augmented view of the input image to predict the class assignment generated by the target prompt together with an alternative augmented view of the same image. The proposed SwapPrompt can be easily deployed on vision-language models without additional requirement, and experimental results show that it achieves state-of-the-art test-time adaptation performance on ImageNet and nine other datasets. It is also shown that SwapPrompt can even achieve comparable performance with supervised prompt adaptation methods.
****
Does a sparse ReLU network training problem always admit an optimum ?
TUNG LE, Remi Gribonval, Elisa Riccietti
Given a training set, a loss function, and a neural network architecture, it is often taken for granted that optimal network parameters exist, and a common practice is to apply available optimization algorithms to search for them. In this work, we show that the existence of an optimal solution is not always guaranteed, especially in the context of sparse ReLU neural networks.In particular, we first show that optimization problems involving deep networks with certain sparsity patterns do not always have optimal parameters, and that optimization algorithms may then diverge. Via a new topological relation between sparse ReLU neural networks and their linear counterparts, we derive --using existing tools from real algebraic geometry-- an algorithm to verify that a given sparsity pattern suffers from this issue. Then, the existence of a global optimum is proved for every concrete optimization problem involving a shallow sparse ReLU neural network of output dimension one. Overall, the analysis is based on the investigation of two topological properties of the space of functions implementable as sparse ReLU neural networks: a best approximation property, and a closedness property, both in the uniform norm. This is studied both for (finite) domains corresponding to practical training on finite training sets, and for more general domains such as the unit cube. This allows us to provide conditions for the guaranteed existence of an optimum given a sparsity pattern. The results apply not only to several sparsity patterns proposed in recent works on network pruning/sparsification, but also to classical dense neural networks, including architectures not covered by existing results.
****
Knowledge Diffusion for Distillation
Tao Huang, Yuan Zhang, Mingkai Zheng, Shan You, Fei Wang, Chen Qian, Chang Xu
The representation gap between teacher and student is an emerging topic in knowledge distillation (KD). To reduce the gap and improve the performance, current methods often resort to complicated training schemes, loss functions, and feature alignments, which are task-specific and feature-specific. In this paper, we state that the essence of these methods is to discard the noisy information and dis

till the valuable information in the feature, and propose a novel KD method dubbed DiffKD, to explicitly denoise and match features using diffusion models. Our approach is based on the observation that student features typically contain more noises than teacher features due to the smaller capacity of student model. To address this, we propose to denoise student features using a diffusion model trained by teacher features. This allows us to perform better distillation between the refined clean feature and teacher feature. Additionally, we introduce a light-weight diffusion model with a linear autoencoder to reduce the computation cost and an adaptive noise matching module to improve the denoising performance. Extensive experiments demonstrate that DiffKD is effective across various types of features and achieves state-of-the-art performance consistently on image classification, object detection, and semantic segmentation tasks. Code is available at https://github.com/hunto/DiffKD.
****

BayesTune: Bayesian Sparse Deep Model Fine-tuning

Minyoung Kim, Timothy Hospedales

Deep learning practice is increasingly driven by powerful foundation models (FM), pre-trained at scale and then fine-tuned for specific tasks of interest. A key property of this workflow is the efficacy of performing sparse or parameter-efficient fine-tuning, meaning that by updating only a tiny fraction of the whole FM parameters on a downstream task can lead to surprisingly good performance, often even superior to a full model update. However, it is not clear what is the optimal and principled way to select which parameters to update. Although a growing number of sparse fine-tuning ideas have been proposed, they are mostly not satisfactory, relying on hand-crafted heuristics or heavy approximation. In this paper we propose a novel Bayesian sparse fine-tuning algorithm: we place a (sparse) Laplace prior for each parameter of the FM, with the mean equal to the initial value and the scale parameter having a hyper-prior that encourages small scale. Roughly speaking, the posterior means of the scale parameters indicate how important it is to update the corresponding parameter away from its initial value when solving the downstream task. Given the sparse prior, most scale parameters are small a posteriori, and the few large-valued scale parameters identify those FM parameters that crucially need to be updated away from their initial values. Based on this, we can threshold the scale parameters to decide which parameters to update or freeze, leading to a principled sparse fine-tuning strategy. To efficiently infer the posterior distribution of the scale parameters, we adopt the Langevin MCMC sampler, requiring only two times the complexity of the vanilla SGD. Tested on popular NLP benchmarks as well as the VTAB vision tasks, our approach shows significant improvement over the state-of-the-arts (e.g., 1% point higher than the best SOTA when fine-tuning RoBERTa for GLUE and SuperGLUE benchmarks).
****

Exploring Loss Functions for Time-based Training Strategy in Spiking Neural Networks

Yaoyu Zhu, Wei Fang, Xiaodong Xie, Tiejun Huang, Zhaofei Yu

Spiking Neural Networks (SNNs) are considered promising brain-inspired energy-efficient models due to their event-driven computing paradigm.The spatiotemporal spike patterns used to convey information in SNNs consist of both rate coding and temporal coding, where the temporal coding is crucial to biological-plausible learning rules such as spike-timing-dependent-plasticity.The time-based training strategy is proposed to better utilize the temporal information in SNNs and learn in an asynchronous fashion.However, some recent works train SNNs by the time-based scheme with rate-coding-dominated loss functions.In this paper, we first map rate-based loss functions to time-based counterparts and explain why they are also applicable to the time-based training scheme.After that, we infer that loss functions providing adequate positive overall gradients help training by theoretical analysis.Based on this, we propose the enhanced counting loss to replace the commonly used mean square counting loss.In addition, we transfer the training of scale factor in weight standardization into thresholds.Experiments show that our approach outperforms previous time-based training methods in most datasets.

Our work provides insights for training SNNs with time-based schemes and offers a fresh perspective on the correlation between rate coding and temporal coding. Our code is available at https://github.com/zhuyaoyu/SNN-temporal-training-losses.
****
Learning Rate Free Sampling in Constrained Domains
Louis Sharrock, Lester Mackey, Christopher Nemeth
We introduce a suite of new particle-based algorithms for sampling in constrained domains which are entirely learning rate free. Our approach leverages coin betting ideas from convex optimisation, and the viewpoint of constrained sampling as a mirrored optimisation problem on the space of probability measures. Based on this viewpoint, we also introduce a unifying framework for several existing constrained sampling algorithms, including mirrored Langevin dynamics and mirrored Stein variational gradient descent. We demonstrate the performance of our algorithms on a range of numerical examples, including sampling from targets on the simplex, sampling with fairness constraints, and constrained sampling problems in post-selection inference. Our results indicate that our algorithms achieve competitive performance with existing constrained sampling methods, without the need to tune any hyperparameters.
****
Volume Feature Rendering for Fast Neural Radiance Field Reconstruction
Kang Han, Wei Xiang, Lu Yu
Neural radiance fields (NeRFs) are able to synthesize realistic novel views from multi-view images captured from distinct positions and perspectives. In NeRF's rendering pipeline, neural networks are used to represent a scene independently or transform queried learnable feature vector of a point to the expected color or density. With the aid of geometry guides either in the form of occupancy grids or proposal networks, the number of color neural network evaluations can be reduced from hundreds to dozens in the standard volume rendering framework. However, many evaluations of the color neural network are still a bottleneck for fast NeRF reconstruction. This paper revisits volume feature rendering (VFR) for the purpose of fast NeRF reconstruction. The VFR integrates the queried feature vectors of a ray into one feature vector, which is then transformed to the final pixel color by a color neural network. This fundamental change to the standard volume rendering framework requires only one single color neural network evaluation to render a pixel, which substantially lowers the high computational complexity of the rendering framework attributed to a large number of color neural network evaluations. Consequently, we can use a comparably larger color neural network to achieve a better rendering quality while maintaining the same training and rendering time costs. This approach achieves the state-of-the-art rendering quality on both synthetic and real-world datasets while requiring less training time compared with existing methods.
****
Offline RL with Discrete Proxy Representations for Generalizability in POMDPs
Pengjie Gu, Xinyu Cai, Dong Xing, Xinrun Wang, Mengchen Zhao, Bo An
Offline Reinforcement Learning (RL) has demonstrated promising results in various applications by learning policies from previously collected datasets, reducing the need for online exploration and interactions. However, real-world scenarios usually involve partial observability, which brings crucial challenges of the deployment of offline RL methods: i) the policy trained on data with full observability is not robust against the masked observations during execution, and ii) the information of which parts of observations are masked is usually unknown during training. In order to address these challenges, we present Offline RL with DiscrEte pRoxy representations (ORDER), a probabilistic framework which leverages novel state representations to improve the robustness against diverse masked observabilities. Specifically, we propose a discrete representation of the states and use a proxy representation to recover the states from masked partial observable trajectories. The training of ORDER can be compactly described as the following three steps. i) Learning the discrete state representations on data with full observations, ii) Training the decision module based on the discrete representa

tions, and iii) Training the proxy discrete representations on the data with various partial observations, aligning with the discrete representations. We conduct extensive experiments to evaluate ORDER, showcasing its effectiveness in offline RL for diverse partially observable scenarios and highlighting the significance of discrete proxy representations in generalization performance.ORDER is a flexible framework to employ any offline RL algorithms and we hope that ORDER can pave the way for the deployment of RL policy against various partial observabilities in the real world.
****
Meta-AdaM: An Meta-Learned Adaptive Optimizer with Momentum for Few-Shot Learning
Siyuan Sun, Hongyang Gao
We introduce Meta-AdaM, a meta-learned adaptive optimizer with momentum, designed for few-shot learning tasks that pose significant challenges to deep learning models due to the limited number of labeled examples. Meta-learning has been successfully employed to address these challenges by transferring meta-learned prior knowledge to new tasks. Most existing works focus on meta-learning an optimal model initialization or an adaptive learning rate learner for rapid convergence. However, these approaches either neglect to consider weight-update history for the adaptive learning rate learner or fail to effectively integrate momentum for fast convergence, as seen in many-shot learning settings. To tackle these limitations, we propose a meta-learned learning rate learner that utilizes weight-update history as input to predict more appropriate learning rates for rapid convergence. Furthermore, for the first time, our approach incorporates momentum into the optimization process of few-shot learning via a double look-ahead mechanism, enabling rapid convergence similar to many-shot settings. Extensive experimental results on benchmark datasets demonstrate the effectiveness of the proposed Meta-AdaM.
****
Fairness Continual Learning Approach to Semantic Scene Understanding in Open-World Environments
Thanh-Dat Truong, Hoang-Quan Nguyen, Bhiksha Raj, Khoa Luu
Continual semantic segmentation aims to learn new classes while maintaining the information from the previous classes. Although prior studies have shown impressive progress in recent years, the fairness concern in the continual semantic segmentation needs to be better addressed. Meanwhile, fairness is one of the most vital factors in deploying the deep learning model, especially in human-related or safety applications. In this paper, we present a novel Fairness Continual Learning approach to the semantic segmentation problem.In particular, under the fairness objective, a new fairness continual learning framework is proposed based on class distributions.Then, a novel Prototypical Contrastive Clustering loss is proposed to address the significant challenges in continual learning, i.e., catastrophic forgetting and background shift. Our proposed loss has also been proven as a novel, generalized learning paradigm of knowledge distillation commonly used in continual learning. Moreover, the proposed Conditional Structural Consistency loss further regularized the structural constraint of the predicted segmentation. Our proposed approach has achieved State-of-the-Art performance on three standard scene understanding benchmarks, i.e., ADE20K, Cityscapes, and Pascal VOC, and promoted the fairness of the segmentation model.
****
Post Hoc Explanations of Language Models Can Improve Language Models
Satyapriya Krishna, Jiaqi Ma, Dylan Slack, Asma Ghandeharioun, Sameer Singh, Himabindu Lakkaraju
Large Language Models (LLMs) have demonstrated remarkable capabilities in performing complex tasks. Moreover, recent research has shown that incorporating human-annotated rationales (e.g., Chain-of-Thought prompting) during in-context learning can significantly enhance the performance of these models, particularly on tasks that require reasoning capabilities. However, incorporating such rationales poses challenges in terms of scalability as this requires a high degree of human involvement. In this work, we present a novel framework, Amplifying Model Perf

ormance by Leveraging In-Context Learning with Post Hoc Explanations (AMPLIFY), which addresses the aforementioned challenges by automating the process of ratio nale generation. To this end, we leverage post hoc explanation methods which out put attribution scores (explanations) capturing the influence of each of the inp ut features on model predictions. More specifically, we construct automated natu ral language rationales that embed insights from post hoc explanations to provid e corrective signals to LLMs. Extensive experimentation with real-world datasets demonstrates that our framework, AMPLIFY, leads to prediction accuracy improvem ents of about 10-25% over a wide range of tasks, including those where prior app roaches which rely on human-annotated rationales such as Chain-of-Thought prompt ing fall short. Our work makes one of the first attempts at highlighting the pot ential of post hoc explanations as valuable tools for enhancing the effectivenes s of LLMs. Furthermore, we conduct additional empirical analyses and ablation st udies to demonstrate the impact of each of the components of AMPLIFY, which, in turn, lead to critical insights for refining in context learning.
****
Understanding Diffusion Objectives as the ELBO with Simple Data Augmentation
Diederik Kingma, Ruiqi Gao
To achieve the highest perceptual quality, state-of-the-art diffusion models are optimized with objectives that typically look very different from the maximum l ikelihood and the Evidence Lower Bound (ELBO) objectives. In this work, we revea l that diffusion model objectives are actually closely related to the ELBO.Speci fically, we show that all commonly used diffusion model objectives equate to a w eighted integral of ELBOs over different noise levels, where the weighting depen ds on the specific objective used. Under the condition of monotonic weighting, t he connection is even closer: the diffusion objective then equals the ELBO, comb ined with simple data augmentation, namely Gaussian noise perturbation. We show that this condition holds for a number of state-of-the-art diffusion models. In experiments, we explore new monotonic weightings and demonstrate their effective ness, achieving state-of-the-art FID scores on the high-resolution ImageNet benc hmark.
****
Response Length Perception and Sequence Scheduling: An LLM-Empowered LLM Infere nce Pipeline
Zangwei Zheng, Xiaozhe Ren, Fuzhao Xue, Yang Luo, Xin Jiang, Yang You
Large language models (LLMs) have revolutionized the field of AI, demonstrating unprecedented capacity across various tasks. However, the inference process for LLMs comes with significant computational costs. In this paper, we propose an ef ficient LLM inference pipeline that harnesses the power of LLMs. Our approach be gins by tapping into the potential of LLMs to accurately perceive and predict th e response length with minimal overhead. By leveraging this information, we intr oduce an efficient sequence scheduling technique that groups queries with simila r response lengths into micro-batches. We evaluate our approach on real-world in struction datasets using the LLaMA-based model, and our results demonstrate an i mpressive 86% improvement in inference throughput without compromising effective ness. Notably, our method is orthogonal to other inference acceleration techniqu es, making it a valuable addition to many existing toolkits (e.g., FlashAttentio n, Quantization) for LLM inference.
****
When Demonstrations meet Generative World Models: A Maximum Likelihood Framework for Offline Inverse Reinforcement Learning
Siliang Zeng, Chenliang Li, Alfredo Garcia, Mingyi Hong
Offline inverse reinforcement learning (Offline IRL) aims to recover the structu re of rewards and environment dynamics that underlie observed actions in a fixed , finite set of demonstrations from an expert agent. Accurate models of expertis e in executing a task has applications in safety-sensitive applications such as clinical decision making and autonomous driving. However, the structure of an ex pert's preferences implicit in observed actions is closely linked to the expert' s model of the environment dynamics (i.e. the ``world''). Thus, inaccurate model s of the world obtained from finite data with limited coverage could compound in

accuracy in estimated rewards. To address this issue, we propose a bi-level opti
mization formulation of the estimation task wherein the upper level is likelihoo
d maximization based upon a conservative model of the expert's policy (lower lev
el). The policy model is conservative in that it maximizes reward subject to a p
enalty that is increasing in the uncertainty of the estimated model of the world
. We propose a new algorithmic framework to solve the bi-level optimization prob
lem formulation and provide statistical and computational guarantees of performa
nce for the associated optimal reward estimator. Finally, we demonstrate that t
he proposed algorithm outperforms the state-of-the-art offline IRL and imitation
 learning benchmarks by a large margin, over the continuous control tasks in MuJ
oCo and different datasets in the D4RL benchmark.
****
Neural Priming for Sample-Efficient Adaptation
Matthew Wallingford, Vivek Ramanujan, Alex Fang, Aditya Kusupati, Roozbeh Mottag
hi, Aniruddha Kembhavi, Ludwig Schmidt, Ali Farhadi
We propose Neural Priming, a technique for adapting large pretrained models to d
istribution shifts and downstream tasks given few or no labeled examples. Presen
ted with class names or unlabeled test samples, Neural Priming enables the model
 to recall and conditions its parameters on relevant data seen throughout pretra
ining, thereby priming it for the test distribution. Neural Priming can be perfo
rmed at test time in even for pretraining datasets as large as LAION-2B. Perform
ing lightweight updates on the recalled data significantly improves accuracy acr
oss a variety of distribution shift and transfer learning benchmarks. Concretely
, in the zero-shot setting, we see a 2.45% improvement in accuracy on ImageNet a
nd 3.81% accuracy improvement on average across standard transfer learning bench
marks. Further, using our test time inference scheme, we see a 1.41% accuracy im
provement on ImageNetV2. These results demonstrate the effectiveness of Neural P
riming in addressing the common challenge of limited labeled data and changing d
istributions. Code and models are open-sourced at https://www.github.com/RAIVNLa
b/neural-priming.
****
Derandomized novelty detection with FDR control via conformal e-values
Meshi Bashari, Amir Epstein, Yaniv Romano, Matteo Sesia
Conformal inference provides a general distribution-free method to rigorously ca
librate the output of any machine learning algorithm for novelty detection. Whil
e this approach has many strengths, it has the limitation of being randomized, i
n the sense that it may lead to different results when analyzing twice the same
data and this can hinder the interpretation of any findings. We propose to make
conformal inferences more stable by leveraging suitable conformal e-values inste
ad of p-values to quantify statistical significance. This solution allows the ev
idence gathered from multiple analyses of the same data to be aggregated effecti
vely while provably controlling the false discovery rate. Further, we show that
the proposed method can reduce randomness without much loss of power compared to
 standard conformal inference, partly thanks to an innovative way of weighting c
onformal e-values based on additional side information carefully extracted from
the same data. Simulations with synthetic and real data confirm this solution ca
n be effective at eliminating random noise in the inferences obtained with state
-of-the-art alternative techniques, sometimes also leading to higher power.
****
ZipLM: Inference-Aware Structured Pruning of Language Models
Eldar Kurti■, Elias Frantar, Dan Alistarh
The breakthrough performance of large language models (LLMs) comes with major co
mputational footprints and high deployment costs. In this paper, we progress tow
ards resolving this problem by proposing a novel structured compression approach
 for LLMs, called ZipLM. ZipLM achieves state-of-the-art accuracy-vs-speedup, wh
ile matching a set of desired target runtime speedups in any given inference env
ironment. Specifically, given a model, a dataset, an inference environment, as w
ell as a set of speedup targets, ZipLM iteratively identifies and removes compon
ents with the worst loss-runtime trade-off. Unlike prior methods that specialize
 in either the post-training/one-shot or the gradual compression setting, and on

ly for specific families of models such as BERT (encoder) or GPT (decoder), ZipLM produces state-of-the-art compressed models across all these settings. Furthermore, ZipLM achieves superior results for a fraction of the computational cost relative to prior distillation and pruning techniques, making it a cost-effective approach for generating an entire family of smaller, faster, and highly accurate models, guaranteed to meet the desired inference specifications. In particular, ZipLM outperforms all prior BERT-base distillation and pruning techniques, such as CoFi, MiniLM, and TinyBERT. Moreover, it matches the performance of the heavily optimized MobileBERT model, obtained via extensive architecture search, by simply pruning the baseline BERT-large model. When compressing GPT2, ZipLM outperforms DistilGPT2 while being 60\% smaller and 30\% faster. Our code is available at: https://github.com/IST-DASLab/ZipLM.
****
Private (Stochastic) Non-Convex Optimization Revisited: Second-Order Stationary Points and Excess Risks
Daogao Liu, Arun Ganesh, Sewoong Oh, Abhradeep Guha Thakurta
We reconsider the challenge of non-convex optimization under differential privacy constraint. Building upon the previous variance-reduced algorithm SpiderBoost, we propose a novel framework that employs two types of gradient oracles: one that estimates the gradient at a single point and a more cost-effective option that calculates the gradient difference between two points. Our framework can ensure continuous accuracy of gradient estimations and subsequently enhances the rates of identifying second-order stationary points.Additionally, we consider a more challenging task by attempting to locate the global minima of a non-convex objective via the exponential mechanism without almost any assumptions. Our preliminary results suggest that the regularized exponential mechanism can effectively emulate previous empirical and population risk bounds, negating the need for smoothness assumptions for algorithms with polynomial running time. Furthermore, with running time factors excluded, the exponential mechanism demonstrates promising population risk bound performance, and we provide a nearly matching lower bound.
****
Revealing the unseen: Benchmarking video action recognition under occlusion
Shresth Grover, Vibhav Vineet, Yogesh Rawat
In this work, we study the effect of occlusion on video action recognition. Tofacilitate this study, we propose three benchmark datasets and experiment withseven different video action recognition models. These datasets include two synthetic benchmarks, UCF-101-O and K-400-O, which enabled understanding the effects of fundamental properties of occlusion via controlled experiments. We also propose a real-world occlusion dataset, UCF-101-Y-OCC, which helps in further validating the findings of this study. We find several interesting insights such as 1) transformers are more robust than CNN counterparts, 2) pretraining make modelsrobust against occlusions, and 3) augmentation helps, but does not generalize well to real-world occlusions. In addition, we propose a simple transformer based compositional model, termed as CTx-Net, which generalizes well under this distribution shift. We observe that CTx-Net outperforms models which are trained using occlusions as augmentation, performing significantly better under natural occlusions. We believe this benchmark will open up interesting future research in robust video action recognition
****
TrojLLM: A Black-box Trojan Prompt Attack on Large Language Models
Jiaqi Xue, Mengxin Zheng, Ting Hua, Yilin Shen, Yepeng Liu, Ladislau Bölöni, Qian Lou
Large Language Models (LLMs) are progressively being utilized as machine learning services and interface tools for various applications. However, the security implications of LLMs, particularly in relation to adversarial and Trojan attacks, remain insufficiently examined. In this paper, we propose TrojLLM, an automatic and black-box framework to effectively generate universal and stealthy triggers. When these triggers are incorporated into the input data, the LLMs' outputs can be maliciously manipulated.   Moreover, the framework also supports embedding

Trojans within discrete prompts, enhancing the overall effectiveness and precision of the triggers' attacks. Specifically, we propose a trigger discovery algorithm for generating universal triggers for various inputs by querying victim LLM-based APIs using few-shot data samples. Furthermore, we introduce a novel progressive Trojan poisoning algorithm designed to generate poisoned prompts that retain efficacy and transferability across a diverse range of models. Our experiments and results demonstrate TrojLLM's capacity to effectively insert Trojans into text prompts in real-world black-box LLM APIs including GPT-3.5 and GPT-4, while maintaining exceptional performance on clean test sets. Our work sheds light on the potential security risks in current models and offers a potential defensive approach. The source code of TrojLLM is available at https://github.com/UCF-ML-Research/TrojLLM.
****

Minimax Forward and Backward Learning of Evolving Tasks with Performance Guarantees

Veronica Alvarez, Santiago Mazuelas, Jose A. Lozano

For a sequence of classification tasks that arrive over time, it is common that tasks are evolving in the sense that consecutive tasks often have a higher similarity. The incremental learning of a growing sequence of tasks holds promise to enable accurate classification even with few samples per task by leveraging information from all the tasks in the sequence (forward and backward learning). However, existing techniques developed for continual learning and concept drift adaptation are either designed for tasks with time-independent similarities or only aim to learn the last task in the sequence. This paper presents incremental minimax risk classifiers (IMRCs) that effectively exploit forward and backward learning and account for evolving tasks. In addition, we analytically characterize the performance improvement provided by forward and backward learning in terms of the tasks' expected quadratic change and the number of tasks. The experimental evaluation shows that IMRCs can result in a significant performance improvement, especially for reduced sample sizes.
****

Online Label Shift: Optimal Dynamic Regret meets Practical Algorithms

Dheeraj Baby, Saurabh Garg, Tzu-Ching Yen, Sivaraman Balakrishnan, Zachary Lipton, Yu-Xiang Wang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Self-supervised video pretraining yields robust and more human-aligned visual representations

Nikhil Parthasarathy, S. M. Ali Eslami, Joao Carreira, Olivier Henaff

Humans learn powerful representations of objects and scenes by observing how they evolve over time. Yet, outside of specific tasks that require explicit temporal understanding, static image pretraining remains the dominant paradigm for learning visual foundation models. We question this mismatch, and ask whether video pretraining can yield visual representations that bear the hallmarks of human perception: generalisation across tasks, robustness to perturbations, and consistency with human judgements. To that end we propose a novel procedure for curating videos, and develop a contrastive framework which learns from the complex transformations therein. This simple paradigm for distilling knowledge from videos, called VITO, yields general representations that far outperform prior video pretraining methods on image understanding tasks, and image pretraining methods on video understanding tasks. Moreover, VITO representations are significantly more robust to natural and synthetic deformations than image-, video-, and adversarially-trainedones. Finally, VITO's predictions are strongly aligned with human judgements, surpassing models that were specifically trained for that purpose. Together, these results suggest that video pretraining could be a simple way of learning unified, robust, and human-aligned representations of the visual world.
****

Slot-guided Volumetric Object Radiance Fields
DI QI, Tong Yang, Xiangyu Zhang
We present a novel framework for 3D object-centric representation learning. Our approach effectively decomposes complex scenes into individual objects from a single image in an unsupervised fashion. This method, called \underline{s}lot-guided \underline{V}olumetric \underline{O}bject \underline{R}adiance \underline{F}ields~(sVORF), composes volumetric object radiance fields with object slots as a guidance to implement unsupervised 3D scene decomposition. Specifically, sVORF obtains object slots from a single image via a transformer module, maps these slots to volumetric object radiance fields with a hypernetwork and composes object radiance fields with the guidance of object slots at a 3D location. Moreover, sVORF significantly reduces memory requirement due to small-sized pixel rendering during training. We demonstrate the effectiveness of our approach by showing top results in scene decomposition and generation tasks of complex synthetic datasets (e.g., Room-Diverse). Furthermore, we also confirm the potential of sVORF to segment objects in real-world scenes (e.g., the LLFF dataset). We hope our approach can provide preliminary understanding of the physical world and help ease future research in 3D object-centric representation learning.
****
Riemannian SAM: Sharpness-Aware Minimization on Riemannian Manifolds
Jihun Yun, Eunho Yang
Contemporary advances in the field of deep learning have embarked upon an exploration of the underlying geometric properties of data, thus encouraging the investigation of techniques that consider general manifolds, for example, hyperbolic or orthogonal neural networks. However, the optimization algorithms for training such geometric deep learning models still remain highly under-explored. In this paper, we introduce Riemannian SAM by generalizing conventional Euclidean SAM to Riemannian manifolds. We successfully formulate the sharpness-aware minimization on Riemannian manifolds, leading to one of a novel instantiation, Lorentz SAM. In addition, SAM variants proposed in previous studies such as Fisher SAM can be derived as special examples under our Riemannian SAM framework. We provide the convergence analysis of Riemannian SAM under a less aggressively decaying ascent learning rate than Euclidean SAM. Our analysis serves as a theoretically sound contribution encompassing a diverse range of manifolds, also providing the guarantees for SAM variants such as Fisher SAM, whose convergence analyses are absent. Lastly, we illustrate the superiority of Riemannian SAM in terms of generalization over previous Riemannian optimization algorithms through experiments on knowledge graph completion and machine translation tasks.
****
ODE-based Recurrent Model-free Reinforcement Learning for POMDPs
Xuanle Zhao, Duzhen Zhang, Han Liyuan, Tielin Zhang, Bo Xu
Neural ordinary differential equations (ODEs) are widely recognized as the standard for modeling physical mechanisms, which help to perform approximate inference in unknown physical or biological environments. In partially observable (PO) environments, how to infer unseen information from raw observations puzzled the agents. By using a recurrent policy with a compact context, context-based reinforcement learning provides a flexible way to extract unobservable information from historical transitions. To help the agent extract more dynamics-related information, we present a novel ODE-based recurrent model combines with model-free reinforcement learning (RL) framework to solve partially observable Markov decision processes (POMDPs). We experimentally demonstrate the efficacy of our methods across various PO continuous control and meta-RL tasks. Furthermore, our experiments illustrate that our method is robust against irregular observations, owing to the ability of ODEs to model irregularly-sampled time series.
****
Deep Contract Design via Discontinuous Networks
Tonghan Wang, Paul Duetting, Dmitry Ivanov, Inbal Talgam-Cohen, David C. Parkes
Contract design involves a principal who establishes contractual agreements about payments for outcomes that arise from the actions of an agent. In this paper, we initiate the study of deep learning for the automated design of optimal contr

acts. We introduce a novel representation: the Discontinuous ReLU (DeLU) network, which models the principal's utility as a discontinuous piecewise affine function of the design of a contract where each piece corresponds to the agent taking a particular action. DeLU networks implicitly learn closed-form expressions for the incentive compatibility constraints of the agent and the utility maximization objective of the principal, and support parallel inference on each piece through linear programming or interior-point methods that solve for optimal contracts. We provide empirical results that demonstrate success in approximating the principal's utility function with a small number of training samples and scaling to find approximately optimal contracts on problems with a large number of actions and outcomes.
****
Temporal Continual Learning with Prior Compensation for Human Motion Prediction
Jianwei Tang, Jiangxin Sun, Xiaotong Lin, lifang zhang, Wei-Shi Zheng, Jian-Fang Hu
Human Motion Prediction (HMP) aims to predict future poses at different moments according to past motion sequences. Previous approaches have treated the prediction of various moments equally, resulting in two main limitations: the learning of short-term predictions is hindered by the focus on long-term predictions, and the incorporation of prior information from past predictions into subsequent predictions is limited. In this paper, we introduce a novel multi-stage training framework called Temporal Continual Learning (TCL) to address the above challenges. To better preserve prior information, we introduce the Prior Compensation Factor (PCF). We incorporate it into the model training to compensate for the lost prior information. Furthermore, we derive a more reasonable optimization objective through theoretical derivation. It is important to note that our TCL framework can be easily integrated with different HMP backbone models and adapted to various datasets and applications. Extensive experiments on four HMP benchmark datasets demonstrate the effectiveness and flexibility of TCL. The code is available at https://github.com/hyqlat/TCL.
****
Kernel Quadrature with Randomly Pivoted Cholesky
Ethan Epperly, Elvira Moreno
This paper presents new quadrature rules for functions in a reproducing kernel Hilbert space using nodes drawn by a sampling algorithm known as randomly pivoted Cholesky. The resulting computational procedure compares favorably to previous kernel quadrature methods, which either achieve low accuracy or require solving a computationally challenging sampling problem. Theoretical and numerical results show that randomly pivoted Cholesky is fast and achieves comparable quadrature error rates to more computationally expensive quadrature schemes based on continuous volume sampling, thinning, and recombination. Randomly pivoted Cholesky is easily adapted to complicated geometries with arbitrary kernels, unlocking new potential for kernel quadrature.
****
Analyzing the Sample Complexity of Self-Supervised Image Reconstruction Methods
Tobit Klug, Dogukan Atik, Reinhard Heckel
Supervised training of deep neural networks on pairs of clean image and noisy measurement achieves state-of-the-art performance for many image reconstruction tasks, but such training pairs are difficult to collect. Self-supervised methods enable training based on noisy measurements only, without clean images. In this work, we investigate the cost of self-supervised training in terms of sample complexity for a class of self-supervised methods that enable the computation of unbiased estimates of gradients of the supervised loss, including noise2noise methods. We analytically show that a model trained with such self-supervised training is as good as the same model trained in a supervised fashion, but self-supervised training requires more examples than supervised training. We then study self-supervised denoising and accelerated MRI empirically and characterize the cost of self-supervised training in terms of the number of additional samples required, and find that the performance gap between self-supervised and supervised training vanishes as a function of the training examples, at a problem-dependent rate

, as predicted by our theory.
****
FIRAL: An Active Learning Algorithm for Multinomial Logistic Regression
Youguang Chen, George Biros
We investigate theory and algorithms for pool-based active learning for multiclass classification using multinomial logistic regression. Using finite sample analysis, we prove that the Fisher Information Ratio (FIR) lower and upper bounds the excess risk. Based on our theoretical analysis, we propose an active learning algorithm that employs regret minimization to minimize the FIR. To verify our derived excess risk bounds, we conduct experiments on synthetic datasets. Furthermore, we compare FIRAL with five other methods and found that our scheme outperforms them: it consistently produces the smallest classification error in the multiclass logistic regression setting, as demonstrated through experiments on MNIST, CIFAR-10, and 50-class ImageNet.
****
AMDP: An Adaptive Detection Procedure for False Discovery Rate Control in High-Dimensional Mediation Analysis
Jiarong Ding, Xuehu ZHU
High-dimensional mediation analysis is often associated with a multiple testing problem for detecting significant mediators. Assessing the uncertainty of this detecting process via false discovery rate (FDR) has garnered great interest. To control the FDR in multiple testing, two essential steps are involved: ranking and selection. Existing approaches either construct p-values without calibration or disregard the joint information across tests, leading to conservation in FDR control or non-optimal ranking rules for multiple hypotheses. In this paper, we develop an adaptive mediation detection procedure (referred to as "AMDP") to identify relevant mediators while asymptotically controlling the FDR in high-dimensional mediation analysis. AMDP produces the optimal rule for ranking hypotheses and proposes a data-driven strategy to determine the threshold for mediator selection. This novel method captures information from the proportions of composite null hypotheses and the distribution of p-values, which turns the high dimensionality into an advantage instead of a limitation. The numerical studies on synthetic and real data sets illustrate the performances of AMDP compared with existing approaches.
****
Strong and Precise Modulation of Human Percepts via Robustified ANNs
Guy Gaziv, Michael Lee, James J DiCarlo
The visual object category reports of artificial neural networks (ANNs) are notoriously sensitive to tiny, adversarial image perturbations. Because human category reports (aka human percepts) are thought to be insensitive to those same small-norm perturbations -- and locally stable in general -- this argues that ANNs are incomplete scientific models of human visual perception. Consistent with this, we show that when small-norm image perturbations are generated by standard ANN models, human object category percepts are indeed highly stable. However, in this very same "human-presumed-stable" regime, we find that robustified ANNs reliably discover low-norm image perturbations that strongly disrupt human percepts. These previously undetectable human perceptual disruptions are massive in amplitude, approaching the same level of sensitivity seen in robustified ANNs. Further, we show that robustified ANNs support precise perceptual state interventions: they guide the construction of low-norm image perturbations that strongly alter human category percepts toward specific prescribed percepts. In sum, these contemporary models of biological visual processing are now accurate enough to guide strong and precise interventions on human perception.
****
MomentDiff: Generative Video Moment Retrieval from Random to Real
Pandeng Li, Chen-Wei Xie, Hongtao Xie, Liming Zhao, Lei Zhang, Yun Zheng, Deli Zhao, Yongdong Zhang
Video moment retrieval pursues an efficient and generalized solution to identify the specific temporal segments within an untrimmed video that correspond to a given language description.To achieve this goal, we provide a generative diffusio

n-based framework called MomentDiff, which simulates a typical human retrieval process from random browsing to gradual localization.Specifically, we first diffuse the real span to random noise, and learn to denoise the random noise to the original span with the guidance of similarity between text and video.This allows the model to learn a mapping from arbitrary random locations to real moments, enabling the ability to locate segments from random initialization.Once trained, MomentDiff could sample random temporal segments as initial guesses and iteratively refine them to generate an accurate temporal boundary.Different from discriminative works (e.g., based on learnable proposals or queries), MomentDiff with random initialized spans could resist the temporal location biases from datasets.To evaluate the influence of the temporal location biases, we propose two ``anti-bias'' datasets with location distribution shifts, named Charades-STA-Len and Charades-STA-Mom.The experimental results demonstrate that our efficient framework consistently outperforms state-of-the-art methods on three public benchmarks, and exhibits better generalization and robustness on the proposed anti-bias datasets. The code, model, and anti-bias evaluation datasets will be released publicly.
****
Experimental Designs for Heteroskedastic Variance
Justin Weltz, Tanner Fiez, Alexander Volfovsky, Eric Laber, Blake Mason, houssam nassif, Lalit Jain
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
NeuralGF: Unsupervised Point Normal Estimation by Learning Neural Gradient Function
Qing Li, Huifang Feng, Kanle Shi, Yue Gao, Yi Fang, Yu-Shen Liu, Zhizhong Han
Normal estimation for 3D point clouds is a fundamental task in 3D geometry processing. The state-of-the-art methods rely on priors of fitting local surfaces learned from normal supervision. However, normal supervision in benchmarks comes from synthetic shapes and is usually not available from real scans, thereby limiting the learned priors of these methods. In addition, normal orientation consistency across shapes remains difficult to achieve without a separate post-processing procedure. To resolve these issues, we propose a novel method for estimating oriented normals directly from point clouds without using ground truth normals as supervision. We achieve this by introducing a new paradigm for learning neural gradient functions, which encourages the neural network to fit the input point clouds and yield unit-norm gradients at the points. Specifically, we introduce loss functions to facilitate query points to iteratively reach the moving targets and aggregate onto the approximated surface, thereby learning a global surface representation of the data. Meanwhile, we incorporate gradients into the surface approximation to measure the minimum signed deviation of queries, resulting in a consistent gradient field associated with the surface. These techniques lead to our deep unsupervised oriented normal estimator that is robust to noise, outliers and density variations. Our excellent results on widely used benchmarks demonstrate that our method can learn more accurate normals for both unoriented and oriented normal estimation tasks than the latest methods. The source code and pre-trained model are publicly available.
****
Gacs-Korner Common Information Variational Autoencoder
Michael Kleinman, Alessandro Achille, Stefano Soatto, Jonathan Kao
We propose a notion of common information that allows one to quantify and separate the information that is shared between two random variables from the information that is unique to each. Our notion of common information is defined by an optimization problem over a family of functions and recovers the G\'acs-K\"orner common information as a special case. Importantly, our notion can be approximated empirically using samples from the underlying data distribution. We then provide a method to partition and quantify the common and unique information using a s

imple modification of a traditional variational auto-encoder. Empirically, we demonstrate that our formulation allows us to learn semantically meaningful common and unique factors of variation even on high-dimensional data such as images and videos. Moreover, on datasets where ground-truth latent factors are known, we show that we can accurately quantify the common information between the random variables.
****
LEACE: Perfect linear concept erasure in closed form
Nora Belrose, David Schneider-Joseph, Shauli Ravfogel, Ryan Cotterell, Edward Raff, Stella Biderman
Concept erasure aims to remove specified features from a representation. It can improve fairness (e.g. preventing a classifier from using gender or race) and interpretability (e.g. removing a concept to observe changes in model behavior). We introduce LEAst-squares Concept Erasure (LEACE), a closed-form method which provably prevents all linear classifiers from detecting a concept while changing the representation as little as possible, as measured by a broad class of norms. We apply LEACE to large language models with a novel procedure called concept scrubbing, which erases target concept information from every layer in the network. We demonstrate our method on two tasks: measuring the reliance of language models on part-of-speech information, and reducing gender bias in BERT embeddings. Our code is available at https://github.com/EleutherAI/concept-erasure.
****
Robust low-rank training via approximate orthonormal constraints
Dayana Savostianova, Emanuele Zangrando, Gianluca Ceruti, Francesco Tudisco
With the growth of model and data sizes, a broad effort has been made to design pruning techniques that reduce the resource demand of deep learning pipelines, while retaining model performance. In order to reduce both inference and training costs, a prominent line of work uses low-rank matrix factorizations to represent the network weights. Although able to retain accuracy, we observe that low-rank methods tend to compromise model robustness against adversarial perturbations. By modeling robustness in terms of the condition number of the neural network, we argue that this loss of robustness is due to the exploding singular values of the low-rank weight matrices. Thus, we introduce a robust low-rank training algorithm that maintains the network's weights on the low-rank matrix manifold while simultaneously enforcing approximate orthonormal constraints. The resulting model reduces both training and inference costs while ensuring well-conditioning and thus better adversarial robustness, without compromising model accuracy. This is shown by extensive numerical evidence and by our main approximation theorem that shows the computed robust low-rank network well-approximates the ideal full model, provided a highly performing low-rank sub-network exists.
****
Symmetry-Informed Geometric Representation for Molecules, Proteins, and Crystalline Materials
Shengchao Liu, weitao Du, Yanjing Li, Zhuoxinran Li, Zhiling Zheng, Chenru Duan, Zhi-Ming Ma, Omar Yaghi, Animashree Anandkumar, Christian Borgs, Jennifer Chayes, Hongyu Guo, Jian Tang
Artificial intelligence for scientific discovery has recently generated significant interest within the machine learning and scientific communities, particularly in the domains of chemistry, biology, and material discovery. For these scientific problems, molecules serve as the fundamental building blocks, and machine learning has emerged as a highly effective and powerful tool for modeling their geometric structures. Nevertheless, due to the rapidly evolving process of the field and the knowledge gap between science ({\eg}, physics, chemistry, \& biology) and machine learning communities, a benchmarking study on geometrical representation for such data has not been conducted. To address such an issue, in this paper, we first provide a unified view of the current symmetry-informed geometric methods, classifying them into three main categories: invariance, equivariance with spherical frame basis, and equivariance with vector frame basis. Then we propose a platform, coined Geom3D, which enables benchmarking the effectiveness of geometric strategies. Geom3D contains 16 advanced symmetry-informed geometric

representation models and 14 geometric pretraining methods over 52 diverse tasks, including small molecules, proteins, and crystalline materials. We hope that Geom3D can, on the one hand, eliminate barriers for machine learning researchers interested in exploring scientific problems; and, on the other hand, provide valuable guidance for researchers in computational chemistry, structural biology, and materials science, aiding in the informed selection of representation techniques for specific applications. The source code is available on \href{https://github.com/chao1224/Geom3D}{the GitHub repository}.
****

Feature Selection in the Contrastive Analysis Setting
Ethan Weinberger, Ian Covert, Su-In Lee
Contrastive analysis (CA) refers to the exploration of variations uniquely enriched in a target dataset as compared to a corresponding background dataset generated from sources of variation that are irrelevant to a given task. For example, a biomedical data analyst may wish to find a small set of genes to use as a proxy for variations in genomic data only present among patients with a given disease (target) as opposed to healthy control subjects (background). However, as of yet the problem of feature selection in the CA setting has received little attention from the machine learning community. In this work we present contrastive feature selection (CFS),a method for performing feature selection in the CA setting. We motivate our approach with a novel information-theoretic analysis of representation learning in the CA setting, and we empirically validate CFS on a semi-synthetic dataset and four real-world biomedical datasets. We find that our method consistently outperforms previously proposed state-of-the-art supervised and fully unsupervised feature selection methods not designed for the CA setting. An open-source implementation of our method is available at https://github.com/suinleelab/CFS.
****

GeoDE: a Geographically Diverse Evaluation Dataset for Object Recognition
Vikram V. Ramaswamy, Sing Yu Lin, Dora Zhao, Aaron Adcock, Laurens van der Maaten, Deepti Ghadiyaram, Olga Russakovsky
Current dataset collection methods typically scrape large amounts of data from the web. While this technique is extremely scalable, data collected in this way tends to reinforce stereotypical biases, can contain personally identifiable information, and typically originates from Europe and North America. In this work, we rethink the dataset collection paradigm and introduce GeoDE, a geographically diverse dataset with 61,940 images from 40 classes and 6 world regions, and no personally identifiable information, collected by soliciting images from people across the world. We analyse GeoDE to understand differences in images collected in this manner compared to web-scraping. Despite the smaller size of this dataset, we demonstrate its use as both an evaluation and training dataset, allowing us to highlight shortcomings in current models, as well as demonstrate improved performance even when training on this small dataset. We release the full dataset and code at https://geodiverse-data-collection.cs.princeton.edu/
****

Last-Iterate Convergent Policy Gradient Primal-Dual Methods for Constrained MDPs
Dongsheng Ding, Chen-Yu Wei, Kaiqing Zhang, Alejandro Ribeiro
We study the problem of computing an optimal policy of an infinite-horizon discounted constrained Markov decision process (constrained MDP). Despite the popularity of Lagrangian-based policy search methods used in practice, the oscillation of policy iterates in these methods has not been fully understood, bringing out issues such as violation of constraints and sensitivity to hyper-parameters. To fill this gap, we employ the Lagrangian method to cast a constrained MDP into a constrained saddle-point problem in which max/min players correspond to primal/dual variables, respectively, and develop two single-time-scale policy-based primal-dual algorithms with non-asymptotic convergence of their policy iterates to an optimal constrained policy. Specifically, we first propose a regularized policy gradient primal-dual (RPG-PD) method that updates the policy using an entropy-regularized policy gradient, and the dual variable via a quadratic-regularized gradient ascent, simultaneously. We prove that the policy primal-dual iterates of

RPG-PD converge to a regularized saddle point with a sublinear rate, while the policy iterates converge sublinearly to an optimal constrained policy. We further instantiate RPG-PD in large state or action spaces by including function approximation in policy parametrization, and establish similar sublinear last-iterate policy convergence. Second, we propose an optimistic policy gradient primal-dual (OPG-PD) method that employs the optimistic gradient method to update primal/dual variables, simultaneously. We prove that the policy primal-dual iterates of OPG-PD converge to a saddle point that contains an optimal constrained policy, with a linear rate. To the best of our knowledge, this work appears to be the first non-asymptotic policy last-iterate convergence result for single-time-scale algorithms in constrained MDPs. We further validate the merits and the effectiveness of our methods in computational experiments.
****
Unleashing the Power of Randomization in Auditing Differentially Private ML
Krishna Pillutla, Galen Andrew, Peter Kairouz, H. Brendan McMahan, Alina Oprea, Sewoong Oh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Worst-case Performance of Popular Approximate Nearest Neighbor Search Implementations: Guarantees and Limitations
Piotr Indyk, Haike Xu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
On the Overlooked Structure of Stochastic Gradients
Zeke Xie, Qian-Yuan Tang, Mingming Sun, Ping Li
Stochastic gradients closely relate to both optimization and generalization of deep neural networks (DNNs). Some works attempted to explain the success of stochastic optimization for deep learning by the arguably heavy-tail properties of gradient noise, while other works presented theoretical and empirical evidence against the heavy-tail hypothesis on gradient noise. Unfortunately, formal statistical tests for analyzing the structure and heavy tails of stochastic gradients in deep learning are still under-explored. In this paper, we mainly make two contributions. First, we conduct formal statistical tests on the distribution of stochastic gradients and gradient noise across both parameters and iterations. Our statistical tests reveal that dimension-wise gradients usually exhibit power-law heavy tails, while iteration-wise gradients and stochastic gradient noise caused by minibatch training usually do not exhibit power-law heavy tails. Second, we further discover that the covariance spectra of stochastic gradients have the power-law structures overlooked by previous studies and present its theoretical implications for training of DNNs. While previous studies believed that the anisotropic structure of stochastic gradients matters to deep learning, they did not expect the gradient covariance can have such an elegant mathematical structure. Our work challenges the existing belief and provides novel insights on the structure of stochastic gradients in deep learning.
****
ECG-QA: A Comprehensive Question Answering Dataset Combined With Electrocardiogram
Jungwoo Oh, Gyubok Lee, Seongsu Bae, Joon-myoung Kwon, Edward Choi
Question answering (QA) in the field of healthcare has received much attention due to significant advancements in natural language processing. However, existing healthcare QA datasets primarily focus on medical images, clinical notes, or structured electronic health record tables. This leaves the vast potential of combining electrocardiogram (ECG) data with these systems largely untapped. To address this gap, we present ECG-QA, the first QA dataset specifically designed for E

CG analysis. The dataset comprises a total of 70 question templates that cover a wide range of clinically relevant ECG topics, each validated by an ECG expert to ensure their clinical utility. As a result, our dataset includes diverse ECG interpretation questions, including those that require a comparative analysis of two different ECGs. In addition, we have conducted numerous experiments to provide valuable insights for future research directions. We believe that ECG-QA will serve as a valuable resource for the development of intelligent QA systems capable of assisting clinicians in ECG interpretations.
****
Provable convergence guarantees for black-box variational inference
Justin Domke, Robert Gower, Guillaume Garrigos
Black-box variational inference is widely used in situations where there is no proof that its stochastic optimization succeeds. We suggest this is due to a theoretical gap in existing stochastic optimization proofs—namely the challenge of gradient estimators with unusual noise bounds, and a composite non-smooth objective. For dense Gaussian variational families, we observe that existing gradient estimators based on reparameterization satisfy a quadratic noise bound and give novel convergence guarantees for proximal and projected stochastic gradient descent using this bound. This provides rigorous guarantees that methods similar to those used in practice converge on realistic inference problems.
****
Seeing is not Believing: Robust Reinforcement Learning against Spurious Correlation
Wenhao Ding, Laixi Shi, Yuejie Chi, DING ZHAO
Robustness has been extensively studied in reinforcement learning (RL) to handle various forms of uncertainty such as random perturbations, rare events, and malicious attacks. In this work, we consider one critical type of robustness against spurious correlation, where different portions of the state do not have correlations induced by unobserved confounders. These spurious correlations are ubiquitous in real-world tasks, for instance, a self-driving car usually observes heavy traffic in the daytime and light traffic at night due to unobservable human activity. A model that learns such useless or even harmful correlation could catastrophically fail when the confounder in the test case deviates from the training one. Although motivated, enabling robustness against spurious correlation poses significant challenges since the uncertainty set, shaped by the unobserved confounder and causal structure, is difficult to characterize and identify. Existing robust algorithms that assume simple and unstructured uncertainty sets are therefore inadequate to address this challenge. To solve this issue, we propose Robust State-Confounded Markov Decision Processes (RSC-MDPs) and theoretically demonstrate its superiority in avoiding learning spurious correlations compared with other robust RL counterparts. We also design an empirical algorithm to learn the robust optimal policy for RSC-MDPs, which outperforms all baselines in eight realistic self-driving and manipulation tasks.
****
IBA: Towards Irreversible Backdoor Attacks in Federated Learning
Thuy Dung Nguyen, Tuan A. Nguyen, Anh Tran, Khoa D Doan, Kok-Seng Wong
Federated learning (FL) is a distributed learning approach that enables machine learning models to be trained on decentralized data without compromising end devices' personal, potentially sensitive data. However, the distributed nature and uninvestigated data intuitively introduce new security vulnerabilities, including backdoor attacks. In this scenario, an adversary implants backdoor functionality into the global model during training, which can be activated to cause the desired misbehaviors for any input with a specific adversarial pattern. Despite having remarkable success in triggering and distorting model behavior, prior backdoor attacks in FL often hold impractical assumptions, limited imperceptibility, and durability. Specifically, the adversary needs to control a sufficiently large fraction of clients or know the data distribution of other honest clients. In many cases, the trigger inserted is often visually apparent, and the backdoor effect is quickly diluted if the adversary is removed from the training process. To address these limitations, we propose a novel backdoor attack framework in FL,

the Irreversible Backdoor Attack (IBA), that jointly learns the optimal and vis
ually stealthy trigger and then gradually implants the backdoor into a global mo
del. This approach allows the adversary to execute a backdoor attack that can ev
ade both human and machine inspections. Additionally, we enhance the efficiency
and durability of the proposed attack by selectively poisoning the model's param
eters that are least likely updated by the main task's learning process and cons
training the poisoned model update to the vicinity of the global model. Finally,
 we evaluate the proposed attack framework on several benchmark datasets, includ
ing MNIST, CIFAR-10, and Tiny ImageNet, and achieved high success rates while si
multaneously bypassing existing backdoor defenses and achieving a more durable b
ackdoor effect compared to other backdoor attacks. Overall, IBA offers a more ef
fective, stealthy, and durable approach to backdoor attacks in FL. The code asso
ciated with this paper is available on GitHub.
****
Spontaneous symmetry breaking in generative diffusion models
Gabriel Raya, Luca Ambrogioni
Generative diffusion models have recently emerged as a leading approach for gene
rating high-dimensional data. In this paper, we show that the dynamics of these
models exhibit a spontaneous symmetry breaking that divides the generative dynam
ics into two distinct phases: 1) A linear steady-state dynamics around a central
 fixed-point and 2) an attractor dynamics directed towards the data manifold. Th
ese two "phases'' are separated by the change in stability of the central fixed-
point, with the resulting window of instability being responsible for the divers
ity of the generated samples. Using both theoretical and empirical evidence, we
show that an accurate simulation of the early dynamics does not significantly co
ntribute to the final generation, since early fluctuations are reverted to the c
entral fixed point. To leverage this insight, we propose a Gaussian late initial
ization scheme, which significantly improves model performance, achieving up to
3x FID improvements on fast samplers, while also increasing sample diversity (e.
g., racial composition of generated CelebA images). Our work offers a new way to
 understand the generative dynamics of diffusion models that has the potential t
o bring about higher performance and less biased fast-samplers.
****
RL-based Stateful Neural Adaptive Sampling and Denoising for Real-Time Path Trac
ing
Antoine Scardigli, Lukas Cavigelli, Lorenz K. Müller
Monte-Carlo path tracing is a powerful technique for realistic image synthesis b
ut suffers from high levels of noise at low sample counts, limiting its use in r
eal-time applications. To address this, we propose a framework with end-to-end t
raining of a sampling importance network, a latent space encoder network, and a
denoiser network. Our approach uses reinforcement learning to optimize the sampl
ing importance network, thus avoiding explicit numerically approximated gradient
s. Our method does not aggregate the sampled values per pixel by averaging but k
eeps all sampled values which are then fed into the latent space encoder. The en
coder replaces handcrafted spatiotemporal heuristics by learned representations
in a latent space. Finally, a neural denoiser is trained to refine the output im
age. Our approach increases visual quality on several challenging datasets and r
educes rendering times for equal quality by a factor of 1.6x compared to the pre
vious state-of-the-art, making it a promising solution for real-time application
s.
****
A Data-Free Approach to Mitigate Catastrophic Forgetting in Federated Class Incr
emental Learning for Vision Tasks
Sara Babakniya, Zalan Fabian, Chaoyang He, Mahdi Soltanolkotabi, Salman Avestime
hr
Deep learning models often suffer from forgetting previously learned information
 when trained on new data. This problem is exacerbated in federated learning (FL
), where the data is distributed and can change independently for each user. Man
y solutions are proposed to resolve this catastrophic forgetting in a centralize
d setting. However, they do not apply directly to FL because of its unique compl

exities, such as privacy concerns and resource limitations. To overcome these challenges, this paper presents a framework for \textbf{federated class incremental learning} that utilizes a generative model to synthesize samples from past distributions. This data can be later exploited alongside the training data to mitigate catastrophic forgetting. To preserve privacy, the generative model is trained on the server using data-free methods at the end of each task without requesting data from clients. Moreover, our solution does not demand the users to store old data or models, which gives them the freedom to join/leave the training at any time. Additionally, we introduce SuperImageNet, a new regrouping of the ImageNet dataset specifically tailored for federated continual learning. We demonstrate significant improvements compared to existing baselines through extensive experiments on multiple datasets.
****
On the Connection between Pre-training Data Diversity and Fine-tuning Robustness
Vivek Ramanujan, Thao Nguyen, Sewoong Oh, Ali Farhadi, Ludwig Schmidt
Pre-training has been widely adopted in deep learning to improve model performance, especially when the training data for a target task is limited. In our work, we seek to understand the implications of this training strategy on the generalization properties of downstream models. More specifically, we ask the following question: how do properties of the pre-training distribution affect the robustness of a fine-tuned model? The properties we explore include the label space, label semantics, image diversity, data domains, and data quantity of the pre-training distribution. We find that the primary factor influencing downstream effective robustness (Taori et al., 2020) is data quantity, while other factors have limited significance. For example, reducing the number of ImageNet pre-training classes by 4x while increasing the number of images per class by 4x (that is, keeping total data quantity fixed) does not impact the robustness of fine-tuned models. We demonstrate our findings on pre-training distributions drawn from various natural and synthetic data sources, primarily using the iWildCam-WILDS distribution shift as a test for robustness.
****
Structured Neural Networks for Density Estimation and Causal Inference
Asic Chen, Ruian (Ian) Shi, Xiang Gao, Ricardo Baptista, Rahul G. Krishnan
Injecting structure into neural networks enables learning functions that satisfy invariances with respect to subsets of inputs. For instance, when learning generative models using neural networks, it is advantageous to encode the conditional independence structure of observed variables, often in the form of Bayesian networks. We propose the Structured Neural Network (StrNN), which injects structure through masking pathways in a neural network. The masks are designed via a novel relationship we explore between neural network architectures and binary matrix factorization, to ensure that the desired independencies are respected. We devise and study practical algorithms for this otherwise NP-hard design problem based on novel objectives that control the model architecture. We demonstrate the utility of StrNN in three applications: (1) binary and Gaussian density estimation with StrNN, (2) real-valued density estimation with Structured Autoregressive Flows (StrAFs) and Structured Continuous Normalizing Flows (StrCNF), and (3) interventional and counterfactual analysis with StrAFs for causal inference. Our work opens up new avenues for learning neural networks that enable data-efficient generative modeling and the use of normalizing flows for causal effect estimation.
****
Decision-Aware Actor-Critic with Function Approximation and Theoretical Guarantees
Sharan Vaswani, Amirreza Kazemi, Reza Babanezhad Harikandeh, Nicolas Le Roux
Actor-critic (AC) methods are widely used in reinforcement learning (RL), and benefit from the flexibility of using any policy gradient method as the actor and value-based method as the critic. The critic is usually trained by minimizing the TD error, an objective that is potentially decorrelated with the true goal of achieving a high reward with the actor. We address this mismatch by designing a joint objective for training the actor and critic in a decision-aware fashion. W

e use the proposed objective to design a generic, AC algorithm that can easily handle any function approximation. We explicitly characterize the conditions under which the resulting algorithm guarantees monotonic policy improvement, regardless of the choice of the policy and critic parameterization. Instantiating the generic algorithm results in an actor that involves maximizing a sequence of surrogate functions (similar to TRPO, PPO), and a critic that involves minimizing a closely connected objective. Using simple bandit examples, we provably establish the benefit of the proposed critic objective over the standard squared error. Finally, we empirically demonstrate the benefit of our decision-aware actor-critic framework on simple RL problems.
****
Label-Retrieval-Augmented Diffusion Models for Learning from Noisy Labels
Jian Chen, Ruiyi Zhang, Tong Yu, Rohan Sharma, Zhiqiang Xu, Tong Sun, Changyou Chen
Learning from noisy labels is an important and long-standing problem in machine learning for real applications. One of the main research lines focuses on learning a label corrector to purify potential noisy labels. However, these methods typically rely on strict assumptions and are limited to certain types of label noise. In this paper, we reformulate the label-noise problem from a generative-model perspective, i.e., labels are generated by gradually refining an initial random guess. This new perspective immediately enables existing powerful diffusion models to seamlessly learn the stochastic generative process. Once the generative uncertainty is modeled, we can perform classification inference using maximum likelihood estimation of labels. To mitigate the impact of noisy labels, we propose the Label-Retrieval-Augmented (LRA) diffusion model, which leverages neighbor consistency to effectively construct pseudo-clean labels for diffusion training. Our model is flexible and general, allowing easy incorporation of different types of conditional information, e.g., use of pre-trained models, to further boost model performance. Extensive experiments are conducted for evaluation. Our model achieves new state-of-the-art (SOTA) results on all the standard real-world benchmark datasets. Remarkably, by incorporating conditional information from the powerful CLIP model, our method can boost the current SOTA accuracy by 10-20 absolute points in many cases. Code is available: https://anonymous.4open.science/r/LRA-diffusion-5F2F
****
Cheaply Estimating Inference Efficiency Metrics for Autoregressive Transformer Models
Deepak Narayanan, Keshav Santhanam, Peter Henderson, Rishi Bommasani, Tony Lee, Percy S. Liang
Large language models (LLMs) are highly capable but also computationally expensive. Characterizing the fundamental tradeoff between inference efficiency and model capabilities is thus important, but requires an efficiency metric that is comparable across models from different providers.Unfortunately, raw runtimes measured through black-box APIs do not satisfy this property: model providers can implement software and hardware optimizations orthogonal to the model, and shared infrastructure introduces performance contention.We propose a new metric for inference efficiency called idealized runtime, that puts models on equal footing as though they were served on uniform hardware and software without performance contention, and a cost model to efficiently estimate this metric for autoregressive Transformer models.We also propose variants of the idealized runtime that incorporate the number and type of accelerators needed to serve the model.Using these metrics, we compare ten LLMs developed in 2022 to provide the first analysis of inference efficiency-capability tradeoffs; we make several observations from this analysis, including the fact that the superior inference runtime performance of certain APIs is often a byproduct of optimizations within the API rather than the underlying model.Our code is open sourced at https://github.com/stanford-crfm/helm-efficiency.
****
Differentiable sorting for censored time-to-event data.
Andre Vauvelle, Benjamin Wild, Roland Eils, Spiros Denaxas

Survival analysis is a crucial semi-supervised task in machine learning with significant real-world applications, especially in healthcare. The most common approach to survival analysis, Cox’s partial likelihood, can be interpreted as a ranking model optimized on a lower bound of the concordance index. We follow these connections further, with listwise ranking losses that allow for a relaxation of the pairwise independence assumption. Given the inherent transitivity of ranking, we explore differentiable sorting networks as a means to introduce a stronger transitive inductive bias during optimization. Despite their potential, current differentiable sorting methods cannot account for censoring, a crucial aspect of many real-world datasets. We propose a novel method, Diffsurv, to overcome this limitation by extending differentiable sorting methods to handle censored tasks. Diffsurv predicts matrices of possible permutations that accommodate the label uncertainty introduced by censored samples. Our experiments reveal that Diffsurv outperforms established baselines in various simulated and real-world risk prediction scenarios. Furthermore, we demonstrate the algorithmic advantages of Diffsurv by presenting a novel method for top-k risk prediction that surpasses current methods.
****
Multi-Agent Meta-Reinforcement Learning: Sharper Convergence Rates with Task Similarity
Weichao Mao, Haoran Qiu, Chen Wang, Hubertus Franke, Zbigniew Kalbarczyk, Ravishankar Iyer, Tamer Basar
Multi-agent reinforcement learning (MARL) has primarily focused on solving a single task in isolation, while in practice the environment is often evolving, leaving many related tasks to be solved. In this paper, we investigate the benefits of meta-learning in solving multiple MARL tasks collectively. We establish the first line of theoretical results for meta-learning in a wide range of fundamental MARL settings, including learning Nash equilibria in two-player zero-sum Markov games and Markov potential games, as well as learning coarse correlated equilibria in general-sum Markov games. Under natural notions of task similarity, we show that meta-learning achieves provable sharper convergence to various game-theoretical solution concepts than learning each task separately. As an important intermediate step, we develop multiple MARL algorithms with initialization-dependent convergence guarantees. Such algorithms integrate optimistic policy mirror descents with stage-based value updates, and their refined convergence guarantees (nearly) recover the best known results even when a good initialization is unknown. To our best knowledge, such results are also new and might be of independent interest. We further provide numerical simulations to corroborate our theoretical findings.
****
Feature Learning for Interpretable, Performant Decision Trees
Jack Good, Torin Kovach, Kyle Miller, Artur Dubrawski
Decision trees are regarded for high interpretability arising from their hierarchical partitioning structure built on simple decision rules. However, in practice, this is not realized because axis-aligned partitioning of realistic data results in deep trees, and because ensemble methods are used to mitigate overfitting. Even then, model complexity and performance remain sensitive to transformation of the input, and extensive expert crafting of features from the raw data is common. We propose the first system to alternate sparse feature learning with differentiable decision tree construction to produce small, interpretable trees with good performance. We benchmark against conventional tree-based models and demonstrate several notions of interpretation of a model and its predictions.
****
Particle-based Variational Inference with Generalized Wasserstein Gradient Flow
Ziheng Cheng, Shiyue Zhang, Longlin Yu, Cheng Zhang
Particle-based variational inference methods (ParVIs) such as Stein variational gradient descent (SVGD) update the particles based on the kernelized Wasserstein gradient flow for the Kullback-Leibler (KL) divergence. However, the design of kernels is often non-trivial and can be restrictive for the flexibility of the method. Recent works show that functional gradient flow approximations with quadr

atic form regularization terms can improve performance. In this paper, we propose a ParVI framework, called generalized Wasserstein gradient descent (GWG), based on a generalized Wasserstein gradient flow of the KL divergence, which can be viewed as a functional gradient method with a broader class of regularizers induced by convex functions. We show that GWG exhibits strong convergence guarantees. We also provide an adaptive version that automatically chooses Wasserstein metric to accelerate convergence. In experiments, we demonstrate the effectiveness and efficiency of the proposed framework on both simulated and real data problems.
****
PAC Learning Linear Thresholds from Label Proportions
Anand Brahmbhatt, Rishi Saket, Aravindan Raghuveer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A new perspective on building efficient and expressive 3D equivariant graph neural networks
weitao Du, Yuanqi Du, Limei Wang, Dieqiao Feng, Guifeng Wang, Shuiwang Ji, Carla P. Gomes, Zhi-Ming Ma
Geometric deep learning enables the encoding of physical symmetries in modeling 3D objects. Despite rapid progress in encoding 3D symmetries into Graph Neural Networks (GNNs), a comprehensive evaluation of the expressiveness of these network architectures through a local-to-global analysis lacks today. In this paper, we propose a local hierarchy of 3D isomorphism to evaluate the expressive power of equivariant GNNs and investigate the process of representing global geometric information from local patches. Our work leads to two crucial modules for designing expressive and efficient geometric GNNs; namely local substructure encoding (\textbf{LSE}) and frame transition encoding (\textbf{FTE}). To demonstrate the applicability of our theory, we propose LEFTNet which effectively implements these modules and achieves state-of-the-art performance on both scalar-valued and vector-valued molecular property prediction tasks. We further point out future design space for 3D equivariant graph neural networks. Our codes are available at \url{https://github.com/yuanqidu/LeftNet}.
****
Scalable Fair Influence Maximization
Xiaobin Rui, Zhixiao Wang, Jiayu Zhao, Lichao Sun, Wei Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Calibrate and Boost Logical Expressiveness of GNN Over Multi-Relational and Temporal Graphs
Dingmin Wang, Yeyuan Chen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Task Arithmetic in the Tangent Space: Improved Editing of Pre-Trained Models
Guillermo Ortiz-Jimenez, Alessandro Favero, Pascal Frossard
Task arithmetic has recently emerged as a cost-effective and scalable approach to edit pre-trained models directly in weight space: By adding the fine-tuned weights of different tasks, the model's performance can be improved on these tasks, while negating them leads to task forgetting. Yet, our understanding of the effectiveness of task arithmetic and its underlying principles remains limited. We present a comprehensive study of task arithmetic in vision-language models and show that weight disentanglement is the crucial factor that makes it effective. T

his property arises during pre-training and manifests when distinct directions in weight space govern separate, localized regions in function space associated with the tasks. Notably, we show that fine-tuning models in their tangent space by linearizing them amplifies weight disentanglement. This leads to substantial performance improvements across multiple task arithmetic benchmarks and diverse models. Building on these findings, we provide theoretical and empirical analyses of the neural tangent kernel (NTK) of these models and establish a compelling link between task arithmetic and the spatial localization of the NTK eigenfunctions. Overall, our work uncovers novel insights into the fundamental mechanisms of task arithmetic and offers a more reliable and effective approach to edit pre-trained models through the NTK linearization.
****
ParaFuzz: An Interpretability-Driven Technique for Detecting Poisoned Samples in NLP
Lu Yan, Zhuo Zhang, Guanhong Tao, Kaiyuan Zhang, Xuan Chen, Guangyu Shen, Xiangyu Zhang
Backdoor attacks have emerged as a prominent threat to natural language processing (NLP) models, where the presence of specific triggers in the input can lead poisoned models to misclassify these inputs to predetermined target classes. Current detection mechanisms are limited by their inability to address more covert backdoor strategies, such as style-based attacks. In this work, we propose an innovative test-time poisoned sample detection framework that hinges on the interpretability of model predictions, grounded in the semantic meaning of inputs.We contend that triggers (e.g., infrequent words) are not supposed to fundamentally alter the underlying semantic meanings of poisoned samples as they want to stay stealthy. Based on this observation, we hypothesize that while the model's predictions for paraphrased clean samples should remain stable, predictions for poisoned samples should revert to their true labels upon the mutations applied to triggers during the paraphrasing process.We employ ChatGPT, a state-of-the-art large language model, as our paraphraser and formulate the trigger-removal task as a prompt engineering problem. We adopt fuzzing, a technique commonly used for unearthing software vulnerabilities, to discover optimal paraphrase prompts that can effectively eliminate triggers while concurrently maintaining input semantics.Experiments on 4 types of backdoor attacks, including the subtle style backdoors, and 4 distinct datasets demonstrate that our approach surpasses baseline methods, including STRIP, RAP, and ONION, in precision and recall.
****
DiViNeT: 3D Reconstruction from Disparate Views using Neural Template Regularization
Aditya Vora, Akshay Gadi Patil, Hao Zhang
We present a volume rendering-based neural surface reconstruction method that takes as few as three disparate RGB images as input. Our key idea is to regularize the reconstruction, which is severely ill-posed and leaving significant gaps between the sparse views, by learning a set of neural templates that act as surface priors. Our method, coined DiViNet, operates in two stages. The first stage learns the templates, in the form of 3D Gaussian functions, across different scenes, without 3D supervision. In the reconstruction stage, our predicted templates serve as anchors to help “stitch” the surfaces over sparse regions. We demonstrate that our approach is not only able to complete the surface geometry but also reconstructs surface details to a reasonable extent from few disparate input views. On the DTU and BlendedMVS datasets, our approach achieves the best reconstruction quality among existing methods in the presence of such sparse views and performs on par, if not better, with competing methods when dense views are employed as inputs.
****
Efficient Model-Free Exploration in Low-Rank MDPs
Zak Mhammedi, Adam Block, Dylan J Foster, Alexander Rakhlin
A major challenge in reinforcement learning is to develop practical, sample-efficient algorithms for exploration in high-dimensional domains where generalization and function approximation is required. Low-Rank Markov Decision Processes---w

here transition probabilities admit a low-rank factorization based on an unknown feature embedding---offer a simple, yet expressive framework for RL with function approximation, yet existing algorithms either (1) are computationally intractable, or (2) require restrictive statistical assumptions such as latent variable structure or access to model-based function approximation. In this work, we propose the first provably sample-efficient algorithm for exploration in Low-Rank MDPs that is both computationally efficient and model-free, allowing for general function approximation while requiring no structural assumptions beyond a reachability condition that we show is substantially weaker than that assumed in prior work. Our algorithm, SpanRL, uses the notion of a barycentric spanner for the feature embedding as an efficiently computable basis for exploration, performing efficient spanner computation by interleaving representation learning and policy optimization subroutines. Our analysis---which is appealingly simple and modular---carefully combines several techniques, including a new approach to error-tolerant barycentric spanner computation, and a new analysis of a certain minimax representation learning objective found in prior work.
****
Convex-Concave Zero-Sum Markov Stackelberg Games
Denizalp Goktas, Arjun Prakash, Amy Greenwald
Zero-sum Markov Stackelberg games can be used to model myriad problems, in domains ranging from economics to human robot interaction. In this paper, we develop policy gradient methods that solve these games in continuous state and action settings using noisy gradient estimates computed from observed trajectories of play. When the games are convex-concave, we prove that our algorithms converge to Stackelberg equilibrium in polynomial time. We also show that reach-avoid problems are naturally modeled as convex-concave zero-sum Markov Stackelberg games, and that Stackelberg equilibrium policies are more effective than their Nash counterparts in these problems.
****
Near-Linear Time Algorithm for the Chamfer Distance
Ainesh Bakshi, Piotr Indyk, Rajesh Jayaram, Sandeep Silwal, Erik Waingarten
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Double Pessimism is Provably Efficient for Distributionally Robust Offline Reinforcement Learning: Generic Algorithm and Robust Partial Coverage
Jose Blanchet, Miao Lu, Tong Zhang, Han Zhong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
StyleDrop: Text-to-Image Synthesis of Any Style
Kihyuk Sohn, Lu Jiang, Jarred Barber, Kimin Lee, Nataniel Ruiz, Dilip Krishnan, Huiwen Chang, Yuanzhen Li, Irfan Essa, Michael Rubinstein, Yuan Hao, Glenn Entis, Irina Blok, Daniel Castro Chin
Pre-trained large text-to-image models synthesize impressive images with an appropriate use of text prompts. However, ambiguities inherent in natural language, and out-of-distribution effects make it hard to synthesize arbitrary image styles, leveraging a specific design pattern, texture or material. In this paper, we introduce StyleDrop, a method that enables the synthesis of images that faithfully follow a specific style using a text-to-image model. StyleDrop is extremely versatile and captures nuances and details of a user-provided style, such as color schemes, shading, design patterns, and local and global effects. StyleDrop works by efficiently learning a new style by fine-tuning very few trainable parameters (less than 1\% of total model parameters), and improving the quality via iterative training with either human or automated feedback. Better yet, StyleDrop is able to deliver impressive results even when the user supplies only a single i

mage specifying the desired style. An extensive study shows that, for the task of style tuning text-to-image models, StyleDrop on Muse convincingly outperforms other methods, including DreamBooth and textual inversion on Imagen or Stable Diffusion. More results are available at our project website: https://styledrop.github.io.
****
Random Cuts are Optimal for Explainable k-Medians
Konstantin Makarychev, Liren Shan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Order Matters in the Presence of Dataset Imbalance for Multilingual Learning
Dami Choi, Derrick Xin, Hamid Dadkhahi, Justin Gilmer, Ankush Garg, Orhan Firat, Chih-Kuan Yeh, Andrew M. Dai, Behrooz Ghorbani
In this paper, we empirically study the optimization dynamics of multi-task learning, particularly focusing on those that govern a collection of tasks with significant data imbalance. We present a simple yet effective method of pre-training on high-resource tasks, followed by fine-tuning on a mixture of high/low-resource tasks. We provide a thorough empirical study and analysis of this method's benefits showing that it achieves consistent improvements relative to the performance trade-off profile of standard static weighting. We analyze under what data regimes this method is applicable and show its improvements empirically in neural machine translation (NMT) and multi-lingual language modeling.
****
Optimizing Prompts for Text-to-Image Generation
Yaru Hao, Zewen Chi, Li Dong, Furu Wei
Well-designed prompts can guide text-to-image models to generate amazing images. However, the performant prompts are often model-specific and misaligned with user input. Instead of laborious human engineering, we propose prompt adaptation, a general framework that automatically adapts original user input to model-preferred prompts. Specifically, we first perform supervised fine-tuning with a pretrained language model on a small collection of manually engineered prompts. Then we use reinforcement learning to explore better prompts. We define a reward function that encourages the policy to generate more aesthetically pleasing images while preserving the original user intentions. Experimental results on Stable Diffusion show that our method outperforms manual prompt engineering in terms of both automatic metrics and human preference ratings. Moreover, reinforcement learning further boosts performance, especially on out-of-domain prompts.
****
Improved Bayes Risk Can Yield Reduced Social Welfare Under Competition
Meena Jagadeesan, Michael Jordan, Jacob Steinhardt, Nika Haghtalab
As the scale of machine learning models increases, trends such as scaling laws anticipate consistent downstream improvements in predictive accuracy. However, these trends take the perspective of a single model-provider in isolation, while in reality providers often compete with each other for users. In this work, we demonstrate that competition can fundamentally alter the behavior of these scaling trends, even causing overall predictive accuracy across users to be non-monotonic or decreasing with scale. We define a model of competition for classification tasks, and use data representations as a lens for studying the impact of increases in scale. We find many settings where improving data representation quality (as measured by Bayes risk) decreases the overall predictive accuracy across users (i.e., social welfare) for a marketplace of competing model-providers. Our examples range from closed-form formulas in simple settings to simulations with pretrained representations on CIFAR-10. At a conceptual level, our work suggests that favorable scaling trends for individual model-providers need not translate to downstream improvements in social welfare in marketplaces with multiple model providers.
****

Inverse Dynamics Pretraining Learns Good Representations for Multitask Imitation
David Brandfonbrener, Ofir Nachum, Joan Bruna
In recent years, domains such as natural language processing and image recogniti on have popularized the paradigm of using large datasets to pretrain representat ions that can be effectively transferred to downstream tasks. In this work we ev aluate how such a paradigm should be done in imitation learning, where both pret raining and finetuning data are trajectories collected by experts interacting wi th an unknown environment. Namely, we consider a setting where the pretraining c orpus consists of multitask demonstrations and the task for each demonstration i s set by an unobserved latent context variable. The goal is to use the pretraini ng corpus to learn a low dimensional representation of the high dimensional (e.g ., visual) observation space which can be transferred to a novel context for fin etuning on a limited dataset of demonstrations. Among a variety of possible pret raining objectives, we argue that inverse dynamics modeling -- i.e., predicting an action given the observations appearing before and after it in the demonstra tion -- is well-suited to this setting. We provide empirical evidence of this cl aim through evaluations on a variety of simulated visuomotor manipulation proble ms. While previous work has attempted various theoretical explanations regarding the benefit of inverse dynamics modeling, we find that these arguments are insu fficient to explain the empirical advantages often observed in our settings, and so we derive a novel analysis using a simple but general environment model.
****
Exploiting hidden structures in non-convex games for convergence to Nash equilib rium
Iosif Sakos, Emmanouil-Vasileios Vlatakis-Gkaragkounis, Panayotis Mertikopoulos, Georgios Piliouras
A wide array of modern machine learning applications – from adversarial models t o multi-agent reinforcement learning – can be formulated as non-cooperative game s whose Nash equilibria represent the system’s desired operational states. Despi te having a highly non-convex loss landscape, many cases of interest possess a l atent convex structure that could potentially be leveraged to yield convergence to an equilibrium. Driven by this observation, our paper proposes a flexible fir st-order method that successfully exploits such “hidden structures” and achieves convergence under minimal assumptions for the transformation connecting the pla yers’ control variables to the game’s latent, convex-structured layer. The propo sed method – which we call preconditioned hidden gradient descent (PHGD) – hinge s on a judiciously chosen gradient preconditioning scheme related to natural gra dient methods. Importantly, we make no separability assumptions for the game’s h idden structure, and we provide explicit convergence rate guarantees for both de terministic and stochastic environments.
****
Secure Out-of-Distribution Task Generalization with Energy-Based Models
Shengzhuang Chen, Long-Kai Huang, Jonathan Richard Schwarz, Yilun Du, Ying Wei
The success of meta-learning on out-of-distribution (OOD) tasks in the wild has proved to be hit-and-miss.To safeguard the generalization capability of the meta -learned prior knowledge to OOD tasks, in particularly safety-critical applicati ons, necessitates detection of an OOD task followed by adaptation of the task to wards the prior. Nonetheless, the reliability of estimated uncertainty on OOD ta sks by existing Bayesian meta-learning methods is restricted by incomplete cover age of the feature distribution shift and insufficient expressiveness of the met a-learned prior. Besides, they struggle to adapt an OOD task, running parallel t o the line of cross-domain task adaptation solutions which are vulnerable to ove rfitting.To this end, we build a single coherent framework that supports both de tection and adaptation of OOD tasks, while remaining compatible with off-the-she lf meta-learning backbones. The proposed Energy-Based Meta-Learning (EBML) frame work learns to characterize any arbitrary meta-training task distribution with t he composition of two expressive neural-network-based energy functions. We deplo y the sum of the two energy functions, being proportional to the joint distribut ion of a task, as a reliable score for detecting OOD tasks; during meta-testing, we adapt the OOD task to in-distribution tasks by energy minimization.Experimen

ts on four regression and classification datasets demonstrate the effectiveness of our proposal.
****
Autodecoding Latent 3D Diffusion Models
Evangelos Ntavelis, Aliaksandr Siarohin, Kyle Olszewski, Chaoyang Wang, Luc V Gool, Sergey Tulyakov
Diffusion-based methods have shown impressive visual results in the text-to-image domain. They first learn a latent space using an autoencoder, then run a denoising process on the bottleneck to generate new samples. However, learning an autoencoder requires substantial data in the target domain. Such data is scarce for 3D generation, prohibiting the learning of large-scale diffusion models for 3D synthesis. We present a novel approach to the generation of static and articulated 3D assets that has a 3D autodecoder at its core. The 3D autodecoder framework embeds properties learned from the target dataset in the latent space, which can then be decoded into a volumetric representation for rendering view-consistent appearance and geometry. We then identify the appropriate intermediate volumetric latent space, and introduce robust normalization and de-normalization operations to learn a 3D diffusion from 2D images or monocular videos of rigid or articulated objects. Our approach is flexible enough to use either existing camera supervision or no camera information at all -- instead efficiently learning it during training. Our evaluations demonstrate that our generation results outperform state-of-the-art alternatives on various benchmark datasets and metrics, including multi-view image datasets of synthetic objects, real in-the-wild videos of moving people, and a large-scale, real video dataset of static objects.
****
Physion++: Evaluating Physical Scene Understanding that Requires Online Inference of Different Physical Properties
Hsiao-Yu Tung, Mingyu Ding, Zhenfang Chen, Daniel Bear, Chuang Gan, Josh Tenenbaum, Dan Yamins, Judith Fan, Kevin Smith
General physical scene understanding requires more than simply localizing and recognizing objects -- it requires knowledge that objects can have different latent properties (e.g., mass or elasticity), and that those properties affect the outcome of physical events. While there has been great progress in physical and video prediction models in recent years, benchmarks to test their performance typically do not require an understanding that objects have individual physical properties, or at best test only those properties that are directly observable (e.g., size or color). This work proposes a novel dataset and benchmark, termed Physion++, that rigorously evaluates visual physical prediction in artificial systems under circumstances where those predictions rely on accurate estimates of the latent physical properties of objects in the scene. Specifically, we test scenarios where accurate prediction relies on estimates of properties such as mass, friction, elasticity, and deformability, and where the values of those properties can only be inferred by observing how objects move and interact with other objects or fluids. We evaluate the performance of a number of state-of-the-art prediction models that span a variety of levels of learning vs. built-in knowledge, and compare that performance to a set of human predictions. We find that models that have been trained using standard regimes and datasets do not spontaneously learn to make inferences about latent properties, but also that models that encode objectness and physical states tend to make better predictions. However, there is still a huge gap between all models and human performance, and all models' predictions correlate poorly with those made by humans, suggesting that no state-of-the-art model is learning to make physical predictions in a human-like way. These results show that current deep learning models that succeed in some settings nevertheless fail to achieve human-level physical prediction in other cases, especially those where latent property inference is required. Project page: https://dingmyu.github.io/physion_v2/
****
HA-ViD: A Human Assembly Video Dataset for Comprehensive Assembly Knowledge Understanding
Hao Zheng, Regina Lee, Yuqian Lu

Understanding comprehensive assembly knowledge from videos is critical for futur istic ultra-intelligent industry. To enable technological breakthrough, we prese nt HA-ViD – the first human assembly video dataset that features representative industrial assembly scenarios, natural procedural knowledge acquisition process, and consistent human-robot shared annotations. Specifically, HA-ViD captures di verse collaboration patterns of real-world assembly, natural human behaviors and learning progression during assembly, and granulate action annotations to subje ct, action verb, manipulated object, target object, and tool. We provide 3222 mu lti-view and multi-modality videos), 1.5M frames, 96K temporal labels and 2M spa tial labels. We benchmark four foundational video understanding tasks: action re cognition, action segmentation, object detection and multi-object tracking. Impo rtantly, we analyze their performance and the further reasoning steps for compre hending knowledge in assembly progress, process efficiency, task collaboration, skill parameters and human intention. Details of HA-ViD is available at: https:/ /iai-hrc.github.io/ha-vid.
****
Classical Simulation of Quantum Circuits: Parallel Environments and Benchmark
Xiao-Yang Liu, Zeliang Zhang
Google's quantum supremacy announcement has received broad questions from acade mia and industry due to the debatable estimate of 10,000 years' running time for the classical simulation task on the Summit supercomputer. Has quantum supremac y already come? Or will it come in one or two decades later? To avoid hasty adve rtisements of quantum supremacy by tech giants or quantum startups and eliminat e the cost of dedicating a team to the classical simulation task, we advocate an open-source approach to maintain a trustable benchmark performance. In this pa per, we take a reinforcement learning approach for the classical simulation of q uantum circuits and demonstrate its great potential by reporting an estimated si mulation time of less than 4 days, a speedup of 5.40x over the state-of-the-art method. Specifically, we formulate the classical simulation task as a tensor ne twork contraction ordering problem using the K-spin Ising model and employ a nov el Hamiltonina-based reinforcement learning algorithm. Then, we establish standa rd criteria to evaluate the performance of classical simulation of quantum circu its. We develop a dozen of massively parallel environments to simulate quantum circuits. We open-source our parallel gym environments and benchmarks. We hope the AI/ML community and quantum physics community will collaborate to maintain r eference curves for validating an unequivocal first demonstration of empirical q uantum supremacy.
****
A General Framework for Robust G-Invariance in G-Equivariant Networks
Sophia Sanborn, Nina Miolane
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
EHRSHOT: An EHR Benchmark for Few-Shot Evaluation of Foundation Models
Michael Wornow, Rahul Thapa, Ethan Steinberg, Jason Fries, Nigam Shah
While the general machine learning (ML) community has benefited from public data sets, tasks, and models, the progress of ML in healthcare has been hampered by a lack of such shared assets. The success of foundation models creates new challe nges for healthcare ML by requiring access to shared pretrained models to valida te performance benefits. We help address these challenges through three contribu tions. First, we publish a new dataset, EHRSHOT, which contains de-identified st ructured data from the electronic health records (EHRs) of 6,739 patients from S tanford Medicine. Unlike MIMIC-III/IV and other popular EHR datasets, EHRSHOT is longitudinal and not restricted to ICU/ED patients. Second, we publish the weig hts of CLMBR-T-base, a 141M parameter clinical foundation model pretrained on th e structured EHR data of 2.57M patients. We are one of the first to fully releas e such a model for coded EHR data; in contrast, most prior models released for c linical data (e.g. GatorTron, ClinicalBERT) only work with unstructured text an

d cannot process the rich, structured data within an EHR. We provide an end-to-end pipeline for the community to validate and build upon its performance. Third, we define 15 few-shot clinical prediction tasks, enabling evaluation of foundation models on benefits such as sample efficiency and task adaptation. Our model and dataset are available via a research data use agreement from here: https://stanfordaimi.azurewebsites.net/. Code to reproduce our results is available here: https://github.com/som-shahlab/ehrshot-benchmark.
****
SEVA: Leveraging sketches to evaluate alignment between human and machine visual abstraction
Kushin Mukherjee, Holly Huey, Xuanchen Lu, Yael Vinker, Rio Aguina-Kang, Ariel Shamir, Judith Fan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
State2Explanation: Concept-Based Explanations to Benefit Agent Learning and User Understanding
Devleena Das, Sonia Chernova, Been Kim
As more non-AI experts use complex AI systems for daily tasks, there has been an increasing effort to develop methods that produce explanations of AI decision making that are understandable by non-AI experts. Towards this effort, leveraging higher-level concepts and producing concept-based explanations have become a popular method. Most concept-based explanations have been developed for classification techniques, and we posit that the few existing methods for sequential decision making are limited in scope. In this work, we first contribute a desiderata for defining ``concepts'' in sequential decision making settings. Additionally, inspired by the Protege Effect which states explaining knowledge often reinforces one's self-learning, we explore how concept-based explanations of an RL agent's decision making can in turn improve the agent's learning rate, as well as improve end-user understanding of the agent's decision making. To this end, we contribute a unified framework, State2Explanation (S2E), that involves learning a joint embedding model between state-action pairs and concept-based explanations, and leveraging such learned model to both (1) inform reward shaping during an agent's training, and (2) provide explanations to end-users at deployment for improved task performance. Our experimental validations, in Connect 4 and Lunar Lander, demonstrate the success of S2E in providing a dual-benefit, successfully informing reward shaping and improving agent learning rate, as well as significantly improving end user task performance at deployment time.
****
Machine learning detects terminal singularities
Tom Coates, Alexander Kasprzyk, Sara Veneziale
Algebraic varieties are the geometric shapes defined by systems of polynomial equations; they are ubiquitous across mathematics and science. Amongst these algebraic varieties are Q-Fano varieties: positively curved shapes which have Q-factorial terminal singularities. Q-Fano varieties are of fundamental importance in geometry as they are `atomic pieces’ of more complex shapes – the process of breaking a shape into simpler pieces in this sense is called the Minimal Model Programme.Despite their importance, the classification of Q-Fano varieties remains unknown. In this paper we demonstrate that machine learning can be used to understand this classification. We focus on eight-dimensional positively-curved algebraic varieties that have toric symmetry and Picard rank two, and develop a neural network classifier that predicts with 95% accuracy whether or not such an algebraic variety is Q-Fano. We use this to give a first sketch of the landscape of Q-Fano varieties in dimension eight.How the neural network is able to detect Q-Fano varieties with such accuracy remains mysterious, and hints at some deep mathematical theory waiting to be uncovered. Furthermore, when visualised using the quantum period, an invariant that has played an important role in recent theoretical developments, we observe that the classification as revealed by ML appears to

fall within a bounded region, and is stratified by the Fano index. This suggests that it may be possible to state and prove conjectures on completeness in the future.Inspired by the ML analysis, we formulate and prove a new global combinatorial criterion for a positively curved toric variety of Picard rank two to have terminal singularities. Together with the first sketch of the landscape of Q-Fano varieties in higher dimensions, this gives strong new evidence that machine learning can be an essential tool in developing mathematical conjectures and accelerating theoretical discovery.
****
Efficient Diffusion Policies For Offline Reinforcement Learning
Bingyi Kang, Xiao Ma, Chao Du, Tianyu Pang, Shuicheng Yan
Offline reinforcement learning (RL) aims to learn optimal policies from offline datasets, where the parameterization of policies is crucial but often overlooked. Recently, Diffsuion-QL significantly boosts the performance of offline RL by representing a policy with a diffusion model, whose success relies on a parametrized Markov Chain with hundreds of steps for sampling. However, Diffusion-QL suffers from two critical limitations. 1) It is computationally inefficient to forward and backward through the whole Markov chain during training. 2) It is incompatible with maximum likelihood-based RL algorithms (e.g., policy gradient methods) as the likelihood of diffusion models is intractable. Therefore, we propose efficient diffusion policy (EDP) to overcome these two challenges. EDP approximately constructs actions from corrupted ones at training to avoid running the sampling chain. We conduct extensive experiments on the D4RL benchmark. The results show that EDP can reduce the diffusion policy training time from 5 days to 5 hours on gym-locomotion tasks. Moreover, we show that EDP is compatible with various offline RL algorithms (TD3, CRR, and IQL) and achieves new state-of-the-art on D4RL by large margins over previous methods.
****
Selective Sampling and Imitation Learning via Online Regression
Ayush Sekhari, Karthik Sridharan, Wen Sun, Runzhe Wu
We consider the problem of Imitation Learning (IL) by actively querying noisy expert for feedback. While imitation learning has been empirically successful, much of prior work assumes access to noiseless expert feedback which is not practical in many applications. In fact, when one only has access to noisy expert feedback, algorithms that rely on purely offline data (non-interactive IL) can be shown to need a prohibitively large number of samples to be successful. In contrast, in this work, we provide an interactive algorithm for IL that uses selective sampling to actively query the noisy expert for feedback. Our contributions are twofold: First, we provide a new selective sampling algorithm that works with general function classes and multiple actions, and obtains the best-known bounds for the regret and the number of queries. Next, we extend this analysis to the problem of IL with noisy expert feedback and provide a new IL algorithm that makes limited queries. Our algorithm for selective sampling leverages function approximation, and relies on an online regression oracle w.r.t.~the given model class to predict actions, and to decide whether to query the expert for its label. On the theoretical side, the regret bound of our algorithm is upper bounded by the regret of the online regression oracle, while the query complexity additionally depends on the eluder dimension of the model class. We complement this with a lower bound that demonstrates that our results are tight. We extend our selective sampling algorithm for IL with general function approximation and provide bounds on both the regret and the number of queries made to the noisy expert. A key novelty here is that our regret and query complexity bounds only depend on the number of times the optimal policy (and not the noisy expert, or the learner) go to states that have a small margin.
****
CamoPatch: An Evolutionary Strategy for Generating Camoflauged Adversarial Patches
Phoenix Williams, Ke Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

MADLAD-400: A Multilingual And Document-Level Large Audited Dataset
Sneha Kudugunta, Isaac Caswell, Biao Zhang, Xavier Garcia, Derrick Xin, Aditya Kusupati, Romi Stella, Ankur Bapna, Orhan Firat
We introduce MADLAD-400, a manually audited, general domain 3T token monolingual dataset based on CommonCrawl, spanning 419 languages. We discuss the limitations revealed by self-auditing MADLAD-400, and the role data auditing had in the dataset creation process. We then train and release a 10.7B-parameter multilingual machine translation model on 250 billion tokens covering over 450 languages using publicly available data, and find that it is competitive with models that are significantly larger, and report the results on different domains. In addition, we train a 8B-parameter language model, and assess the results on few-shot translation. We make the baseline models available to the research community.
****

Learning Regularized Monotone Graphon Mean-Field Games
Fengzhuo Zhang, Vincent Tan, Zhaoran Wang, Zhuoran Yang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

From Cloze to Comprehension: Retrofitting Pre-trained Masked Language Models to Pre-trained Machine Reader
Weiwen Xu, Xin Li, Wenxuan Zhang, Meng Zhou, Wai Lam, Luo Si, Lidong Bing
We present Pre-trained Machine Reader (PMR), a novel method for retrofitting pre-trained masked language models (MLMs) to pre-trained machine reading comprehension (MRC) models without acquiring labeled data.PMR can resolve the discrepancy between model pre-training and downstream fine-tuning of existing MLMs.To build the proposed PMR, we constructed a large volume of general-purpose and high-quality MRC-style training data by using Wikipedia hyperlinks and designed a Wiki Anchor Extraction task to guide the MRC-style pre-training.Apart from its simplicity, PMR effectively solves extraction tasks, such as Extractive Question Answering and Named Entity Recognition. PMR shows tremendous improvements over existing approaches, especially in low-resource scenarios.When applied to the sequence classification task in the MRC formulation, PMR enables the extraction of high-quality rationales to explain the classification process, thereby providing greater prediction explainability. PMR also has the potential to serve as a unified model for tackling various extraction and classification tasks in the MRC formulation.
****

DaTaSeg: Taming a Universal Multi-Dataset Multi-Task Segmentation Model
Xiuye Gu, Yin Cui, Jonathan Huang, Abdullah Rashwan, Xuan Yang, Xingyi Zhou, Golnaz Ghiasi, Weicheng Kuo, Huizhong Chen, Liang-Chieh Chen, David Ross
Observing the close relationship among panoptic, semantic and instance segmentation tasks, we propose to train a universal multi-dataset multi-task segmentation model: DaTaSeg. We use a shared representation (mask proposals with class predictions) for all tasks. To tackle task discrepancy, we adopt different merge operations and post-processing for different tasks. We also leverage weak-supervision, allowing our segmentation model to benefit from cheaper bounding box annotations. To share knowledge across datasets, we use text embeddings from the same semantic embedding space as classifiers and share all network parameters among datasets. We train DaTaSeg on ADE semantic, COCO panoptic, and Objects365 detection datasets. DaTaSeg improves performance on all datasets, especially small-scale datasets, achieving 54.0 mIoU on ADE semantic and 53.5 PQ on COCO panoptic. DaTaSeg also enables weakly-supervised knowledge transfer on ADE panoptic and Objects365 instance segmentation. Experiments show DaTaSeg scales with the number of training datasets and enables open-vocabulary segmentation through direct transfer. In addition, we annotate an Objects365 instance segmentation set of 1,000 ima

ges and release it as a public evaluation benchmark on https://laoreja.github.io/dataseg.
****
SituatedGen: Incorporating Geographical and Temporal Contexts into Generative Commonsense Reasoning
Yunxiang Zhang, Xiaojun Wan
Recently, commonsense reasoning in text generation has attracted much attention. Generative commonsense reasoning is the task that requires machines, given a group of keywords, to compose a single coherent sentence with commonsense plausibility. While existing datasets targeting generative commonsense reasoning focus on everyday scenarios, it is unclear how well machines reason under specific geographical and temporal contexts. We formalize this challenging task as SituatedGen, where machines with commonsense should generate a pair of contrastive sentences given a group of keywords including geographical or temporal entities. We introduce a corresponding English dataset consisting of 8,268 contrastive sentence pairs, which are built upon several existing commonsense reasoning benchmarks with minimal manual labor. Experiments show that state-of-the-art generative language models struggle to generate sentences with commonsense plausibility and still lag far behind human performance. Our dataset is publicly available at https://github.com/yunx-z/situated_gen.
****
Multi-scale Diffusion Denoised Smoothing
Jongheon Jeong, Jinwoo Shin
Along with recent diffusion models, randomized smoothing has become one of a few tangible approaches that offers adversarial robustness to models at scale, e.g., those of large pre-trained models. Specifically, one can perform randomized smoothing on any classifier via a simple "denoise-and-classify" pipeline, so-called denoised smoothing, given that an accurate denoiser is available - such as diffusion model. In this paper, we present scalable methods to address the current trade-off between certified robustness and accuracy in denoised smoothing. Our key idea is to "selectively" apply smoothing among multiple noise scales, coined multi-scale smoothing, which can be efficiently implemented with a single diffusion model. This approach also suggests a new objective to compare the collective robustness of multi-scale smoothed classifiers, and questions which representation of diffusion model would maximize the objective. To address this, we propose to further fine-tune diffusion model (a) to perform consistent denoising whenever the original image is recoverable, but (b) to generate rather diverse outputs otherwise. Our experiments show that the proposed multi-scale smoothing scheme, combined with diffusion fine-tuning, not only allows strong certified robustness at high noise scales but also maintains accuracy close to non-smoothed classifiers. Code is available at https://github.com/jh-jeong/smoothing-multiscale.
****
PDE-Refiner: Achieving Accurate Long Rollouts with Neural PDE Solvers
Phillip Lippe, Bas Veeling, Paris Perdikaris, Richard Turner, Johannes Brandstetter
Time-dependent partial differential equations (PDEs) are ubiquitous in science and engineering. Recently, mostly due to the high computational cost of traditional solution techniques, deep neural network based surrogates have gained increased interest. The practical utility of such neural PDE solvers relies on their ability to provide accurate, stable predictions over long time horizons, which is a notoriously hard problem. In this work, we present a large-scale analysis of common temporal rollout strategies, identifying the neglect of non-dominant spatial frequency information, often associated with high frequencies in PDE solutions, as the primary pitfall limiting stable, accurate rollout performance. Based on these insights, we draw inspiration from recent advances in diffusion models to introduce PDE-Refiner; a novel model class that enables more accurate modeling of all frequency components via a multistep refinement process. We validate PDE-Refiner on challenging benchmarks of complex fluid dynamics, demonstrating stable and accurate rollouts that consistently outperform state-of-the-art models, including neural, numerical, and hybrid neural-numerical architectures. We furth

er demonstrate that PDE-Refiner greatly enhances data efficiency, since the denoising objective implicitly induces a novel form of spectral data augmentation. Finally, PDE-Refiner's connection to diffusion models enables an accurate and efficient assessment of the model's predictive uncertainty, allowing us to estimate when the surrogate becomes inaccurate.
****

Accelerating Exploration with Unlabeled Prior Data
Qiyang Li, Jason Zhang, Dibya Ghosh, Amy Zhang, Sergey Levine
Learning to solve tasks from a sparse reward signal is a major challenge for standard reinforcement learning (RL) algorithms. However, in the real world, agents rarely need to solve sparse reward tasks entirely from scratch. More often, we might possess prior experience to draw on that provides considerable guidance about which actions and outcomes are possible in the world, which we can use to explore more effectively for new tasks. In this work, we study how prior data without reward labels may be used to guide and accelerate exploration for an agent solving a new sparse reward task. We propose a simple approach that learns a reward model from online experience, labels the unlabeled prior data with optimistic rewards, and then uses it concurrently alongside the online data for downstream policy and critic optimization. This general formula leads to rapid exploration in several challenging sparse-reward domains where tabula rasa exploration is insufficient, including the AntMaze domain, Adroit hand manipulation domain, and a visual simulated robotic manipulation domain. Our results highlight the ease of incorporating unlabeled prior data into existing online RL algorithms, and the (perhaps surprising) effectiveness of doing so.
****

Towards a Unified Framework of Contrastive Learning for Disentangled Representations
Stefan Matthes, Zhiwei Han, Hao Shen
Contrastive learning has recently emerged as a promising approach for learning data representations that discover and disentangle the explanatory factors of the data.Previous analyses of such approaches have largely focused on individual contrastive losses, such as noise-contrastive estimation (NCE) and InfoNCE, and rely on specific assumptions about the data generating process.This paper extends the theoretical guarantees for disentanglement to a broader family of contrastive methods, while also relaxing the assumptions about the data distribution.Specifically, we prove identifiability of the true latents for four contrastive losses studied in this paper, without imposing common independence assumptions.The theoretical findings are validated on several benchmark datasets.Finally, practical limitations of these methods are also investigated.
****

An Improved Relaxation for Oracle-Efficient Adversarial Contextual Bandits
Kiarash Banihashem, MohammadTaghi Hajiaghayi, Suho Shin, Max Springer
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Sequential Subset Matching for Dataset Distillation
JIAWEI DU, Qin Shi, Joey Tianyi Zhou
Dataset distillation is a newly emerging task that synthesizes a small-size dataset used in training deep neural networks (DNNs) for reducing data storage and model training costs. The synthetic datasets are expected to capture the essence of the knowledge contained in real-world datasets such that the former yields a similar performance as the latter. Recent advancements in distillation methods have produced notable improvements in generating synthetic datasets. However, current state-of-the-art methods treat the entire synthetic dataset as a unified entity and optimize each synthetic instance equally . This static optimization approach may lead to performance degradation in dataset distillation. Specifically, we argue that static optimization can give rise to a coupling issue within the synthetic data, particularly when a larger amount of synthetic data is being opt

imized. This coupling issue, in turn, leads to the failure of the distilled data set to extract the high-level features learned by the deep neural network (DNN) in the latter epochs.In this study, we propose a new dataset distillation strategy called Sequential Subset Matching (SeqMatch), which tackles this problem by adaptively optimizing the synthetic data to encourage sequential acquisition of knowledge during dataset distillation. Our analysis indicates that SeqMatch effectively addresses the coupling issue by sequentially generating the synthetic instances, thereby enhancing its performance significantly. Our proposed SeqMatch outperforms state-of-the-art methods in various datasets, including SVNH, CIFAR-10, CIFAR-100, and Tiny ImageNet.
****

SLaM: Student-Label Mixing for Distillation with Unlabeled Examples
Vasilis Kontonis, Fotis Iliopoulos, Khoa Trinh, Cenk Baykal, Gaurav Menghani, Erik Vee
Knowledge distillation with unlabeled examples is a powerful training paradigm for generating compact and lightweight student models in applications where the amount of labeled data is limited but one has access to a large pool of unlabeled data. In this setting, a large teacher model generates "soft" pseudo-labels for the unlabeled dataset which are then used for training the student model. Despite its success in a wide variety of applications, a shortcoming of this approach is that the teacher's pseudo-labels are often noisy, leading to impaired student performance. In this paper, we present a principled method for knowledge distillation with unlabeled examples that we call Student-Label Mixing (SLaM) and we show that it consistently improves over prior approaches by evaluating it on several standard benchmarks. Finally, we show that SLaM comes with theoretical guarantees; along the way we give an algorithm improving the best-known sample complexity for learning halfspaces with margin under random classification noise, and provide the first convergence analysis for so-called ``forward loss-adjustment" methods.
****

Mitigating the Popularity Bias of Graph Collaborative Filtering: A Dimensional Collapse Perspective
Yifei Zhang, Hao Zhu, yankai Chen, Zixing Song, Piotr Koniusz, Irwin King
Graph-based Collaborative Filtering (GCF) is widely used in personalized recommendation systems. However, GCF suffers from a fundamental problem where features tend to occupy the embedding space inefficiently (by spanning only a low-dimensional subspace). Such an effect is characterized in GCF by the embedding space being dominated by a few of popular items with the user embeddings highly concentrated around them. This enhances the so-called Matthew effect of the popularity bias where popular items are highly recommend whereas remaining items are ignored. In this paper, we analyze the above effect in GCF and reveal that the simplified graph convolution operation (typically used in GCF) shrinks the singular space of the feature matrix. As typical approaches (i.e., optimizing the uniformity term) fail to prevent the embedding space degradation, we propose a decorrelation-enhanced GCF objective that promotes feature diversity by leveraging the so-called principle of redundancy reduction in embeddings. However, unlike conventional methods that use the Euclidean geometry to relax hard constraints for decorrelation, we exploit non-Euclidean geometry. Such a choice helps maintain the range space of the matrix and obtain small condition number, which prevents the embedding space degradation. Our method outperforms contrastive-based GCF models on several benchmark datasets and improves the performance for unpopular items.
****

The Rise of AI Language Pathologists: Exploring Two-level Prompt Learning for Few-shot Weakly-supervised Whole Slide Image Classification
Linhao Qu, xiaoyuan luo, Kexue Fu, Manning Wang, Zhijian Song
This paper introduces the novel concept of few-shot weakly supervised learning for pathology Whole Slide Image (WSI) classification, denoted as FSWC. A solution is proposed based on prompt learning and the utilization of a large language model, GPT-4. Since a WSI is too large and needs to be divided into patches for processing, WSI classification is commonly approached as a Multiple Instance Learn

ing (MIL) problem. In this context, each WSI is considered a bag, and the obtain
ed patches are treated as instances. The objective of FSWC is to classify both b
ags and instances with only a limited number of labeled bags. Unlike conventiona
l few-shot learning problems, FSWC poses additional challenges due to its weak b
ag labels within the MIL framework. Drawing inspiration from the recent achievem
ents of vision-language models (V-L models) in downstream few-shot classificatio
n tasks, we propose a two-level prompt learning MIL framework tailored for patho
logy, incorporating language prior knowledge. Specifically, we leverage CLIP to
extract instance features for each patch, and introduce a prompt-guided pooling
strategy to aggregate these instance features into a bag feature. Subsequently,
we employ a small number of labeled bags to facilitate few-shot prompt learning
based on the bag features. Our approach incorporates the utilization of GPT-4 in
 a question-and-answer mode to obtain language prior knowledge at both the insta
nce and bag levels, which are then integrated into the instance and bag level la
nguage prompts. Additionally, a learnable component of the language prompts is t
rained using the available few-shot labeled data. We conduct extensive experimen
ts on three real WSI datasets encompassing breast cancer, lung cancer, and cervi
cal cancer, demonstrating the notable performance of the proposed method in bag
and instance classification. All codes will be made publicly accessible.
****
State Sequences Prediction via Fourier Transform for Representation Learning
Mingxuan Ye, Yufei Kuang, Jie Wang, Yang Rui, Wengang Zhou, Houqiang Li, Feng Wu
While deep reinforcement learning (RL) has been demonstrated effective in solvin
g complex control tasks, sample efficiency remains a key challenge due to the la
rge amounts of data required for remarkable performance. Existing research explo
res the application of representation learning for data-efficient RL, e.g., lear
ning predictive representations by predicting long-term future states. However,
many existing methods do not fully exploit the structural information inherent i
n sequential state signals, which can potentially improve the quality of long-te
rm decision-making but is difficult to discern in the time domain. To tackle thi
s problem, we propose State Sequences Prediction via Fourier Transform (SPF), a
novel method that exploits the frequency domain of state sequences to extract th
e underlying patterns in time series data for learning expressive representation
s efficiently. Specifically, we theoretically analyze the existence of structura
l information in state sequences, which is closely related to policy performance
 and signal regularity, and then propose to predict the Fourier transform of inf
inite-step future state sequences to extract such information. One of the appeal
ing features of SPF is that it is simple to implement while not requiring storag
e of infinite-step future states as prediction targets. Experiments demonstrate
that the proposed method outperforms several state-of-the-art algorithms in term
s of both sample efficiency and performance.
****
Beyond Myopia: Learning from Positive and Unlabeled Data through Holistic Predic
tive Trends
Wang Xinrui, Wenhai Wan, Chuanxing Geng, Shao-Yuan Li, Songcan Chen
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Fast Approximation of Similarity Graphs with Kernel Density Estimation
Peter Macgregor, He Sun
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
An Efficient Dataset Condensation Plugin and Its Application to Continual Learni
ng
Enneng Yang, Li Shen, Zhenyi Wang, Tongliang Liu, Guibing Guo

Dataset condensation (DC) distills a large real-world dataset into a small synthetic dataset, with the goal of training a network from scratch on the latter that performs similarly to the former. State-of-the-art (SOTA) DC methods have achieved satisfactory results through techniques such as accuracy, gradient, training trajectory, or distribution matching. However, these works all perform matching in the high-dimension pixel spaces, ignoring that natural images are usually locally connected and have lower intrinsic dimensions, resulting in low condensation efficiency. In this work, we propose a simple-yet-efficient dataset condensation plugin that matches the raw and synthetic datasets in a low-dimensional manifold. Specifically, our plugin condenses raw images into two low-rank matrices instead of parameterized image matrices. Our plugin can be easily incorporated into existing DC methods, thereby containing richer raw dataset information at limited storage costs to improve the downstream applications' performance. We verify on multiple public datasets that when the proposed plugin is combined with SOTA DC methods, the performance of the network trained on synthetic data is significantly improved compared to traditional DC methods. Moreover, when applying the DC methods as a plugin to continual learning tasks, we observed that our approach effectively mitigates catastrophic forgetting of old tasks under limited memory buffer constraints and avoids the problem of raw data privacy leakage.
****
Bootstrapped Training of Score-Conditioned Generator for Offline Design of Biological Sequences
Minsu Kim, Federico Berto, Sungsoo Ahn, Jinkyoo Park
We study the problem of optimizing biological sequences, e.g., proteins, DNA, and RNA, to maximize a black-box score function that is only evaluated in an offline dataset. We propose a novel solution, bootstrapped training of score-conditioned generator (BootGen) algorithm. Our algorithm repeats a two-stage process. In the first stage, our algorithm trains the biological sequence generator with rank-based weights to enhance the accuracy of sequence generation based on high scores. The subsequent stage involves bootstrapping, which augments the training dataset with self-generated data labeled by a proxy score function. Our key idea is to align the score-based generation with a proxy score function, which distills the knowledge of the proxy score function to the generator. After training, we aggregate samples from multiple bootstrapped generators and proxies to produce a diverse design. Extensive experiments show that our method outperforms competitive baselines on biological sequential design tasks. We provide reproducible source code: https://github.com/kaist-silab/bootgen.
****
Statistically Valid Variable Importance Assessment through Conditional Permutations
Ahmad CHAMMA, Denis A. Engemann, Bertrand Thirion
Variable importance assessment has become a crucial step in machine-learning applications when using complex learners, such as deep neural networks, on large-scale data. Removal-based importance assessment is currently the reference approach, particularly when statistical guarantees are sought to justify variable inclusion. It is often implemented with variable permutation schemes. On the flip side, these approaches risk misidentifying unimportant variables as important in the presence of correlations among covariates. Here we develop a systematic approach for studying Conditional Permutation Importance (CPI) that is model agnostic and computationally lean, as well as reusable benchmarks of state-of-the-art variable importance estimators. We show theoretically and empirically that \textit{CPI} overcomes the limitations of standard permutation importance by providing accurate type-I error control. When used with a deep neural network, \textit{CPI} consistently showed top accuracy across benchmarks. An experiment on real-world data analysis in a large-scale medical dataset showed that \textit{CPI} provides a more parsimonious selection of statistically significant variables. Our results suggest that \textit{CPI} can be readily used as drop-in replacement for permutation-based methods.
****
Towards Better Dynamic Graph Learning: New Architecture and Unified Library

Le Yu, Leilei Sun, Bowen Du, Weifeng Lv
We propose DyGFormer, a new Transformer-based architecture for dynamic graph learning. DyGFormer is conceptually simple and only needs to learn from nodes' historical first-hop interactions by: (1) a neighbor co-occurrence encoding scheme that explores the correlations of the source node and destination node based on their historical sequences; (2) a patching technique that divides each sequence into multiple patches and feeds them to Transformer, allowing the model to effectively and efficiently benefit from longer histories. We also introduce DyGLib, a unified library with standard training pipelines, extensible coding interfaces, and comprehensive evaluating protocols to promote reproducible, scalable, and credible dynamic graph learning research. By performing exhaustive experiments on thirteen datasets for dynamic link prediction and dynamic node classification tasks, we find that DyGFormer achieves state-of-the-art performance on most of the datasets, demonstrating its effectiveness in capturing nodes' correlations and long-term temporal dependencies. Moreover, some results of baselines are inconsistent with previous reports, which may be caused by their diverse but less rigorous implementations, showing the importance of DyGLib. All the used resources are publicly available at https://github.com/yule-BUAA/DyGLib.
****
Implicit Manifold Gaussian Process Regression
Bernardo Fichera, Slava Borovitskiy, Andreas Krause, Aude G Billard
Gaussian process regression is widely used because of its ability to provide well-calibrated uncertainty estimates and handle small or sparse datasets. However, it struggles with high-dimensional data. One possible way to scale this technique to higher dimensions is to leverage the implicit low-dimensional manifold upon which the data actually lies, as postulated by the manifold hypothesis. Prior work ordinarily requires the manifold structure to be explicitly provided though, i.e. given by a mesh or be known to be one of the well-known manifolds like the sphere. In contrast, in this paper we propose a Gaussian process regression technique capable of inferring implicit structure directly from data (labeled and unlabeled) in a fully differentiable way. For the resulting model, we discuss its convergence to the Matérn Gaussian process on the assumed manifold. Our technique scales up to hundreds of thousands of data points, and improves the predictive performance and calibration of the standard Gaussian process regression in some high-dimensional settings.
****
UDC-SIT: A Real-World Dataset for Under-Display Cameras
Kyusu Ahn, Byeonghyun Ko, HyunGyu Lee, Chanwoo Park, Jaejin Lee
Under Display Camera (UDC) is a novel imaging system that mounts a digital camera lens beneath a display panel with the panel covering the camera. However, the display panel causes severe degradation to captured images, such as low transmittance, blur, noise, and flare. The restoration of UDC-degraded images is challenging because of the unique luminance and diverse patterns of flares. Existing UDC dataset studies focus on unrealistic or synthetic UDC degradation rather than real-world UDC images. In this paper, we propose a real-world UDC dataset called UDC-SIT. To obtain the non-degraded and UDC-degraded images for the same scene, we propose an image-capturing system and an image alignment technique that exploits discrete Fourier transform (DFT) to align a pair of captured images. UDC-SIT also includes comprehensive annotations missing from other UDC datasets, such as light source, day/night, indoor/outdoor, and flare components (e.g., shimmers, streaks, and glares). We compare UDC-SIT with four existing representative UDC datasets and present the problems with existing UDC datasets. To show UDC-SIT's effectiveness, we compare UDC-SIT and a representative synthetic UDC dataset using four representative learnable image restoration models. The result indicates that the models trained with the synthetic UDC dataset are impractical because the synthetic UDC dataset does not reflect the actual characteristics of UDC-degraded images. UDC-SIT can enable further exploration in the UDC image restoration area and provide better insights into the problem. UDC-SIT is available at: https://github.com/mcrl/UDC-SIT.
****

The Crucial Role of Normalization in Sharpness-Aware Minimization
Yan Dai, Kwangjun Ahn, Suvrit Sra
Sharpness-Aware Minimization (SAM) is a recently proposed gradient-based optimiz er (Foret et al., ICLR 2021) that greatly improves the prediction performance of deep neural networks. Consequently, there has been a surge of interest in expla ining its empirical success. We focus, in particular, on understanding the role played by normalization, a key component of the SAM updates. We theoretically an d empirically study the effect of normalization in SAM for both convex and non-c onvex functions, revealing two key roles played by normalization: i) it helps in stabilizing the algorithm; and ii) it enables the algorithm to drift along a co ntinuum (manifold) of minima -- a property identified by recent theoretical work s that is the key to better performance. We further argue that these two propert ies of normalization make SAM robust against the choice of hyper-parameters, sup porting the practicality of SAM. Our conclusions are backed by various experimen ts.
****
Policy Space Diversity for Non-Transitive Games
Jian Yao, Weiming Liu, Haobo Fu, Yaodong Yang, Stephen McAleer, Qiang Fu, Wei Ya ng
Policy-Space Response Oracles (PSRO) is an influential algorithm framework for a pproximating a Nash Equilibrium (NE) in multi-agent non-transitive games. Many p revious studies have been trying to promote policy diversity in PSRO. A major we akness with existing diversity metrics is that a more diverse (according to thei r diversity metrics) population does not necessarily mean (as we proved in the p aper) a better approximation to a NE. To alleviate this problem, we propose a ne w diversity metric, the improvement of which guarantees a better approximation t o a NE. Meanwhile, we develop a practical and well-justified method to optimize our diversity metric using only state-action samples. By incorporating our diver sity regularization into the best response solving of PSRO, we obtain a new PSRO variant, \textit{Policy Space Diversity} PSRO (PSD-PSRO). We present the conver gence property of PSD-PSRO. Empirically, extensive experiments on single-state g ames, Leduc, and Goofspiel demonstrate that PSD-PSRO is more effective in produc ing significantly less exploitable policies than state-of-the-art PSRO variants.
****
COOM: A Game Benchmark for Continual Reinforcement Learning
Tristan Tomilin, Meng Fang, Yudi Zhang, Mykola Pechenizkiy
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Video-Mined Task Graphs for Keystep Recognition in Instructional Videos
Kumar Ashutosh, Santhosh Kumar Ramakrishnan, Triantafyllos Afouras, Kristen Grau man
Procedural activity understanding requires perceiving human actions in terms of a broader task, where multiple keysteps are performed in sequence across a long video to reach a final goal state---such as the steps of a recipe or the steps o f a DIY fix-it task. Prior work largely treats keystep recognition in isolation of this broader structure, or else rigidly confines keysteps to align with a pa rticular sequential script. We propose discovering a task graph automatically f rom how-to videos to represent probabilistically how people tend to execute keys teps, then leverage this graph to regularize keystep recognition in novel videos . On multiple datasets of real-world instructional video, we show the impact: m ore reliable zero-shot keystep localization and improved video representation le arning, exceeding the state of the art.
****
Affinity-Aware Graph Networks
Ameya Velingker, Ali Sinop, Ira Ktena, Petar Veli■kovi■, Sreenivas Gollapudi
Graph Neural Networks (GNNs) have emerged as a powerful technique for learning o n relational data. Owing to the relatively limited number of message passing ste

ps they perform—and hence a smaller receptive field—there has been significant interest in improving their expressivity by incorporating structural aspects of the underlying graph. In this paper, we explore the use of affinity measures as features in graph neural networks, in particular measures arising from random walks, including effective resistance, hitting and commute times. We propose message passing networks based on these features and evaluate their performance on a variety of node and graph property prediction tasks. Our architecture has low computational complexity, while our features are invariant to the permutations of the underlying graph. The measures we compute allow the network to exploit the connectivity properties of the graph, thereby allowing us to outperform relevant benchmarks for a wide variety of tasks, often with significantly fewer message passing steps. On one of the largest publicly available graph regression datasets, OGB-LSC-PCQM4Mv1, we obtain the best known single-model validation MAE at the time of writing.

****

Eliminating Catastrophic Overfitting Via Abnormal Adversarial Examples Regularization

Runqi Lin, Chaojian Yu, Tongliang Liu

Single-step adversarial training (SSAT) has demonstrated the potential to achieve both efficiency and robustness. However, SSAT suffers from catastrophic overfitting (CO), a phenomenon that leads to a severely distorted classifier, making it vulnerable to multi-step adversarial attacks. In this work, we observe that some adversarial examples generated on the SSAT-trained network exhibit anomalous behaviour, that is, although these training samples are generated by the inner maximization process, their associated loss decreases instead, which we named abnormal adversarial examples (AAEs). Upon further analysis, we discover a close relationship between AAEs and classifier distortion, as both the number and outputs of AAEs undergo a significant variation with the onset of CO. Given this observation, we re-examine the SSAT process and uncover that before the occurrence of CO, the classifier already displayed a slight distortion, indicated by the presence of few AAEs. Furthermore, the classifier directly optimizing these AAEs will accelerate its distortion, and correspondingly, the variation of AAEs will sharply increase as a result. In such a vicious circle, the classifier rapidly becomes highly distorted and manifests as CO within a few iterations. These observations motivate us to eliminate CO by hindering the generation of AAEs. Specifically, we design a novel method, termed Abnormal Adversarial Examples Regularization (AAER), which explicitly regularizes the variation of AAEs to hinder the classifier from becoming distorted. Extensive experiments demonstrate that our method can effectively eliminate CO and further boost adversarial robustness with negligible additional computational overhead. Our implementation can be found at https://github.com/tmllab/2023NeurIPSAAER.

****

Hardware Resilience Properties of Text-Guided Image Classifiers

Syed Talal Wasim, Kabila Haile Soboka, Abdulrahman Mahmoud, Salman H. Khan, David Brooks, Gu-Yeon Wei

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

Reliable Off-Policy Learning for Dosage Combinations

Jonas Schweisthal, Dennis Frauen, Valentyn Melnychuk, Stefan Feuerriegel

Decision-making in personalized medicine such as cancer therapy or critical care must often make choices for dosage combinations, i.e., multiple continuous treatments. Existing work for this task has modeled the effect of multiple treatments independently, while estimating the joint effect has received little attention but comes with non-trivial challenges. In this paper, we propose a novel method for reliable off-policy learning for dosage combinations. Our method proceeds along three steps: (1) We develop a tailored neural network that estimates the individualized dose-response function while accounting for the joint effect of mul

tiple dependent dosages. (2) We estimate the generalized propensity score using conditional normalizing flows in order to detect regions with limited overlap in the shared covariate-treatment space. (3) We present a gradient-based learning algorithm to find the optimal, individualized dosage combinations. Here, we ensure reliable estimation of the policy value by avoiding regions with limited overlap. We finally perform an extensive evaluation of our method to show its effectiveness. To the best of our knowledge, ours is the first work to provide a method for reliable off-policy learning for optimal dosage combinations.
****

A Unified Algorithm Framework for Unsupervised Discovery of Skills based on Determinantal Point Process
Jiayu Chen, Vaneet Aggarwal, Tian Lan
Learning rich skills under the option framework without supervision of external rewards is at the frontier of reinforcement learning research. Existing works mainly fall into two distinctive categories: variational option discovery that maximizes the diversity of the options through a mutual information loss (while ignoring coverage) and Laplacian-based methods that focus on improving the coverage of options by increasing connectivity of the state space (while ignoring diversity). In this paper, we show that diversity and coverage in unsupervised option discovery can indeed be unified under the same mathematical framework. To be specific, we explicitly quantify the diversity and coverage of the learned options through a novel use of Determinantal Point Process (DPP) and optimize these objectives to discover options with both superior diversity and coverage. Our proposed algorithm, ODPP, has undergone extensive evaluation on challenging tasks created with Mujoco and Atari. The results demonstrate that our algorithm outperforms state-of-the-art baselines in both diversity- and coverage-driven categories.
****

Discovering Intrinsic Spatial-Temporal Logic Rules to Explain Human Actions
Chengzhi Cao, Chao Yang, Ruimao Zhang, Shuang Li
We propose an interpretable model to uncover the behavioral patterns of human movements by analyzing their trajectories. Our approach is based on the belief that human actions are driven by intentions and are influenced by environmental factors such as spatial relationships with surrounding objects. To model this, we use a set of spatial-temporal logic rules that include intention variables as principles. These rules are automatically discovered and used to capture the dynamics of human actions. To learn the model parameters and rule content, we design an EM learning algorithm that treats the unknown rule content as a latent variable. In the E-step, we evaluate the posterior over the latent rule content, and in the M-step, we optimize the rule generator and model parameters by maximizing the expected log-likelihood. Our model has wide-ranging applications in areas such as sports analytics, robotics, and autonomous cars. We demonstrate the model's superior interpretability and prediction performance on both pedestrian and NBA basketball player datasets, achieving promising results.
****

DiT-3D: Exploring Plain Diffusion Transformers for 3D Shape Generation
Shentong Mo, Enze Xie, Ruihang Chu, Lanqing Hong, Matthias Niessner, Zhenguo Li
Recent Diffusion Transformers (i.e., DiT) have demonstrated their powerful effectiveness in generating high-quality 2D images. However, it is unclear how the Transformer architecture performs equally well in 3D shape generation, as previous 3D diffusion methods mostly adopted the U-Net architecture. To bridge this gap, we propose a novel Diffusion Transformer for 3D shape generation, named DiT-3D, which can directly operate the denoising process on voxelized point clouds using plain Transformers. Compared to existing U-Net approaches, our DiT-3D is more scalable in model size and produces much higher quality generations.Specifically, the DiT-3D adopts the design philosophy of DiT but modifies it by incorporating 3D positional and patch embeddings to aggregate input from voxelized point clouds.To reduce the computational cost of self-attention in 3D shape generation, we incorporate 3D window attention into Transformer blocks, as the increased 3D token length resulting from the additional dimension of voxels can lead to high computation.Finally, linear and devoxelization layers are used to predict the den

oised point clouds. In addition, we empirically observe that the pre-trained DiT-2D checkpoint on ImageNet can significantly improve DiT-3D on ShapeNet.Experimental results on the ShapeNet dataset demonstrate that the proposed DiT-3D achieves state-of-the-art performance in high-fidelity and diverse 3D point cloud generation.
****
DESSERT: An Efficient Algorithm for Vector Set Search with Vector Set Queries
Joshua Engels, Benjamin Coleman, Vihan Lakshman, Anshumali Shrivastava
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Dynamic Sparsity Is Channel-Level Sparsity Learner
Lu Yin, Gen Li, Meng Fang, Li Shen, Tianjin Huang, Zhangyang "Atlas" Wang, Vlado Menkovski, Xiaolong Ma, Mykola Pechenizkiy, Shiwei Liu
Sparse training has received an upsurging interest in machine learning due to its tantalizing saving potential for both the entire training process as well as the inference. Dynamic sparse training (DST) as a leading approach can train deep neural networks at high sparsity from scratch to match the performance of their dense counterparts. However, most if not all DST prior arts demonstrate their effectiveness on unstructured sparsity with highly irregular sparse patterns, which receives limited support in common hardware. This limitation hinders the usage of DST in practice. In this paper, we propose Channel-aware dynamic sparse (Chase), that for the first time seamlessly translates the promise of unstructured dynamic sparsity to GPU-friendly channel-level sparsity (not fine-grained N:M or group sparsity) during one end-to-end training process, without any ad-hoc operations. The resulting small sparse networks can be directly accelerated by commodity hardware, without using any particularly sparsity-aware hardware accelerators. This appealing outcome is partially motivated by a hidden phenomenon of dynamic sparsity: off-the-shelf unstructured DST implicitly involves biased parameter reallocation across channels, with a large fraction of channels (up to 60%) being sparser than others. By progressively identifying and removing these channels during training, our approach transfers unstructured sparsity to channel-wise sparsity. Our experimental results demonstrate that Chase achieves 1.7x inference throughput speedup on common GPU devices without compromising accuracy with ResNet-50 on ImageNet. We release our code in https://github.com/luuyin/chase.
****
Bayesian nonparametric (non-)renewal processes for analyzing neural spike train variability
David Liu, Mate Lengyel
Neural spiking activity is generally variable, non-stationary, and exhibits complex dependencies on covariates, such as sensory input or behavior. These dependencies have been proposed to be signatures of specific computations, and so characterizing them with quantitative rigor is critical for understanding neural computations. Approaches based on point processes provide a principled statistical framework for modeling neural spiking activity. However, currently, they only allow the instantaneous mean, but not the instantaneous variability, of responses to depend on covariates. To resolve this limitation, we propose a scalable Bayesian approach generalizing modulated renewal processes using sparse variational Gaussian processes. We leverage pathwise conditioning for computing nonparametric priors over conditional interspike interval distributions and rely on automatic relevance determination to detect lagging interspike interval dependencies beyond renewal order. After systematically validating our method on synthetic data, we apply it to two foundational datasets of animal navigation: head direction cells in freely moving mice and hippocampal place cells in rats running along a linear track. Our model exhibits competitive or better predictive power compared to state-of-the-art baselines, and outperforms them in terms of capturing interspike interval statistics. These results confirm the importance of modeling covariate-dependent spiking variability, and further analyses of our fitted models rev

eal rich patterns of variability modulation beyond the temporal resolution of fl exible count-based approaches.
****
Evaluating Self-Supervised Learning for Molecular Graph Embeddings
Hanchen Wang, Jean Kaddour, Shengchao Liu, Jian Tang, Joan Lasenby, Qi Liu
Graph Self-Supervised Learning (GSSL) provides a robust pathway for acquiring em beddings without expert labelling, a capability that carries profound implicatio ns for molecular graphs due to the staggering number of potential molecules and the high cost of obtaining labels. However, GSSL methods are designed not for op timisation within a specific domain but rather for transferability across a vari ety of downstream tasks. This broad applicability complicates their evaluation. Addressing this challenge, we present "Molecular Graph Representation Evaluation " (MOLGRAPHEVAL), generating detailed profiles of molecular graph embeddings wit h interpretable and diversified attributes. MOLGRAPHEVAL offers a suite of probi ng tasks grouped into three categories: (i) generic graph, (ii) molecular substr ucture, and (iii) embedding space properties. By leveraging MOLGRAPHEVAL to benc hmark existing GSSL methods against both current downstream datasets and our sui te of tasks, we uncover significant inconsistencies between inferences drawn sol ely from existing datasets and those derived from more nuanced probing. These fi ndings suggest that current evaluation methodologies fail to capture the entiret y of the landscape.
****
SEEDS: Exponential SDE Solvers for Fast High-Quality Sampling from Diffusion Mod els
Martin Gonzalez, Nelson Fernandez Pinto, Thuy Tran, elies Gherbi, Hatem Hajri, N ader Masmoudi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Robust Multi-Agent Reinforcement Learning via Adversarial Regularization: Theore tical Foundation and Stable Algorithms
Alexander Bukharin, Yan Li, Yue Yu, Qingru Zhang, Zhehui Chen, Simiao Zuo, Chao Zhang, Songan Zhang, Tuo Zhao
Multi-Agent Reinforcement Learning (MARL) has shown promising results across sev eral domains. Despite this promise, MARL policies often lack robustness and are therefore sensitive to small changes in their environment. This presents a serio us concern for the real world deployment of MARL algorithms, where the testing e nvironment may slightly differ from the training environment. In this work we sh ow that we can gain robustness by controlling a policy’s Lipschitz constant, and under mild conditions, establish the existence of a Lipschitz and close-to-opti mal policy. Motivated by these insights, we propose a new robust MARL framework, ERNIE, that promotes the Lipschitz continuity of the policies with respect to t he state observations and actions by adversarial regularization. The ERNIE frame work provides robustness against noisy observations, changing transition dynamic s, and malicious actions of agents. However, ERNIE’s adversarial regularization may introduce some training instability. To reduce this instability, we reformul ate adversarial regularization as a Stackelberg game. We demonstrate the effecti veness of the proposed framework with extensive experiments in traffic light con trol and particle environments. In addition, we extend ERNIE to mean-field MARL with a formulation based on distributionally robust optimization that outperform s its non-robust counterpart and is of independent interest. Our code is availab le at https://github.com/abukharin3/ERNIE.
****
Private estimation algorithms for stochastic block models and mixture models
Hongjie Chen, Vincent Cohen-Addad, Tommaso d’Orsi, Alessandro Epasto, Jacob Imol a, David Steurer, Stefan Tiegel
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
UP-NeRF: Unconstrained Pose Prior-Free Neural Radiance Field
Injae Kim, Minhyuk Choi, Hyunwoo J. Kim
Neural Radiance Field (NeRF) has enabled novel view synthesis with high fidelity given images and camera poses. Subsequent works even succeeded in eliminating the necessity of pose priors by jointly optimizing NeRF and camera pose. However, these works are limited to relatively simple settings such as photometrically consistent and occluder-free image collections or a sequence of images from a video. So they have difficulty handling unconstrained images with varying illumination and transient occluders. In this paper, we propose UP-NeRF (Unconstrained Pose-prior-free Neural Radiance Fields) to optimize NeRF with unconstrained image collections without camera pose prior. We tackle these challenges with surrogate tasks that optimize color-insensitive feature fields and a separate module for transient occluders to block their influence on pose estimation. In addition, we introduce a candidate head to enable more robust pose estimation and transient-aware depth supervision to minimize the effect of incorrect prior. Our experiments verify the superior performance of our method compared to the baselines including BARF and its variants in a challenging internet photo collection, Phototourism dataset. The code of UP-NeRF is available at https://github.com/mlvlab/UP-NeRF.
****
Nearly Optimal Bounds for Cyclic Forgetting
William Swartworth, Deanna Needell, Rachel Ward, Mark Kong, Halyun Jeong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
ProteinInvBench: Benchmarking Protein Inverse Folding on Diverse Tasks, Models, and Metrics
Zhangyang Gao, Cheng Tan, Yijie Zhang, Xingran Chen, Lirong Wu, Stan Z. Li
Protein inverse folding has attracted increasing attention in recent years. However, we observe that current methods are usually limited to the CATH dataset and the recovery metric. The lack of a unified framework for ensembling and comparing different methods hinders the comprehensive investigation. In this paper, we propose ProteinBench, a new benchmark for protein design, which comprises extended protein design tasks, integrated models, and diverse evaluation metrics. We broaden the application of methods originally designed for single-chain protein design to new scenarios of multi-chain and \textit{de novo} protein design. Recent impressive methods, including GraphTrans, StructGNN, GVP, GCA, AlphaDesign, ProteinMPNN, PiFold and KWDesign are integrated into our framework. In addition to the recovery, we also evaluate the confidence, diversity, sc-TM, efficiency, and robustness to thoroughly revisit current protein design approaches and inspire future work. As a result, we establish the first comprehensive benchmark for protein design, which is publicly available at \url{https://github.com/A4Bio/OpenCPD}.
****
Understanding and Improving Feature Learning for Out-of-Distribution Generalization
Yongqiang Chen, Wei Huang, Kaiwen Zhou, Yatao Bian, Bo Han, James Cheng
A common explanation for the failure of out-of-distribution (OOD) generalization is that the model trained with empirical risk minimization (ERM) learns spurious features instead of invariant features. However, several recent studies challenged this explanation and found that deep networks may have already learned sufficiently good features for OOD generalization. Despite the contradictions at first glance, we theoretically show that ERM essentially learns both spurious and invariant features, while ERM tends to learn spurious features faster if the spurious correlation is stronger. Moreover, when fed the ERM learned features to the

 OOD objectives, the invariant feature learning quality significantly affects th
e final OOD performance, as OOD objectives rarely learn new features. Therefore,
 ERM feature learning can be a bottleneck to OOD generalization. To alleviate th
e reliance, we propose Feature Augmented Training (FeAT), to enforce the model t
o learn richer features ready for OOD generalization. FeAT iteratively augments
the model to learn new features while retaining the already learned features. In
 each round, the retention and augmentation operations are performed on differen
t subsets of the training data that capture distinct features. Extensive experim
ents show that FeAT effectively learns richer features thus boosting the perform
ance of various OOD objectives.
****
Train Hard, Fight Easy: Robust Meta Reinforcement Learning
Ido Greenberg, Shie Mannor, Gal Chechik, Eli Meirom
A major challenge of reinforcement learning (RL) in real-world applications is t
he variation between environments, tasks or clients. Meta-RL (MRL) addresses thi
s issue by learning a meta-policy that adapts to new tasks. Standard MRL methods
 optimize the average return over tasks, but often suffer from poor results in t
asks of high risk or difficulty. This limits system reliability since test tasks
 are not known in advance. In this work, we define a robust MRL objective with a
 controlled robustness level. Optimization of analogous robust objectives in RL
is known to lead to both biased gradients and data inefficiency. We prove that t
he gradient bias disappears in our proposed MRL framework. The data inefficiency
 is addressed via the novel Robust Meta RL algorithm (RoML). RoML is a meta-algo
rithm that generates a robust version of any given MRL algorithm, by identifying
 and over-sampling harder tasks throughout training. We demonstrate that RoML ac
hieves robust returns on multiple navigation and continuous control benchmarks.
****
Towards Semi-Structured Automatic ICD Coding via Tree-based Contrastive Learning
Chang Lu, Chandan Reddy, Ping Wang, Yue Ning
Automatic coding of International Classification of Diseases (ICD) is a multi-la
bel text categorization task that involves extracting disease or procedure codes
 from clinical notes. Despite the application of state-of-the-art natural langua
ge processing (NLP) techniques, there are still challenges including limited ava
ilability of data due to privacy constraints and the high variability of clinica
l notes caused by different writing habits of medical professionals and various
pathological features of patients. In this work, we investigate the semi-structu
red nature of clinical notes and propose an automatic algorithm to segment them
into sections. To address the variability issues in existing ICD coding models w
ith limited data, we introduce a contrastive pre-training approach on sections u
sing a soft multi-label similarity metric based on tree edit distance. Additiona
lly, we design a masked section training strategy to enable ICD coding models to
 locate sections related to ICD codes. Extensive experimental results demonstrat
e that our proposed training strategies effectively enhance the performance of e
xisting ICD coding methods.
****
Stable Vectorization of Multiparameter Persistent Homology using Signed Barcodes
 as Measures
David Loiseaux, Luis Scoccola, Mathieu Carrière, Magnus Bakke Botnan, Steve OUDO
T
Persistent homology (PH) provides topological descriptors for geometric data, s
uch as weighted graphs, which are interpretable, stable to perturbations, and in
variant under, e.g., relabeling. Most applications of PH focus on the one-parame
ter case---where the descriptors summarize the changes in topology of data as it
 is filtered by a single quantity of interest---and there is now a wide array of
 methods enabling the use of one-parameter PH descriptors in data science, which
 rely on the stable vectorization of these descriptors as elements of a Hilbert
space. Although the multiparameter PH (MPH) of data that is filtered by several
quantities of interest encodes much richer information than its one-parameter co
unterpart, the scarceness of stability results for MPH descriptors has so far li
mited the available options for the stable vectorization of MPH. In this paper,

we aim to bring together the best of both worlds by showing how the interpretation of signed barcodes---a recent family of MPH descriptors---as signed Radon measures leads to natural extensions of vectorization strategies from one parameter to multiple parameters. The resulting feature vectors are easy to define and to compute, and provably stable. While, as a proof of concept, we focus on simple choices of signed barcodes and vectorizations, we already see notable performance improvements when comparing our feature vectors to state-of-the-art topology-based methods on various types of data.
****
Subclass-Dominant Label Noise: A Counterexample for the Success of Early Stopping
Yingbin Bai, Zhongyi Han, Erkun Yang, Jun Yu, Bo Han, Dadong Wang, Tongliang Liu
In this paper, we empirically investigate a previously overlooked and widespread type of label noise, subclass-dominant label noise (SDN). Our findings reveal that, during the early stages of training, deep neural networks can rapidly memorize mislabeled examples in SDN. This phenomenon poses challenges in effectively selecting confident examples using conventional early stopping techniques. To address this issue, we delve into the properties of SDN and observe that long-trained representations are superior at capturing the high-level semantics of mislabeled examples, leading to a clustering effect where similar examples are grouped together. Based on this observation, we propose a novel method called NoiseCluster that leverages the geometric structures of long-trained representations to identify and correct SDN. Our experiments demonstrate that NoiseCluster outperforms state-of-the-art baselines on both synthetic and real-world datasets, highlighting the importance of addressing SDN in learning with noisy labels. The code is available at https://github.com/tmllab/2023NeurIPSSDN.
****
OpenMask3D: Open-Vocabulary 3D Instance Segmentation
Ayca Takmaz, Elisabetta Fedele, Robert Sumner, Marc Pollefeys, Federico Tombari, Francis Engelmann
We introduce the task of open-vocabulary 3D instance segmentation. Current approaches for 3D instance segmentation can typically only recognize object categories from a pre-defined closed set of classes that are annotated in the training datasets. This results in important limitations for real-world applications where one might need to perform tasks guided by novel, open-vocabulary queries related to a wide variety of objects. Recently, open-vocabulary 3D scene understanding methods have emerged to address this problem by learning queryable features for each point in the scene. While such a representation can be directly employed to perform semantic segmentation, existing methods cannot separate multiple object instances. In this work, we address this limitation, and propose OpenMask3D, which is a zero-shot approach for open-vocabulary 3D instance segmentation. Guided by predicted class-agnostic 3D instance masks, our model aggregates per-mask features via multi-view fusion of CLIP-based image embeddings. Experiments and ablation studies on ScanNet200 and Replica show that OpenMask3D outperforms other open-vocabulary methods, especially on the long-tail distribution. Qualitative experiments further showcase OpenMask3D’s ability to segment object properties based on free-form queries describing geometry, affordances, and materials.
****
Model-Based Reparameterization Policy Gradient Methods: Theory and Practical Algorithms
Shenao Zhang, Boyi Liu, Zhaoran Wang, Tuo Zhao
ReParameterization (RP) Policy Gradient Methods (PGMs) have been widely adopted for continuous control tasks in robotics and computer graphics. However, recent studies have revealed that, when applied to long-term reinforcement learning problems, model-based RP PGMs may experience chaotic and non-smooth optimization landscapes with exploding gradient variance, which leads to slow convergence. This is in contrast to the conventional belief that reparameterization methods have low gradient estimation variance in problems such as training deep generative models. To comprehend this phenomenon, we conduct a theoretical examination of model-based RP PGMs and search for solutions to the optimization difficulties. Spec

ifically, we analyze the convergence of the model-based RP PGMs and pinpoint the smoothness of function approximators as a major factor that affects the quality of gradient estimation. Based on our analysis, we propose a spectral normalization method to mitigate the exploding variance issue caused by long model unrolls. Our experimental results demonstrate that proper normalization significantly reduces the gradient variance of model-based RP PGMs. As a result, the performance of the proposed method is comparable or superior to other gradient estimators, such as the Likelihood Ratio (LR) gradient estimator. Our code is available at https://github.com/agentification/RP_PGM.
****
Bitstream-Corrupted Video Recovery: A Novel Benchmark Dataset and Method
Tianyi Liu, Kejun Wu, Yi Wang, Wenyang Liu, Kim-Hui Yap, Lap-Pui Chau
The past decade has witnessed great strides in video recovery by specialist technologies, like video inpainting, completion, and error concealment. However, they typically simulate the missing content by manual-designed error masks, thus failing to fill in the realistic video loss in video communication (e.g., telepresence, live streaming, and internet video) and multimedia forensics. To address this, we introduce the bitstream-corrupted video (BSCV) benchmark, the first benchmark dataset with more than 28,000 video clips, which can be used for bitstream-corrupted video recovery in the real world. The BSCV is a collection of 1) a proposed three-parameter corruption model for video bitstream, 2) a large-scale dataset containing rich error patterns, multiple corruption levels, and flexible dataset branches, and 3) a new video recovery framework that serves as a benchmark. We evaluate state-of-the-art video inpainting methods on the BSCV dataset, demonstrating existing approaches' limitations and our framework's advantages in solving the bitstream-corrupted video recovery problem. The benchmark and dataset are released at https://github.com/LIUTIGHE/BSCV-Dataset.
****
Structured Neural-PI Control with End-to-End Stability and Output Tracking Guarantees
Wenqi Cui, Yan Jiang, Baosen Zhang, Yuanyuan Shi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DäRF: Boosting Radiance Fields from Sparse Input Views with Monocular Depth Adaptation
Jiuhn Song, Seonghoon Park, Honggyu An, Seokju Cho, Min-Seop Kwak, Sungjin Cho, Seungryong Kim
Neural radiance field (NeRF) shows powerful performance in novel view synthesis and 3D geometry reconstruction, but it suffers from critical performance degradation when the number of known viewpoints is drastically reduced. Existing works attempt to overcome this problem by employing external priors, but their success is limited to certain types of scenes or datasets. Employing monocular depth estimation (MDE) networks, pretrained on large-scale RGB-D datasets, with powerful generalization capability may be a key to solving this problem: however, using MDE in conjunction with NeRF comes with a new set of challenges due to various ambiguity problems exhibited by monocular depths. In this light, we propose a novel framework, dubbed DäRF, that achieves robust NeRF reconstruction with a handful of real-world images by combining the strengths of NeRF and monocular depth estimation through online complementary training. Our framework imposes the MDE network's powerful geometry prior to NeRF representation at both seen and unseen viewpoints to enhance its robustness and coherence. In addition, we overcome the ambiguity problems of monocular depths through patch-wise scale-shift fitting and geometry distillation, which adapts the MDE network to produce depths aligned accurately with NeRF geometry. Experiments show our framework achieves state-of-the-art results both quantitatively and qualitatively, demonstrating consistent and reliable performance in both indoor and outdoor real-world datasets.
****

Enhancing Knowledge Transfer for Task Incremental Learning with Data-free Subnetwork
Qiang Gao, Xiaojun Shan, Yuchen Zhang, Fan Zhou
As there exist competitive subnetworks within a dense network in concert with Lottery Ticket Hypothesis, we introduce a novel neuron-wise task incremental learning method, namely Data-free Subnetworks (DSN), which attempts to enhance the elastic knowledge transfer across the tasks that sequentially arrive. Specifically, DSN primarily seeks to transfer knowledge to the new coming task from the learned tasks by selecting the affiliated weights of a small set of neurons to be activated, including the reused neurons from prior tasks via neuron-wise masks. And it also transfers possibly valuable knowledge to the earlier tasks via data-free replay. Especially, DSN inherently relieves the catastrophic forgetting and the unavailability of past data or possible privacy concerns. The comprehensive experiments conducted on four benchmark datasets demonstrate the effectiveness of the proposed DSN in the context of task-incremental learning by comparing it to several state-of-the-art baselines. In particular, DSN enables the knowledge transfer to the earlier tasks, which is often overlooked by prior efforts.
****
rPPG-Toolbox: Deep Remote PPG Toolbox
Xin Liu, Girish Narayanswamy, Akshay Paruchuri, Xiaoyu Zhang, Jiankai Tang, Yuzhe Zhang, Roni Sengupta, Shwetak Patel, Yuntao Wang, Daniel McDuff
Camera-based physiological measurement is a fast growing field of computer vision. Remote photoplethysmography (rPPG) utilizes imaging devices (e.g., cameras) to measure the peripheral blood volume pulse (BVP) via photoplethysmography, and enables cardiac measurement via webcams and smartphones. However, the task is non-trivial with important pre-processing, modeling and post-processing steps required to obtain state-of-the-art results. Replication of results and benchmarking of new models is critical for scientific progress; however, as with many other applications of deep learning, reliable codebases are not easy to find or use. We present a comprehensive toolbox, rPPG-Toolbox, unsupervised and supervised rPPG models with support for public benchmark datasets, data augmentation and systematic evaluation: https://github.com/ubicomplab/rPPG-Toolbox.
****
Proximity-Informed Calibration for Deep Neural Networks
Miao Xiong, Ailin Deng, Pang Wei W. Koh, Jiaying Wu, Shen Li, Jianqing Xu, Bryan Hooi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Toolformer: Language Models Can Teach Themselves to Use Tools
Timo Schick, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, Thomas Scialom
Language models (LMs) exhibit remarkable abilities to solve new tasks from just a few examples or textual instructions, especially at scale. They also, paradoxically, struggle with basic functionality, such as arithmetic or factual lookup, where much simpler and smaller specialized models excel. In this paper, we show that LMs can teach themselves to use external tools via simple APIs and achieve the best of both worlds. We introduce Toolformer, a model trained to decide which APIs to call, when to call them, what arguments to pass, and how to best incorporate the results into future token prediction. This is done in a self-supervised way, requiring nothing more than a handful of demonstrations for each API. We incorporate a range of tools, including a calculator, a Q&A system, a search engine, a translation system, and a calendar. Toolformer achieves substantially improved zero-shot performance across a variety of downstream tasks, often competitive with much larger models, without sacrificing its core language modeling abilities.
****
The probability flow ODE is provably fast

Sitan Chen, Sinho Chewi, Holden Lee, Yuanzhi Li, Jianfeng Lu, Adil Salim
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Faster Discrete Convex Function Minimization with Predictions: The M-Convex Case
Taihei Oki, Shinsaku Sakaue
Recent years have seen a growing interest in accelerating optimization algorithms with machine-learned predictions. Sakaue and Oki (NeurIPS 2022) have developed a general framework that warm-starts the L-convex function minimization method with predictions, revealing the idea's usefulness for various discrete optimization problems. In this paper, we present a framework for using predictions to accelerate M-convex function minimization, thus complementing previous research and extending the range of discrete optimization algorithms that can benefit from predictions. Our framework is particularly effective for an important subclass called laminar convex minimization, which appears in many operations research applications. Our methods can improve time complexity bounds upon the best worst-case results by using predictions and even have potential to go beyond a lower-bound result.
****
Using Imperfect Surrogates for Downstream Inference: Design-based Supervised Learning for Social Science Applications of Large Language Models
Naoki Egami, Musashi Hinck, Brandon Stewart, Hanying Wei
In computational social science (CSS), researchers analyze documents to explain social and political phenomena. In most scenarios, CSS researchers first obtain labels for documents and then explain labels using interpretable regression analyses in the second step. One increasingly common way to annotate documents cheaply at scale is through large language models (LLMs). However, like other scalable ways of producing annotations, such surrogate labels are often imperfect and biased. We present a new algorithm for using imperfect annotation surrogates for downstream statistical analyses while guaranteeing statistical properties—like asymptotic unbiasedness and proper uncertainty quantification—which are fundamental to CSS research. We show that direct use of surrogate labels in downstream statistical analyses leads to substantial bias and invalid confidence intervals, even with high surrogate accuracy of 80-90\%. To address this, we build on debiased machine learning to propose the design-based supervised learning (DSL) estimator. DSL employs a doubly-robust procedure to combine surrogate labels with a smaller number of high-quality, gold-standard labels. Our approach guarantees valid inference for downstream statistical analyses, even when surrogates are arbitrarily biased and without requiring stringent assumptions, by controlling the probability of sampling documents for gold-standard labeling. Both our theoretical analysis and experimental results show that DSL provides valid statistical inference while achieving root mean squared errors comparable to existing alternatives that focus only on prediction without inferential guarantees.
****
Individual Arbitrariness and Group Fairness
Carol Long, Hsiang Hsu, Wael Alghamdi, Flavio Calmon
Machine learning tasks may admit multiple competing models that achieve similar performance yet produce conflicting outputs for individual samples---a phenomenon known as predictive multiplicity. We demonstrate that fairness interventions in machine learning optimized solely for group fairness and accuracy can exacerbate predictive multiplicity. Consequently, state-of-the-art fairness interventions can mask high predictive multiplicity behind favorable group fairness and accuracy metrics. We argue that a third axis of ``arbitrariness'' should be considered when deploying models to aid decision-making in applications of individual-level impact.To address this challenge, we propose an ensemble algorithm applicable to any fairness intervention that provably ensures more consistent predictions.
****

ASPEN: Breaking Operator Barriers for Efficient Parallelization of Deep Neural N
etworks
Jongseok Park, Kyungmin Bin, Gibum Park, Sangtae Ha, Kyunghan Lee
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
####################################
Parallel Submodular Function Minimization
Deeparnab Chakrabarty, Andrei Graur, Haotian Jiang, Aaron Sidford
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
####################################
Emergent Communication for Rules Reasoning
Yuxuan Guo, Yifan Hao, Rui Zhang, Enshuai Zhou, Zidong Du, xishan zhang, Xinkai
Song, Yuanbo Wen, Yongwei Zhao, Xuehai Zhou, Jiaming Guo, Qi Yi, Shaohui Peng, D
i Huang, Ruizhi Chen, Qi Guo, Yunji Chen
Research on emergent communication between deep-learning-based agents has receiv
ed extensive attention due to its inspiration for linguistics and artificial int
elligence. However, previous attempts have hovered around emerging communicati
on under perception-oriented environmental settings, that forces agents to desc
ribe low-level perceptual features intra image or symbol contexts. In this work
, inspired by the classic human reasoning test (namely Raven's Progressive Matri
x), we propose the Reasoning Game, a cognition-oriented environment that encoura
ges agents to reason and communicate high-level rules, rather than perceived low
-level contexts. Moreover, we propose 1) an unbiased dataset (namely rule-RAVEN
) as a benchmark to avoid overfitting, 2) and a two-stage curriculum agent train
ing method as a baseline for more stable convergence in the Reasoning Game, whe
re contexts and semantics are bilaterally drifting. Experimental results show t
hat, in the Reasoning Game, a semantically stable and compositional language eme
rges to solve reasoning problems. The emerged language helps agents apply the e
xtracted rules to the generalization of unseen context attributes, and to the tr
ansfer between different context attributes or even tasks.
####################################
A Regularized Conditional GAN for Posterior Sampling in Image Recovery Problems
Matthew Bendel, Rizwan Ahmad, Philip Schniter
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
####################################
Would I have gotten that reward? Long-term credit assignment by counterfactual c
ontribution analysis
Alexander Meulemans, Simon Schug, Seijin Kobayashi, nathaniel daw, Gregory Wayne
To make reinforcement learning more sample efficient, we need better credit assi
gnment methods that measure an action’s influence on future rewards. Building up
on Hindsight Credit Assignment (HCA), we introduce Counterfactual Contribution A
nalysis (COCOA), a new family of model-based credit assignment algorithms. Our a
lgorithms achieve precise credit assignment by measuring the contribution of act
ions upon obtaining subsequent rewards, by quantifying a counterfactual query: ‘
Would the agent still have reached this reward if it had taken another action?’.
 We show that measuring contributions w.r.t. rewarding states, as is done in HCA
, results in spurious estimates of contributions, causing HCA to degrade towards
 the high-variance REINFORCE estimator in many relevant environments. Instead, w
e measure contributions w.r.t. rewards or learned representations of the rewardi
ng objects, resulting in gradient estimates with lower variance. We run experime
nts on a suite of problems specifically designed to evaluate long-term credit as
signment capabilities. By using dynamic programming, we measure ground-truth pol

icy gradients and show that the improved performance of our new model-based credit assignment methods is due to lower bias and variance compared to HCA and common baselines. Our results demonstrate how modeling action contributions towards rewarding outcomes can be leveraged for credit assignment, opening a new path towards sample-efficient reinforcement learning.
****
HOH: Markerless Multimodal Human-Object-Human Handover Dataset with Large Object Count
Noah Wiederhold, Ava Megyeri, DiMaggio Paris, Sean Banerjee, Natasha Banerjee
We present the HOH (Human-Object-Human) Handover Dataset, a large object count dataset with 136 objects, to accelerate data-driven research on handover studies, human-robot handover implementation, and artificial intelligence (AI) on handover parameter estimation from 2D and 3D data of two-person interactions. HOH contains multi-view RGB and depth data, skeletons, fused point clouds, grasp type and handedness labels, object, giver hand, and receiver hand 2D and 3D segmentations, giver and receiver comfort ratings, and paired object metadata and aligned 3D models for 2,720 handover interactions spanning 136 objects and 20 giver-receiver pairs—40 with role-reversal—organized from 40 participants. We also show experimental results of neural networks trained using HOH to perform grasp, orientation, and trajectory prediction. As the only fully markerless handover capture dataset, HOH represents natural human-human handover interactions, overcoming challenges with markered datasets that require specific suiting for body tracking, and lack high-resolution hand tracking. To date, HOH is the largest handover dataset in terms of object count, participant count, pairs with role reversal accounted for, and total interactions captured.
****
Tight Risk Bounds for Gradient Descent on Separable Data
Matan Schliserman, Tomer Koren
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Video Prediction Models as Rewards for Reinforcement Learning
Alejandro Escontrela, Ademi Adeniji, Wilson Yan, Ajay Jain, Xue Bin Peng, Ken Goldberg, Youngwoon Lee, Danijar Hafner, Pieter Abbeel
Specifying reward signals that allow agents to learn complex behaviors is a long-standing challenge in reinforcement learning.A promising approach is to extract preferences for behaviors from unlabeled videos, which are widely available on the internet. We present Video Prediction Rewards (VIPER), an algorithm that leverages pretrained video prediction models as action-free reward signals for reinforcement learning. Specifically, we first train an autoregressive transformer on expert videos and then use the video prediction likelihoods as reward signals for a reinforcement learning agent. VIPER enables expert-level control without programmatic task rewards across a wide range of DMC, Atari, and RLBench tasks. Moreover, generalization of the video prediction model allows us to derive rewards for an out-of-distribution environment where no expert data is available, enabling cross-embodiment generalization for tabletop manipulation. We see our work as starting point for scalable reward specification from unlabeled videos that will benefit from the rapid advances in generative modeling. Source code and datasets are available on the project website: https://ViperRL.com
****
Provably (More) Sample-Efficient Offline RL with Options
Xiaoyan Hu, Ho-fung Leung
The options framework yields empirical success in long-horizon planning problems of reinforcement learning (RL). Recent works show that options help improve the sample efficiency in online RL. However, these results are no longer applicable to scenarios where exploring the environment online is risky, e.g., automated driving and healthcare. In this paper, we provide the first analysis of the sample complexity for offline RL with options, where the agent learns from a dataset

without further interaction with the environment. We derive a novel information-theoretic lower bound, which generalizes the one for offline learning with actions. We propose the PEssimistic Value Iteration for Learning with Options (PEVIO) algorithm and establish near-optimal suboptimality bounds for two popular data-collection procedures, where the first one collects state-option transitions and the second one collects state-action transitions. We show that compared to offline RL with actions, using options not only enjoys a faster finite-time convergence rate (to the optimal value) but also attains a better performance when either the options are carefully designed or the offline data is limited. Based on these results, we analyze the pros and cons of the data-collection procedures.
****

Rewrite Caption Semantics: Bridging Semantic Gaps for Language-Supervised Semantic Segmentation

Yun Xing, Jian Kang, Aoran Xiao, Jiahao Nie, Ling Shao, Shijian Lu

Vision-Language Pre-training has demonstrated its remarkable zero-shot recognition ability and potential to learn generalizable visual representations from languagesupervision. Taking a step ahead, language-supervised semantic segmentation enables spatial localization of textual inputs by learning pixel grouping solely from image-text pairs. Nevertheless, the state-of-the-art suffers from a clear semantic gap between visual and textual modalities: plenty of visual concepts appeared in images are missing in their paired captions. Such semantic misalignment circulates in pre-training, leading to inferior zero-shot performance in dense predictions due to insufficient visual concepts captured in textual representations. To close such semantic gap, we propose Concept Curation (CoCu), a pipeline that leverages CLIP to compensate for the missing semantics. For each image-text pair, we establish a concept archive that maintains potential visually-matched concepts with our proposed vision-driven expansion and text-to-vision-guided ranking. Relevant concepts can thus be identified via cluster-guided sampling and fed into pre-training, thereby bridging the gap between visual and textual semantics. Extensive experiments over a broad suite of 8 segmentation benchmarks show that CoCu achieves superb zero-shot transfer performance and greatly boosts language-supervised segmentation baseline by a large margin, suggesting the value of closing semantic gap in pre-training data.
****

Approximate Allocation Matching for Structural Causal Bandits with Unobserved Confounders

Lai Wei, Muhammad Qasim Elahi, Mahsa Ghasemi, Murat Kocaoglu

Structural causal bandit provides a framework for online decision-making problems when causal information is available. It models the stochastic environment with a structural causal model (SCM) that governs the causal relations between random variables. In each round, an agent applies an intervention (or no intervention) by setting certain variables to some constants and receives a stochastic reward from a non-manipulable variable. Though the causal structure is given, the observational and interventional distributions of these random variables are unknown beforehand, and they can only be learned through interactions with the environment. Therefore, to maximize the expected cumulative reward, it is critical to balance the explore-versus-exploit tradeoff. We assume each random variable takes a finite number of distinct values, and consider a semi-Markovian setting, where random variables are affected by unobserved confounders. Using the canonical SCM formulation to discretize the domains of unobserved variables, we efficiently integrate samples to reduce model uncertainty. This gives the decision maker a natural advantage over those in a classical multi-armed bandit setup. We provide a logarithmic asymptotic regret lower bound for the structural causal bandit problem. Inspired by the lower bound, we design an algorithm that can utilize the causal structure to accelerate the learning process and take informative and rewarding interventions. We establish that our algorithm achieves a logarithmic regret and demonstrate that it outperforms the existing methods via simulations.
****

Human-Guided Complexity-Controlled Abstractions

Andi Peng, Mycal Tucker, Eoin Kenny, Noga Zaslavsky, Pulkit Agrawal, Julie A Sha

h
Neural networks often learn task-specific latent representations that fail to generalize to novel settings or tasks. Conversely, humans learn discrete representations (i.e., concepts or words) at a variety of abstraction levels (e.g., "bird" vs. "sparrow'") and use the appropriate abstraction based on tasks. Inspired by this, we train neural models to generate a spectrum of discrete representations, and control the complexity of the representations (roughly, how many bits are allocated for encoding inputs) by tuning the entropy of the distribution over representations. In finetuning experiments, using only a small number of labeled examples for a new task, we show that (1) tuning the representation to a task-appropriate complexity level supports the greatest finetuning performance, and (2) in a human-participant study, users were able to identify the appropriate complexity level for a downstream task via visualizations of discrete representations. Our results indicate a promising direction for rapid model finetuning by leveraging human insight.
****
Scenario Diffusion: Controllable Driving Scenario Generation With Diffusion
Ethan Pronovost, Meghana Reddy Ganesina, Noureldin Hendy, Zeyu Wang, Andres Morales, Kai Wang, Nick Roy
Automated creation of synthetic traffic scenarios is a key part of scaling the safety validation of autonomous vehicles (AVs). In this paper, we propose Scenario Diffusion, a novel diffusion-based architecture for generating traffic scenarios that enables controllable scenario generation. We combine latent diffusion, object detection and trajectory regression to generate distributions of synthetic agent poses, orientations and trajectories simultaneously. This distribution is conditioned on the map and sets of tokens describing the desired scenario to provide additional control over the generated scenario. We show that our approach has sufficient expressive capacity to model diverse traffic patterns and generalizes to different geographical regions.
****
Label-Only Model Inversion Attacks via Knowledge Transfer
Bao-Ngoc Nguyen, Keshigeyan Chandrasegaran, Milad Abdollahzadeh, Ngai-Man (Man) Cheung
In a model inversion (MI) attack, an adversary abuses access to a machine learning (ML) model to infer and reconstruct private training data. Remarkable progress has been made in the white-box and black-box setups, where the adversary has access to the complete model or the model's soft output respectively. However, there is very limited study in the most challenging but practically important setup: Label-only MI attacks, where the adversary only has access to the model's predicted label (hard label) without confidence scores nor any other model information. In this work, we propose LOKT, a novel approach for label-only MI attacks. Our idea is based on transfer of knowledge from the opaque target model to surrogate models. Subsequently, using these surrogate models, our approach can harness advanced white-box attacks. We propose knowledge transfer based on generative modelling, and introduce a new model, Target model-assisted ACGAN (T-ACGAN), for effective knowledge transfer. Our method casts the challenging label-only MI into the more tractable white-box setup. We provide analysis to support that surrogate models based on our approach serve as effective proxies for the target model for MI. Our experiments show that our method significantly outperforms existing SOTA Label-only MI attack by more than 15% across all MI benchmarks. Furthermore, our method compares favorably in terms of query budget. Our study highlights rising privacy threats for ML models even when minimal information (i.e., hard labels) is exposed. Our study highlights rising privacy threats for ML models even when minimal information (i.e., hard labels) is exposed. Our code, demo, models and reconstructed data are available at our project page:https://ngoc-nguyen-0.github.io/lokt/
****
On the Adversarial Robustness of Out-of-distribution Generalization Models
Xin Zou, Weiwei Liu
Out-of-distribution (OOD) generalization has attracted increasing research atten

tion in recent years, due to its promising experimental results in real-world ap plications. Interestingly, we find that existing OOD generalization methods are vulnerable to adversarial attacks. This motivates us to study OOD adversarial ro bustness. We first present theoretical analyses of OOD adversarial robustness in two different complementary settings. Motivated by the theoretical results, we design two algorithms to improve the OOD adversarial robustness. Finally, we con duct experiments to validate the effectiveness of our proposed algorithms.
****

Utilitarian Algorithm Configuration

Devon Graham, Kevin Leyton-Brown, Tim Roughgarden

We present the first nontrivial procedure for configuring heuristic algorithms t o maximize the utility provided to their end users while also offering theoretic al guarantees about performance. Existing procedures seek configurations that mi nimize expected runtime. However, very recent theoretical work argues that expec ted runtime minimization fails to capture algorithm designers' preferences. Here we show that the utilitarian objective also confers significant algorithmic ben efits. Intuitively, this is because mean runtime is dominated by extremely long runs even when they are incredibly rare; indeed, even when an algorithm never gi ves rise to such long runs, configuration procedures that provably minimize mean runtime must perform a huge number of experiments to demonstrate this fact. In contrast, utility is bounded and monotonically decreasing in runtime, allowing f or meaningful empirical bounds on a configuration's performance. This paper buil ds on this idea to describe effective and theoretically sound configuration proc edures. We prove upper bounds on the runtime of these procedures that are simila r to theoretical lower bounds, while also demonstrating their performance empiri cally.
****

Double Randomized Underdamped Langevin with Dimension-Independent Convergence Gu arantee

Yuanshi Liu, Cong Fang, Tong Zhang

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Non-Asymptotic Analysis of a UCB-based Top Two Algorithm

Marc Jourdan, Rémy Degenne

A Top Two sampling rule for bandit identification is a method which selects the next arm to sample from among two candidate arms, a leader and a challenger. Due to their simplicity and good empirical performance, they have received increase d attention in recent years. However, for fixed-confidence best arm identificati on, theoretical guarantees for Top Two methods have only been obtained in the as ymptotic regime, when the error level vanishes. In this paper, we derive the fir st non-asymptotic upper bound on the expected sample complexity of a Top Two alg orithm, which holds for any error level. Our analysis highlights sufficient prop erties for a regret minimization algorithm to be used as leader. These propertie s are satisfied by the UCB algorithm, and our proposed UCB-based Top Two algorit hm simultaneously enjoys non-asymptotic guarantees and competitive empirical per formance.
****

Statistical and Computational Trade-off in Multi-Agent Multi-Armed Bandits

Filippo Vannella, Alexandre Proutiere, Jaeseong Jeong

We study the problem of regret minimization in Multi-Agent Multi-Armed Bandits (MAMABs) where the rewards are defined through a factor graph. We derive an insta nce-specific regret lower bound and characterize the minimal expected number of times each global action should be explored. Unfortunately, this bound and the c orresponding optimal exploration process are obtained by solving a combinatorial optimization problem with a set of variables and constraints exponentially grow ing with the number of agents. We approximate the regret lower bound problem via Mean Field techniques to reduce the number of variables and constraints. By tun

ing the latter, we explore the trade-off between achievable regret and complexit y. We devise Efficient Sampling for MAMAB (ESM), an algorithm whose regret asymp totically matches the corresponding approximated lower bound. We assess the regr et and computational complexity of ESM numerically, using both synthetic and rea l-world experiments in radio communications networks.
****
On permutation symmetries in Bayesian neural network posteriors: a variational p erspective
Simone Rossi, Ankit Singh, Thomas Hannagan
The elusive nature of gradient-based optimization in neural networks is tied to their loss landscape geometry, which is poorly understood. However recent work h as brought solid evidence that there is essentially no loss barrier between the local solutions of gradient descent, once accounting for weight-permutations tha t leave the network's computation unchanged. This raises questions for approxima te inference in Bayesian neural networks (BNNs), where we are interested in marg inalizing over multiple points in the loss landscape.In this work, we first exte nd the formalism of marginalized loss barrier and solution interpolation to BNNs , before proposing a matching algorithm to search for linearly connected solutio ns. This is achieved by aligning the distributions of two independent approximat e Bayesian solutions with respect to permutation matrices. Building on the work of Ainsworth et al. (2023), we frame the problem as a combinatorial optimization one, using an approximation to the sum of bilinear assignment problem. We then experiment on a variety of architectures and datasets, finding nearly zero margi nalized loss barriers for linearly connected solutions.
****
Hierarchical Decomposition of Prompt-Based Continual Learning: Rethinking Obscur ed Sub-optimality
Liyuan Wang, Jingyi Xie, Xingxing Zhang, Mingyi Huang, Hang Su, Jun Zhu
Prompt-based continual learning is an emerging direction in leveraging pre-train ed knowledge for downstream continual learning, and has almost reached the perfo rmance pinnacle under supervised pre-training. However, our empirical research r eveals that the current strategies fall short of their full potential under the more realistic self-supervised pre-training, which is essential for handling vas t quantities of unlabeled data in practice. This is largely due to the difficult y of task-specific knowledge being incorporated into instructed representations via prompt parameters and predicted by uninstructed representations at test time . To overcome the exposed sub-optimality, we conduct a theoretical analysis of t he continual learning objective in the context of pre-training, and decompose it into hierarchical components: within-task prediction, task-identity inference, and task-adaptive prediction. Following these empirical and theoretical insights , we propose Hierarchical Decomposition (HiDe-)Prompt, an innovative approach th at explicitly optimizes the hierarchical components with an ensemble of task-spe cific prompts and statistics of both uninstructed and instructed representations , further with the coordination of a contrastive regularization strategy. Our ex tensive experiments demonstrate the superior performance of HiDe-Prompt and its robustness to pre-training paradigms in continual learning (e.g., up to 15.01% a nd 9.61% lead on Split CIFAR-100 and Split ImageNet-R, respectively).
****
3D molecule generation by denoising voxel grids
Pedro O. O. Pinheiro, Joshua Rackers, Joseph Kleinhenz, Michael Maser, Omar Mahm ood, Andrew Watkins, Stephen Ra, Vishnu Sresht, Saeed Saremi
We propose a new score-based approach to generate 3D molecules represented as at omic densities on regular grids.First, we train a denoising neural network that learns to map from a smooth distribution of noisy molecules to the distribution of real molecules.Then, we follow the neural empirical Bayes framework [Saremi a nd Hyvarinen, 2019] and generate molecules in two steps: (i) sample noisy densit y grids from a smooth distribution via underdamped Langevin Markov chain Monte C arlo, and (ii) recover the "clean" molecule by denoising the noisy grid with a s ingle step.Our method, VoxMol, generates molecules in a fundamentally different way than the current state of the art (ie, diffusion models applied to atom poin

t clouds). It differs in terms of the data representation, the noise model, the network architecture and the generative modeling algorithm.Our experiments show that VoxMol captures the distribution of drug-like molecules better than state of the art, while being faster to generate samples.
****
Accessing Higher Dimensions for Unsupervised Word Translation
Sida Wang
The striking ability of unsupervised word translation has been demonstrated recently with the help of low-dimensional word vectors / pretraining, which is used by all successful methods and assumed to be necessary. We test and challenge this assumption by developing a method that can also make use of high dimensional signal. Freed from the limits of low dimensions, we show that relying on low-dimensional vectors and their incidental properties miss out on better denoising methods and signals in high dimensions, thus stunting the potential of the data. Our results show that unsupervised translation can be achieved more easily and robustly than previously thought -- less than 80MB and minutes of CPU time is required to achieve over 50\% accuracy for English to Finnish, Hungarian, and Chinese translations when trained in the same domain; even under domain mismatch, the method still works fully unsupervised on English NewsCrawl to Chinese Wikipedia and English Europarl to Spanish Wikipedia, among others. These results challenge prevailing assumptions on the necessity and superiority of low-dimensional vectors and show that the higher dimension signal can be used rather than thrown away.
****
Inverse Reinforcement Learning with the Average Reward Criterion
Feiyang Wu, Jingyang Ke, Anqi Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DisDiff: Unsupervised Disentanglement of Diffusion Probabilistic Models
Tao Yang, Yuwang Wang, Yan Lu, Nanning Zheng
Targeting to understand the underlying explainable factors behind observations and modeling the conditional generation process on these factors, we connect disentangled representation learning to diffusion probabilistic models (DPMs) to take advantage of the remarkable modeling ability of DPMs. We propose a new task, disentanglement of (DPMs): given a pre-trained DPM, without any annotations of the factors, the task is to automatically discover the inherent factors behind the observations and disentangle the gradient fields of DPM into sub-gradient fields, each conditioned on the representation of each discovered factor. With disentangled DPMs, those inherent factors can be automatically discovered, explicitly represented and clearly injected into the diffusion process via the sub-gradient fields. To tackle this task, we devise an unsupervised approach, named DisDiff, and for the first time achieving disentangled representation learning in the framework of DPMs. Extensive experiments on synthetic and real-world datasets demonstrate the effectiveness of DisDiff.
****
Information-guided Planning: An Online Approach for Partially Observable Problems
Matheus Aparecido Do Carmo Alves, Amokh Varma, Yehia Elkhatib, Leandro Soriano Marcolino
This paper presents IB-POMCP, a novel algorithm for online planning under partial observability. Our approach enhances the decision-making process by using estimations of the world belief's entropy to guide a tree search process and surpass the limitations of planning in scenarios with sparse reward configurations. By performing what we denominate as an information-guided planning process, the algorithm, which incorporates a novel I-UCB function, shows significant improvements in reward and reasoning time compared to state-of-the-art baselines in several benchmark scenarios, along with theoretical convergence guarantees.

****
Bayesian Metric Learning for Uncertainty Quantification in Image Retrieval
Frederik Warburg, Marco Miani, Silas Brack, Søren Hauberg
We propose a Bayesian encoder for metric learning. Rather than relying on neural
 amortization as done in prior works, we learn a distribution over the network w
eights with the Laplace Approximation. We first prove that the contrastive loss
is a negative log-likelihood on the spherical space. We propose three methods th
at ensure a positive definite covariance matrix. Lastly, we present a novel deco
mposition of the Generalized Gauss-Newton approximation. Empirically, we show th
at our Laplacian Metric Learner (LAM) yields well-calibrated uncertainties, reli
ably detects out-of-distribution examples, and has state-of-the-art predictive p
erformance.
****
Neural Modulation for Flash Memory: An Unsupervised Learning Framework for Impro
ved Reliability
Jonathan Zedaka, Elisha Halperin, Evgeny Blaichman, Amit Berman
Recent years have witnessed a significant increase in the storage density of NAN
D flash memory, making it a critical component in modern electronic devices. How
ever, with the rise in storage capacity comes an increased likelihood of errors
in data storage and retrieval. The growing number of errors poses ongoing challe
nges for system designers and engineers, in terms of the characterization, model
ing, and optimization of NAND-based systems. We present a novel approach for mod
eling and preventing errors by utilizing the capabilities of generative and unsu
pervised machine learning methods. As part of our research, we constructed and t
rained a neural modulator that translates information bits into programming oper
ations on each memory cell in NAND devices. Our modulator, tailored explicitly f
or flash memory channels, provides a smart writing scheme that reduces programmi
ng errors as well as compensates for data degradation over time. Specifically, t
he modulator is based on an auto-encoder architecture with an additional channel
 model embedded between the encoder and the decoder. A conditional generative ad
versarial network (cGAN) was used to construct the channel model. Optimized for
the end-of-life work-point, the learned memory system outperforms the prior art
by up to 56\% in raw bit error rate (RBER) and extends the lifetime of the flash
 memory block by up to 25\%.
****
Privacy Amplification via Compression: Achieving the Optimal Privacy-Accuracy-Co
mmunication Trade-off in Distributed Mean Estimation
Wei-Ning Chen, Dan Song, Ayfer Ozgur, Peter Kairouz
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Normalization Layers Are All That Sharpness-Aware Minimization Needs
Maximilian Mueller, Tiffany Vlaar, David Rolnick, Matthias Hein
Sharpness-aware minimization (SAM) was proposed to reduce sharpness of minima an
d has been shown to enhance generalization performance in various settings. In t
his work we show that perturbing only the affine normalization parameters (typic
ally comprising 0.1% of the total parameters) in the adversarial step of SAM can
 outperform perturbing all of the parameters. This finding generalizesto differe
nt SAM variants and both ResNet (Batch Normalization) and Vision Transformer (La
yer Normalization) architectures. We consider alternative sparse perturbation ap
proaches and find that these do not achieve similar performance enhancement at s
uch extreme sparsity levels, showing that this behaviour is unique to the normal
ization layers. Although our findings reaffirm the effectivenessof SAM in improv
ing generalization performance, they cast doubt on whether this is solely caused
 by reduced sharpness.
****
Robust Bayesian Satisficing
Artun Saday, Y. Cahit Y■ld■r■m, Cem Tekin

Distributional shifts pose a significant challenge to achieving robustness in co ntemporary machine learning. To overcome this challenge, robust satisficing (RS) seeks a robust solution to an unspecified distributional shift while achieving a utility above a desired threshold. This paper focuses on the problem of RS in contextual Bayesian optimization when there is a discrepancy between the true an d reference distributions of the context. We propose a novel robust Bayesian sat isficing algorithm called RoBOS for noisy black-box optimization. Our algorithm guarantees sublinear lenient regret under certain assumptions on the amount of d istribution shift. In addition, we define a weaker notion of regret called robus t satisficing regret, in which our algorithm achieves a sublinear upper bound in dependent of the amount of distribution shift. To demonstrate the effectiveness of our method, we apply it to various learning problems and compare it to other approaches, such as distributionally robust optimization.
****
Neural Ideal Large Eddy Simulation: Modeling Turbulence with Neural Stochastic D ifferential Equations
Anudhyan Boral, Zhong Yi Wan, Leonardo Zepeda-Núñez, James Lottes, Qing Wang, Yi -Fan Chen, John Anderson, Fei Sha
We introduce a data-driven learning framework that assimilates two powerful idea s: ideal large eddy simulation (LES) from turbulence closure modeling and neural stochastic differential equations (SDE) for stochastic modeling. The ideal LES models the LES flow by treating each full-order trajectory as a random realizati on of the underlying dynamics, as such, the effect of small-scales is marginaliz ed to obtain the deterministic evolution of the LES state. However, ideal LES is analytically intractable. In our work, we use a latent neural SDE to model the evolution of the stochastic process and an encoder-decoder pair for transforming between the latent space and the desired ideal flow field. This stands in sharp contrast to other types of neural parameterization of closure models where each trajectory is treated as a deterministic realization of the dynamics. We show t he effectiveness of our approach (niLES – neural ideal LES) on two challenging c haotic dynamical systems: Kolmogorov flow at a Reynolds number of 20,000 and flo w past a cylinder at Reynolds number 500. Compared to competing methods, our met hod can handle non-uniform geometries using unstructured meshes seamlessly. In p articular, niLES leads to trajectories with more accurate statistics and enhance s stability, particularly for long-horizon rollouts. (Source codes and datasets will be made publicly available.)
****
On the Generalization Error of Stochastic Mirror Descent for Quadratically-Bound ed Losses: an Improved Analysis
Ta Duy Nguyen, Alina Ene, Huy Nguyen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
StableFDG: Style and Attention Based Learning for Federated Domain Generalizatio n
Jungwuk Park, Dong-Jun Han, Jinho Kim, Shiqiang Wang, Christopher Brinton, Jaeky un Moon
Traditional federated learning (FL) algorithms operate under the assumption that the data distributions at training (source domains) and testing (target domain) are the same. The fact that domain shifts often occur in practice necessitates equipping FL methods with a domain generalization (DG) capability. However, exis ting DG algorithms face fundamental challenges in FL setups due to the lack of s amples/domains in each client’s local dataset. In this paper, we propose StableF DG, a style and attention based learning strategy for accomplishing federated do main generalization, introducing two key contributions. The first is style-based learning, which enables each client to explore novel styles beyond the original source domains in its local dataset, improving domain diversity based on the pr oposed style sharing, shifting, and exploration strategies. Our second contribut

ion is an attention-based feature highlighter, which captures the similarities between the features of data samples in the same class, and emphasizes the important/common characteristics to better learn the domain-invariant characteristics of each class in data-poor FL scenarios. Experimental results show that StableFDG outperforms existing baselines on various DG benchmark datasets, demonstrating its efficacy.
****
MG-ViT: A Multi-Granularity Method for Compact and Efficient Vision Transformers
Yu Zhang, Yepeng Liu, Duoqian Miao, Qi Zhang, Yiwei Shi, Liang Hu
Vision Transformer (ViT) faces obstacles in wide application due to its huge computational cost. Almost all existing studies on compressing ViT adopt the manner of splitting an image with a single granularity, with very few exploration of splitting an image with multi-granularity. As we know, important information often randomly concentrate in few regions of an image, necessitating multi-granularity attention allocation to an image. Enlightened by this, we introduce the multi-granularity strategy to compress ViT, which is simple but effective. We propose a two-stage multi-granularity framework, MG-ViT, to balance ViT’s performance and computational cost. In single-granularity inference stage, an input image is split into a small number of patches for simple inference. If necessary, multi-granularity inference stage will be instigated, where the important patches are further subsplit into multi-finer-grained patches for subsequent inference. Moreover, prior studies on compression only for classification, while we extend the multi-granularity strategy to hierarchical ViT for downstream tasks such as detection and segmentation. Extensive experiments Prove the effectiveness of the multi-granularity strategy. For instance, on ImageNet, without any loss of performance, MG-ViT reduces 47\% FLOPs of LV-ViT-S and 56\% FLOPs of DeiT-S.
****
Recurrent Temporal Revision Graph Networks
Yizhou Chen, Anxiang Zeng, Qingtao Yu, Kerui Zhang, Cao Yuanpeng, Kangle Wu, Guangda Huzhang, Han Yu, Zhiming Zhou
Temporal graphs offer more accurate modeling of many real-world scenarios than static graphs. However, neighbor aggregation, a critical building block of graph networks, for temporal graphs, is currently straightforwardly extended from that of static graphs. It can be computationally expensive when involving all historical neighbors during such aggregation. In practice, typically only a subset of the most recent neighbors are involved. However, such subsampling leads to incomplete and biased neighbor information. To address this limitation, we propose a novel framework for temporal neighbor aggregation that uses the recurrent neural network with node-wise hidden states to integrate information from all historical neighbors for each node to acquire the complete neighbor information. We demonstrate the superior theoretical expressiveness of the proposed framework as well as its state-of-the-art performance in real-world applications. Notably, it achieves a significant +9.4% improvement on averaged precision in a real-world Ecommerce dataset over existing methods on 2-layer models.
****
Recursion in Recursion: Two-Level Nested Recursion for Length Generalization with Scalability
Jishnu Ray Chowdhury, Cornelia Caragea
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
xTrimoGene: An Efficient and Scalable Representation Learner for Single-Cell RNA-Seq Data
Jing Gong, Minsheng Hao, Xingyi Cheng, Xin Zeng, Chiming Liu, Jianzhu Ma, Xuegong Zhang, Taifeng Wang, Le Song
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
ANPL: Towards Natural Programming with Interactive Decomposition
Di Huang, Ziyuan Nan, Xing Hu, Pengwei Jin, Shaohui Peng, Yuanbo Wen, Rui Zhang, Zidong Du, Qi Guo, Yewen Pu, Yunji Chen
Though LLMs are capable of generating plausible programs, it’s challenging to interact with the LLMs further to revise the program, especially if the user’s specific requirements are different from the initial proposal. In this paper, we introduce ANPL, an interactive programming system that ensures users can always refine the generated code towards their specific programmatic intents via structureddecompositions. Borrowing the paradigm of sketching from program synthesis, an ANPL program consists of a set of input-outputs that it must satisfy, a “sketch” — control/data flow expressed in precise code (e.g. Python), and “holes” — sub-modules to be implemented by the LLM specified with natural language. The user revises an ANPL program by either modifying the sketch, changing the language used to describe the holes, or providing additional input-outputs to a particular hole, turning it into a sub-ANPL program that can be solved recursively. This workflow allows the users to offload programming burdens to the LLM as much as possible while retaining the ability to pinpoint and resolve bugs locally, without exposing the rest of the program to the LLM. We deploy ANPL on the Abstraction and Reasoning Corpus (ARC), a set of unique tasks that are challenging for state-of-the-art AI systems, showing it outperforms baseline programming systems that (a) without the ability to decompose tasks interactively and (b) without the guarantee that the modules can be correctly composed together. Additional evaluations on APPS, HumanEval, and real-world programming tasks have validated that the ANPL framework is applicable to multiple programming domains. We release the ANPL solutions to the ARC tasks as a dataset, providing insights into how humans decompose novel tasks programmatically.
****
Anonymous and Copy-Robust Delegations for Liquid Democracy
Markus Utke, Ulrike Schmidt-Kraepelin
Liquid democracy with ranked delegations is a novel voting scheme that unites the practicability of representative democracy with the idealistic appeal of direct democracy: Every voter decides between casting their vote on a question at hand or delegating their voting weight to some other, trusted agent. Delegations are transitive, and since voters may end up in a delegation cycle, they are encouraged to indicate not only a single delegate, but a set of potential delegates and a ranking among them. Based on the delegation preferences of all voters, a delegation rule selects one representative per voter. Previous work has revealed a trade-off between two properties of delegation rules called anonymity and copy-robustness. To overcome this issue we study two fractional delegation rules: Mixed Borda branching, which generalizes a rule satisfying copy-robustness, and the random walk rule, which satisfies anonymity. Using the Markov chain tree theorem, we show that the two rules are in fact equivalent, and simultaneously satisfy generalized versions of the two properties. Combining the same theorem with Fulkerson's algorithm, we develop a polynomial-time algorithm for computing the outcome of the studied delegation rule. This algorithm is of independent interest, having applications in semi-supervised learning and graph theory.
****
Framework and Benchmarks for Combinatorial and Mixed-variable Bayesian Optimization
Kamil Dreczkowski, Antoine Grosnit, Haitham Bou Ammar
This paper introduces a modular framework for Mixed-variable and Combinatorial Bayesian Optimization (MCBO) to address the lack of systematic benchmarking and standardized evaluation in the field. Current MCBO papers often introduce non-diverse or non-standard benchmarks to evaluate their methods, impeding the proper assessment of different MCBO primitives and their combinations. Additionally, papers introducing a solution for a single MCBO primitive often omit benchmarking against baselines that utilize the same methods for the remaining primitives. This omission is primarily due to the significant implementation overhead involved

, resulting in a lack of controlled assessments and an inability to showcase the
 merits of a contribution effectively.To overcome these challenges, our proposed
 framework enables an effortless combination of Bayesian Optimization components
, and provides a diverse set of synthetic and real-world benchmarking tasks. Lev
eraging this flexibility, we implement 47 novel MCBO algorithms and benchmark th
em against seven existing MCBO solvers and five standard black-box optimization
algorithms on ten tasks, conducting over 4000 experiments. Our findings reveal a
 superior combination of MCBO primitives outperforming existing approaches and i
llustrate the significance of model fit and the use of a trust region. We make o
ur MCBO library available under the MIT license at \url{https://github.com/huawe
i-noah/HEBO/tree/master/MCBO}.
****
KD-Zero: Evolving Knowledge Distiller for Any Teacher-Student Pairs
Lujun Li, Peijie Dong, Anggeng Li, Zimian Wei, Ya Yang
Knowledge distillation (KD) has emerged as an effective technique for compressin
g models that can enhance the lightweight model. Conventional KD methods propos
e various designs to allow student model to imitate the teacher better. However
, these handcrafted KD designs heavily rely on expert knowledge and may be sub-o
ptimal for various teacher-student pairs. In this paper, we present a novel fra
mework, KD-Zero, which utilizes evolutionary search to automatically discover pr
omising distiller from scratch for any teacher-student architectures. Specific
ally, we first decompose the generalized distiller into knowledge transformation
s, distance functions, and loss weights. Then, we construct our distiller sear
ch space by selecting advanced operations for these three components. With shar
pness and represent gap as fitting objectives, we evolve candidate populations a
nd generate better distillers by crossover and mutation. To ensure efficient se
arching, we employ the loss-rejection protocol, search space shrinkage, and prox
y settings during the search process. In this manner, the discovered distiller
can address the capacity gap and cross-architecture challenges for any teacher-s
tudent pairs in the final distillation stage. Comprehensive experiments reveal
that KD-Zero consistently outperforms other state-of-the-art methods across dive
rse architectures on classification, detection, and segmentation tasks. Noticea
bly, we provide some practical insights in designing the distiller by analyzing
the distiller discovered. Codes are available in supplementary materials.
****
Minimum-Risk Recalibration of Classifiers
Zeyu Sun, Dogyoon Song, Alfred Hero
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
DynPoint: Dynamic Neural Point For View Synthesis
Kaichen Zhou, Jia-Xing Zhong, Sangyun Shin, Kai Lu, Yiyuan Yang, Andrew Markham,
 Niki Trigoni
The introduction of neural radiance fields has greatly improved the effectivenes
s of view synthesis for monocular videos. However, existing algorithms face diff
iculties when dealing with uncontrolled or lengthy scenarios, and require extens
ive training time specific to each new scenario.To tackle these limitations, we
propose DynPoint, an algorithm designed to facilitate the rapid synthesis of nov
el views for unconstrained monocular videos. Rather than encoding the entirety o
f the scenario information into a latent representation, DynPoint concentrates o
n predicting the explicit 3D correspondence between neighboring frames to realiz
e information aggregation.Specifically, this correspondence prediction is achiev
ed through the estimation of consistent depth and scene flow information across
frames.Subsequently, the acquired correspondence is utilized to aggregate inform
ation from multiple reference frames to a target frame, by constructing hierarch
ical neural point clouds. The resulting framework enables swift and accurate vie
w synthesis for desired views of target frames. The experimental results obtaine
d demonstrate the considerable acceleration of training time achieved - typicall

y an order of magnitude - by our proposed method while yielding comparable outcomes compared to prior approaches. Furthermore, our method exhibits strong robustness in handling long-duration videos without learning a canonical representation of video content.
****

Data-driven Optimal Filtering for Linear Systems with Unknown Noise Covariances
Shahriar Talebi, Amirhossein Taghvaei, Mehran Mesbahi
This paper examines learning the optimal filtering policy, known as the Kalman gain, for a linear system with unknown noise covariance matrices using noisy output data. The learning problem is formulated as a stochastic policy optimiza- tion problem, aiming to minimize the output prediction error. This formulation provides a direct bridge between data-driven optimal control and, its dual, op- timal filtering. Our contributions are twofold. Firstly, we conduct a thorough convergence analysis of the stochastic gradient descent algorithm, adopted for the filtering problem, accounting for biased gradients and stability constraints. Secondly, we carefully leverage a combination of tools from linear system theory and high-dimensional statistics to derive bias-variance error bounds that scale logarithmically with problem dimension, and, in contrast to subspace methods, the length of output trajectories only affects the bias term.
****

PPi: Pretraining Brain Signal Model for Patient-independent Seizure Detection
Zhizhang Yuan, Daoze Zhang, YANG YANG, Junru Chen, Yafeng Li
Automated seizure detection is of great importance to epilepsy diagnosis and treatment. An emerging method used in seizure detection, stereoelectroencephalography (SEEG), can provide detailed and stereoscopic brainwave information. However, modeling SEEG in clinical scenarios will face challenges like huge domain shift between different patients and dramatic pattern evolution among different brain areas. In this study, we propose a Pretraining-based model for Patient-independent seizure detection (PPi) to address these challenges. Firstly, we design two novel self-supervised tasks which can extract rich information from abundant SEEG data while preserving the unique characteristics between brain signals recorded from different brain areas. Then two techniques channel background subtraction and brain region enhancement are proposed to effectively tackle the domain shift problem. Extensive experiments show that PPi outperforms the SOTA baselines on two public datasets and a real-world clinical dataset collected by ourselves, which demonstrates the effectiveness and practicability of PPi. Finally, visualization analysis illustrates the rationality of the two domain generalization techniques.
****

Unsupervised Polychromatic Neural Representation for CT Metal Artifact Reduction
Qing Wu, Lixuan Chen, Ce Wang, Hongjiang Wei, S. Kevin Zhou, Jingyi Yu, Yuyao Zhang
Emerging neural reconstruction techniques based on tomography (e.g., NeRF, NeAT, and NeRP) have started showing unique capabilities in medical imaging. In this work, we present a novel Polychromatic neural representation (Polyner) to tackle the challenging problem of CT imaging when metallic implants exist within the human body. CT metal artifacts arise from the drastic variation of metal's attenuation coefficients at various energy levels of the X-ray spectrum, leading to a nonlinear metal effect in CT measurements. Recovering CT images from metal-affected measurements hence poses a complicated nonlinear inverse problem where empirical models adopted in previous metal artifact reduction (MAR) approaches lead to signal loss and strongly aliased reconstructions. Polyner instead models the MAR problem from a nonlinear inverse problem perspective. Specifically, we first derive a polychromatic forward model to accurately simulate the nonlinear CT acquisition process. Then, we incorporate our forward model into the implicit neural representation to accomplish reconstruction. Lastly, we adopt a regularizer to preserve the physical properties of the CT images across different energy levels while effectively constraining the solution space. Our Polyner is an unsupervised method and does not require any external training data. Experimenting with multiple datasets shows that our Polyner achieves comparable or better performanc

e than supervised methods on in-domain datasets while demonstrating significant performance improvements on out-of-domain datasets. To the best of our knowledge, our Polyner is the first unsupervised MAR method that outperforms its supervised counterparts. The code for this work is available at: https://github.com/iwuqing/Polyner.
****
DAMEX: Dataset-aware Mixture-of-Experts for visual understanding of mixture-of-datasets
Yash Jain, Harkirat Behl, Zsolt Kira, Vibhav Vineet
Construction of a universal detector poses a crucial question: How can we most effectively train a model on a large mixture of datasets?   The answer lies in learning dataset-specific features and ensembling their knowledge but do all this in a single model.   Previous methods achieve this by having separate detection heads on a common backbone but that results in a significant increase in parameters.   In this work, we present Mixture-of-Experts as a solution, highlighting that MoE are much more than a scalability tool.   We propose Dataset-Aware Mixture-of-Experts, DAMEX where we train the experts to become an `expert' of a dataset by learning to route each dataset tokens to its mapped expert.   Experiments on Universal Object-Detection Benchmark show that we outperform the existing state-of-the-art by average +10.2 AP score and improve over our non-MoE baseline by average +2.0 AP score. We also observe consistent gains while mixing datasets with (1) limited availability, (2) disparate domains and (3) divergent label sets.   Further, we qualitatively show that DAMEX is robust against expert representation collapse. Code is available at https://github.com/jinga-lala/DAMEX
****
FourierGNN: Rethinking Multivariate Time Series Forecasting from a Pure Graph Perspective
Kun Yi, Qi Zhang, Wei Fan, Hui He, Liang Hu, Pengyang Wang, Ning An, Longbing Cao, Zhendong Niu
Multivariate time series (MTS) forecasting has shown great importance in numerous industries. Current state-of-the-art graph neural network (GNN)-based forecasting methods usually require both graph networks (e.g., GCN) and temporal networks (e.g., LSTM) to capture inter-series (spatial) dynamics and intra-series (temporal) dependencies, respectively. However, the uncertain compatibility of the two networks puts an extra burden on handcrafted model designs. Moreover, the separate spatial and temporal modeling naturally violates the unified spatiotemporal inter-dependencies in real world, which largely hinders the forecasting performance. To overcome these problems, we explore an interesting direction of directly applying graph networks and rethink MTS forecasting from a pure graph perspective. We first define a novel data structure, hypervariate graph, which regards each series value (regardless of variates or timestamps) as a graph node, and represents sliding windows as space-time fully-connected graphs. This perspective considers spatiotemporal dynamics unitedly and reformulates classic MTS forecasting into the predictions on hypervariate graphs. Then, we propose a novel architecture Fourier Graph Neural Network (FourierGNN) by stacking our proposed Fourier Graph Operator (FGO) to perform matrix multiplications in Fourier space. FourierGNN accommodates adequate expressiveness and achieves much lower complexity, which can effectively and efficiently accomplish {the forecasting}. Besides, our theoretical analysis reveals FGO's equivalence to graph convolutions in the time domain, which further verifies the validity of FourierGNN. Extensive experiments on seven datasets have demonstrated our superior performance with higher efficiency and fewer parameters compared with state-of-the-art methods. Code is available at this repository: https://github.com/aikunyi/FourierGNN.
****
Representation Equivalent Neural Operators: a Framework for Alias-free Operator Learning
Francesca Bartolucci, Emmanuel de Bézenac, Bogdan Raonic, Roberto Molinaro, Siddhartha Mishra, Rima Alaifari
Recently, operator learning, or learning mappings between infinite-dimensional function spaces, has garnered significant attention, notably in relation to learn

ing partial differential equations from data. Conceptually clear when outlined on paper, neural operators necessitate discretization in the transition to computer implementations. This step can compromise their integrity, often causing them to deviate from the underlying operators. This research offers a fresh take on neural operators with a framework Representation equivalent Neural Operators (ReNO) designed to address these issues. At its core is the concept of operator aliasing, which measures inconsistency between neural operators and their discrete representations. We explore this for widely-used operator learning techniques. Our findings detail how aliasing introduces errors when handling different discretizations and grids and loss of crucial continuous structures. More generally, this framework not only sheds light on existing challenges but, given its constructive and broad nature, also potentially offers tools for developing new neural operators.
****

Unsupervised Anomaly Detection with Rejection

Lorenzo Perini, Jesse Davis

Anomaly detection aims at detecting unexpected behaviours in the data. Because anomaly detection is usually an unsupervised task, traditional anomaly detectors learn a decision boundary by employing heuristics based on intuitions, which are hard to verify in practice. This introduces some uncertainty, especially close to the decision boundary, that may reduce the user trust in the detector's predictions. A way to combat this is by allowing the detector to reject predictions with high uncertainty (Learning to Reject). This requires employing a confidence metric that captures the distance to the decision boundary and setting a rejection threshold to reject low-confidence predictions. However, selecting a proper metric and setting the rejection threshold without labels are challenging tasks. In this paper, we solve these challenges by setting a constant rejection threshold on the stability metric computed by ExCeeD. Our insight relies on a theoretical analysis of such a metric. Moreover, setting a constant threshold results in strong guarantees: we estimate the test rejection rate, and derive a theoretical upper bound for both the rejection rate and the expected prediction cost. Experimentally, we show that our method outperforms some metric-based methods.
****

4D Panoptic Scene Graph Generation

Jingkang Yang, Jun CEN, WENXUAN PENG, Shuai Liu, Fangzhou Hong, Xiangtai Li, Kaiyang Zhou, Qifeng Chen, Ziwei Liu

We are living in a three-dimensional space while moving forward through a fourth dimension: time. To allow artificial intelligence to develop a comprehensive understanding of such a 4D environment, we introduce 4D Panoptic Scene Graph (PSG-4D), a new representation that bridges the raw visual data perceived in a dynamic 4D world and high-level visual understanding. Specifically, PSG-4D abstracts rich 4D sensory data into nodes, which represent entities with precise location and status information, and edges, which capture the temporal relations. To facilitate research in this new area, we build a richly annotated PSG-4D dataset consisting of 3K RGB-D videos with a total of 1M frames, each of which is labeled with 4D panoptic segmentation masks as well as fine-grained, dynamic scene graphs. To solve PSG-4D, we propose PSG4DFormer, a Transformer-based model that can predict panoptic segmentation masks, track masks along the time axis, and generate the corresponding scene graphs via a relation component. Extensive experiments on the new dataset show that our method can serve as a strong baseline for future research on PSG-4D. In the end, we provide a real-world application example to demonstrate how we can achieve dynamic scene understanding by integrating a large language model into our PSG-4D system.
****

ChatGPT-Powered Hierarchical Comparisons for Image Classification

Zhiyuan Ren, Yiyang Su, Xiaoming Liu

The zero-shot open-vocabulary setting poses challenges for image classification. Fortunately, utilizing a vision-language model like CLIP, pre-trained on image-textpairs, allows for classifying images by comparing embeddings. Leveraging largelanguage models (LLMs) such as ChatGPT can further enhance CLIP’s accuracyby in

corporating class-specific knowledge in descriptions. However, CLIP stillexhibits a bias towards certain classes and generates similar descriptions for similarclasses, disregarding their differences. To address this problem, we present anovel image classification framework via hierarchical comparisons. By recursivelycomparing and grouping classes with LLMs, we construct a class hierarchy. Withsuch a hierarchy, we can classify an image by descending from the top to the bottomof the hierarchy, comparing image and text embeddings at each level. Throughextensive experiments and analyses, we demonstrate that our proposed approach isintuitive, effective, and explainable. Code will be released upon publication.
****

Module-wise Adaptive Distillation for Multimodality Foundation Models
Chen Liang, Jiahui Yu, Ming-Hsuan Yang, Matthew Brown, Yin Cui, Tuo Zhao, Boqing Gong, Tianyi Zhou
Pre-trained multimodal foundation models have demonstrated remarkable generalizability but pose challenges for deployment due to their large sizes. One effective approach to reducing their sizes is layerwise distillation, wherein small student models are trained to match the hidden representations of large teacher models at each layer. Motivated by our observation that certain architecture components, referred to as modules, contribute more significantly to the student's performance than others, we propose to track the contributions of individual modules by recording the loss decrement after distillation each module and choose the module with a greater contribution to distill more frequently. Such an approach can be naturally formulated as a multi-armed bandit (MAB) problem, where modules and loss decrements are considered as arms and rewards, respectively. We then develop a modified-Thompson sampling algorithm named OPTIMA to address the nonstationarity of module contributions resulting from model updating. Specifically, we leverage the observed contributions in recent history to estimate the changing contribution of each module and select modules based on these estimations to maximize the cumulative contribution. We evaluate the effectiveness of OPTIMA through distillation experiments on various multimodal understanding and image captioning tasks, using the CoCa-Large model \citep{yu2022coca} as the teacher model.
****

Evaluating Cognitive Maps and Planning in Large Language Models with CogEval
Ida Momennejad, Hosein Hasanbeig, Felipe Vieira Frujeri, Hiteshi Sharma, Nebojsa Jojic, Hamid Palangi, Robert Ness, Jonathan Larson
Recently an influx of studies claims emergent cognitive abilities in large language models (LLMs). Yet, most rely on anecdotes, overlook contamination of training sets, or lack systematic Evaluation involving multiple tasks, control conditions, multiple iterations, and statistical robustness tests. Here we make two major contributions. First, we propose CogEval, a cognitive science-inspired protocol for the systematic evaluation of cognitive capacities in LLMs. The CogEval protocol can be followed for the evaluation of various abilities. Second, here we follow CogEval to systematically evaluate cognitive maps and planning ability across eight LLMs (OpenAI GPT-4, GPT-3.5-turbo-175B, davinci-003-175B, Google Bard, Cohere-xlarge-52.4B, Anthropic Claude-1-52B, LLaMA-13B, and Alpaca-7B). We base our task prompts on human experiments, which offer both established construct validity for evaluating planning, and are absent from LLM training sets. We find that, while LLMs show apparent competence in a few planning tasks with simpler structures, systematic evaluation reveals striking failure modes in planning tasks, including hallucinations of invalid trajectories and falling in loops. These findings do not support the idea of emergent out-of-the-box planning ability in LLMs. This could be because LLMs do not understand the latent relational structures underlying planning problems, known as cognitive maps, and fail at unrolling goal-directed trajectories based on the underlying structure. Implications for application and future directions are discussed.
****

Unsupervised Image Denoising with Score Function
Yutong Xie, Mingze Yuan, Bin Dong, Quanzheng Li
Though achieving excellent performance in some cases, current unsupervised learning methods for single image denoising usually have constraints in applications.

In this paper, we propose a new approach which is more general and applicable to complicated noise models. Utilizing the property of score function, the gradient of logarithmic probability, we define a solving system for denoising. Once the score function of noisy images has been estimated, the denoised result can be obtained through the solving system. Our approach can be applied to multiple noise models, such as the mixture of multiplicative and additive noise combined with structured correlation. Experimental results show that our method is comparable when the noise model is simple, and has good performance in complicated cases where other methods are not applicable or perform poorly.
****
Iterative Reachability Estimation for Safe Reinforcement Learning
Milan Ganai, Zheng Gong, Chenning Yu, Sylvia Herbert, Sicun Gao
Ensuring safety is important for the practical deployment of reinforcement learning (RL). Various challenges must be addressed, such as handling stochasticity in the environments, providing rigorous guarantees of persistent state-wise safety satisfaction, and avoiding overly conservative behaviors that sacrifice performance. We propose a new framework, Reachability Estimation for Safe Policy Optimization (RESPO), for safety-constrained RL in general stochastic settings. In the feasible set where there exist violation-free policies, we optimize for rewards while maintaining persistent safety. Outside this feasible set, our optimization produces the safest behavior by guaranteeing entrance into the feasible set whenever possible with the least cumulative discounted violations. We introduce a class of algorithms using our novel reachability estimation function to optimize in our proposed framework and in similar frameworks such as those concurrently handling multiple hard and soft constraints. We theoretically establish that our algorithms almost surely converge to locally optimal policies of our safe optimization framework. We evaluate the proposed methods on a diverse suite of safe RL environments from Safety Gym, PyBullet, and MuJoCo, and show the benefits in improving both reward performance and safety compared with state-of-the-art baselines.
****
DoReMi: Optimizing Data Mixtures Speeds Up Language Model Pretraining
Sang Michael Xie, Hieu Pham, Xuanyi Dong, Nan Du, Hanxiao Liu, Yifeng Lu, Percy S. Liang, Quoc V Le, Tengyu Ma, Adams Wei Yu
The mixture proportions of pretraining data domains (e.g., Wikipedia, books, web text) greatly affect language model (LM) performance. In this paper, we propose Domain Reweighting with Minimax Optimization (DoReMi), which first trains a small proxy model using group distributionally robust optimization (Group DRO) over domains to produce domain weights (mixture proportions) without knowledge of downstream tasks. We then resample a dataset with these domain weights and train a larger, full-sized model. In our experiments, we use DoReMi on a 280M-parameter proxy model to set the domain weights for training an 8B-parameter model (30x larger) more efficiently. On The Pile, DoReMi improves perplexity across all domains, even when it downweights a domain. DoReMi improves average few-shot downstream accuracy by 6.5% points over a baseline model trained using The Pile's default domain weights and reaches the baseline accuracy with 2.6x fewer training steps. On the GLaM dataset, DoReMi, which has no knowledge of downstream tasks, even matches the performance of using domain weights tuned on downstream tasks.
****
OpenSTL: A Comprehensive Benchmark of Spatio-Temporal Predictive Learning
Cheng Tan, Siyuan Li, Zhangyang Gao, Wenfei Guan, Zedong Wang, Zicheng Liu, Lirong Wu, Stan Z. Li
Spatio-temporal predictive learning is a learning paradigm that enables models to learn spatial and temporal patterns by predicting future frames from given past frames in an unsupervised manner. Despite remarkable progress in recent years, a lack of systematic understanding persists due to the diverse settings, complex implementation, and difficult reproducibility. Without standardization, comparisons can be unfair and insights inconclusive. To address this dilemma, we propose OpenSTL, a comprehensive benchmark for spatio-temporal predictive learning that categorizes prevalent approaches into recurrent-based and recurrent-free mode

ls. OpenSTL provides a modular and extensible framework implementing various sta
te-of-the-art methods. We conduct standard evaluations on datasets across variou
s domains, including synthetic moving object trajectory, human motion, driving s
cenes, traffic flow, and weather forecasting. Based on our observations, we prov
ide a detailed analysis of how model architecture and dataset properties affect
spatio-temporal predictive learning performance. Surprisingly, we find that recu
rrent-free models achieve a good balance between efficiency and performance than
 recurrent models. Thus, we further extend the common MetaFormers to boost recur
rent-free spatial-temporal predictive learning. We open-source the code and mode
ls at https://github.com/chengtan9907/OpenSTL.
****
Guiding The Last Layer in Federated Learning with Pre-Trained Models
Gwen Legate, Nicolas Bernier, Lucas Page-Caccia, Edouard Oyallon, Eugene Belilov
sky
Federated Learning (FL) is an emerging paradigm that allows a model to be traine
d across a number of participants without sharing data. Recent works have begun
to consider the effects of using pre-trained models as an initialization point f
or existing FL algorithms; however, these approaches ignore the vast body of eff
icient transfer learning literature from the centralized learning setting. Here
we revisit the problem of FL from a pre-trained model considered in prior work a
nd expand it to a set of computer vision transfer learning problems. We first ob
serve that simply fitting a linear classification head can be efficient in many
cases. We then show that in the FL setting, fitting a classifier using the Neare
st Class Means (NCM) can be done exactly and orders of magnitude more efficient
ly than existing proposals, while obtaining strong performance. Finally, we demo
nstrate that using a two-stage approach of obtaining the classifier and then fin
e-tuning the model can yield rapid convergence and improved generalization in th
e federated setting. We demonstrate the potential our method has to reduce commu
nication and compute costs while achieving better model performance.
****
Unbiased learning of deep generative models with structured discrete representat
ions
Henry C Bendekgey, Gabe Hope, Erik Sudderth
By composing graphical models with deep learning architectures, we learn generat
ive models with the strengths of both frameworks. The structured variational aut
oencoder (SVAE) inherits structure and interpretability from graphical models, a
nd flexible likelihoods for high-dimensional data from deep learning, but poses
substantial optimization challenges. We propose novel algorithms for learning S
VAEs, and are the first to demonstrate the SVAE's ability to handle multimodal u
ncertainty when data is missing by incorporating discrete latent variables. Our
 memory-efficient implicit differentiation scheme makes the SVAE tractable to le
arn via gradient descent, while demonstrating robustness to incomplete optimizat
ion. To more rapidly learn accurate graphical model parameters, we derive a meth
od for computing natural gradients without manual derivations, which avoids bias
es found in prior work. These optimization innovations enable the first compari
sons of the SVAE to state-of-the-art time series models, where the SVAE performs
 competitively while learning interpretable and structured discrete data represe
ntations.
****
GLEMOS: Benchmark for Instantaneous Graph Learning Model Selection
Namyong Park, Ryan Rossi, Xing Wang, Antoine Simoulin, Nesreen K. Ahmed, Christo
s Faloutsos
The choice of a graph learning (GL) model (i.e., a GL algorithm and its hyperpar
ameter settings) has a significant impact on the performance of downstream tasks
. However, selecting the right GL model becomes increasingly difficult and time
consuming as more and more GL models are developed. Accordingly, it is of great
significance and practical value to equip users of GL with the ability to perfor
m a near-instantaneous selection of an effective GL model without manual interve
ntion. Despite the recent attempts to tackle this important problem, there has b
een no comprehensive benchmark environment to evaluate the performance of GL mod

el selection methods. To bridge this gap, we present GLEMOS in this work, a comp rehensive benchmark for instantaneous GL model selection that makes the followin g contributions. (i) GLEMOS provides extensive benchmark data for fundamental GL tasks, i.e., link prediction and node classification, including the performance s of 366 models on 457 graphs on these tasks. (ii) GLEMOS designs multiple evalu ation settings, and assesses how effectively representative model selection tech niques perform in these different settings. (iii) GLEMOS is designed to be easil y extended with new models, new graphs, and new performance records. (iv) Based on the experimental results, we discuss the limitations of existing approaches a nd highlight future research directions. To promote research on this significant problem, we make the benchmark data and code publicly available at https://namy ongpark.github.io/glemos.
****
STEVE-1: A Generative Model for Text-to-Behavior in Minecraft
Shalev Lifshitz, Keiran Paster, Harris Chan, Jimmy Ba, Sheila McIlraith
Constructing AI models that respond to text instructions is challenging, especia lly for sequential decision-making tasks. This work introduces a methodology, in spired by unCLIP, for instruction-tuning generative models of behavior without r elying on a large dataset of instruction-labeled trajectories. Using this method ology, we create an instruction-tuned Video Pretraining (VPT) model called STEVE -1, which can follow short-horizon open-ended text and visual instructions in Mi necraft. STEVE-1 is trained in two steps: adapting the pretrained VPT model to f ollow commands in MineCLIP's latent space, then training a prior to predict late nt codes from text. This allows us to finetune VPT through self-supervised behav ioral cloning and hindsight relabeling, reducing the need for costly human text annotations, and all for only $60 of compute. By leveraging pretrained models li ke VPT and MineCLIP and employing best practices from text-conditioned image gen eration, STEVE-1 sets a new bar for open-ended instruction following in Minecraf t with low-level controls (mouse and keyboard) and raw pixel inputs, far outperf orming previous baselines and robustly completing 12 of 13 tasks in our early-ga me evaluation suite. We provide experimental evidence highlighting key factors f or downstream performance, including pretraining, classifier-free guidance, and data scaling. All resources, including our model weights, training scripts, and evaluation tools are made available for further research.
****
Thrust: Adaptively Propels Large Language Models with External Knowledge
Xinran Zhao, Hongming Zhang, Xiaoman Pan, Wenlin Yao, Dong Yu, Jianshu Chen
Although large-scale pre-trained language models (PTLMs) are shown to encode ric h knowledge in their model parameters, the inherent knowledge in PTLMs can be op aque or static, making external knowledge necessary. However, the existing infor mation retrieval techniques could be costly and may even introduce noisy and som etimes misleading knowledge. To address these challenges, we propose the instanc e-level adaptive propulsion of external knowledge (IAPEK), where we only conduct the retrieval when necessary. To achieve this goal, we propose to model whether a PTLM contains enough knowledge to solve an instance with a novel metric, Thru st, which leverages the representation distribution of a small amount of seen in stances. Extensive experiments demonstrate that Thrust is a good measurement of models' instance-level knowledgeability. Moreover, we can achieve higher cost-ef ficiency with the Thrust score as the retrieval indicator than the naive usage o f external knowledge on 88% of the evaluated tasks with 26% average performance improvement. Such findings shed light on the real-world practice of knowledge-en hanced LMs with a limited budget for knowledge seeking due to computation latenc y or costs.
****
OneNet: Enhancing Time Series Forecasting Models under Concept Drift by Online E nsembling
yifan zhang, Qingsong Wen, xue wang, Weiqi Chen, Liang Sun, Zhang Zhang, Liang W ang, Rong Jin, Tieniu Tan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Holistic Evaluation of Text-to-Image Models
Tony Lee, Michihiro Yasunaga, Chenlin Meng, Yifan Mai, Joon Sung Park, Agrim Gup ta, Yunzhi Zhang, Deepak Narayanan, Hannah Teufel, Marco Bellagente, Minguk Kang , Taesung Park, Jure Leskovec, Jun-Yan Zhu, Fei-Fei Li, Jiajun Wu, Stefano Ermon , Percy S. Liang
The stunning qualitative improvement of text-to-image models has led to their wi despread attention and adoption. However, we lack a comprehensive quantitative u nderstanding of their capabilities and risks. To fill this gap, we introduce a n ew benchmark, Holistic Evaluation of Text-to-Image Models (HEIM). Whereas previo us evaluations focus mostly on image-text alignment and image quality, we identi fy 12 aspects, including text-image alignment, image quality, aesthetics, origin ality, reasoning, knowledge, bias, toxicity, fairness, robustness, multilinguali ty, and efficiency. We curate 62 scenarios encompassing these aspects and evalua te 26 state-of-the-art text-to-image models on this benchmark. Our results revea l that no single model excels in all aspects, with different models demonstratin g different strengths. We release the generated images and human evaluation resu lts for full transparency at https://crfm.stanford.edu/heim/latest and the code at https://github.com/stanford-crfm/helm, which is integrated with the HELM code base
****
Shape Non-rigid Kinematics (SNK): A Zero-Shot Method for Non-Rigid Shape Matchin g via Unsupervised Functional Map Regularized Reconstruction
Souhaib Attaiki, Maks Ovsjanikov
We present Shape Non-rigid Kinematics (SNK), a novel zero-shot method for non-ri gid shape matching that eliminates the need for extensive training or ground tru th data.SNK operates on a single pair of shapes, and employs a reconstruction-ba sed strategy using an encoder-decoder architecture, which deforms the source sha pe to closely match the target shape. During the process, an unsupervised functi onal map is predicted and converted into a point-to-point map, serving as a supe rvisory mechanism for the reconstruction. To aid in training, we have designed a new decoder architecture that generates smooth, realistic deformations. SNK dem onstrates competitive results on traditional benchmarks, simplifying the shape-m atching process without compromising accuracy. Our code can be found online: htt ps://github.com/pvnieo/SNK
****
Frequency Domain-Based Dataset Distillation
Donghyeok Shin, Seungjae Shin, Il-chul Moon
This paper presents FreD, a novel parameterization method for dataset distillati on, which utilizes the frequency domain to distill a small-sized synthetic datas et from a large-sized original dataset. Unlike conventional approaches that focu s on the spatial domain, FreD employs frequency-based transforms to optimize the frequency representations of each data instance. By leveraging the concentratio n of spatial domain information on specific frequency components, FreD intellige ntly selects a subset of frequency dimensions for optimization, leading to a sig nificant reduction in the required budget for synthesizing an instance. Through the selection of frequency dimensions based on the explained variance, FreD demo nstrates both theoretical and empirical evidence of its ability to operate effic iently within a limited budget, while better preserving the information of the o riginal dataset compared to conventional parameterization methods. Furthermore, Based on the orthogonal compatibility of FreD with existing methods, we confirm that FreD consistently improves the performances of existing distillation method s over the evaluation scenarios with different benchmark datasets. We release th e code at https://github.com/sdh0818/FreD.
****
Three-Way Trade-Off in Multi-Objective Learning: Optimization, Generalization an d Conflict-Avoidance
Lisha Chen, Heshan Fernando, Yiming Ying, Tianyi Chen

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
OV-PARTS: Towards Open-Vocabulary Part Segmentation
Meng Wei, Xiaoyu Yue, Wenwei Zhang, Shu Kong, Xihui Liu, Jiangmiao Pang
Segmenting and recognizing diverse object parts is a crucial ability in applications spanning various computer vision and robotic tasks. While significant progress has been made in object-level Open-Vocabulary Semantic Segmentation (OVSS), i.e., segmenting objects with arbitrary text, the corresponding part-level research poses additional challenges. Firstly, part segmentation inherently involves intricate boundaries, while limited annotated data compounds the challenge. Secondly, part segmentation introduces an open granularity challenge due to the diverse and often ambiguous definitions of parts in the open world. Furthermore, the large-scale vision and language models, which play a key role in the open vocabulary setting, struggle to recognize parts as effectively as objects. To comprehensively investigate and tackle these challenges, we propose an Open-Vocabulary Part Segmentation (OV-PARTS) benchmark. OV-PARTS includes refined versions of two publicly available datasets: Pascal-Part-116 and ADE20K-Part-234. And it covers three specific tasks: Generalized Zero-Shot Part Segmentation, Cross-Dataset Part Segmentation, and Few-Shot Part Segmentation, providing insights into analogical reasoning, open granularity and few-shot adapting abilities of models. Moreover, we analyze and adapt two prevailing paradigms of existing object-level OVSS methods for OV-PARTS. Extensive experimental analysis is conducted to inspire future research in leveraging foundational models for OV-PARTS. The code and dataset are available at https://github.com/kellyiss/OV_PARTS.
****
Large Language Models are Visual Reasoning Coordinators
Liangyu Chen, Bo Li, Sheng Shen, Jingkang Yang, Chunyuan Li, Kurt Keutzer, Trevor Darrell, Ziwei Liu
Visual reasoning requires multimodal perception and commonsense cognition of the world. Recently, multiple vision-language models (VLMs) have been proposed with excellent commonsense reasoning ability in various domains. However, how to harness the collective power of these complementary VLMs is rarely explored. Existing methods like ensemble still struggle to aggregate these models with the desired higher-order communications. In this work, we propose Cola, a novel paradigm that coordinates multiple VLMs for visual reasoning. Our key insight is that a large language model (LLM) can efficiently coordinate multiple VLMs by facilitating natural language communication that leverages their distinct and complementary capabilities. Extensive experiments demonstrate that our instruction tuning variant, Cola-FT, achieves state-of-the-art performance on visual question answering (VQA), outside knowledge VQA, visual entailment, and visual spatial reasoning tasks. Moreover, we show that our in-context learning variant, Cola-Zero, exhibits competitive performance in zero and few-shot settings, without finetuning. Through systematic ablation studies and visualizations, we validate that a coordinator LLM indeed comprehends the instruction prompts as well as the separate functionalities of VLMs; it then coordinates them to enable impressive visual reasoning capabilities.
****
Boosting Adversarial Transferability by Achieving Flat Local Maxima
Zhijin Ge, Hongying Liu, Wang Xiaosen, Fanhua Shang, Yuanyuan Liu
Transfer-based attack adopts the adversarial examples generated on the surrogate model to attack various models, making it applicable in the physical world and attracting increasing interest. Recently, various adversarial attacks have emerged to boost adversarial transferability from different perspectives. In this work, inspired by the observation that flat local minima are correlated with good generalization, we assume and empirically validate that adversarial examples at a flat local region tend to have good transferability by introducing a penalized gradient norm to the original loss function. Since directly optimizing the gradi

ent regularization norm is computationally expensive and intractable for generat ing adversarial examples, we propose an approximation optimization method to sim plify the gradient update of the objective function. Specifically, we randomly s ample an example and adopt a first-order procedure to approximate the curvature of the second-order Hessian matrix, which makes computing more efficient by inte rpolating two Jacobian matrices. Meanwhile, in order to obtain a more stable gra dient direction, we randomly sample multiple examples and average the gradients of these examples to reduce the variance due to random sampling during the itera tive process. Extensive experimental results on the ImageNet-compatible dataset show that the proposed method can generate adversarial examples at flat local re gions, and significantly improve the adversarial transferability on either norma lly trained models or adversarially trained models than the state-of-the-art att acks. Our codes are available at: https://github.com/Trustworthy-AI-Group/PGN.
****

From Trainable Negative Depth to Edge Heterophily in Graphs
Yuchen Yan, Yuzhong Chen, Huiyuan Chen, Minghua Xu, Mahashweta Das, Hao Yang, Ha nghang Tong
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

ADGym: Design Choices for Deep Anomaly Detection
Minqi Jiang, Chaochuan Hou, Ao Zheng, Songqiao Han, Hailiang Huang, Qingsong Wen , Xiyang Hu, Yue Zhao
Deep learning (DL) techniques have recently found success in anomaly detection (AD) across various fields such as finance, medical services, and cloud computing . However, most of the current research tends to view deep AD algorithms as a wh ole, without dissecting the contributions of individual design choices like loss functions and network architectures. This view tends to diminish the value of p reliminary steps like data preprocessing, as more attention is given to newly de signed loss functions, network architectures, and learning paradigms. In this pa per, we aim to bridge this gap by asking two key questions: (i) Which design cho ices in deep AD methods are crucial for detecting anomalies? (ii) How can we aut omatically select the optimal design choices for a given AD dataset, instead of relying on generic, pre-existing solutions? To address these questions, we intro duce ADGym, a platform specifically crafted for comprehensive evaluation and aut omatic selection of AD design elements in deep methods. Our extensive experiment s reveal that relying solely on existing leading methods is not sufficient. In c ontrast, models developed using ADGym significantly surpass current state-of-the -art techniques.
****

Low-shot Object Learning with Mutual Exclusivity Bias
Anh Thai, Ahmad Humayun, Stefan Stojanov, Zixuan Huang, Bikram Boote, James M. R ehg
This paper introduces Low-shot Object Learning with Mutual Exclusivity Bias (LSM E), the first computational framing of mutual exclusivity bias, a phenomenon com monly observed in infants during word learning. We provide a novel dataset, comp rehensive baselines, and a SOTA method to enable the ML community to tackle this challenging learning task. The goal of LSME is to analyze an RGB image of a sce ne containing multiple objects and correctly associate a previously-unknown obje ct instance with a provided category label. This association is then used to per form low-shot learning to test category generalization. We provide a data genera tion pipeline for the LSME problem and conduct a thorough analysis of the factor s that contribute to its difficulty. Additionally, we evaluate the performance o f multiple baselines, including state-of-the-art foundation models. Finally, we present a baseline approach that outperforms state-of-the-art models in terms of low-shot accuracy. Code and data are available at https://github.com/rehg-lab/L SME.
****

GPT-ST: Generative Pre-Training of Spatio-Temporal Graph Neural Networks
Zhonghang Li, Lianghao Xia, Yong Xu, Chao Huang
In recent years, there has been a rapid development of spatio-temporal predictio n techniques in response to the increasing demands of traffic management and tra vel planning. While advanced end-to-end models have achieved notable success in improving predictive performance, their integration and expansion pose significa nt challenges. This work aims to address these challenges by introducing a spati o-temporal pre-training framework that seamlessly integrates with downstream bas elines and enhances their performance. The framework is built upon two key desig ns: (i) We propose a spatio-temporal mask autoencoder as a pre-training model fo r learning spatio-temporal dependencies. The model incorporates customized param eter learners and hierarchical spatial pattern encoding networks. These modules are specifically designed to capture spatio-temporal customized representations and intra- and inter-cluster region semantic relationships, which have often bee n neglected in existing approaches. (ii) We introduce an adaptive mask strategy as part of the pre-training mechanism. This strategy guides the mask autoencoder in learning robust spatio-temporal representations and facilitates the modeling of different relationships, ranging from intra-cluster to inter-cluster, in an easy-to-hard training manner. Extensive experiments conducted on representative benchmarks demonstrate the effectiveness of our proposed method. We have made ou r model implementation publicly available at https://github.com/HKUDS/GPT-ST.
****
Direct Preference-based Policy Optimization without Reward Modeling
Gaon An, Junhyeok Lee, Xingdong Zuo, Norio Kosaka, Kyung-Min Kim, Hyun Oh Song
Preference-based reinforcement learning (PbRL) is an approach that enables RL ag ents to learn from preference, which is particularly useful when formulating a r eward function is challenging. Existing PbRL methods generally involve a two-ste p procedure: they first learn a reward model based on given preference data and then employ off-the-shelf reinforcement learning algorithms using the learned re ward model. However, obtaining an accurate reward model solely from preference i nformation, especially when the preference is from human teachers, can be diffic ult. Instead, we propose a PbRL algorithm that directly learns from preference w ithout requiring any reward modeling. To achieve this, we adopt a contrastive le arning framework to design a novel policy scoring metric that assigns a high sco re to policies that align with the given preferences. We apply our algorithm to offline RL tasks with actual human preference labels and show that our algorithm outperforms or is on par with the existing PbRL methods. Notably, on high-dimen sional control tasks, our algorithm surpasses offline RL methods that learn with ground-truth reward information. Finally, we show that our algorithm can be suc cessfully applied to fine-tune large language models.
****
On the Identifiability and Interpretability of Gaussian Process Models
Jiawen Chen, Wancen Mu, Yun Li, Didong Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Enhancing Motion Deblurring in High-Speed Scenes with Spike Streams
Shiyan Chen, Jiyuan Zhang, Yajing Zheng, Tiejun Huang, Zhaofei Yu
Traditional cameras produce desirable vision results but struggle with motion bl ur in high-speed scenes due to long exposure windows. Existing frame-based deblu rring algorithms face challenges in extracting useful motion cues from severely blurred images. Recently, an emerging bio-inspired vision sensor known as the sp ike camera has achieved an extremely high frame rate while preserving rich spati al details, owing to its novel sampling mechanism. However, typical binary spike streams are relatively low-resolution, degraded image signals devoid of color i nformation, making them unfriendly to human vision. In this paper, we propose a novel approach that integrates the two modalities from two branches, leveraging spike streams as auxiliary visual cues for guiding deblurring in high-speed moti

on scenes. We propose the first spike-based motion deblurring model with bidirec tional information complementarity. We introduce a content-aware motion magnitud e attention module that utilizes learnable mask to extract relevant information from blurry images effectively, and we incorporate a transposed cross-attention fusion module to efficiently combine features from both spike data and blurry RG B images.Furthermore, we build two extensive synthesized datasets for training a nd validation purposes, encompassing high-temporal-resolution spikes, blurry ima ges, and corresponding sharp images. The experimental results demonstrate that o ur method effectively recovers clear RGB images from highly blurry scenes and ou tperforms state-of-the-art deblurring algorithms in multiple settings.
****
Faith and Fate: Limits of Transformers on Compositionality
Nouha Dziri, Ximing Lu, Melanie Sclar, Xiang (Lorraine) Li, Liwei Jiang, Bill Yu chen Lin, Sean Welleck, Peter West, Chandra Bhagavatula, Ronan Le Bras, Jena Hwa ng, Soumya Sanyal, Xiang Ren, Allyson Ettinger, Zaid Harchaoui, Yejin Choi
Transformer large language models (LLMs) have sparked admiration for their excep tional performance on tasks that demand intricate multi-step reasoning. Yet, the se models simultaneously show failures on surprisingly trivial problems. This be gs the question: Are these errors incidental, or do they signal more substantial limitations?In an attempt to demystify transformer LLMs, we investigate the lim its of these models across three representative compositional tasks---multi-digi t multiplication, logic grid puzzles, and a classic dynamic programming problem. These tasks require breaking problems down into sub-steps and synthesizing thes e steps into a precise answer. We formulate compositional tasks as computation graphs to systematically quantify the level of complexity, and break down reason ing steps into intermediate sub-procedures. Our empirical findings suggest that transformer LLMs solve compositional tasks by reducing multi-step compositional reasoning into linearized subgraph matching, without necessarily developing syst ematic problem-solving skills. To round off our empirical study, we provide the oretical arguments on abstract multi-step reasoning problems that highlight how autoregressive generations' performance can rapidly decay with increased task co mplexity.
****
Towards a fuller understanding of neurons with Clustered Compositional Explanati ons
Biagio La Rosa, Leilani Gilpin, Roberto Capobianco
Compositional Explanations is a method for identifying logical formulas of conce pts that approximate the neurons' behavior. However, these explanations are link ed to the small spectrum of neuron activations (i.e., the highest ones) used to check the alignment, thus lacking completeness. In this paper, we propose a gene ralization, called Clustered Compositional Explanations, that combines Composit ional Explanations with clustering and a novel search heuristic to approximate a broader spectrum of the neuron behavior. We define and address the problems con nected to the application of these methods to multiple ranges of activations, an alyze the insights retrievable by using our algorithm, and propose desiderata qu alities that can be used to study the explanations returned by different algorit hms.
****
TpuGraphs: A Performance Prediction Dataset on Large Tensor Computational Graphs
Mangpo Phothilimthana, Sami Abu-El-Haija, Kaidi Cao, Bahare Fatemi, Michael Burr ows, Charith Mendis, Bryan Perozzi
Precise hardware performance models play a crucial role in code optimizations. T hey can assist compilers in making heuristic decisions or aid autotuners in iden tifying the optimal configuration for a given program. For example, the autotune r for XLA, a machine learning compiler, discovered 10–20\% speedup on state-of-t he-art models serving substantial production traffic at Google. Although there e xist a few datasets for program performance prediction, they target small sub-pr ograms such as basic blocks or kernels. This paper introduces TpuGraphs, a perfo rmance prediction dataset on full tensor programs, represented as computational graphs, running on Tensor Processing Units (TPUs). Each graph in the dataset rep

resents the main computation of a machine learning workload, e.g., a training ep och or an inference step. Each data sample contains a computational graph, a com pilation configuration, and the execution time of the graph when compiled with t he configuration. The graphs in the dataset are collected from open-source machi ne learning programs, featuring popular model architectures (e.g., ResNet, Effic ientNet, Mask R-CNN, and Transformer). TpuGraphs provides 25x more graphs than t he largest graph property prediction dataset (with comparable graph sizes), and 770x larger graphs on average compared to existing performance prediction datase ts on machine learning programs. This graph-level prediction task on large graph s introduces new challenges in learning, ranging from scalability, training effi ciency, to model quality.
****
ScaleLong: Towards More Stable Training of Diffusion Model via Scaling Network L ong Skip Connection
Zhongzhan Huang, Pan Zhou, Shuicheng Yan, Liang Lin
In diffusion models, UNet is the most popular network backbone, since its long s kip connects (LSCs) to connect distant network blocks can aggregate long-distant information and alleviate vanishing gradient. Unfortunately, UNet often suffers from unstable training in diffusion models which can be alleviated by scaling i ts LSC coefficients smaller. However, theoretical understandings of the instabil ity of UNet in diffusion models and also the performance improvement of LSC scal ing remain absent yet. To solve this issue, we theoretically show that the coeff icients of LSCs in UNet have big effects on the stableness of the forward and ba ckward propagation and robustness of UNet. Specifically, the hidden feature and gradient of UNet at any layer can oscillate and their oscillation ranges are act ually large which explains the instability of UNet training. Moreover, UNet is a lso provably sensitive to perturbed input, and predicts an output distant from t he desired output, yielding oscillatory loss and thus oscillatory gradient. Besi des, we also observe the theoretical benefits of the LSC coefficient scaling of UNet in the stableness of hidden features and gradient and also robustness. Fina lly,   inspired by our theory, we propose an effective coefficient scaling frame work ScaleLong that scales the coefficients of LSC in UNet and better improve the training stability of UNet. Experimental results on CIFAR10, CelebA, ImageN et and COCO show that our methods are superior to stabilize training, and yield about 1.5x training acceleration on different diffusion models with UNet or UViT backbones.
****
Faster approximate subgraph counts with privacy
Dung Nguyen, Mahantesh Halappanavar, Venkatesh Srinivasan, Anil Vullikanti
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Recasting Continual Learning as Sequence Modeling
Soochan Lee, Jaehyeon Son, Gunhee Kim
In this work, we aim to establish a strong connection between two significant bo dies of machine learning research: continual learning and sequence modeling.That is, we propose to formulate continual learning as a sequence modeling problem, allowing advanced sequence models to be utilized for continual learning.Under th is formulation, the continual learning process becomes the forward pass of a seq uence model.By adopting the meta-continual learning (MCL) framework, we can trai n the sequence model at the meta-level, on multiple continual learning episodes. As a specific example of our new formulation, we demonstrate the application of Transformers and their efficient variants as MCL methods.Our experiments on seve n benchmarks, covering both classification and regression, show that sequence mo dels can be an attractive solution for general MCL.
****
Multiply Robust Federated Estimation of Targeted Average Treatment Effects
Larry Han, Zhu Shen, Jose Zubizarreta

Federated or multi-site studies have distinct advantages over single-site studies, including increased generalizability, the ability to study underrepresented populations, and the opportunity to study rare exposures and outcomes. However, these studies are complicated by the need to preserve the privacy of each individual's data, heterogeneity in their covariate distributions, and different data structures between sites. We propose a novel federated approach to derive valid causal inferences for a target population using multi-site data. We adjust for covariate shift and accommodate covariate mismatch between sites by developing a multiply-robust and privacy-preserving nuisance function estimation approach. Our methodology incorporates transfer learning to estimate ensemble weights to combine information from source sites. We show that these learned weights are efficient and optimal under different scenarios. We showcase the finite sample advantages of our approach in terms of efficiency and robustness compared to existing state-of-the-art approaches. We apply our approach to study the treatment effect of percutaneous coronary intervention (PCI) on the duration of hospitalization for patients experiencing acute myocardial infarction (AMI) with data from the Centers for Medicare \& Medicaid Services (CMS).
****
Learning Motion Refinement for Unsupervised Face Animation
Jiale Tao, Shuhang Gu, Wen Li, Lixin Duan
Unsupervised face animation aims to generate a human face video based on theappearance of a source image, mimicking the motion from a driving video. Existingmethods typically adopted a prior-based motion model (e.g., the local affine motion model or the local thin-plate-spline motion model). While it is able to capturethe coarse facial motion, artifacts can often be observed around the tiny motionin local areas (e.g., lips and eyes), due to the limited ability of these methods to model the finer facial motions. In this work, we design a new unsupervisedface animation approach to learn simultaneously the coarse and finer motions. Inparticular, while exploiting the local affine motion model to learn the global coarsefacial motion, we design a novel motion refinement module to compensate forthe local affine motion model for modeling finer face motions in local areas. Themotion refinement is learned from the dense correlation between the source anddriving images. Specifically, we first construct a structure correlation volume basedon the keypoint features of the source and driving images. Then, we train a modelto generate the tiny facial motions iteratively from low to high resolution. Thelearned motion refinements are combined with the coarse motion to generate the new image. Extensive experiments on widely used benchmarks demonstrate thatour method achieves the best results among state-of-the-art baselines.
****
Masked Space-Time Hash Encoding for Efficient Dynamic Scene Reconstruction
Feng Wang, Zilong Chen, Guokang Wang, Yafei Song, Huaping Liu
In this paper, we propose the Masked Space-Time Hash encoding (MSTH), a novel method for efficiently reconstructing dynamic 3D scenes from multi-view or monocular videos. Based on the observation that dynamic scenes often contain substantial static areas that result in redundancy in storage and computations, MSTH represents a dynamic scene as a weighted combination of a 3D hash encoding and a 4D hash encoding. The weights for the two components are represented by a learnable mask which is guided by an uncertainty-based objective to reflect the spatial and temporal importance of each 3D position. With this design, our method can reduce the hash collision rate by avoiding redundant queries and modifications on static areas, making it feasible to represent a large number of space-time voxels by hash tables with small size.Besides, without the requirements to fit the large numbers of temporally redundant features independently, our method is easier to optimize and converge rapidly with only twenty minutes of training for a 300-frame dynamic scene. We evaluate our method on extensive dynamic scenes. As a result, MSTH obtains consistently better results than previous state-of-the-art methods with only 20 minutes of training time and 130 MB of memory storage.
****
Bifurcations and loss jumps in RNN training
Lukas Eisenmann, Zahra Monfared, Niclas Göring, Daniel Durstewitz

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Neural Foundations of Mental Simulation: Future Prediction of Latent Representations on Dynamic Scenes
Aran Nayebi, Rishi Rajalingham, Mehrdad Jazayeri, Guangyu Robert Yang
Humans and animals have a rich and flexible understanding of the physical world, which enables them to infer the underlying dynamical trajectories of objects and events, plausible future states, and use that to plan and anticipate the consequences of actions.However, the neural mechanisms underlying these computations are unclear.We combine a goal-driven modeling approach with dense neurophysiological data and high-throughput human behavioral readouts that contain thousands of comparisons to directly impinge on this question.Specifically, we construct and evaluate several classes of sensory-cognitive networks to predict the future state of rich, ethologically-relevant environments, ranging from self-supervised end-to-end models with pixel-wise or object-slot objectives, to models that future predict in the latent space of purely static image-pretrained or dynamic video-pretrained foundation models.We find that ``scale is \emph{not} all you need'', and that many state-of-the-art machine learning models fail to perform well on our neural and behavioral benchmarks for future prediction.In fact, only one class of models matches these data well overall.We find that neural responses are currently best predicted by models trained to predict the future state of their environment in the \emph{latent} space of pretrained foundation models optimized for \emph{dynamic} scenes in a self-supervised manner.These models also approach the neurons' ability to predict the environmental state variables that are visually hidden from view, despite not being explicitly trained to do so.Finally, we find that not all foundation model latents are equal.Notably, models that future predict in the latent space of video foundation models that are optimized to support a \emph{diverse} range of egocentric sensorimotor tasks, reasonably match \emph{both} human behavioral error patterns and neural dynamics across all environmental scenarios that we were able to test.Overall, these findings suggest that the neural mechanisms and behaviors of primate mental simulation have strong inductive biases associated with them, and are thus far most consistent with being optimized to future predict on \emph{reusable} visual representations that are useful for Embodied AI more generally.
****
Learning Visual Prior via Generative Pre-Training
Jinheng Xie, Kai Ye, Yudong Li, Yuexiang Li, Kevin Qinghong Lin, Yefeng Zheng, Linlin Shen, Mike Zheng Shou
Various stuff and things in visual data possess specific traits, which can be learned by deep neural networks and are implicitly represented as the visual prior, e.g., object location and shape, in the model. Such prior potentially impacts many vision tasks. For example, in conditional image synthesis, spatial conditions failing to adhere to the prior can result in visually inaccurate synthetic results. This work aims to explicitly learn the visual prior and enable the customization of sampling. Inspired by advances in language modeling, we propose to learn Visual prior via Generative Pre-Training, dubbed VisorGPT. By discretizing visual locations, e.g., bounding boxes, human pose, and instance masks, into sequences, VisorGPT can model visual prior through likelihood maximization. Besides, prompt engineering is investigated to unify various visual locations and enable customized sampling of sequential outputs from the learned prior. Experimental results demonstrate the effectiveness of VisorGPT in modeling visual prior and extrapolating to novel scenes, potentially motivating that discrete visual locations can be integrated into the learning paradigm of current language models to further perceive visual world. Code is available at https://sierkinhane.github.io/visor-gpt.
****
Operator Learning with Neural Fields: Tackling PDEs on General Geometries

Louis Serrano, Lise Le Boudec, Armand Kassaï Koupaï, Thomas X Wang, Yuan Yin, Je
an-Noël Vittaut, Patrick Gallinari
Machine learning approaches for solving partial differential equations require l
earning mappings between function spaces. While convolutional or graph neural ne
tworks are constrained to discretized functions, neural operators present a prom
ising milestone toward mapping functions directly. Despite impressive results th
ey still face challenges with respect to the domain geometry and typically rely
on some form of discretization. In order to alleviate such limitations, we prese
nt CORAL, a new method that leverages coordinate-based networks for solving PDEs
on general geometries. CORAL is designed to remove constraints on the input mes
h, making it applicable to any spatial sampling and geometry. Its ability extend
s to diverse problem domains, including PDE solving, spatio-temporal forecasting
, and inverse problems like geometric design. CORAL demonstrates robust performa
nce across multiple resolutions and performs well in both convex and non-convex
domains, surpassing or performing on par with state-of-the-art models.
****
Learning Dense Flow Field for Highly-accurate Cross-view Camera Localization
Zhenbo Song, ze xianghui, Jianfeng Lu, Yujiao Shi
This paper addresses the problem of estimating the 3-DoF camera pose for a groun
d-level image with respect to a satellite image that encompasses the local surro
undings. We propose a novel end-to-end approach that leverages the learning of d
ense pixel-wise flow fields in pairs of ground and satellite images to calculate
the camera pose. Our approach differs from existing methods by constructing the
feature metric at the pixel level, enabling full-image supervision for learning
distinctive geometric configurations and visual appearances across views. Speci
fically, our method employs two distinct convolution networks for ground and sat
ellite feature extraction. Then, we project the ground feature map to the bird's
eye view (BEV) using a fixed camera height assumption to achieve preliminary ge
ometric alignment. To further establish the content association between the BEV
and satellite features, we introduce a residual convolution block to refine the
projected BEV feature. Optical flow estimation is performed on the refined BEV f
eature map and the satellite feature map using flow decoder networks based on RA
FT. After obtaining dense flow correspondences, we apply the least square method
to filter matching inliers and regress the ground camera pose. Extensive experi
ments demonstrate significant improvements compared to state-of-the-art methods.
Notably, our approach reduces the median localization error by 89\%, 19\%, 80\%
, and 35\% on the KITTI, Ford multi-AV, VIGOR, and Oxford RobotCar datasets, res
pectively.
****
Spatially Resolved Gene Expression Prediction from Histology Images via Bi-modal
Contrastive Learning
Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, Gar
y Bader
Histology imaging is an important tool in medical diagnosis and research, enabli
ng the examination of tissue structure and composition at the microscopic level.
Understanding the underlying molecular mechanisms of tissue architecture is cri
tical in uncovering disease mechanisms and developing effective treatments.Gene
expression profiling provides insight into the molecular processes underlying ti
ssue architecture, but the process can be time-consuming and expensive. We prese
nt BLEEP (Bi-modaL Embedding for Expression Prediction), a bi-modal embedding fr
amework capable of generating spatially resolved gene expression profiles of who
le-slide Hematoxylin and eosin (H&E) stained histology images. BLEEP uses contra
stive learning to construct a low-dimensional joint embedding space from a refer
ence dataset using paired image and expression profiles at micrometer resolution
. With this approach, the gene expression of any query image patch can be impute
d using the expression profiles from the reference dataset. We demonstrate BLEEP
’s effectiveness in gene expression prediction by benchmarking its performance o
n a human liver tissue dataset captured using the 10x Visium platform, where it
achieves significant improvements over existing methods. Our results demonstrate
the potential of BLEEP to provide insights into the molecular mechanisms underl

ying tissue architecture, with important implications in diagnosis and research of various diseases. The proposed approach can significantly reduce the time and cost associated with gene expression profiling, opening up new avenues for high-throughput analysis of histology images for both research and clinical applications.
****

Data Augmentations for Improved (Large) Language Model Generalization

Amir Feder, Yoav Wald, Claudia Shi, Suchi Saria, David Blei

The reliance of text classifiers on spurious correlations can lead to poor generalization at deployment, raising concerns about their use in safety-critical domains such as healthcare. In this work, we propose to use counterfactual data augmentation, guided by knowledge of the causal structure of the data, to simulate interventions on spurious features and to learn more robust text classifiers. We show that this strategy is appropriate in prediction problems where the label is spuriously correlated with an attribute. Under the assumptions of such problems, we discuss the favorable sample complexity of counterfactual data augmentation, compared to importance re-weighting. Pragmatically, we match examples using auxiliary data, based on diff-in-diff methodology, and use a large language model (LLM) to represent a conditional probability of text. Through extensive experimentation on learning caregiver-invariant predictors of clinical diagnoses from medical narratives and on semi-synthetic data, we demonstrate that our method for simulating interventions improves out-of-distribution (OOD) accuracy compared to baseline invariant learning algorithms.
****

Adversarial Counterfactual Environment Model Learning

Xiong-Hui Chen, Yang Yu, Zhengmao Zhu, ZhiHua Yu, Chen Zhenjun, Chenghe Wang, Yinan Wu, Rong-Jun Qin, Hongqiu Wu, Ruijin Ding, Huang Fangsheng

An accurate environment dynamics model is crucial for various downstream tasks in sequential decision-making, such as counterfactual prediction, off-policy evaluation, and offline reinforcement learning. Currently, these models were learned through empirical risk minimization (ERM) by step-wise fitting of historical transition data. This way was previously believed unreliable over long-horizon rollouts because of the compounding errors, which can lead to uncontrollable inaccuracies in predictions. In this paper, we find that the challenge extends beyond just long-term prediction errors: we reveal that even when planning with one step, learned dynamics models can also perform poorly due to the selection bias of behavior policies during data collection. This issue will significantly mislead the policy optimization process even in identifying single-step optimal actions, further leading to a greater risk in sequential decision-making scenarios.To tackle this problem, we introduce a novel model-learning objective called adversarial weighted empirical risk minimization (AWRM). AWRM incorporates an adversarial policy that exploits the model to generate a data distribution that weakens the model's prediction accuracy, and subsequently, the model is learned under this adversarial data distribution.We implement a practical algorithm, GALILEO, for AWRM and evaluate it on two synthetic tasks, three continuous-control tasks, and \textit{a real-world application}. The experiments demonstrate that GALILEO can accurately predict counterfactual actions and improve various downstream tasks, including offline policy evaluation and improvement, as well as online decision-making.
****

Finding Safe Zones of Markov Decision Processes Policies

Lee Cohen, Yishay Mansour, Michal Moshkovitz

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Zero-One Laws of Graph Neural Networks

Sam Adam-Day, Iliant, Ismail Ceylan

Graph neural networks (GNNs) are the de facto standard deep learning architectur

es for machine learning on graphs. This has led to a large body of work analyzi
ng the capabilities and limitations of these models, particularly pertaining to
 their representation and extrapolation capacity. We offer a novel theoretical
perspective on the representation and extrapolation capacity of GNNs, by answeri
ng the question: how do GNNs behave as the number of graph nodes become very lar
ge? Under mild assumptions, we show that when we draw graphs of increasing size
from the Erd■s–Rényi model, the probability that such graphs are mapped to a par
ticular output by a class of GNN classifiers tends to either zero or one. This c
lass includes the popular graph convolutional network architecture. The result e
stablishes `zero-one laws' for these GNNs, and analogously to other convergence
laws, entails theoretical limitations on their capacity. We empirically verify
 our results, observing that the theoretical asymptotic limits are evident alrea
dy on relatively small graphs.
****
Towards Revealing the Mystery behind Chain of Thought: A Theoretical Perspective
Guhao Feng, Bohang Zhang, Yuntian Gu, Haotian Ye, Di He, Liwei Wang
Recent studies have discovered that Chain-of-Thought prompting (CoT) can dramati
cally improve the performance of Large Language Models (LLMs), particularly when
 dealing with complex tasks involving mathematics or reasoning. Despite the enor
mous empirical success, the underlying mechanisms behind CoT and how it unlocks
the potential of LLMs remain elusive. In this paper, we take a first step toward
s theoretically answering these questions. Specifically, we examine the expressi
vity of LLMs with CoT in solving fundamental mathematical and decision-making pr
oblems. By using circuit complexity theory, we first give impossibility results
showing that bounded-depth Transformers are unable to directly produce correct a
nswers for basic arithmetic/equation tasks unless the model size grows super-pol
ynomially with respect to the input length. In contrast, we then prove by constr
uction that autoregressive Transformers of constant size suffice to solve both t
asks by generating CoT derivations using a commonly used math language format. M
oreover, we show LLMs with CoT can handle a general class of decision-making pro
blems known as Dynamic Programming, thus justifying their power in tackling comp
lex real-world tasks. Finally, an extensive set of experiments show that, while
Transformers always fail to directly predict the answers, they can consistently
learn to generate correct solutions step-by-step given sufficient CoT demonstrat
ions.
****
Divide, Evaluate, and Refine: Evaluating and Improving Text-to-Image Alignment w
ith Iterative VQA Feedback
Jaskirat Singh, Liang Zheng
The field of text-conditioned image generation has made unparalleled progress wi
th the recent advent of latent diffusion models. While revolutionary, as the com
plexity of given text input increases, the current state of art diffusion models
 may still fail in generating images that accurately convey the semantics of the
 given prompt. Furthermore, such misalignments are often left undetected by pret
rained multi-modal models such as CLIP. To address these problems, in this pape
r, we explore a simple yet effective decompositional approach towards both evalu
ation and improvement of text-to-image alignment. In particular, we first intro
duce a Decompositional-Alignment-Score which given a complex caption decomposes
it into a set of disjoint assertions. The alignment of each assertion with gener
ated images is then measured using a VQA model. Finally, alignment scores for di
fferent assertions are combined aposteriori to give the final text-to-image alig
nment score. Experimental analysis reveals that the proposed alignment metric sh
ows a significantly higher correlation with human ratings as opposed to traditio
nal CLIP, BLIP scores. Furthermore, we also find that the assertion level alignm
ent scores also provide useful feedback which can then be used in a simple itera
tive procedure to gradually increase the expressivity of different assertions in
 the final image outputs. Human user studies indicate that the proposed approach
 surpasses previous state-of-the-art by 8.7% in overall text-to-image alignment
accuracy.
****

Distributed Personalized Empirical Risk Minimization
Yuyang Deng, Mohammad Mahdi Kamani, Pouria Mahdavinia, Mehrdad Mahdavi
This paper advocates a new paradigm Personalized Empirical Risk Minimization (PERM) to facilitate learning from heterogeneous data sources without imposing stringent constraints on computational resources shared by participating devices. In PERM, we aim at learning a distinct model for each client by personalizing the aggregation of local empirical losses by effectively estimating the statistical discrepancy among data distributions, which entails optimal statistical accuracy for all local distributions and overcomes the data heterogeneity issue. To learn personalized models at scale, we propose a distributed algorithm that replaces the standard model averaging with model shuffling to simultaneously optimize PERM objectives for all devices. This also allows to learn distinct model architectures (e.g., neural networks with different number of parameters) for different clients, thus confining to underlying memory and compute resources of individual clients. We rigorously analyze the convergence of proposed algorithm and conduct experiments that corroborates the effectiveness of proposed paradigm.
****
Training-free Diffusion Model Adaptation for Variable-Sized Text-to-Image Synthesis
Zhiyu Jin, Xuli Shen, Bin Li, Xiangyang Xue
Diffusion models (DMs) have recently gained attention with state-of-the-art performance in text-to-image synthesis. Abiding by the tradition in deep learning, DMs are trained and evaluated on the images with fixed sizes. However, users are demanding for various images with specific sizes and various aspect ratio. This paper focuses on adapting text-to-image diffusion models to handle such variety while maintaining visual fidelity. First we observe that, during the synthesis, lower resolution images suffer from incomplete object portrayal, while higher resolution images exhibit repetitively disordered presentation. Next, we establish a statistical relationship indicating that attention entropy changes with token quantity, suggesting that models aggregate spatial information in proportion to image resolution. The subsequent interpretation on our observations is that objects are incompletely depicted due to limited spatial information for low resolutions, while repetitively disorganized presentation arises from redundant spatial information for high resolutions. From this perspective, we propose a scaling factor to alleviate the change of attention entropy and mitigate the defective pattern observed. Extensive experimental results validate the efficacy of the proposed scaling factor, enabling models to achieve better visual effects, image quality, and text alignment. Notably, these improvements are achieved without additional training or fine-tuning techniques.
****
Enhancing Sharpness-Aware Optimization Through Variance Suppression
Bingcong Li, Georgios Giannakis
Sharpness-aware minimization (SAM) has well documented merits in enhancing generalization of deep neural networks, even without sizable data augmentation. Embracing the geometry of the loss function, where neighborhoods of 'flat minima' heighten generalization ability, SAM seeks 'flat valleys' by minimizing the maximum loss caused by an adversary perturbing parameters within the neighborhood.Although critical to account for sharpness of the loss function, such an 'over-friendly adversary' can curtail the outmost level of generalization. The novel approach of this contribution fosters stabilization of adversaries through variance suppression (VaSSO) to avoid such friendliness. VaSSO's provable stability safeguards its numerical improvement over SAM in model-agnostic tasks, including image classification and machine translation. In addition, experiments confirm that VaSSO endows SAM with robustness against high levels of label noise. Code is available at https://github.com/BingcongLi/VaSSO.
****
Efficient Algorithms for Generalized Linear Bandits with Heavy-tailed Rewards
Bo Xue, Yimu Wang, Yuanyu Wan, Jinfeng Yi, Lijun Zhang
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Multinomial Logistic Regression: Asymptotic Normality on Null Covariates in High -Dimensions
Kai Tan, Pierre C Bellec
This paper investigates the asymptotic distribution of the maximum-likelihood es timate (MLE) in multinomial logistic models in the high-dimensional regime where dimension and sample size are of the same order. While classical large-sample t heory provides asymptotic normality of the MLE under certain conditions, such cl assical results are expected to fail in high-dimensions as documented for the bi nary logistic case in the seminal work of Sur and Candès [2019]. We address this issue in classification problems with 3 or more classes, by developing asymptot ic normality and asymptotic chi-square results for the multinomial logistic MLE (also known as cross-entropy minimizer) on null covariates. Our theory leads to a new methodology to test the significance of a given feature. Extensive simulat ion studies on synthetic data corroborate these asymptotic results and confirm t he validity of proposed p-values for testing the significance of a given feature .
****
Why think step by step? Reasoning emerges from the locality of experience
Ben Prystawski, Michael Li, Noah Goodman
Humans have a powerful and mysterious capacity to reason. Working through a set of mental steps enables us to make inferences we would not be capable of making directly even though we get no additional data from the world. Similarly, when l arge language models generate intermediate steps (a chain of thought) before ans wering a question, they often produce better answers than they would directly. W e investigate why and how chain-of-thought reasoning is useful in language model s, testing the hypothesis that reasoning is effective when training data consist s of overlapping local clusters of variables that influence each other strongly. These training conditions enable the chaining of accurate local inferences to e stimate relationships between variables that were not seen together in training. We prove that there will exist a "reasoning gap", where reasoning through inter mediate variables reduces bias, for the simple case of an autoregressive density estimator trained on local samples from a chain-structured probabilistic model. We then test our hypothesis experimentally in more complex models, training an autoregressive language model on samples from Bayes nets but only including a su bset of variables in each sample. We test language models’ ability to match cond itional probabilities with and without intermediate reasoning steps, finding tha t intermediate steps are only helpful when the training data is locally structur ed with respect to dependencies between variables. The combination of locally st ructured observations and reasoning is much more data-efficient than training on all variables. Our results illustrate how the effectiveness of reasoning step b y step is rooted in the local statistical structure of the training data.
****
Analyzing Generalization of Neural Networks through Loss Path Kernels
Yilan Chen, Wei Huang, Hao Wang, Charlotte Loh, Akash Srivastava, Lam Nguyen, Li ly Weng
Deep neural networks have been increasingly used in real-world applications, mak ing it critical to ensure their ability to adapt to new, unseen data. In this pa per, we study the generalization capability of neural networks trained with (sto chastic) gradient flow. We establish a new connection between the loss dynamics of gradient flow and general kernel machines by proposing a new kernel, called l oss path kernel. This kernel measures the similarity between two data points by evaluating the agreement between loss gradients along the path determined by the gradient flow. Based on this connection, we derive a new generalization upper b ound that applies to general neural network architectures. This new bound is tig ht and strongly correlated with the true generalization error. We apply our resu lts to guide the design of neural architecture search (NAS) and demonstrate favo

rable performance compared with state-of-the-art NAS algorithms through numerica
l experiments.
****
Operation-Level Early Stopping for Robustifying Differentiable NAS
Shen Jiang, Zipeng Ji, Guanghui Zhu, Chunfeng Yuan, Yihua Huang
Differentiable NAS (DARTS) is a simple and efficient neural architecture search method that has been extensively adopted in various machine learning tasks.% Nevertheless, DARTS still encounters several robustness issues, mainly the domination of skip connections.% The resulting architectures are full of parametric-free operations, leading to performance collapse.% Existing methods suggest that the skip connection has additional advantages in optimization compared to other parametric operations and propose to alleviate the domination of skip connections by eliminating these additional advantages.% In this paper, we analyze this issue from a simple and straightforward perspective and propose that the domination of skip connections results from parametric operations overfitting the training data while architecture parameters are trained on the validation data, leading to undesired behaviors.% Based on this observation, we propose the operation-level early stopping (OLES) method to overcome this issue and robustify DARTS without introducing any computation overhead.% Extensive experimental results can verify our hypothesis and the effectiveness of OLES.
****
Training on Foveated Images Improves Robustness to Adversarial Attacks
Muhammad Shah, Aqsa Kashaf, Bhiksha Raj
Deep neural networks (DNNs) have been shown to be vulnerable to adversarial attacks-- subtle, perceptually indistinguishable perturbations of inputs that change the response of the model. In the context of vision, we hypothesize that an important contributor to the robustness of human visual perception is constant exposure to low-fidelity visual stimuli in our peripheral vision. To investigate this hypothesis, we develop RBlur, an image transform that simulates the loss in fidelity of peripheral vision by blurring the image and reducing its color saturation based on the distance from a given fixation point. We show that compared to DNNs trained on the original images, DNNs trained on images transformed by RBlur are substantially more robust to adversarial attacks, as well as other, non-adversarial, corruptions, achieving up to 25% higher accuracy on perturbed data.
****
Label Poisoning is All You Need
Rishi Jha, Jonathan Hayase, Sewoong Oh
In a backdoor attack, an adversary injects corrupted data into a model's training dataset in order to gain control over its predictions on images with a specific attacker-defined trigger. A typical corrupted training example requires altering both the image, by applying the trigger, and the label. Models trained on clean images, therefore, were considered safe from backdoor attacks. However, in some common machine learning scenarios, the training labels are provided by potentially malicious third-parties. This includes crowd-sourced annotation and knowledge distillation. We, hence, investigate a fundamental question: can we launch a successful backdoor attack by only corrupting labels? We introduce a novel approach to design label-only backdoor attacks, which we call FLIP, and demonstrate its strengths on three datasets (CIFAR-10, CIFAR-100, and Tiny-ImageNet) and four architectures (ResNet-32, ResNet-18, VGG-19, and Vision Transformer). With only 2% of CIFAR-10 labels corrupted, FLIP achieves a near-perfect attack success rate of 99.4% while suffering only a 1.8% drop in the clean test accuracy. Our approach builds upon the recent advances in trajectory matching, originally introduced for dataset distillation.
****
Learning Trajectories are Generalization Indicators
Jingwen Fu, Zhizheng Zhang, Dacheng Yin, Yan Lu, Nanning Zheng
This paper explores the connection between learning trajectories of Deep Neural Networks (DNNs) and their generalization capabilities when optimized using (stochastic) gradient descent algorithms. Instead of concentrating solely on the generalization error of the DNN post-training, we present a novel perspective for an

alyzing generalization error by investigating the contribution of each update step to the change in generalization error. This perspective enable a more direct comprehension of how the learning trajectory influences generalization error. Building upon this analysis, we propose a new generalization bound that incorporates more extensive trajectory information.Our proposed generalization bound depends on the complexity of learning trajectory and the ratio between the bias and diversity of training set. Experimental observations reveal that our method effectively captures the generalization error throughout the training process. Furthermore, our approach can also track changes in generalization error when adjustments are made to learning rates and label noise levels. These results demonstrate that learning trajectory information is a valuable indicator of a model's generalization capabilities.
****
CoDet: Co-occurrence Guided Region-Word Alignment for Open-Vocabulary Object Detection
Chuofan Ma, Yi Jiang, Xin Wen, Zehuan Yuan, Xiaojuan Qi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Rewarded soups: towards Pareto-optimal alignment by interpolating weights fine-tuned on diverse rewards
Alexandre Rame, Guillaume Couairon, Corentin Dancette, Jean-Baptiste Gaya, Mustafa Shukor, Laure Soulier, Matthieu Cord
Foundation models are first pre-trained on vast unsupervised datasets and then fine-tuned on labeled data. Reinforcement learning, notably from human feedback (RLHF), can further align the network with the intended usage. Yet the imperfections in the proxy reward may hinder the training and lead to suboptimal results; the diversity of objectives in real-world tasks and human opinions exacerbate the issue. This paper proposes embracing the heterogeneity of diverse rewards by following a multi-policy strategy. Rather than focusing on a single a priori reward, we aim for Pareto-optimal generalization across the entire space of preferences. To this end, we propose rewarded soup, first specializing multiple networks independently (one for each proxy reward) and then interpolating their weights linearly. This succeeds empirically because we show that the weights remain linearly connected when fine-tuned on diverse rewards from a shared pre-trained initialization. We demonstrate the effectiveness of our approach for text-to-text (summarization, Q&A, helpful assistant, review), text-image (image captioning, text-to-image generation, visual grounding), and control (locomotion) tasks. We hope to enhance the alignment of deep models, and how they interact with the world in all its diversity.
****
Optimal Block-wise Asymmetric Graph Construction for Graph-based Semi-supervised Learning
Zixing Song, Yifei Zhang, Irwin King
Graph-based semi-supervised learning (GSSL) serves as a powerful tool to model the underlying manifold structures of samples in high-dimensional spaces. It involves two phases: constructing an affinity graph from available data and inferring labels for unlabeled nodes on this graph. While numerous algorithms have been developed for label inference, the crucial graph construction phase has received comparatively less attention, despite its significant influence on the subsequent phase. In this paper, we present an optimal asymmetric graph structure for the label inference phase with theoretical motivations. Unlike existing graph construction methods, we differentiate the distinct roles that labeled nodes and unlabeled nodes could play. Accordingly, we design an efficient block-wise graph learning algorithm with a global convergence guarantee. Other benefits induced by our method, such as enhanced robustness to noisy node features, are explored as well. Finally, we perform extensive experiments on synthetic and real-world datasets to demonstrate its superiority to the state-of-the-art graph construction m

ethods in GSSL.
****
On the Complexity of Differentially Private Best-Arm Identification with Fixed C
onfidence
Achraf Azize, Marc Jourdan, Aymen Al Marjani, Debabrota Basu
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
COCO-Counterfactuals: Automatically Constructed Counterfactual Examples for Imag
e-Text Pairs
Tiep Le, VASUDEV LAL, Phillip Howard
Counterfactual examples have proven to be valuable in the field of natural langu
age processing (NLP) for both evaluating and improving the robustness of languag
e models to spurious correlations in datasets. Despite their demonstrated utilit
y for NLP, multimodal counterfactual examples have been relatively unexplored du
e to the difficulty of creating paired image-text data with minimal counterfactu
al changes. To address this challenge, we introduce a scalable framework for aut
omatic generation of counterfactual examples using text-to-image diffusion model
s. We use our framework to create COCO-Counterfactuals, a multimodal counterfact
ual dataset of paired image and text captions based on the MS-COCO dataset. We v
alidate the quality of COCO-Counterfactuals through human evaluations and show t
hat existing multimodal models are challenged by our counterfactual image-text p
airs. Additionally, we demonstrate the usefulness of COCO-Counterfactuals for im
proving out-of-domain generalization of multimodal vision-language models via tr
aining data augmentation. We make our code and the COCO-Counterfactuals dataset
publicly available.
****
BasisFormer: Attention-based Time Series Forecasting with Learnable and Interpre
table Basis
Zelin Ni, Hang Yu, Shizhan Liu, Jianguo Li, Weiyao Lin
Bases have become an integral part of modern deep learning-based models for time
 series forecasting due to their ability to act as feature extractors or future
references. To be effective, a basis must be tailored to the specific set of tim
e series data and exhibit distinct correlation with each time series within the
set. However, current state-of-the-art methods are limited in their ability to s
atisfy both of these requirements simultaneously. To address this challenge, we
propose BasisFormer, an end-to-end time series forecasting architecture that lev
erages learnable and interpretable bases. This architecture comprises three comp
onents: First, we acquire bases through adaptive self-supervised learning, which
 treats the historical and future sections of the time series as two distinct vi
ews and employs contrastive learning. Next, we design a Coef module that calcula
tes the similarity coefficients between the time series and bases in the histori
cal view via bidirectional cross-attention. Finally, we present a Forecast modul
e that selects and consolidates the bases in the future view based on the simila
rity coefficients, resulting in accurate future predictions. Through extensive e
xperiments on six datasets, we demonstrate that BasisFormer outperforms previous
 state-of-the-art methods by 11.04% and 15.78% respectively for univariate and m
ultivariate forecasting tasks. Code isavailable at: https://github.com/nzl511619
0/Basisformer.
****
Towards Foundation Models for Scientific Machine Learning: Characterizing Scalin
g and Transfer Behavior
Shashank Subramanian, Peter Harrington, Kurt Keutzer, Wahid Bhimji, Dmitriy Moro
zov, Michael W. Mahoney, Amir Gholami
Pre-trained machine learning (ML) models have shown great performance for awide
range of applications, in particular in natural language processing (NLP)and com
puter vision (CV). Here, we study how pre-training could be used forscientific m
achine learning (SciML) applications, specifically in the context oftransfer lea

rning. We study the transfer behavior of these models as (i) the pretrainedmodel size is scaled, (ii) the downstream training dataset size is scaled,(iii) the physics parameters are systematically pushed out of distribution, and (iv)how a single model pre-trained on a mixture of different physics problems canbe adapted to various downstream applications. We find that—when fine-tunedappropriately—transfer learning can help reach desired accuracy levels with ordersof magnitude fewer downstream examples (across different tasks that can even beout-of-distribution) than training from scratch, with consistent behaviour across awide range of downstream examples. We also find that fine-tuning these modelsyields more performance gains as model size increases, compared to training fromscratch on new downstream tasks. These results hold for a broad range of PDElearning tasks. All in all, our results demonstrate the potential of the "pre-train andfine-tune" paradigm for SciML problems, demonstrating a path towards buildingSciML foundation models. Our code is available as open-source.
****
Hyperbolic Space with Hierarchical Margin Boosts Fine-Grained Learning from Coarse Labels
Shu-Lin Xu, Yifan Sun, Faen Zhang, Anqi Xu, Xiu-Shen Wei, Yi Yang
Learning fine-grained embeddings from coarse labels is a challenging task due to limited label granularity supervision, i.e., lacking the detailed distinctions required for fine-grained tasks. The task becomes even more demanding when attempting few-shot fine-grained recognition, which holds practical significance in various applications. To address these challenges, we propose a novel method that embeds visual embeddings into a hyperbolic space and enhances their discriminative ability with a hierarchical cosine margins manner. Specifically, the hyperbolic space offers distinct advantages, including the ability to capture hierarchical relationships and increased expressive power, which favors modeling fine-grained objects. Based on the hyperbolic space, we further enforce relatively large/small similarity margins between coarse/fine classes, respectively, yielding the so-called hierarchical cosine margins manner. While enforcing similarity margins in the regular Euclidean space has become popular for deep embedding learning, applying it to the hyperbolic space is non-trivial and validating the benefit for coarse-to-fine generalization is valuable. Extensive experiments conducted on five benchmark datasets showcase the effectiveness of our proposed method, yielding state-of-the-art results surpassing competing methods.
****
PromptIR: Prompting for All-in-One Image Restoration
Vaishnav Potlapalli, Syed Waqas Zamir, Salman H. Khan, Fahad Shahbaz Khan
Image restoration involves recovering a high-quality clean image from its degraded version. Deep learning-based methods have significantly improved image restoration performance, however, they have limited generalization ability to different degradation types and levels. This restricts their real-world application since it requires training individual models for each specific degradation and knowing the input degradation type to apply the relevant model. We present a prompt-based learning approach, PromptIR, for All-In-One image restoration that can effectively restore images from various types and levels of degradation. In particular, our method uses prompts to encode degradation-specific information, which is then used to dynamically guide the restoration network. This allows our method to generalize to different degradation types and levels, while still achieving state-of-the-art results on image denoising, deraining, and dehazing. Overall, PromptIR offers a generic and efficient plugin module with few lightweight prompts that can be used to restore images of various types and levels of degradation with no prior information on the corruptions present in the image. Our code and pre-trained models are available here: https://github.com/va1shn9v/PromptIR
****
Creating a Public Repository for Joining Private Data
James Cook, Milind Shyani, Nina Mishra
How can one publish a dataset with sensitive attributes in a way that both preserves privacy and enables joins with other datasets on those same sensitive attributes? This problem arises in many contexts, e.g., a hospital and an airline may

want to jointly determine whether people who take long-haul flights are more likely to catch respiratory infections. If they join their data by a common keyed user identifier such as email address, they can determine the answer, though it breaks privacy. This paper shows how the hospital can generate a private sketch and how the airline can privately join with the hospital's sketch by email address. The proposed solution satisfies pure differential privacy and gives approximate answers to linear queries and optimization problems over those joins. Whereas prior work such as secure function evaluation requires sender/receiver interaction, a distinguishing characteristic of the proposed approach is that it is non-interactive. Consequently, the sketch can be published to a repository for any organization to join with, facilitating data discovery. The accuracy of the method is demonstrated through both theoretical analysis and extensive empirical evidence.
****
What Truly Matters in Trajectory Prediction for Autonomous Driving?
Tran Phong, Haoran Wu, Cunjun Yu, Panpan Cai, Sifa Zheng, David Hsu
Trajectory prediction plays a vital role in the performance of autonomous driving systems, and prediction accuracy, such as average displacement error (ADE) or final displacement error (FDE), is widely used as a performance metric. However, a significant disparity exists between the accuracy of predictors on fixed datasets and driving performance when the predictors are used downstream for vehicle control, because of a dynamics gap. In the real world, the prediction algorithm influences the behavior of the ego vehicle, which, in turn, influences the behaviors of other vehicles nearby. This interaction results in predictor-specific dynamics that directly impacts prediction results. In fixed datasets, since other vehicles' responses are predetermined, this interaction effect is lost, leading to a significant dynamics gap. This paper studies the overlooked significance of this dynamics gap. We also examine several other factors contributing to the disparity between prediction performance and driving performance. The findings highlight the trade-off between the predictor's computational efficiency and prediction accuracy in determining real-world driving performance. In summary, an interactive, task-driven evaluation protocol for trajectory prediction is crucial to capture its effectiveness for autonomous driving. Source code along with experimental settings is available online (https://whatmatters23.github.io/).
****
AUDIT: Audio Editing by Following Instructions with Latent Diffusion Models
Yuancheng Wang, Zeqian Ju, Xu Tan, Lei He, Zhizheng Wu, Jiang Bian, sheng zhao
Audio editing is applicable for various purposes, such as adding background sound effects, replacing a musical instrument, and repairing damaged audio. Recently, some diffusion-based methods achieved zero-shot audio editing by using a diffusion and denoising process conditioned on the text description of the output audio. However, these methods still have some problems: 1) they have not been trained on editing tasks and cannot ensure good editing effects; 2) they can erroneously modify audio segments that do not require editing; 3) they need a complete description of the output audio, which is not always available or necessary in practical scenarios. In this work, we propose AUDIT, an instruction-guided audio editing model based on latent diffusion models. Specifically, \textbf{AUDIT} has three main design features: 1) we construct triplet training data (instruction, input audio, output audio) for different audio editing tasks and train a diffusion model using instruction and input (to be edited) audio as conditions and generating output (edited) audio; 2) it can automatically learn to only modify segments that need to be edited by comparing the difference between the input and output audio; 3) it only needs edit instructions instead of full target audio descriptions as text input. AUDIT achieves state-of-the-art results in both objective and subjective metrics for several audio editing tasks (e.g., adding, dropping, replacement, inpainting, super-resolution). Demo samples are available at https://audit-demopage.github.io/.
****
An Optimization-based Approach To Node Role Discovery in Networks: Approximating Equitable Partitions

Michael Scholkemper, Michael T Schaub
Similar to community detection, partitioning the nodes of a complex network acco
rding to their structural roles aims to identify fundamental building blocks of
a network, which can be used, e.g., to find simplified descriptions of the netwo
rk connectivity, to derive reduced order models for dynamical processes unfoldin
g on processes, or as ingredients for various network analysis and graph mining
tasks. In this work, we offer a fresh look on the problem of role extraction and
 its differences to community detection and present a definition of node roles a
nd two associated optimization problems (cost functions) grounded in ideas relat
ed to graph-isomorphism tests, the Weisfeiler-Leman algorithm and equitable part
itions. We present theoretical guarantees and validate our approach via a novel
“role-infused partition benchmark”, a network model from which we can sample net
works in which nodes are endowed with different roles in a stochastic way.
****
Robust Model Reasoning and Fitting via Dual Sparsity Pursuit
Xingyu Jiang, Jiayi Ma
In this paper, we contribute to solving a threefold problem: outlier rejection,
true model reasoning and parameter estimation with a unified optimization modeli
ng. To this end, we first pose this task as a sparse subspace recovering problem
, to search a maximum of independent bases under an over-embedded data space. Th
en we convert the objective into a continuous optimization paradigm that estimat
es sparse solutions for both bases and errors. Wherein a fast and robust solver
is proposed to accurately estimate the sparse subspace parameters and error entr
ies, which is implemented by a proximal approximation method under the alternati
ng optimization framework with the ``optimal'' sub-gradient descent. Extensive e
xperiments regarding known and unknown model fitting on synthetic and challengin
g real datasets have demonstrated the superiority of our method against the stat
e-of-the-art. We also apply our method to multi-class multi-model fitting and lo
op closure detection, and achieve promising results both in accuracy and efficie
ncy. Code is released at: https://github.com/StaRainJ/DSP.
****
Evaluating the Robustness of Interpretability Methods through Explanation Invari
ance and Equivariance
Jonathan Crabbé, Mihaela van der Schaar
Interpretability methods are valuable only if their explanations faithfully desc
ribe the explained model. In this work, we consider neural networks whose predic
tions are invariant under a specific symmetry group. This includes popular archi
tectures, ranging from convolutional to graph neural networks. Any explanation t
hat faithfully explains this type of model needs to be in agreement with this in
variance property. We formalize this intuition through the notion of explanation
 invariance and equivariance by leveraging the formalism from geometric deep lea
rning. Through this rigorous formalism, we derive (1) two metrics to measure the
 robustness of any interpretability method with respect to the model symmetry gr
oup; (2) theoretical robustness guarantees for some popular interpretability met
hods and (3) a systematic approach to increase the invariance of any interpretab
ility method with respect to a symmetry group. By empirically measuring our metr
ics for explanations of models associated with various modalities and symmetry g
roups, we derive a set of 5 guidelines to allow users and developers of interpre
tability methods to produce robust explanations.
****
Back-Modality: Leveraging Modal Transformation for Data Augmentation
Zhi Li, Yifan Liu, Yin Zhang
We introduce Back-Modality, a novel data augmentation schema predicated on modal
 transformation. Data from an initial modality undergoes transformation to an in
termediate modality, followed by a reverse transformation. This framework serves
 dual roles. On one hand, it operates as a general data augmentation strategy. O
n the other hand, it allows for other augmentation techniques, suitable for the
intermediate modality, to enhance the initial modality. For instance, data augme
ntation methods applicable to pure text can be employed to augment images, there
by facilitating the cross-modality of data augmentation techniques. To validate

the viability and efficacy of our framework, we proffer three instantiations of Back-Modality: back-captioning, back-imagination, and back-speech. Comprehensive evaluations across tasks such as image classification, sentiment classification, and textual entailment demonstrate that our methods consistently enhance performance under data-scarce circumstances.
****
Gradient-Based Feature Learning under Structured Data
Alireza Mousavi-Hosseini, Denny Wu, Taiji Suzuki, Murat A. Erdogdu
Recent works have demonstrated that the sample complexity of gradient-based learning of single index models, i.e. functions that depend on a 1-dimensional projection of the input data, is governed by their information exponent. However, these results are only concerned with isotropic data, while in practice the input often contains additional structure which can implicitly guide the algorithm. In this work, we investigate the effect of a spiked covariance structure and reveal several interesting phenomena. First, we show that in the anisotropic setting, the commonly used spherical gradient dynamics may fail to recover the true direction, even when the spike is perfectly aligned with the target direction. Next, we show that appropriate weight normalization that is reminiscent of batch normalization can alleviate this issue. Further, by exploiting the alignment between the (spiked) input covariance and the target, we obtain improved sample complexity compared to the isotropic case. In particular, under the spiked model with a suitably large spike, the sample complexity of gradient-based training can be made independent of the information exponent while also outperforming lower bounds for rotationally invariant kernel methods.
****
Does Invariant Graph Learning via Environment Augmentation Learn Invariance?
Yongqiang Chen, Yatao Bian, Kaiwen Zhou, Binghui Xie, Bo Han, James Cheng
Invariant graph representation learning aims to learn the invariance among data from different environments for out-of-distribution generalization on graphs. As the graph environment partitions are usually expensive to obtain, augmenting the environment information has become the de facto approach. However, the usefulness of the augmented environment information has never been verified. In this work, we find that it is fundamentally impossible to learn invariant graph representations via environment augmentation without additional assumptions. Therefore, we develop a set of minimal assumptions, including variation sufficiency and variation consistency, for feasible invariant graph learning. We then propose a new framework Graph invAriant Learning Assistant (GALA). GALA incorporates an assistant model that needs to be sensitive to graph environment changes or distribution shifts. The correctness of the proxy predictions by the assistant model hence can differentiate the variations in spurious subgraphs. We show that extracting the maximally invariant subgraph to the proxy predictions provably identifies the underlying invariant subgraph for successful OOD generalization under the established minimal assumptions. Extensive experiments on datasets including DrugOOD with various graph distribution shifts confirm the effectiveness of GALA.
****
Mitigating Test-Time Bias for Fair Image Retrieval
Fanjie Kong, Shuai Yuan, Weituo Hao, Ricardo Henao
We address the challenge of generating fair and unbiased image retrieval results given neutral textual queries (with no explicit gender or race connotations), while maintaining the utility (performance) of the underlying vision-language (VL) model. Previous methods aim to disentangle learned representations of images and text queries from gender and racial characteristics. However, we show these are inadequate at alleviating bias for the desired equal representation result, as there usually exists test-time bias in the target retrieval set. So motivated, we introduce a straightforward technique, Post-hoc Bias Mitigation (PBM), that post-processes the outputs from the pre-trained vision-language model. We evaluate our algorithm on real-world image search datasets, Occupation 1 and 2, as well as two large-scale image-text datasets, MS-COCO and Flickr30k. Our approach achieves the lowest bias, compared with various existing bias-mitigation methods, in text-based image retrieval result while maintaining satisfactory retrieval pe

rformance. The source code is publicly available at \url{https://github.com/timqqt/FairTextbasedImageRetrieval}.
****
Lower Bounds on Adaptive Sensing for Matrix Recovery
Praneeth Kacham, David Woodruff
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Transient Neural Radiance Fields for Lidar View Synthesis and 3D Reconstruction
Anagh Malik, Parsa Mirdehghan, Sotiris Nousias, Kyros Kutulakos, David Lindell
Neural radiance fields (NeRFs) have become a ubiquitous tool for modeling scene appearance and geometry from multiview imagery. Recent work has also begun to explore how to use additional supervision from lidar or depth sensor measurements in the NeRF framework. However, previous lidar-supervised NeRFs focus on rendering conventional camera imagery and use lidar-derived point cloud data as auxiliary supervision; thus, they fail to incorporate the underlying image formation model of the lidar. Here, we propose a novel method for rendering transient NeRFs that take as input the raw, time-resolved photon count histograms measured by a single-photon lidar system, and we seek to render such histograms from novel views. Different from conventional NeRFs, the approach relies on a time-resolved version of the volume rendering equation to render the lidar measurements and capture transient light transport phenomena at picosecond timescales. We evaluate our method on a first-of-its-kind dataset of simulated and captured transient multiview scans from a prototype single-photon lidar. Overall, our work brings NeRFs to a new dimension of imaging at transient timescales, newly enabling rendering of transient imagery from novel views. Additionally, we show that our approach recovers improved geometry and conventional appearance compared to point cloud-based supervision when training on few input viewpoints. Transient NeRFs may be especially useful for applications which seek to simulate raw lidar measurements for downstream tasks in autonomous driving, robotics, and remote sensing.
****
An Exploration-by-Optimization Approach to Best of Both Worlds in Linear Bandits
Shinji Ito, Kei Takemura
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
The expressive power of pooling in Graph Neural Networks
Filippo Maria Bianchi, Veronica Lachi
In Graph Neural Networks (GNNs), hierarchical pooling operators generate local summaries of the data by coarsening the graph structure and the vertex features. Considerable attention has been devoted to analyzing the expressive power of message-passing (MP) layers in GNNs, while a study on how graph pooling affects the expressiveness of a GNN is still lacking. Additionally, despite the recent advances in the design of pooling operators, there is not a principled criterion to compare them. In this work, we derive sufficient conditions for a pooling operator to fully preserve the expressive power of the MP layers before it. These conditions serve as a universal and theoretically-grounded criterion for choosing among existing pooling operators or designing new ones. Based on our theoretical findings, we analyze several existing pooling operators and identify those that fail to satisfy the expressiveness conditions. Finally, we introduce an experimental setup to verify empirically the expressive power of a GNN equipped with pooling layers, in terms of its capability to perform a graph isomorphism test.
****
Cal-DETR: Calibrated Detection Transformer
Muhammad Akhtar Munir, Salman H. Khan, Muhammad Haris Khan, Mohsen Ali, Fahad Shahbaz Khan

Albeit revealing impressive predictive performance for several computer vision t
asks, deep neural networks (DNNs) are prone to making overconfident predictions.
 This limits the adoption and wider utilization of DNNs in many safety-critical
applications. There have been recent efforts toward calibrating DNNs, however, a
lmost all of them focus on the classification task. Surprisingly, very little at
tention has been devoted to calibrating modern DNN-based object detectors, espec
ially detection transformers, which have recently demonstrated promising detecti
on performance and are influential in many decision-making systems. In this work
, we address the problem by proposing a mechanism for calibrated detection trans
formers (Cal-DETR), particularly for Deformable-DETR, UP-DETR, and DINO. We purs
ue the train-time calibration route and make the following contributions. First,
 we propose a simple yet effective approach for quantifying uncertainty in trans
former-based object detectors. Second, we develop an uncertainty-guided logit mo
dulation mechanism that leverages the uncertainty to modulate the class logits.
Third, we develop a logit mixing approach that acts as a regularizer with detect
ion-specific losses and is also complementary to the uncertainty-guided logit mo
dulation technique to further improve the calibration performance. Lastly, we co
nduct extensive experiments across three in-domain and four out-domain scenarios
. Results corroborate the effectiveness of Cal-DETR against the competing train-
time methods in calibrating both in-domain and out-domain detections while maint
aining or even improving the detection performance. Our codebase and pre-trained
 models can be accessed at \url{https://github.com/akhtarvision/cal-detr}.
****
Trajectory Alignment: Understanding the Edge of Stability Phenomenon via Bifurca
tion Theory
Minhak Song, Chulhee Yun
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
OBELICS: An Open Web-Scale Filtered Dataset of Interleaved Image-Text Documents
Hugo Laurençon, Lucile Saulnier, Leo Tronchon, Stas Bekman, Amanpreet Singh, Ant
on Lozhkov, Thomas Wang, Siddharth Karamcheti, Alexander Rush, Douwe Kiela, Matt
hieu Cord, Victor Sanh
Large multimodal models trained on natural documents, which interleave images an
d text, outperform models trained on image-text pairs on various multimodal benc
hmarks. However, the datasets used to train these models have not been released,
 and the collection process has not been fully specified. We introduce the OBEL
ICS dataset, an open web-scale filtered dataset of interleaved image-text docume
nts comprising 141 million web pages extracted from Common Crawl, 353 million as
sociated images, and 115 billion text tokens. We describe the dataset creation p
rocess, present comprehensive filtering rules, and provide an analysis of the da
taset's content. To show the viability of OBELICS, we train on the dataset visio
n and language models of 9 and 80 billion parameters, IDEFICS-9B and IDEFICS, an
d obtain competitive performance on different multimodal benchmarks. We release
our dataset, models and code.
****
ID and OOD Performance Are Sometimes Inversely Correlated on Real-world Datasets
Damien Teney, Yong Lin, Seong Joon Oh, Ehsan Abbasnejad
Several studies have compared the in-distribution (ID) and out-of-distribution (
OOD) performance of models in computer vision and NLP. They report a frequent po
sitive correlation and some surprisingly never even observe an inverse correlati
on indicative of a necessary trade-off. The possibility of inverse patterns is i
mportant to determine whether ID performance can serve as a proxy for OOD genera
lization capabilities.This paper shows that inverse correlations between ID and
OOD performance do happen with multiple real-world datasets, not only in artific
ial worst-case settings. We explain theoretically how these cases arise and how
past studies missed them because of improper methodologies that examined a biase
d selection of models.Our observations lead to recommendations that contradict t

hose found in much of the current literature.- High OOD performance sometimes re quires trading off ID performance.- Focusing on ID performance alone may not lea d to optimal OOD performance. It may produce diminishing (eventually negative) r eturns in OOD performance.- In these cases, studies on OOD generalization that u se ID performance for model selection (a common recommended practice) will neces sarily miss the best-performing models, making these studies blind to a whole ra nge of phenomena.
****

On Generalization Bounds for Projective Clustering

Maria Sofia Bucarelli, Matilde Larsen, Chris Schwiegelshohn, Mads Toftrup

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Emergence of Shape Bias in Convolutional Neural Networks through Activation Spar sity

Tianqin Li, Ziqi Wen, Yangfan Li, Tai Sing Lee

Current deep-learning models for object recognition are known to be heavily bias ed toward texture. In contrast, human visual systems are known to be biased towa rd shape and structure. What could be the design principles in human visual syst ems that led to this difference? How could we introduce more shape bias into the deep learning models? In this paper, we report that sparse coding, a ubiquitous principle in the brain, can in itself introduce shape bias into the network. W e found that enforcing the sparse coding constraint using a non-differential Top -K operation can lead to the emergence of structural encoding in neurons in con volutional neural networks, resulting in a smooth decomposition of objects into parts and subparts and endowing the networks with shape bias. We demonstrated this emergence of shape bias and its functional benefits for different network s tructures with various datasets. For object recognition convolutional neural net works, the shape bias leads to greater robustness against style and pattern chan ge distraction. For the image synthesis generative adversary networks, the emer ged shape bias leads to more coherent and decomposable structures in the synthes ized images. Ablation studies suggest that sparse codes tend to encode structure s, whereas the more distributed codes tend to favor texture. Our code is host at the github repository: https://topk-shape-bias.github.io/
****

Resilient Constrained Learning

Ignacio Hounie, Alejandro Ribeiro, Luiz F. O. Chamon

When deploying machine learning solutions, they must satisfy multiple requiremen ts beyond accuracy, such as fairness, robustness, or safety. These requirements are imposed during training either implicitly, using penalties, or explicitly, u sing constrained optimization methods based on Lagrangian duality. Either way, s pecifying requirements is hindered by the presence of compromises and limited pr ior knowledge about the data. Furthermore, their impact on performance can often only be evaluated by actually solving the learning problem. This paper presents a constrained learning approach that adapts the requirements while simultaneous ly solving the learning task. To do so, it relaxes the learning constraints in a way that contemplates how much they affect the task at hand by balancing the pe rformance gains obtained from the relaxation against a user-defined cost of that relaxation. We call this approach resilient constrained learning after the term used to describe ecological systems that adapt to disruptions by modifying thei r operation. We show conditions under which this balance can be achieved and int roduce a practical algorithm to compute it, for which we derive approximation an d generalization guarantees. We showcase the advantages of this resilient learni ng method in image classification tasks involving multiple potential invariances and in federated learning under distribution shift.
****

Recovering Simultaneously Structured Data via Non-Convex Iteratively Reweighted Least Squares

Christian Kümmerle, Johannes Maly
We propose a new algorithm for the problem of recovering data that adheres to mu
ltiple, heterogenous low-dimensional structures from linear observations. Focuss
ing on data matrices that are simultaneously row-sparse and low-rank, we propose
 and analyze an iteratively reweighted least squares (IRLS) algorithm that is ab
le to leverage both structures. In particular, it optimizes a combination of non
-convex surrogates for row-sparsity and rank, a balancing of which is built into
 the algorithm. We prove locally quadratic convergence of the iterates to a simu
ltaneously structured data matrix in a regime of minimal sample complexity (up t
o constants and a logarithmic factor), which is known to be impossible for a com
bination of convex surrogates. In experiments, we show that the IRLS method exhi
bits favorable empirical convergence, identifying simultaneously row-sparse and
low-rank matrices from fewer measurements than state-of-the-art methods.
****
Error Bounds for Learning with Vector-Valued Random Features
Samuel Lanthaler, Nicholas H. Nelsen
This paper provides a comprehensive error analysis of learning with vector-value
d random features (RF). The theory is developed for RF ridge regression in a ful
ly general infinite-dimensional input-output setting, but nonetheless applies to
 and improves existing finite-dimensional analyses. In contrast to comparable wo
rk in the literature, the approach proposed here relies on a direct analysis of
the underlying risk functional and completely avoids the explicit RF ridge regre
ssion solution formula in terms of random matrices. This removes the need for co
ncentration results in random matrix theory or their generalizations to random o
perators. The main results established in this paper include strong consistency
of vector-valued RF estimators under model misspecification and minimax optimal
convergence rates in the well-specified setting. The parameter complexity (numbe
r of random features) and sample complexity (number of labeled data) required to
 achieve such rates are comparable with Monte Carlo intuition and free from loga
rithmic factors.
****
CoDA: Collaborative Novel Box Discovery and Cross-modal Alignment for Open-vocab
ulary 3D Object Detection
Yang Cao, Zeng Yihan, Hang Xu, Dan Xu
Open-vocabulary 3D Object Detection (OV-3DDet) aims to detect objects from an ar
bitrary list of categories within a 3D scene, which remains seldom explored in t
he literature. There are primarily two fundamental problems in OV-3DDet, i.e., l
ocalizing and classifying novel objects. This paper aims at addressing the two p
roblems simultaneously via a unified framework, under the condition of limited b
ase categories. To localize novel 3D objects, we propose an effective 3D Novel O
bject Discovery strategy, which utilizes both the 3D box geometry priors and 2D
semantic open-vocabulary priors to generate pseudo box labels of the novel objec
ts. To classify novel object boxes, we further develop a cross-modal alignment m
odule based on discovered novel boxes, to align feature spaces between 3D pointc
loud and image/text modalities. Specifically, the alignment process contains a c
lass-agnostic and a class-discriminative alignment, incorporating not only the b
ase objects with annotations but also the increasingly discovered novel objects,
 resulting in an iteratively enhanced alignment. The novel box discovery and cro
ssmodal alignment are jointly learned to collaboratively benefit each other. The
novel object discovery can directly impact the cross-modal alignment, while a be
tter feature alignment can, in turn, boost the localization capability, leading
to a unified OV-3DDet framework, named CoDA, for simultaneous novel object local
ization and classification. Extensive experiments on two challenging datasets (i
.e., SUN-RGBD and ScanNet) demonstrate the effectiveness of our method and also
show a significant mAP improvement upon the best-performing alternative method b
y 80%. Codes and pre-trained models are released on the project page.
****
Don’t blame Dataset Shift! Shortcut Learning due to Gradients and Cross Entropy
Aahlad Manas Puli, Lily Zhang, Yoav Wald, Rajesh Ranganath
Common explanations for shortcut learning assume that the shortcut improves pred

iction only under the training distribution. Thus, models trained in the typical way by minimizing log-loss using gradient descent, which we call default-ERM, should utilize the shortcut. However, even when the stable feature determines the label in the training distribution and the shortcut does not provide any additional information, like in perception tasks, default-ERM exhibits shortcut learning. Why are such solutions preferred when the loss can be driven to zero when using the stable feature alone? By studying a linear perception task, we show that default-ERM's preference for maximizing the margin, even without overparameterization, leads to models that depend more on the shortcut than the stable feature. This insight suggests that default-ERM's implicit inductive bias towards max-margin may be unsuitable for perception tasks. Instead, we consider inductive biases toward uniform margins. We show that uniform margins guarantee sole dependence on the perfect stable feature in the linear perception task and suggest alternative loss functions, termed margin control (MARG-CTRL), that encourage uniform-margin solutions. MARG-CTRL techniques mitigate shortcut learning on a variety of vision and language tasks, showing that changing inductive biases can remove the need for complicated shortcut-mitigating methods in perception tasks.
****
Scan and Snap: Understanding Training Dynamics and Token Composition in 1-layer Transformer
Yuandong Tian, Yiping Wang, Beidi Chen, Simon S. Du
Transformer architecture has shown impressive performance in multiple research domains and has become the backbone of many neural network models. However, there is limited understanding on how it works. In particular, with a simple predictive loss, how the representation emerges from the gradient \emph{training dynamics} remains a mystery. In this paper, for 1-layer transformer with one self-attention layer plus one decoder layer, we analyze its SGD training dynamics for the task of next token prediction in a mathematically rigorous manner. We open the black box of the dynamic process of how the self-attention layer combines input tokens, and reveal the nature of underlying inductive bias. More specifically, with the assumption (a) no positional encoding, (b) long input sequence, and (c) the decoder layer learns faster than the self-attention layer, we prove that self-attention acts as a \emph{discriminative scanning algorithm}: starting from uniform attention, it gradually attends more to distinct key tokens for a specific next token to be predicted, and pays less attention to common key tokens that occur across different next tokens. Among distinct tokens, it progressively drops attention weights, following the order of low to high co-occurrence between the key and the query token in the training set. Interestingly, this procedure does not lead to winner-takes-all, but stops due to a \emph{phase transition} that is controllable by the learning rate of the decoder layer, leaving (almost) fixed token combination. We verify this \textbf{\emph{scan and snap}} dynamics on synthetic and real-world data (WikiText-103).
****
Stein Π-Importance Sampling
Congye Wang, Ye Chen, Heishiro Kanagawa, Chris J. Oates
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
GPT4Tools: Teaching Large Language Model to Use Tools via Self-instruction
Rui Yang, Lin Song, Yanwei Li, Sijie Zhao, Yixiao Ge, Xiu Li, Ying Shan
This paper aims to efficiently enable Large Language Models (LLMs) to use multi-modal tools.The advanced proprietary LLMs, such as ChatGPT and GPT-4, have shown great potential for tool usage through sophisticated prompt engineering.Nevertheless, these models typically rely on prohibitive computational costs and publicly inaccessible data.To address these challenges, we propose the GPT4Tools based on self-instruct to enable open-source LLMs, such as LLaMA and OPT, to use tools.It generates an instruction-following dataset by prompting an advanced teacher with various multi-modal contexts.By using the Low-Rank Adaptation (LoRA) optim

ization, our approach facilitates the open-source LLMs to solve a range of visua
l problems, including visual comprehension and image generation.Moreover, we pro
vide a benchmark to evaluate the ability of LLMs to use tools, which is performe
d in both zero-shot and fine-tuning ways.Extensive experiments demonstrate the e
ffectiveness of our method on various language models, which not only significan
tly improves the accuracy of invoking seen tools, but also enables the zero-shot
 capacity for unseen tools.
****
Reinforcement Learning with Fast and Forgetful Memory
Steven Morad, Ryan Kortvelesy, Stephan Liwicki, Amanda Prorok
Nearly all real world tasks are inherently partially observable, necessitating t
he use of memory in Reinforcement Learning (RL). Most model-free approaches summ
arize the trajectory into a latent Markov state using memory models borrowed fro
m Supervised Learning (SL), even though RL tends to exhibit different training a
nd efficiency characteristics. Addressing this discrepancy, we introduce Fast an
d Forgetful Memory, an algorithm-agnostic memory model designed specifically for
 RL. Our approach constrains the model search space via strong structural priors
 inspired by computational psychology. It is a drop-in replacement for recurrent
 neural networks (RNNs) in recurrent RL algorithms, achieving greater reward tha
n RNNs across various recurrent benchmarks and algorithms without changing any h
yperparameters. Moreover, Fast and Forgetful Memory exhibits training speeds two
 orders of magnitude faster than RNNs, attributed to its logarithmic time and li
near space complexity. Our implementation is available at https://github.com/pro
roklab/ffm.
****
Systematic Visual Reasoning through Object-Centric Relational Abstraction
Taylor Webb, Shanka Subhra Mondal, Jonathan D Cohen
Human visual reasoning is characterized by an ability to identify abstract patte
rns from only a small number of examples, and to systematically generalize those
 patterns to novel inputs. This capacity depends in large part on our ability to
 represent complex visual inputs in terms of both objects and relations. Recent
work in computer vision has introduced models with the capacity to extract objec
t-centric representations, leading to the ability to process multi-object visual
 inputs, but falling short of the systematic generalization displayed by human r
easoning. Other recent models have employed inductive biases for relational abst
raction to achieve systematic generalization of learned abstract rules, but have
 generally assumed the presence of object-focused inputs. Here, we combine these
 two approaches, introducing Object-Centric Relational Abstraction (OCRA), a mod
el that extracts explicit representations of both objects and abstract relations
, and achieves strong systematic generalization in tasks (including a novel data
set, CLEVR-ART, with greater visual complexity) involving complex visual display
s.
****
In-Context Impersonation Reveals Large Language Models' Strengths and Biases
Leonard Salewski, Stephan Alaniz, Isabel Rio-Torto, Eric Schulz, Zeynep Akata
In everyday conversations, humans can take on different roles and adapt their vo
cabulary to their chosen roles. We explore whether LLMs can take on, that is imp
ersonate, different roles when they generate text in-context. We ask LLMs to ass
ume different personas before solving vision and language tasks. We do this by p
refixing the prompt with a persona that is associated either with a social ident
ity or domain expertise. In a multi-armed bandit task, we find that LLMs pretend
ing to be children of different ages recover human-like developmental stages of
exploration. In a language-based reasoning task, we find that LLMs impersonating
 domain experts perform better than LLMs impersonating non-domain experts. Final
ly, we test whether LLMs' impersonations are complementary to visual information
 when describing different categories. We find that impersonation can improve pe
rformance: an LLM prompted to be a bird expert describes birds better than one p
rompted to be a car expert. However, impersonation can also uncover LLMs' biases
: an LLM prompted to be a man describes cars better than one prompted to be a wo
man. These findings demonstrate that LLMs are capable of taking on diverse roles

and that this in-context impersonation can be used to uncover their strengths and hidden biases. Our code is available at https://github.com/ExplainableML/in-context-impersonation.
****
The s-value: evaluating stability with respect to distributional shifts
Suyash Gupta, Dominik Rothenhäusler
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
When Does Optimizing a Proper Loss Yield Calibration?
Jaroslaw Blasiok, Parikshit Gopalan, Lunjia Hu, Preetum Nakkiran
Optimizing proper loss functions is popularly believed to yield predictors with good calibration properties; the intuition being that for such losses, the global optimum is to predict the ground-truth probabilities, which is indeed calibrated. However, typical machine learning models are trained to approximately minimize loss over restricted families of predictors, that are unlikely to contain the ground truth. Under what circumstances does optimizing proper loss over a restricted family yield calibrated models? What precise calibration guarantees does it give? In this work, we provide a rigorous answer to these questions. We replace the global optimality with a local optimality condition stipulating that the (proper) loss of the predictor cannot be reduced much by post-processing its predictions with a certain family of Lipschitz functions. We show that any predictor with this local optimality satisfies smooth calibration as defined in [Kakade and Foster, 2008, B■asiok et al., 2023]. Local optimality is plausibly satisfied by well-trained DNNs, which suggests an explanation for why they are calibrated from proper loss minimization alone. Finally, we show that the connection between local optimality and calibration error goes both ways: nearly calibrated predictors are also nearly locally optimal.
****
Language Is Not All You Need: Aligning Perception with Language Models
Shaohan Huang, Li Dong, Wenhui Wang, Yaru Hao, Saksham Singhal, Shuming Ma, Tengchao Lv, Lei Cui, Owais Khan Mohammed, Barun Patra, Qiang Liu, Kriti Aggarwal, Zewen Chi, Nils Bjorck, Vishrav Chaudhary, Subhojit Som, XIA SONG, Furu Wei
A big convergence of language, multimodal perception, action, and world modeling is a key step toward artificial general intelligence. In this work, we introduce KOSMOS-1, a Multimodal Large Language Model (MLLM) that can perceive general modalities, learn in context (i.e., few-shot), and follow instructions (i.e., zero-shot). Specifically, we train KOSMOS-1 from scratch on web-scale multi-modal corpora, including arbitrarily interleaved text and images, image-caption pairs, and text data. We evaluate various settings, including zero-shot, few-shot, and multimodal chain-of-thought prompting, on a wide range of tasks without any gradient updates or finetuning. Experimental results show that KOSMOS-1 achieves impressive performance on (i) language understanding, generation, and even OCR-free NLP (directly fed with document images), (ii) perception-language tasks, including multimodal dialogue, image captioning, visual question answering, and (iii) vision tasks, such as image recognition with descriptions (specifying classification via text instructions). We also show that MLLMs can benefit from cross-modal transfer, i.e., transfer knowledge from language to multimodal, and from multimodal to language. In addition, we introduce a dataset of Raven IQ test, which diagnoses the nonverbal reasoning capability of MLLMs.
****
Out-of-distribution Detection Learning with Unreliable Out-of-distribution Sources
Haotian Zheng, Qizhou Wang, Zhen Fang, Xiaobo Xia, Feng Liu, Tongliang Liu, Bo Han
Out-of-distribution (OOD) detection discerns OOD data where the predictor cannot make valid predictions as in-distribution (ID) data, thereby increasing the reliability of open-world classification. However, it is typically hard to collect

real out-of-distribution (OOD) data for training a predictor capable of discerning ID and OOD patterns. This obstacle gives rise to data generation-based learning methods, synthesizing OOD data via data generators for predictor training without requiring any real OOD data. Related methods typically pre-train a generator on ID data and adopt various selection procedures to find those data likely to be the OOD cases. However, generated data may still coincide with ID semantics, i.e., mistaken OOD generation remains, confusing the predictor between ID and OOD data. To this end, we suggest that generated data (with mistaken OOD generation) can be used to devise an auxiliary OOD detection task to facilitate real OOD detection. Specifically, we can ensure that learning from such an auxiliary task is beneficial if the ID and the OOD parts have disjoint supports, with the help of a well-designed training procedure for the predictor. Accordingly, we propose a powerful data generation-based learning method named Auxiliary Task-based OOD Learning (ATOL) that can relieve the mistaken OOD generation. We conduct extensive experiments under various OOD detection setups, demonstrating the effectiveness of our method against its advanced counterparts.
****
Robust covariance estimation with missing values and cell-wise contamination
Grégoire Pacreau, Karim Lounici
Large datasets are often affected by cell-wise outliers in the form of missing or erroneous data. However, discarding any samples containing outliers may result in a dataset that is too small to accurately estimate the covariance matrix. Moreover, the robust procedures designed to address this problem require the invertibility of the covariance operator and thus are not effective on high-dimensional data. In this paper, we propose an unbiased estimator for the covariance in the presence of missing values that does not require any imputation step and still achieves near minimax statistical accuracy with the operator norm. We also advocate for its use in combination with cell-wise outlier detection methods to tackle cell-wise contamination in a high-dimensional and low-rank setting, where state-of-the-art methods may suffer from numerical instability and long computation times. To complement our theoretical findings, we conducted an experimental study which demonstrates the superiority of our approach over the state of the art both in low and high dimension settings.
****
Patch Diffusion: Faster and More Data-Efficient Training of Diffusion Models
Zhendong Wang, Yifan Jiang, Huangjie Zheng, Peihao Wang, Pengcheng He, Zhangyang "Atlas" Wang, Weizhu Chen, Mingyuan Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Clustering the Sketch: Dynamic Compression for Embedding Tables
Henry Tsang, Thomas Ahle
Embedding tables are used by machine learning systems to work with categorical features.In modern Recommendation Systems, these tables can be very large, necessitating the development of new methods for fitting them in memory, even during training.We suggest Clustered Compositional Embeddings (CCE) which combines clustering-based compression like quantization to codebooks with dynamic methods like The Hashing Trick and Compositional Embeddings [Shi et al., 2020].Experimentally CCE achieves the best of both worlds: The high compression rate of codebook-based quantization, but \emph{dynamically} like hashing-based methods, so it can be used during training.Theoretically, we prove that CCE is guaranteed to converge to the optimal codebook and give a tight bound for the number of iterations required.
****
Dynamic Personalized Federated Learning with Adaptive Differential Privacy
Xiyuan Yang, Wenke Huang, Mang Ye
Personalized federated learning with differential privacy has been considered a feasible solution to address non-IID distribution of data and privacy leakage ri

sks. However, current personalized federated learning methods suffer from inflexible personalization and convergence difficulties due to two main factors: 1) Firstly, we observe that the prevailing personalization methods mainly achieve this by personalizing a fixed portion of the model, which lacks flexibility. 2) Moreover, we further demonstrate that the default gradient calculation is sensitive to the widely-used clipping operations in differential privacy, resulting in difficulties in convergence. Considering that Fisher information values can serve as an effective measure for estimating the information content of parameters by reflecting the model sensitivity to parameters, we aim to leverage this property to address the aforementioned challenges. In this paper, we propose a novel federated learning method with Dynamic Fisher Personalization and Adaptive Constraint (FedDPA) to handle these challenges. Firstly, by using layer-wise Fisher information to measure the information content of local parameters, we retain local parameters with high Fisher values during the personalization process, which are considered informative, simultaneously prevent these parameters from noise perturbation. Secondly, we introduce an adaptive approach by applying differential constraint strategies to personalized parameters and shared parameters identified in the previous for better convergence. Our method boosts performance through flexible personalization while mitigating the slow convergence caused by clipping operations. Experimental results on CIFAR-10, FEMNIST and SVHN dataset demonstrate the effectiveness of our approach in achieving better performance and robustness against clipping, under personalized federated learning with differential privacy.
****
Bias in Evaluation Processes: An Optimization-Based Model
L. Elisa Celis, Amit Kumar, Anay Mehrotra, Nisheeth K. Vishnoi
Biases with respect to socially-salient attributes of individuals have been well documented in evaluation processes used in settings such as admissions and hiring. We view such an evaluation process as a transformation of a distribution of the true utility of an individual for a task to an observed distribution and model it as a solution to a loss minimization problem subject to an information constraint. Our model has two parameters that have been identified as factors leading to biases: the resource-information trade-off parameter in the information constraint and the risk-averseness parameter in the loss function. We characterize the distributions that arise from our model and study the effect of the parameters on the observed distribution. The outputs of our model enrich the class of distributions that can be used to capture variation across groups in the observed evaluations. We empirically validate our model by fitting real-world datasets and use it to study the effect of interventions in a downstream selection task. These results contribute to an understanding of the emergence of bias in evaluation processes and provide tools to guide the deployment of interventions to mitigate biases.
****
Data Minimization at Inference Time
Cuong Tran, Nando Fioretto
In high-stakes domains such as legal, banking, hiring, and healthcare, learning models frequently rely on sensitive user information for inference, necessitating the complete set of features. This not only poses significant privacy risks for individuals but also demands substantial human effort from organizations to verify information accuracy. This study asks whether it is necessary to use all input features for accurate predictions at inference time. The paper demonstrates that, in a personalized setting, individuals may only need to disclose a small subset of features without compromising decision-making accuracy. The paper also provides an efficient sequential algorithm to determine the appropriate attributes for each individual to provide. Evaluations across various learning tasks show that individuals can potentially report as little as 10\% of their information while maintaining the same accuracy level as a model that employs the full set of user information.
****
Learning Adaptive Tensorial Density Fields for Clean Cryo-ET Reconstruction

YUANHAO WANG, Ramzi Idoughi, Wolfgang Heidrich
We present a novel learning-based framework for reconstructing 3D structures from tilt-series cryo-Electron Tomography (cryo-ET) data. Cryo-ET is a powerful imaging technique that can achieve near-atomic resolutions. Still, it suffers from challenges such as missing-wedge acquisition, large data size, and high noise levels. Our framework addresses these challenges by using an adaptive tensorial-based representation for the 3D density field of the scanned sample. First, we optimize a quadtree structure to partition the volume of interest. Then, we learn a vector-matrix factorization of the tensor representing the density field in each node. Moreover, we use a loss function that combines a differentiable tomographic formation model with three regularization terms: total variation, boundary consistency constraint, and an isotropic Fourier prior. Our framework allows us to query the density at any location using the learned representation and obtain a high-quality 3D tomogram. We demonstrate the superiority of our framework over existing methods using synthetic and real data. Thus, our framework boosts the quality of the reconstruction while reducing the computation time and the memory footprint. The code is available at https://github.com/yuanhaowang1213/adaptivetensordf.
****
Resetting the Optimizer in Deep RL: An Empirical Study
Kavosh Asadi, Rasool Fakoor, Shoham Sabach
We focus on the task of approximating the optimal value function in deep reinforcement learning. This iterative process is comprised of solving a sequence of optimization problems where the loss function changes per iteration. The common approach to solving this sequence of problems is to employ modern variants of the stochastic gradient descent algorithm such as Adam. These optimizers maintain their own internal parameters such as estimates of the first-order and the second-order moments of the gradient, and update them over time. Therefore, information obtained in previous iterations is used to solve the optimization problem in the current iteration. We demonstrate that this can contaminate the moment estimates because the optimization landscape can change arbitrarily from one iteration to the next one. To hedge against this negative effect, a simple idea is to reset the internal parameters of the optimizer when starting a new iteration. We empirically investigate this resetting idea by employing various optimizers in conjunction with the Rainbow algorithm. We demonstrate that this simple modification significantly improves the performance of deep RL on the Atari benchmark.
****
Why Does Sharpness-Aware Minimization Generalize Better Than SGD?
Zixiang Chen, Junkai Zhang, Yiwen Kou, Xiangning Chen, Cho-Jui Hsieh, Quanquan Gu
The challenge of overfitting, in which the model memorizes the training data and fails to generalize to test data, has become increasingly significant in the training of large neural networks. To tackle this challenge, Sharpness-Aware Minimization (SAM) has emerged as a promising training method, which can improve the generalization of neural networks even in the presence of label noise. However, a deep understanding of how SAM works, especially in the setting of nonlinear neural networks and classification tasks, remains largely missing. This paper fills this gap by demonstrating why SAM generalizes better than Stochastic Gradient Descent (SGD) for a certain data model and two-layer convolutional ReLU networks. The loss landscape of our studied problem is nonsmooth, thus current explanations for the success of SAM based on the Hessian information are insufficient. Our result explains the benefits of SAM, particularly its ability to prevent noise learning in the early stages, thereby facilitating more effective learning of features. Experiments on both synthetic and real data corroborate our theory.
****
Grassmann Manifold Flows for Stable Shape Generation
Ryoma Yataka, Kazuki Hirashima, Masashi Shiraishi
Recently, studies on machine learning have focused on methods that use symmetry implicit in a specific manifold as an inductive bias.Grassmann manifolds provide the ability to handle fundamental shapes represented as shape spaces, enabling

stable shape analysis. In this paper, we present a novel approach in which we es tablish the theoretical foundations for learning distributions on the Grassmann manifold via continuous normalization flows, with the explicit goal of generatin g stable shapes.Our approach facilitates more robust generation by effectively e liminating the influence of extraneous transformations, such as rotations and in versions, through learning and generating within a Grassmann manifold designed t o accommodate the essential shape information of the object.The experimental res ults indicated that the proposed method could generate high-quality samples by c apturing the data structure.Furthermore, the proposed method significantly outpe rformed state-of-the-art methods in terms of the log-likelihood or evidence lowe r bound.The results obtained are expected to stimulate further research in this field, leading to advances for stable shape generation and analysis.
****
Marich: A Query-efficient Distributionally Equivalent Model Extraction Attack
Pratik Karmakar, Debabrota Basu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Evaluating Post-hoc Explanations for Graph Neural Networks via Robustness Analys is
Junfeng Fang, Wei Liu, Yuan Gao, Zemin Liu, An Zhang, Xiang Wang, Xiangnan He
This work studies the evaluation of explaining graph neural networks (GNNs), whi ch is crucial to the credibility of post-hoc explainability in practical usage. Conventional evaluation metrics, and even explanation methods -- which mainly fo llow the paradigm of feeding the explanatory subgraph and measuring output diffe rence -- always suffer from the notorious out-of-distribution (OOD) issue. In th is work, we endeavor to confront the issue by introducing a novel evaluation met ric, termed OOD-resistant Adversarial Robustness (OAR). Specifically, we draw in spiration from the notion of adversarial robustness and evaluate post-hoc explan ation subgraphs by calculating their robustness under attack. On top of that, an elaborate OOD reweighting block is inserted into the pipeline to confine the ev aluation process to the original data distribution. For applications involving l arge datasets, we further devise a Simplified version of OAR (SimOAR), which ach ieves a significant improvement in computational efficiency at the cost of a sma ll amount of performance. Extensive empirical studies validate the effectiveness of our OAR and SimOAR.
****
A Unifying Perspective on Multi-Calibration: Game Dynamics for Multi-Objective L earning
Nika Haghtalab, Michael Jordan, Eric Zhao
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Not All Neuro-Symbolic Concepts Are Created Equal: Analysis and Mitigation of Re asoning Shortcuts
Emanuele Marconato, Stefano Teso, Antonio Vergari, Andrea Passerini
Neuro-Symbolic (NeSy) predictive models hold the promise of improved compliance with given constraints, systematic generalization, and interpretability, as they allow to infer labels that are consistent with some prior knowledge by reasonin g over high-level concepts extracted from sub-symbolic inputs. It was recently s hown that NeSy predictors are affected by reasoning shortcuts: they can attain h igh accuracy but by leveraging concepts with \textit{unintended semantics}, thus coming short of their promised advantages. Yet, a systematic characterization o f reasoning shortcuts and of potential mitigation strategies is missing. This wo rk fills this gap by characterizing them as unintended optima of the learning ob jective and identifying four key conditions behind their occurrence. Based on th

is, we derive several natural mitigation strategies, and analyze their efficacy both theoretically and empirically. Our analysis shows reasoning shortcuts are difficult to deal with, casting doubts on the trustworthiness and interpretability of existing NeSy solutions.
****
Contrastive Moments: Unsupervised Halfspace Learning in Polynomial Time
Xinyuan Cao, Santosh Vempala
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Boosting Spectral Clustering on Incomplete Data via Kernel Correction and Affinity Learning
Fangchen Yu, Runze Zhao, Zhan Shi, Yiwen Lu, Jicong Fan, Yicheng Zeng, Jianfeng Mao, Wenye Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DASpeech: Directed Acyclic Transformer for Fast and High-quality Speech-to-Speech Translation
Qingkai Fang, Yan Zhou, Yang Feng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Learning Large-scale Neural Fields via Context Pruned Meta-Learning
Jihoon Tack, Subin Kim, Sihyun Yu, Jaeho Lee, Jinwoo Shin, Jonathan Richard Schwarz
We introduce an efficient optimization-based meta-learning technique for large-scale neural field training by realizing significant memory savings through automated online context point selection. This is achieved by focusing each learning step on the subset of data with the highest expected immediate improvement in model quality, resulting in the almost instantaneous modeling of global structure and subsequent refinement of high-frequency details. We further improve the quality of our meta-learned initialization by introducing a bootstrap correction resulting in the minimization of any error introduced by reduced context sets while simultaneously mitigating the well-known myopia of optimization-based meta-learning. Finally, we show how gradient re-scaling at meta-test time allows the learning of extremely high-quality neural fields in significantly shortened optimization procedures. Our framework is model-agnostic, intuitive, straightforward to implement, and shows significant reconstruction improvements for a wide range of signals. We provide an extensive empirical evaluation on nine datasets across multiple multiple modalities, demonstrating state-of-the-art results while providing additional insight through careful analysis of the algorithmic components constituting our method. Code is available at https://github.com/jihoontack/GradNCP
****
AlberDICE: Addressing Out-Of-Distribution Joint Actions in Offline Multi-Agent RL via Alternating Stationary Distribution Correction Estimation
Daiki E. Matsunaga, Jongmin Lee, Jaeseok Yoon, Stefanos Leonardos, Pieter Abbeel, Kee-Eung Kim
One of the main challenges in offline Reinforcement Learning (RL) is the distribution shift that arises from the learned policy deviating from the data collection policy. This is often addressed by avoiding out-of-distribution (OOD) actions during policy improvement as their presence can lead to substantial performance degradation. This challenge is amplified in the offline Multi-Agent RL (MARL) s

etting since the joint action space grows exponentially with the number of agent s.To avoid this curse of dimensionality, existing MARL methods adopt either valu e decomposition methods or fully decentralized training of individual agents. Ho wever, even when combined with standard conservatism principles, these methods c an still result in the selection of OOD joint actions in offline MARL. To this e nd, we introduce AlberDICE,an offline MARL algorithm that alternatively performs centralized training of individual agents based on stationary distribution opti mization. AlberDICE circumvents the exponential complexity of MARL by computing the best response of one agent at a time while effectively avoiding OOD joint ac tion selection. Theoretically, we show that the alternating optimization procedu re converges to Nash policies. In the experiments, we demonstrate that AlberDICE significantly outperforms baseline algorithms on a standard suite of MARL bench marks.
****

Approximate inference of marginals using the IBIA framework

Shivani Bathla, Vinita Vasudevan

Exact inference of marginals in probabilistic graphical models (PGM) is known to be intractable, necessitating the use of approximate methods. Most of the exist ing variational techniques perform iterative message passing in loopy graphs whi ch is slow to converge for many benchmarks. In this paper, we propose a new algo rithm for marginal inference that is based on the incremental build-infer-approx imate (IBIA) paradigm. Our algorithm converts the PGM into a sequence of linked clique tree forests (SLCTF) with bounded clique sizes, and then uses a heuristic belief update algorithm to infer the marginals. For the special case of Bayesia n networks, we show that if the incremental build step in IBIA uses the topologi cal order of variables then (a) the prior marginals are consistent in all CTFs i n the SLCTF and (b) the posterior marginals are consistent once all evidence va riables are added to the SLCTF. In our approach, the belief propagation step is non-iterative and the accuracy-complexity trade-off is controlled using user-def ined clique size bounds. Results for several benchmark sets from recent UAI comp etitions show that our method gives either better or comparable accuracy than ex isting variational and sampling based methods, with smaller runtimes.
****

HiNeRV: Video Compression with Hierarchical Encoding-based Neural Representation

Ho Man Kwan, Ge Gao, Fan Zhang, Andrew Gower, David Bull

Learning-based video compression is currently a popular research topic, offering the potential to compete with conventional standard video codecs. In this conte xt, Implicit Neural Representations (INRs) have previously been used to represen t and compress image and video content, demonstrating relatively high decoding s peed compared to other methods. However, existing INR-based methods have failed to deliver rate quality performance comparable with the state of the art in vide o compression. This is mainly due to the simplicity of the employed network arch itectures, which limit their representation capability. In this paper, we propos e HiNeRV, an INR that combines light weight layers with novel hierarchical posit ional encodings. We employs depth-wise convolutional, MLP and interpolation laye rs to build the deep and wide network architecture with high capacity. HiNeRV is also a unified representation encoding videos in both frames and patches at the same time, which offers higher performance and flexibility than existing method s. We further build a video codec based on HiNeRV and a refined pipeline for tra ining, pruning and quantization that can better preserve HiNeRV's performance du ring lossy model compression. The proposed method has been evaluated on both UVG and MCL-JCV datasets for video compression, demonstrating significant improveme nt over all existing INRs baselines and competitive performance when compared to learning-based codecs (72.3\% overall bit rate saving over HNeRV and 43.4\% ove r DCVC on the UVG dataset, measured in PSNR).
****

Bicriteria Approximation Algorithms for the Submodular Cover Problem

Wenjing Chen, Victoria Crawford

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Responsible AI (RAI) Games and Ensembles

Yash Gupta, Runtian Zhai, Arun Suggala, Pradeep Ravikumar

Several recent works have studied the societal effects of AI; these include issu es such as fairness, robustness, and safety. In many of these objectives, a lea rner seeks to minimize its worst-case loss over a set of predefined distribution s (known as uncertainty sets), with usual examples being perturbed versions of t he empirical distribution. In other words, the aforementioned problems can be wr itten as min-max problems over these uncertainty sets. In this work, we provide a general framework for studying these problems, which we refer to as Responsibl e AI (RAI) games. We provide two classes of algorithms for solving these games: (a) game-play based algorithms, and (b) greedy stagewise estimation algorithms. The former class is motivated by online learning and game theory, whereas the l atter class is motivated by the classical statistical literature on boosting, an d regression. We empirically demonstrate the applicability and competitive perfo rmance of our techniques for solving several RAI problems, particularly around s ubpopulation shift.
****

May the Force be with You: Unified Force-Centric Pre-Training for 3D Molecular C onformations

Rui Feng, Qi Zhu, Huan Tran, Binghong Chen, Aubrey Toland, Rampi Ramprasad, Chao Zhang

Recent works have shown the promise of learning pre-trained models for 3D molecu lar representation.However, existing pre-training models focus predominantly on equilibrium data and largely overlook off-equilibrium conformations.It is challe nging to extend these methods to off-equilibrium data because their training obj ective relies on assumptions ofconformations being the local energy minima. We a ddress this gap by proposing a force-centric pretraining model for 3D molecular conformations covering both equilibrium and off-equilibrium data.For off-equilib rium data, our model learns directly from their atomic forces. For equilibrium d ata, we introduce zero-force regularization and forced-based denoising technique s to approximate near-equilibrium forces.We obtain a unified pre-trained model f or 3D molecular representation with over 15 million diverse conformations. Exper iments show that, with our pre-training objective, we increase forces accuracy b y around 3 times compared to the un-pre-trained Equivariant Transformer model. B y incorporating regularizations on equilibrium data, we solved the problem of un stable MD simulations in vanilla Equivariant Transformers, achieving state-of-th e-art simulation performance with 2.45 times faster inference time than NequIP. As a powerful molecular encoder, our pre-trained model achieves on-par performa nce with state-of-the-art property prediction tasks.
****

Deep Fractional Fourier Transform

Hu Yu, Jie Huang, Lingzhi LI, man zhou, Feng Zhao

Existing deep learning-based computer vision methods usually operate in the spat ial and frequency domains, which are two orthogonal \textbf{individual} perspect ives for image processing.In this paper, we introduce a new spatial-frequency an alysis tool, Fractional Fourier Transform (FRFT), to provide comprehensive \text bf{unified} spatial-frequency perspectives.The FRFT is a unified continuous spat ial-frequency transform that simultaneously reflects an image's spatial and freq uency representations, making it optimal for processing non-stationary image sig nals.We explore the properties of the FRFT for image processing and present a fa st implementation of the 2D FRFT, which facilitates its widespread use.Based on these explorations, we introduce a simple yet effective operator, Multi-order FR actional Fourier Convolution (MFRFC), which exhibits the remarkable merits of pr ocessing images from more perspectives in the spatial-frequency plane. Our propo sed MFRFC is a general and basic operator that can be easily integrated into var ious tasks for performance improvement.We experimentally evaluate the MFRFC on v arious computer vision tasks, including object detection, image classification,

guided super-resolution, denoising, dehazing, deraining, and low-light enhancement. Our proposed MFRFC consistently outperforms baseline methods by significant margins across all tasks.
****

Survival Permanental Processes for Survival Analysis with Time-Varying Covariates

Hideaki Kim

Survival or time-to-event data with time-varying covariates are common in practice, and exploring the non-stationarity in covariates is essential to accurately analyzing the nonlinear dependence of time-to-event outcomes on covariates. Traditional survival analysis methods such as Cox proportional hazards model have been extended to address the time-varying covariates through a counting process formulation, although sophisticated machine learning methods that can accommodate time-varying covariates have been limited. In this paper, we propose a non-parametric Bayesian survival model to analyze the nonlinear dependence of time-to-event outcomes on time-varying covariates. We focus on a computationally feasible Cox process called permanental process, which assumes the square root of hazard function to be generated from a Gaussian process, and tailor it for survival data with time-varying covariates. We verify that the proposed model holds with the representer theorem, a beneficial property for functional analysis, which offers us a fast Bayesian estimation algorithm that scales linearly with the number of observed events without relying on Markov Chain Monte Carlo computation. We evaluate our algorithm on synthetic and real-world data, and show that it achieves comparable predictive accuracy while being tens to hundreds of times faster than state-of-the-art methods.
****

Learn to Categorize or Categorize to Learn? Self-Coding for Generalized Category Discovery

Sarah Rastegar, Hazel Doughty, Cees Snoek

In the quest for unveiling novel categories at test time, we confront the inherent limitations of traditional supervised recognition models that are restricted by a predefined category set. While strides have been made in the realms of self-supervised and open-world learning towards test-time category discovery, a crucial yet often overlooked question persists: what exactly delineates a category? In this paper, we conceptualize a category through the lens of optimization, viewing it as an optimal solution to a well-defined problem. Harnessing this unique conceptualization, we propose a novel, efficient and self-supervised method capable of discovering previously unknown categories at test time. A salient feature of our approach is the assignment of minimum length category codes to individual data instances, which encapsulates the implicit category hierarchy prevalent in real-world datasets. This mechanism affords us enhanced control over category granularity, thereby equipping our model to handle fine-grained categories adeptly. Experimental evaluations, bolstered by state-of-the-art benchmark comparisons, testify to the efficacy of our solution in managing unknown categories at test time. Furthermore, we fortify our proposition with a theoretical foundation, providing proof of its optimality. Our code is available at: https://github.com/SarahRastegar/InfoSieve.
****

Training Chain-of-Thought via Latent-Variable Inference

Du Phan, Matthew Douglas Hoffman, David Dohan, Sholto Douglas, Tuan Anh Le, Aaron Parisi, Pavel Sountsov, Charles Sutton, Sharad Vikram, Rif A. Saurous

Large language models (LLMs) solve problems more accurately and interpretably when instructed to work out the answer step by step using a "chain-of-thought" (CoT) prompt. One can also improve LLMs' performance on a specific task by supervised fine-tuning, i.e., by using gradient ascent on some tunable parameters to maximize the average log-likelihood of correct answers from a labeled training set. Naively combining CoT with supervised tuning requires supervision not just of the correct answers, but also of detailed rationales that lead to those answers; these rationales are expensive to produce by hand. Instead, we propose a fine-tuning strategy that tries to maximize the \emph{marginal} log-likelihood of gener

ating a correct answer using CoT prompting, approximately averaging over all possible rationales. The core challenge is sampling from the posterior over rationales conditioned on the correct answer; we address it using a simple Markov-chain Monte Carlo (MCMC) expectation-maximization (EM) algorithm inspired by the self-taught reasoner (STaR), memoized wake-sleep, Markovian score climbing, and persistent contrastive divergence. This algorithm also admits a novel control-variate technique that drives the variance of our gradient estimates to zero as the model improves. Applying our technique to GSM8K and the tasks in BIG-Bench Hard, we find that this MCMC-EM fine-tuning technique typically improves the model's accuracy on held-out examples more than STaR or prompt-tuning with or without CoT.
****

VAST: A Vision-Audio-Subtitle-Text Omni-Modality Foundation Model and Dataset
Sihan Chen, Handong Li, Qunbo Wang, Zijia Zhao, Mingzhen Sun, Xinxin Zhu, Jing Liu
Vision and text have been fully explored in contemporary video-text foundational models, while other modalities such as audio and subtitles in videos have not received sufficient attention. In this paper, we resort to establish connections between multi-modality video tracks, including Vision, Audio, and Subtitle, and Text by exploring an automatically generated large-scale omni-modality video caption dataset called VAST-27M. Specifically, we first collect 27 million open-domain video clips and separately train a vision and an audio captioner to generate vision and audio captions. Then, we employ an off-the-shelf Large Language Model (LLM) to integrate the generated captions, together with subtitles and instructional prompts into omni-modality captions. Based on the proposed VAST-27M dataset, we train an omni-modality video-text foundational model named VAST, which can perceive and process vision, audio, and subtitle modalities from video, and better support various tasks including vision-text, audio-text, and multi-modal video-text tasks (retrieval, captioning and QA). Extensive experiments have been conducted to demonstrate the effectiveness of our proposed VAST-27M corpus and VAST foundation model. VAST achieves 22 new state-of-the-art results on various cross-modality benchmarks.
****

Bayesian Learning via Q-Exponential Process
Shuyi Li, Michael O'Connor, Shiwei Lan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Repetition In Repetition Out: Towards Understanding Neural Text Degeneration from the Data Perspective
Huayang Li, Tian Lan, Zihao Fu, Deng Cai, Lemao Liu, Nigel Collier, Taro Watanabe, Yixuan Su
There are a number of diverging hypotheses about the neural text degeneration problem, i.e., generating repetitive and dull loops, which makes this problem both interesting and confusing. In this work, we aim to advance our understanding by presenting a straightforward and fundamental explanation from the data perspective. Our preliminary investigation reveals a strong correlation between the degeneration issue and the presence of repetitions in training data. Subsequent experiments also demonstrate that by selectively dropping out the attention to repetitive words in training data, degeneration can be significantly minimized. Furthermore, our empirical analysis illustrates that prior works addressing the degeneration issue from various standpoints, such as the high-inflow words, the likelihood objective, and the self-reinforcement phenomenon, can be interpreted by one simple explanation. That is, penalizing the repetitions in training data is a common and fundamental factor for their effectiveness. Moreover, our experiments reveal that penalizing the repetitions in training data remains critical even when considering larger model sizes and instruction tuning.
****

Facing Off World Model Backbones: RNNs, Transformers, and S4

Fei Deng, Junyeong Park, Sungjin Ahn

World models are a fundamental component in model-based reinforcement learning (MBRL). To perform temporally extended and consistent simulations of the future in partially observable environments, world models need to possess long-term memory. However, state-of-the-art MBRL agents, such as Dreamer, predominantly employ recurrent neural networks (RNNs) as their world model backbone, which have limited memory capacity. In this paper, we seek to explore alternative world model backbones for improving long-term memory. In particular, we investigate the effectiveness of Transformers and Structured State Space Sequence (S4) models, motivated by their remarkable ability to capture long-range dependencies in low-dimensional sequences and their complementary strengths. We propose S4WM, the first world model compatible with parallelizable SSMs including S4 and its variants. By incorporating latent variable modeling, S4WM can efficiently generate high-dimensional image sequences through latent imagination. Furthermore, we extensively compare RNN-, Transformer-, and S4-based world models across four sets of environments, which we have tailored to assess crucial memory capabilities of world models, including long-term imagination, context-dependent recall, reward prediction, and memory-based reasoning. Our findings demonstrate that S4WM outperforms Transformer-based world models in terms of long-term memory, while exhibiting greater efficiency during training and imagination. These results pave the way for the development of stronger MBRL agents.

****

STARSS23: An Audio-Visual Dataset of Spatial Recordings of Real Scenes with Spatiotemporal Annotations of Sound Events

Kazuki Shimada, Archontis Politis, Parthasaarathy Sudarsanam, Daniel A. Krause, Kengo Uchida, Sharath Adavanne, Aapo Hakala, Yuichiro Koyama, Naoya Takahashi, Shusuke Takahashi, Tuomas Virtanen, Yuki Mitsufuji

While direction of arrival (DOA) of sound events is generally estimated from multichannel audio data recorded in a microphone array, sound events usually derive from visually perceptible source objects, e.g., sounds of footsteps come from the feet of a walker. This paper proposes an audio-visual sound event localization and detection (SELD) task, which uses multichannel audio and video information to estimate the temporal activation and DOA of target sound events. Audio-visual SELD systems can detect and localize sound events using signals from a microphone array and audio-visual correspondence. We also introduce an audio-visual dataset, Sony-TAu Realistic Spatial Soundscapes 2023 (STARSS23), which consists of multichannel audio data recorded with a microphone array, video data, and spatiotemporal annotation of sound events. Sound scenes in STARSS23 are recorded with instructions, which guide recording participants to ensure adequate activity and occurrences of sound events. STARSS23 also serves human-annotated temporal activation labels and human-confirmed DOA labels, which are based on tracking results of a motion capture system. Our benchmark results demonstrate the benefits of using visual object positions in audio-visual SELD tasks. The data is available at https://zenodo.org/record/7880637.

****

Inserting Anybody in Diffusion Models via Celeb Basis

Ge Yuan, Xiaodong Cun, Yong Zhang, Maomao Li, Chenyang Qi, Xintao Wang, Ying Shan, Huicheng Zheng

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.

****

Scaling Open-Vocabulary Object Detection

Matthias Minderer, Alexey Gritsenko, Neil Houlsby

Open-vocabulary object detection has benefited greatly from pretrained vision-language models, but is still limited by the amount of available detection training data. While detection training data can be expanded by using Web image-text pairs as weak supervision, this has not been done at scales comparable to image-level pretraining. Here, we scale up detection data with self-training, which uses

an existing detector to generate pseudo-box annotations on image-text pairs. Major challenges in scaling self-training are the choice of label space, pseudo-annotation filtering, and training efficiency. We present the OWLv2 model and OWL-ST self-training recipe, which address these challenges. OWLv2 surpasses the performance of previous state-of-the-art open-vocabulary detectors already at comparable training scales (~10M examples). However, with OWL-ST, we can scale to over 1B examples, yielding further large improvement: With an L/14 architecture, OWL-ST improves AP on LVIS rare classes, for which the model has seen no human box annotations, from 31.2% to 44.6% (43% relative improvement). OWL-ST unlocks Web-scale training for open-world localization, similar to what has been seen for image classification and language modelling. Code and checkpoints are available on GitHub.
****
Formulating Discrete Probability Flow Through Optimal Transport
Pengze Zhang, Hubery Yin, Chen Li, Xiaohua Xie
Continuous diffusion models are commonly acknowledged to display a deterministic probability flow, whereas discrete diffusion models do not. In this paper, we aim to establish the fundamental theory for the probability flow of discrete diffusion models. Specifically, we first prove that the continuous probability flow is the Monge optimal transport map under certain conditions, and also present an equivalent evidence for discrete cases. In view of these findings, we are then able to define the discrete probability flow in line with the principles of optimal transport. Finally, drawing upon our newly established definitions, we propose a novel sampling method that surpasses previous discrete diffusion models in its ability to generate more certain outcomes. Extensive experiments on the synthetic toy dataset and the CIFAR-10 dataset have validated the effectiveness of our proposed discrete probability flow. Code is released at: https://github.com/PangzeCheung/Discrete-Probability-Flow.
****
Successor-Predecessor Intrinsic Exploration
Changmin Yu, Neil Burgess, Maneesh Sahani, Samuel J Gershman
Exploration is essential in reinforcement learning, particularly in environments where external rewards are sparse. Here we focus on exploration with intrinsic rewards, where the agent transiently augments the external rewards with self-generated intrinsic rewards. Although the study of intrinsic rewards has a long history, existing methods focus on composing the intrinsic reward based on measures of future prospects of states, ignoring the information contained in the retrospective structure of transition sequences. Here we argue that the agent can utilise retrospective information to generate explorative behaviour with structure-awareness, facilitating efficient exploration based on global instead of local information. We propose Successor-Predecessor Intrinsic Exploration (SPIE), an exploration algorithm based on a novel intrinsic reward combining prospective and retrospective information. We show that SPIE yields more efficient and ethologically plausible exploratory behaviour in environments with sparse rewards and bottleneck states than competing methods. We also implement SPIE in deep reinforcement learning agents, and show that the resulting agent achieves stronger empirical performance than existing methods on sparse-reward Atari games.
****
TFLEX: Temporal Feature-Logic Embedding Framework for Complex Reasoning over Temporal Knowledge Graph
Xueyuan Lin, Haihong E, Chengjin Xu, Gengxian Zhou, Haoran Luo, Tianyi Hu, Fenglong Su, Ningyuan Li, Mingzhi Sun
Multi-hop logical reasoning over knowledge graph plays a fundamental role in many artificial intelligence tasks. Recent complex query embedding methods for reasoning focus on static KGs, while temporal knowledge graphs have not been fully explored. Reasoning over TKGs has two challenges: 1. The query should answer entities or timestamps; 2. The operators should consider both set logic on entity set and temporal logic on timestamp set.To bridge this gap, we introduce the multi-hop logical reasoning problem on TKGs and then propose the first temporal complex query embedding named Temporal Feature-Logic Embedding framework (TFLEX) to

answer the temporal complex queries. Specifically, we utilize fuzzy logic to c
ompute the logic part of the Temporal Feature-Logic embedding, thus naturally mo
deling all first-order logic operations on the entity set. In addition, we furt
her extend fuzzy logic on timestamp set to cope with three extra temporal operat
ors (After, Before and Between).Experiments on numerous query patterns demonstra
te the effectiveness of our method.
****
StyleGAN knows Normal, Depth, Albedo, and More
Anand Bhattad, Daniel McKee, Derek Hoiem, David Forsyth
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Are Vision Transformers More Data Hungry Than Newborn Visual Systems?
Lalit Pandey, Samantha Wood, Justin Wood
Vision transformers (ViTs) are top-performing models on many computer vision ben
chmarks and can accurately predict human behavior on object recognition tasks. H
owever, researchers question the value of using ViTs as models of biological lea
rning because ViTs are thought to be more “data hungry” than brains, with ViTs r
equiring more training data than brains to reach similar levels of performance.
To test this assumption, we directly compared the learning abilities of ViTs and
 animals, by performing parallel controlled-rearing experiments on ViTs and newb
orn chicks. We first raised chicks in impoverished visual environments containin
g a single object, then simulated the training data available in those environme
nts by building virtual animal chambers in a video game engine. We recorded the
first-person images acquired by agents moving through the virtual chambers and u
sed those images to train self-supervised ViTs that leverage time as a teaching
signal, akin to biological visual systems. When ViTs were trained “through the e
yes” of newborn chicks, the ViTs solved the same view-invariant object recogniti
on tasks as the chicks. Thus, ViTs were not more data hungry than newborn chicks
: both learned view-invariant object representations in impoverished visual envi
ronments. The flexible and generic attention-based learning mechanism in ViTs—co
mbined with the embodied data streams available to newborn animals—appears suffi
cient to drive the development of animal-like object recognition.
****
How to Scale Your EMA
Dan Busbridge, Jason Ramapuram, Pierre Ablin, Tatiana Likhomanenko, Eeshan Gunes
h Dhekane, Xavier Suau Cuadros, Russell Webb
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Unsupervised Graph Neural Architecture Search with Disentangled Self-Supervision
Zeyang Zhang, Xin Wang, Ziwei Zhang, Guangyao Shen, Shiqi Shen, Wenwu Zhu
The existing graph neural architecture search (GNAS) methods heavily rely on sup
ervised labels during the search process, failing to handle ubiquitous scenarios
 where supervisions are not available. In this paper, we study the problem of un
supervised graph neural architecture search, which remains unexplored in the lit
erature. The key problem is to discover the latent graph factors that drive the
formation of graph data as well as the underlying relations between the factors
and the optimal neural architectures. Handling this problem is challenging given
 that the latent graph factors together with architectures are highly entangled
due to the nature of the graph and the complexity of the neural architecture sea
rch process. To address the challenge, we propose a novel Disentangled Self-supe
rvised Graph Neural Architecture Search (DSGAS) model, which is able to discover
 the optimal architectures capturing various latent graph factors in a self-supe
rvised fashion based on unlabeled graph data. Specifically, we first design a di
sentangled graph super-network capable of incorporating multiple architectures w

ith factor-wise disentanglement, which are optimized simultaneously. Then, we estimate the performance of architectures under different factors by our proposed self-supervised training with joint architecture-graph disentanglement. Finally, we propose a contrastive search with architecture augmentations to discover architectures with factor-specific expertise. Extensive experiments on 11 real-world datasets demonstrate that the proposed model is able to achieve state-of-the-art performance against several baseline methods in an unsupervised manner.
****
Setting the Trap: Capturing and Defeating Backdoors in Pretrained Language Models through Honeypots
Ruixiang (Ryan) Tang, Jiayi Yuan, Yiming Li, Zirui Liu, Rui Chen, Xia Hu
In the field of natural language processing, the prevalent approach involves fine-tuning pretrained language models (PLMs) using local samples. Recent research has exposed the susceptibility of PLMs to backdoor attacks, wherein the adversaries can embed malicious prediction behaviors by manipulating a few training samples. In this study, our objective is to develop a backdoor-resistant tuning procedure that yields a backdoor-free model, no matter whether the fine-tuning dataset contains poisoned samples. To this end, we propose and integrate an \emph{honeypot module} into the original PLM, specifically designed to absorb backdoor information exclusively. Our design is motivated by the observation that lower-layer representations in PLMs carry sufficient backdoor features while carrying minimal information about the original tasks. Consequently, we can impose penalties on the information acquired by the honeypot module to inhibit backdoor creation during the fine-tuning process of the stem network. Comprehensive experiments conducted on benchmark datasets substantiate the effectiveness and robustness of our defensive strategy. Notably, these results indicate a substantial reduction in the attack success rate ranging from 10\% to 40\% when compared to prior state-of-the-art methods.
****
Learning Large-Scale MTP$_2$ Gaussian Graphical Models via Bridge-Block Decomposition
Xiwen WANG, Jiaxi Ying, Daniel Palomar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Collaborative Score Distillation for Consistent Visual Editing
Subin Kim, Kyungmin Lee, June Suk Choi, Jongheon Jeong, Kihyuk Sohn, Jinwoo Shin
Generative priors of large-scale text-to-image diffusion models enable a wide range of new generation and editing applications on diverse visual modalities. However, when adapting these priors to complex visual modalities, often represented as multiple images (e.g., video or 3D scene), achieving consistency across a set of images is challenging. In this paper, we address this challenge with a novel method, Collaborative Score Distillation (CSD). CSD is based on the Stein Variational Gradient Descent (SVGD). Specifically, we propose to consider multiple samples as “particles” in the SVGD update and combine their score functions to distill generative priors over a set of images synchronously. Thus, CSD facilitates the seamless integration of information across 2D images, leading to a consistent visual synthesis across multiple samples. We show the effectiveness of CSD in a variety of editing tasks, encompassing the visual editing of panorama images, videos, and 3D scenes. Our results underline the competency of CSD as a versatile method for enhancing inter-sample consistency, thereby broadening the applicability of text-to-image diffusion models.
****
FLuID: Mitigating Stragglers in Federated Learning using Invariant Dropout
Irene Wang, Prashant Nair, Divya Mahajan
Federated Learning (FL) allows machine learning models to train locally on individual mobile devices, synchronizing model updates via a shared server. This approach safeguards user privacy; however, it also generates a heterogeneous trainin

g environment due to the varying performance capabilities across devices. As a r esult, “straggler” devices with lower performance often dictate the overalltrain ing time in FL. In this work, we aim to alleviate this performance bottleneck du e to stragglers by dynamically balancing the training load across the system. We introduce Invariant Dropout, a method that extracts a sub-model based on the we ight update threshold, thereby minimizing potential impacts on accuracy. Buildin g on this dropout technique, we develop an adaptive training framework, Federate d Learning using Invariant Dropout (FLuID). FLuID offers a lightweight sub-model extraction to regulate computational intensity, thereby reducing the load on st raggler devices without affecting model quality. Our method leverages neuron upd ates from non-straggler devices to construct a tailored sub-model for each strag gler based on client performance profiling. Furthermore, FLuID can dynamically a dapt to changes in stragglers as runtime conditions shift. We evaluate FLuID usi ng five real-world mobile clients. The evaluations show that Invariant Dropout m aintains baseline model efficiency while alleviating the performance bottleneck of stragglers through a dynamic, runtime approach.
****
Learning to Augment Distributions for Out-of-distribution Detection
Qizhou Wang, Zhen Fang, Yonggang Zhang, Feng Liu, Yixuan Li, Bo Han
Open-world classification systems should discern out-of-distribution (OOD) data whose labels deviate from those of in-distribution (ID) cases, motivating recent studies in OOD detection. Advanced works, despite their promising progress, may still fail in the open world, owing to the lacking knowledge about unseen OOD d ata in advance. Although one can access auxiliary OOD data (distinct from unseen ones) for model training, it remains to analyze how such auxiliary data will wo rk in the open world. To this end, we delve into such a problem from a learning theory perspective, finding that the distribution discrepancy between the auxili ary and the unseen real OOD data is the key to affect the open-world detection p erformance. Accordingly, we propose Distributional-Augmented OOD Learning (DAOL) , alleviating the OOD distribution discrepancy by crafting an OOD distribution s et that contains all distributions in a Wasserstein ball centered on the auxilia ry OOD distribution. We justify that the predictor trained over the worst OOD da ta in the ball can shrink the OOD distribution discrepancy, thus improving the o pen-world detection performance given only the auxiliary OOD data. We conduct ex tensive evaluations across representative OOD detection setups, demonstrating th e superiority of our DAOL over its advanced counterparts.
****
Covariance-adaptive best arm identification
El Mehdi Saad, Gilles Blanchard, Nicolas Verzelen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
What a MESS: Multi-Domain Evaluation of Zero-Shot Semantic Segmentation
Benedikt Blumenstiel, Johannes Jakubik, Hilde Kuehne, Michael Vössing
While semantic segmentation has seen tremendous improvements in the past, there are still significant labeling efforts necessary and the problem of limited gene ralization to classes that have not been present during training. To address thi s problem, zero-shot semantic segmentation makes use of large self-supervised vi sion-language models, allowing zero-shot transfer to unseen classes. In this wor k, we build a benchmark for Multi-domain Evaluation of Zero-Shot Semantic Segmen tation (MESS), which allows a holistic analysis of performance across a wide ran ge of domain-specific datasets such as medicine, engineering, earth monitoring, biology, and agriculture. To do this, we reviewed 120 datasets, developed a taxo nomy, and classified the datasets according to the developed taxonomy. We select a representative subset consisting of 22 datasets and propose it as the MESS be nchmark. We evaluate eight recently published models on the proposed MESS benchm ark and analyze characteristics for the performance of zero-shot transfer models . The toolkit is available at https://github.com/blumenstiel/MESS.

****
Swarm Reinforcement Learning for Adaptive Mesh Refinement
Niklas Freymuth, Philipp Dahlinger, Tobias Würth, Simon Reisch, Luise Kärger, Gerhard Neumann
The Finite Element Method, an important technique in engineering, is aided by Adaptive Mesh Refinement (AMR), which dynamically refines mesh regions to allow for a favorable trade-off between computational speed and simulation accuracy. Classical methods for AMR depend on task-specific heuristics or expensive error estimators, hindering their use for complex simulations. Recent learned AMR methods tackle these problems, but so far scale only to simple toy examples. We formulate AMR as a novel Adaptive Swarm Markov Decision Process in which a mesh is modeled as a system of simple collaborating agents that may split into multiple new agents. This framework allows for a spatial reward formulation that simplifies the credit assignment problem, which we combine with Message Passing Networks to propagate information between neighboring mesh elements. We experimentally validate the effectiveness of our approach, Adaptive Swarm Mesh Refinement (ASMR), showing that it learns reliable, scalable, and efficient refinement strategies on a set of challenging problems. Our approach significantly speeds up computation, achieving up to 30-fold improvement compared to uniform refinements in complex simulations. Additionally, we outperform learned baselines and achieve a refinement quality that is on par with a traditional error-based AMR strategy without expensive oracle information about the error signal.
****
Fast Projected Newton-like Method for Precision Matrix Estimation under Total Positivity
Jian-Feng CAI, José Vinícius de Miranda Cardoso , Daniel Palomar, Jiaxi Ying
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
BanditPAM++: Faster k-medoids Clustering
Mo Tiwari, Ryan Kang, Donghyun Lee, Sebastian Thrun, Ilan Shomorony, Martin J. Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Minigrid & Miniworld: Modular & Customizable Reinforcement Learning Environments for Goal-Oriented Tasks
Maxime Chevalier-Boisvert, Bolun Dai, Mark Towers, Rodrigo Perez-Vicente, Lucas Willems, Salem Lahlou, Suman Pal, Pablo Samuel Castro, J Terry
We present the Minigrid and Miniworld libraries which provide a suite of goal-oriented 2D and 3D environments. The libraries were explicitly created with a minimalistic design paradigm to allow users to rapidly develop new environments for a wide range of research-specific needs. As a result, both have received widescale adoption by the RL community, facilitating research in a wide range of areas. In this paper, we outline the design philosophy, environment details, and their world generation API. We also showcase the additional capabilities brought by the unified API between Minigrid and Miniworld through case studies on transfer learning (for both RL agents and humans) between the different observation spaces. The source code of Minigrid and Miniworld can be found at https://github.com/Farama-Foundation/Minigrid and https://github.com/Farama-Foundation/Miniworld along with their documentation at https://minigrid.farama.org/ and https://miniworld.farama.org/.
****
Cross-Domain Policy Adaptation via Value-Guided Data Filtering
Kang Xu, Chenjia Bai, Xiaoteng Ma, Dong Wang, Bin Zhao, Zhen Wang, Xuelong Li, Wei Li

Generalizing policies across different domains with dynamics mismatch poses a si
gnificant challenge in reinforcement learning. For example, a robot learns the p
olicy in a simulator, but when it is deployed in the real world, the dynamics of
 the environment may be different. Given the source and target domain with dynam
ics mismatch, we consider the online dynamics adaptation problem, in which case
the agent can access sufficient source domain data while online interactions wit
h the target domain are limited. Existing research has attempted to solve the pr
oblem from the dynamics discrepancy perspective. In this work, we reveal the lim
itations of these methods and explore the problem from the value difference pers
pective via a novel insight on the value consistency across domains. Specificall
y, we present the Value-Guided Data Filtering (VGDF) algorithm, which selectivel
y shares transitions from the source domain based on the proximity of paired val
ue targets across the two domains. Empirical results on various environments wit
h kinematic and morphology shifts demonstrate that our method achieves superior
performance compared to prior approaches.
****
Connecting Certified and Adversarial Training
Yuhao Mao, Mark Müller, Marc Fischer, Martin Vechev
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Effectively Learning Initiation Sets in Hierarchical Reinforcement Learning
Akhil Bagaria, Ben Abbatematteo, Omer Gottesman, Matt Corsaro, Sreehari Rammohan
, George Konidaris
An agent learning an option in hierarchical reinforcement learning must solve th
ree problems: identify the option's subgoal (termination condition), learn a pol
icy, and learn where that policy will succeed (initiation set). The termination
condition is typically identified first, but the option policy and initiation se
t must be learned simultaneously, which is challenging because the initiation se
t depends on the option policy, which changes as the agent learns. Consequently,
 data obtained from option execution becomes invalid over time, leading to an in
accurate initiation set that subsequently harms downstream task performance. We
highlight three issues---data non-stationarity, temporal credit assignment, and
pessimism---specific to learning initiation sets, and propose to address them us
ing tools from off-policy value estimation and classification. We show that our
method learns higher-quality initiation sets faster than existing methods (in Mi
niGrid and Montezuma's Revenge), can automatically discover promising grasps for
 robot manipulation (in Robosuite), and improves the performance of a state-of-t
he-art option discovery method in a challenging maze navigation task in MuJoCo.
****
Alignment with human representations supports robust few-shot learning
Ilia Sucholutsky, Tom Griffiths
Should we care whether AI systems have representations of the world that are sim
ilar to those of humans? We provide an information-theoretic analysis that sugge
sts that there should be a U-shaped relationship between the degree of represent
ational alignment with humans and performance on few-shot learning tasks. We con
firm this prediction empirically, finding such a relationship in an analysis of
the performance of 491 computer vision models. We also show that highly-aligned
models are more robust to both natural adversarial attacks and domain shifts. Ou
r results suggest that human-alignment is often a sufficient, but not necessary,
 condition for models to make effective use of limited data, be robust, and gene
ralize well.
****
ReMaX: Relaxing for Better Training on Efficient Panoptic Segmentation
Shuyang Sun, WEIJUN WANG, Andrew Howard, Qihang Yu, Philip Torr, Liang-Chieh Che
n
This paper presents a new mechanism to facilitate the training of mask transform
ers for efficient panoptic segmentation, democratizing its deployment. We observ

e that due to the high complexity in the training objective of panoptic segmentation, it will inevitably lead to much higher penalization on false positive. Such unbalanced loss makes the training process of the end-to-end mask-transformer based architectures difficult, especially for efficient models. In this paper, we present ReMaX that adds relaxation to mask predictions and class predictions during the training phase for panoptic segmentation. We demonstrate that via these simple relaxation techniques during training, our model can be consistently improved by a clear margin without any extra computational cost on inference. By combining our method with efficient backbones like MobileNetV3-Small, our method achieves new state-of-the-art results for efficient panoptic segmentation on COCO, ADE20K and Cityscapes. Code and pre-trained checkpoints will be available at https://github.com/google-research/deeplab2.
****
The Behavior and Convergence of Local Bayesian Optimization
Kaiwen Wu, Kyurae Kim, Roman Garnett, Jacob Gardner
A recent development in Bayesian optimization is the use of local optimization strategies, which can deliver strong empirical performance on high-dimensional problems compared to traditional global strategies. The "folk wisdom" in the literature is that the focus on local optimization sidesteps the curse of dimensionality; however, little is known concretely about the expected behavior or convergence of Bayesian local optimization routines. We first study the behavior of the local approach, and find that the statistics of individual local solutions of Gaussian process sample paths are surprisingly good compared to what we would expect to recover from global methods. We then present the first rigorous analysis of such a Bayesian local optimization algorithm recently proposed by Müller et al. (2021), and derive convergence rates in both the noisy and noiseless settings.
****
Contrastive Sampling Chains in Diffusion Models
Junyu Zhang, Daochang Liu, Shichao Zhang, Chang Xu
The past few years have witnessed great success in the use of diffusion models (DMs) to generate high-fidelity images with the help of stochastic differential equations (SDEs). However, discretization error is an inevitable limitation when utilizing numerical solvers to solve SDEs. To address this limitation, we provide a theoretical analysis demonstrating that an appropriate combination of the contrastive loss and score matching serves as an upper bound of the KL divergence between the true data distribution and the model distribution. To obtain this bound, we utilize a contrastive loss to construct a contrastive sampling chain to fine-tuning the pre-trained DM. In this manner, our method reduces the discretization error and thus yields a smaller gap between the true data distribution and our model distribution. Moreover, the presented method can be applied to fine-tuning various pre-trained DMs, both with or without fast sampling algorithms, contributing to better sample quality or slightly faster sampling speeds. To validate the efficacy of our method, we conduct comprehensive experiments. For example, on CIFAR10, when applied to a pre-trained EDM, our method improves the FID from 2.04 to 1.88 with 35 neural function evaluations (NFEs), and reduces NFEs from 35 to 25 to achieve the same 2.04 FID.
****
Effective Robustness against Natural Distribution Shifts for Models with Different Training Data
Zhouxing Shi, Nicholas Carlini, Ananth Balashankar, Ludwig Schmidt, Cho-Jui Hsieh, Alex Beutel, Yao Qin
``Effective robustness'' measures the extra out-of-distribution (OOD) robustness beyond what can be predicted from the in-distribution (ID) performance. Existing effective robustness evaluations typically use a single test set such as ImageNet to evaluate the ID accuracy. This becomes problematic when evaluating models trained on different data distributions, e.g., comparing models trained on ImageNet vs. zero-shot language-image pre-trained models trained on LAION. In this paper, we propose a new evaluation metric to evaluate and compare the effective robustness of models trained on different data. To do this, we control for the accuracy on multiple ID test sets that cover the training distributions for all th

e evaluated models. Our new evaluation metric provides a better estimate of effe ctive robustness when there are models with different training data. It may also explain the surprising effective robustness gains of zero-shot CLIP-like models exhibited in prior works that used ImageNet as the only ID test set, while the gains diminish under our new evaluation. Additional artifacts including interact ive visualizations are provided at https://shizhouxing.github.io/effective-robus tness.
****
Tailoring Self-Attention for Graph via Rooted Subtrees
Siyuan Huang, Yunchong Song, Jiayue Zhou, Zhouhan Lin
Attention mechanisms have made significant strides in graph learning, yet they s till exhibit notable limitations: local attention faces challenges in capturing long-range information due to the inherent problems of the message-passing schem e, while global attention cannot reflect the hierarchical neighborhood structure and fails to capture fine-grained local information. In this paper, we propose a novel multi-hop graph attention mechanism, named Subtree Attention (STA), to a ddress the aforementioned issues. STA seamlessly bridges the fully-attentional s tructure and the rooted subtree, with theoretical proof that STA approximates th e global attention under extreme settings. By allowing direct computation of att ention weights among multi-hop neighbors, STA mitigates the inherent problems in existing graph attention mechanisms. Further we devise an efficient form for ST A by employing kernelized softmax, which yields a linear time complexity. Our re sulting GNN architecture, the STAGNN, presents a simple yet performant STA-based graph neural network leveraging a hop-aware attention strategy. Comprehensive e valuations on ten node classification datasets demonstrate that STA-based models outperform existing graph transformers and mainstream GNNs. The codeis availabl e at https://github.com/LUMIA-Group/SubTree-Attention.
****
Squeeze, Recover and Relabel: Dataset Condensation at ImageNet Scale From A New Perspective
Zeyuan Yin, Eric Xing, Zhiqiang Shen
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Disentangled Wasserstein Autoencoder for T-Cell Receptor Engineering
Tianxiao Li, Hongyu Guo, Filippo Grazioli, Mark Gerstein, Martin Renqiang Min
In protein biophysics, the separation between the functionally important residue s (forming the active site or binding surface) and those that create the overall structure (the fold) is a well-established and fundamental concept. Identifying and modifying those functional sites is critical for protein engineering but co mputationally non-trivial, and requires significant domain knowledge. To automat e this process from a data-driven perspective, we propose a disentangled Wassers tein autoencoder with an auxiliary classifier, which isolates the function-relat ed patterns from the rest with theoretical guarantees. This enables one-pass pro tein sequence editing and improves the understanding of the resulting sequences and editing actions involved. To demonstrate its effectiveness, we apply it to T -cell receptors (TCRs), a well-studied structure-function case. We show that our method can be used to alter the function of TCRs without changing the structura l backbone, outperforming several competing methods in generation quality and ef ficiency, and requiring only 10\% of the running time needed by baseline models. To our knowledge, this is the first approach that utilizes disentangled represe ntations for TCR engineering.
****
Modality-Independent Teachers Meet Weakly-Supervised Audio-Visual Event Parser
Yung-Hsuan Lai, Yen-Chun Chen, Frank Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Uncovering Prototypical Knowledge for Weakly Open-Vocabulary Semantic Segmentation
Fei Zhang, Tianfei Zhou, Boyang Li, Hao He, Chaofan Ma, Tianjiao Zhang, Jiangchao Yao, Ya Zhang, Yanfeng Wang
This paper studies the problem of weakly open-vocabulary semantic segmentation (WOVSS), which learns to segment objects of arbitrary classes using mere image-text pairs. Existing works turn to enhance the vanilla vision transformer by introducing explicit grouping recognition, i.e., employing several group tokens/centroids to cluster the image tokens and perform the group-text alignment. Nevertheless, these methods suffer from a granularity inconsistency regarding the usage of group tokens, which are aligned in the all-to-one v.s. one-to-one manners during the training and inference phases, respectively. We argue that this discrepancy arises from the lack of elaborate supervision for each group token. To bridge this granularity gap, this paper explores explicit supervision for the group tokens from the prototypical knowledge. To this end, this paper proposes the non-learnable prototypical regularization (NPR) where non-learnable prototypes are estimated from source features to serve as supervision and enable contrastive matching of the group tokens. This regularization encourages the group tokens to segment objects with less redundancy and capture more comprehensive semantic regions, leading to increased compactness and richness. Based on NPR, we propose the prototypical guidance segmentation network (PGSeg) that incorporates multi-modal regularization by leveraging prototypical sources from both images and texts at different levels, progressively enhancing the segmentation capability with diverse prototypical patterns. Experimental results show that our proposed method achieves state-of-the-art performance on several benchmark datasets.
****
A Theoretical Analysis of Optimistic Proximal Policy Optimization in Linear Markov Decision Processes
Han Zhong, Tong Zhang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Uncertainty-Aware Instance Reweighting for Off-Policy Learning
Xiaoying Zhang, Junpu Chen, Hongning Wang, Hong Xie, Yang Liu, John C.S. Lui, Hang Li
Off-policy learning, referring to the procedure of policy optimization with access only to logged feedback data, has shown importance in various important real-world applications, such as search engines and recommender systems. While the ground-truth logging policy is usually unknown, previous work simply takes its estimated value for the off-policy learning, ignoring the negative impact from both high bias and high variance resulted from such an estimator. And these impact is often magnified on samples with small and inaccurately estimated logging probabilities. The contribution of this work is to explicitly model the uncertainty in the estimated logging policy, and propose an Uncertainty-aware Inverse Propensity Score estimator (UIPS) for improved off-policy learning, with a theoretical convergence guarantee. Experiment results on the synthetic and real-world recommendation datasets demonstrate that UIPS significantly improves the quality of the discovered policy, when compared against an extensive list of state-of-the-art baselines.
****
Model-free Posterior Sampling via Learning Rate Randomization
Daniil Tiapkin, Denis Belomestny, Daniele Calandriello, Eric Moulines, Remi Munos, Alexey Naumov, Pierre Perrault, Michal Valko, Pierre Ménard
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
TriRE: A Multi-Mechanism Learning Paradigm for Continual Knowledge Retention and Promotion
Preetha Vijayan, Prashant Bhat, Bahram Zonooz, Elahe Arani
Continual learning (CL) has remained a persistent challenge for deep neural networks due to catastrophic forgetting (CF) of previously learned tasks. Several techniques such as weight regularization, experience rehearsal, and parameter isolation have been proposed to alleviate CF. Despite their relative success, these research directions have predominantly remained orthogonal and suffer from several shortcomings, while missing out on the advantages of competing strategies. On the contrary, the brain continually learns, accommodates, and transfers knowledge across tasks by simultaneously leveraging several neurophysiological processes, including neurogenesis, active forgetting, neuromodulation, metaplasticity, experience rehearsal, and context-dependent gating, rarely resulting in CF. Inspired by how the brain exploits multiple mechanisms concurrently, we propose TriRE, a novel CL paradigm that encompasses retaining the most prominent neurons for each task, revising and solidifying the extracted knowledge of current and past tasks, and actively promoting less active neurons for subsequent tasks through rewinding and relearning. Across CL settings, TriRE significantly reduces task interference and surpasses different CL approaches considered in isolation.
****
Implicit Variational Inference for High-Dimensional Posteriors
Anshuk Uppal, Kristoffer Stensbo-Smidt, Wouter Boomsma, Jes Frellsen
In variational inference, the benefits of Bayesian models rely on accurately capturing the true posterior distribution. We propose using neural samplers that specify implicit distributions, which are well-suited for approximating complex multimodal and correlated posteriors in high-dimensional spaces. Our approach introduces novel bounds for approximate inference using implicit distributions by locally linearising the neural sampler. This is distinct from existing methods that rely on additional discriminator networks and unstable adversarial objectives. Furthermore, we present a new sampler architecture that, for the first time, enables implicit distributions over tens of millions of latent variables, addressing computational concerns by using differentiable numerical approximations. We empirically show that our method is capable of recovering correlations across layers in large Bayesian neural networks, a property that is crucial for a network's performance but notoriously challenging to achieve. To the best of our knowledge, no other method has been shown to accomplish this task for such large models. Through experiments in downstream tasks, we demonstrate that our expressive posteriors outperform state-of-the-art uncertainty quantification methods, validating the effectiveness of our training algorithm and the quality of the learned implicit approximation.
****
k-Median Clustering via Metric Embedding: Towards Better Initialization with Differential Privacy
Chenglin Fan, Ping Li, Xiaoyun Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Towards a Unified Analysis of Kernel-based Methods Under Covariate Shift
Xingdong Feng, Xin HE, Caixing Wang, Chao Wang, Jingnan Zhang
Covariate shift occurs prevalently in practice, where the input distributions of the source and target data are substantially different. Despite its practical importance in various learning problems, most of the existing methods only focus on some specific learning tasks and are not well validated theoretically and numerically. To tackle this problem, we propose a unified analysis of general nonparametric methods in a reproducing kernel Hilbert space (RKHS) under covariate shift. Our theoretical results are established for a general loss belonging to a

rich loss function family, which includes many commonly used methods as special cases, such as mean regression, quantile regression, likelihood-based classification, and margin-based classification. Two types of covariate shift problems are the focus of this paper and the sharp convergence rates are established for a general loss function to provide a unified theoretical analysis, which concurs with the optimal results in literature where the squared loss is used. Extensive numerical studies on synthetic and real examples confirm our theoretical findings and further illustrate the effectiveness of our proposed method.
****

Learning Functional Transduction

Mathieu Chalvidal, Thomas Serre, Rufin VanRullen

Research in statistical learning has polarized into two general approaches to perform regression analysis: Transductive methods construct estimates directly based on exemplar data using generic relational principles which might suffer from the curse of dimensionality. Conversely, inductive methods can potentially fit highly complex functions at the cost of compute-intensive solution searches. In this work, we leverage the theory of vector-valued Reproducing Kernel Banach Spaces (RKBS) to propose a hybrid approach: We show that transductive regression systems can be meta-learned with gradient descent to form efficient in-context neural approximators of function defined over both finite and infinite-dimensional spaces (operator regression). Once trained, our Transducer can almost instantaneously capture new functional relationships and produce original image estimates, given a few pairs of input and output examples. We demonstrate the benefit of our meta-learned transductive approach to model physical systems influenced by varying external factors with little data at a fraction of the usual deep learning training costs for partial differential equations and climate modeling applications.
****

Gaussian Membership Inference Privacy

Tobias Leemann, Martin Pawelczyk, Gjergji Kasneci

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Modality-Agnostic Self-Supervised Learning with Meta-Learned Masked Auto-Encoder

Huiwon Jang, Jihoon Tack, Daewon Choi, Jongheon Jeong, Jinwoo Shin

Despite its practical importance across a wide range of modalities, recent advances in self-supervised learning (SSL) have been primarily focused on a few well-curated domains, e.g., vision and language, often relying on their domain-specific knowledge. For example, Masked Auto-Encoder (MAE) has become one of the popular architectures in these domains, but less has explored its potential in other modalities. In this paper, we develop MAE as a unified, modality-agnostic SSL framework. In turn, we argue meta-learning as a key to interpreting MAE as a modality-agnostic learner, and propose enhancements to MAE from the motivation to jointly improve its SSL across diverse modalities, coined MetaMAE as a result. Our key idea is to view the mask reconstruction of MAE as a meta-learning task: masked tokens are predicted by adapting the Transformer meta-learner through the amortization of unmasked tokens. Based on this novel interpretation, we propose to integrate two advanced meta-learning techniques. First, we adapt the amortized latent of the Transformer encoder using gradient-based meta-learning to enhance the reconstruction. Then, we maximize the alignment between amortized and adapted latents through task contrastive learning which guides the Transformer encoder to better encode the task-specific knowledge. Our experiment demonstrates the superiority of MetaMAE in the modality-agnostic SSL benchmark (called DABS), significantly outperforming prior baselines.
****

Dual Self-Awareness Value Decomposition Framework without Individual Global Max for Cooperative MARL

Zhiwei Xu, Bin Zhang, dapeng li, Guangchong Zhou, Zeren Zhang, Guoliang Fan

Value decomposition methods have gained popularity in the field of cooperative multi-agent reinforcement learning. However, almost all existing methods follow the principle of Individual Global Max (IGM) or its variants, which limits their problem-solving capabilities. To address this, we propose a dual self-awareness value decomposition framework, inspired by the notion of dual self-awareness in psychology, that entirely rejects the IGM premise. Each agent consists of an ego policy for action selection and an alter ego value function to solve the credit assignment problem. The value function factorization can ignore the IGM assumption by utilizing an explicit search procedure. On the basis of the above, we also suggest a novel anti-ego exploration mechanism to avoid the algorithm becoming stuck in a local optimum. As the first fully IGM-free value decomposition method, our proposed framework achieves desirable performance in various cooperative tasks.
****
DiffAttack: Evasion Attacks Against Diffusion-Based Adversarial Purification
Mintong Kang, Dawn Song, Bo Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Squared Neural Families: A New Class of Tractable Density Models
Russell Tsuchida, Cheng Soon Ong, Dino Sejdinovic
Flexible models for probability distributions are an essential ingredient in many machine learning tasks. We develop and investigate a new class of probability distributions, which we call a Squared Neural Family (SNEFY), formed by squaring the 2-norm of a neural network and normalising it with respect to a base measure. Following the reasoning similar to the well established connections between infinitely wide neural networks and Gaussian processes, we show that SNEFYs admit closed form normalising constants in many cases of interest, thereby resulting in flexible yet fully tractable density models. SNEFYs strictly generalise classical exponential families, are closed under conditioning, and have tractable marginal distributions. Their utility is illustrated on a variety of density estimation, conditional density estimation, and density estimation with missing data tasks.
****
Michelangelo: Conditional 3D Shape Generation based on Shape-Image-Text Aligned Latent Representation
Zibo Zhao, Wen Liu, Xin Chen, Xianfang Zeng, Rui Wang, Pei Cheng, BIN FU, Tao Chen, Gang Yu, Shenghua Gao
We present a novel alignment-before-generation approach to tackle the challenging task of generating general 3D shapes based on 2D images or texts. Directly learning a conditional generative model from images or texts to 3D shapes is prone to producing inconsistent results with the conditions because 3D shapes have an additional dimension whose distribution significantly differs from that of 2D images and texts. To bridge the domain gap among the three modalities and facilitate multi-modal-conditioned 3D shape generation, we explore representing 3D shapes in a shape-image-text-aligned space. Our framework comprises two models: a Shape-Image-Text-Aligned Variational Auto-Encoder (SITA-VAE) and a conditional Aligned Shape Latent Diffusion Model (ASLDM). The former model encodes the 3D shapes into the shape latent space aligned to the image and text and reconstructs the fine-grained 3D neural fields corresponding to given shape embeddings via the transformer-based decoder. The latter model learns a probabilistic mapping function from the image or text space to the latent shape space. Our extensive experiments demonstrate that our proposed approach can generate higher-quality and more diverse 3D shapes that better semantically conform to the visual or textural conditional inputs, validating the effectiveness of the shape-image-text-aligned space for cross-modality 3D shape generation.
****
Estimating Riemannian Metric with Noise-Contaminated Intrinsic Distance

Jiaming Qiu, Xiongtao Dai
We extend metric learning by studying the Riemannian manifold structure of the underlying data space induced by similarity measures between data points. The key quantity of interest here is the Riemannian metric, which characterizes the Riemannian geometry and defines straight lines and derivatives on the manifold. Being able to estimate the Riemannian metric allows us to gain insights into the underlying manifold and compute geometric features such as the geodesic curves. We model the observed similarity measures as noisy responses generated from a function of the intrinsic geodesic distance between data points. A new local regression approach is proposed to learn the Riemannian metric tensor and its derivatives based on a Taylor expansion for the squared geodesic distances, accommodating different types of data such as continuous, binary, or comparative responses. We develop theoretical foundation for our method by deriving the rates of convergence for the asymptotic bias and variance of the estimated metric tensor. The proposed method is shown to be versatile in simulation studies and real data applications involving taxi trip time in New York City and MNIST digits.
****
Inner Product-based Neural Network Similarity
Wei Chen, Zichen Miao, Qiang Qiu
Analyzing representational similarity among neural networks (NNs) is essential for interpreting or transferring deep models. In application scenarios where numerous NN models are learned, it becomes crucial to assess model similarities in computationally efficient ways. In this paper, we propose a new paradigm for reducing NN representational similarity to filter subspace distance. Specifically, when convolutional filters are decomposed as a linear combination of a set of filter subspace elements, denoted as filter atoms, and have those decomposed atom coefficients shared across networks, NN representational similarity can be significantly simplified as calculating the cosine distance among respective filter atoms, to achieve millions of times computation reduction over popular probing-based methods. We provide both theoretical and empirical evidence that such simplified filter subspace-based similarity preserves a strong linear correlation with other popular probing-based metrics, while being significantly more efficient to obtain and robust to probing data. We further validate the effectiveness of the proposed method in various application scenarios where numerous models exist, such as federated and continual learning as well as analyzing training dynamics. We hope our findings can help further explorations of real-time large-scale representational similarity analysis in neural networks.
****
State-space models with layer-wise nonlinearity are universal approximators with exponential decaying memory
Shida Wang, Beichen Xue
State-space models have gained popularity in sequence modelling due to their simple and efficient network structures. However, the absence of nonlinear activation along the temporal direction limits the model's capacity. In this paper, we prove that stacking state-space models with layer-wise nonlinear activation is sufficient to approximate any continuous sequence-to-sequence relationship. Our findings demonstrate that the addition of layer-wise nonlinear activation enhances the model's capacity to learn complex sequence patterns. Meanwhile, it can be seen both theoretically and empirically that the state-space models do not fundamentally resolve the issue of exponential decaying memory. Theoretical results are justified by numerical verifications.
****
Decoding the Enigma: Benchmarking Humans and AIs on the Many Facets of Working Memory
Ankur Sikarwar, Mengmi Zhang
Working memory (WM), a fundamental cognitive process facilitating the temporary storage, integration, manipulation, and retrieval of information, plays a vital role in reasoning and decision-making tasks. Robust benchmark datasets that capture the multifaceted nature of WM are crucial for the effective development and evaluation of AI WM models. Here, we introduce a comprehensive Working Memory (W

orM) benchmark dataset for this purpose. WorM comprises 10 tasks and a total of 1 million trials, assessing 4 functionalities, 3 domains, and 11 behavioral and neural characteristics of WM. We jointly trained and tested state-of-the-art recurrent neural networks and transformers on all these tasks. We also include human behavioral benchmarks as an upper bound for comparison. Our results suggest that AI models replicate some characteristics of WM in the brain, most notably primacy and recency effects, and neural clusters and correlates specialized for different domains and functionalities of WM. In the experiments, we also reveal some limitations in existing models to approximate human behavior. This dataset serves as a valuable resource for communities in cognitive psychology, neuroscience, and AI, offering a standardized framework to compare and enhance WM models, investigate WM's neural underpinnings, and develop WM models with human-like capabilities. Our source code and data are available at: https://github.com/ZhangLab-DeepNeuroCogLab/WorM
****
Many-body Approximation for Non-negative Tensors
KAZU GHALAMKARI, Mahito Sugiyama, Yoshinobu Kawahara
We present an alternative approach to decompose non-negative tensors, called many-body approximation. Traditional decomposition methods assume low-rankness in the representation, resulting in difficulties in global optimization and target rank selection. We avoid these problems by energy-based modeling of tensors, where a tensor and its mode correspond to a probability distribution and a random variable, respectively. Our model can be globally optimized in terms of the KL divergence minimization by taking the interaction between variables (that is, modes), into account that can be tuned more intuitively than ranks. Furthermore, we visualize interactions between modes as tensor networks and reveal a nontrivial relationship between many-body approximation and low-rank approximation. We demonstrate the effectiveness of our approach in tensor completion and approximation.
****
Breaking the Communication-Privacy-Accuracy Tradeoff with f-Differential Privacy
Richeng Jin, Zhonggen Su, caijun zhong, Zhaoyang Zhang, Tony Quek, Huaiyu Dai
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Multi-Prompt Alignment for Multi-Source Unsupervised Domain Adaptation
Haoran Chen, Xintong Han, Zuxuan Wu, Yu-Gang Jiang
Most existing methods for unsupervised domain adaptation (UDA) rely on a shared network to extract domain-invariant features. However, when facing multiple source domains, optimizing such a network involves updating the parameters of the entire network, making it both computationally expensive and challenging, particularly when coupled with min-max objectives. Inspired by recent advances in prompt learning that adapts high-capacity models for downstream tasks in a computationally economic way, we introduce Multi-Prompt Alignment (MPA), a simple yet efficient framework for multi-source UDA. Given a source and target domain pair, MPA first trains an individual prompt to minimize the domain gap through a contrastive loss. Then, MPA denoises the learned prompts through an auto-encoding process and aligns them by maximizing the agreement of all the reconstructed prompts. Moreover, we show that the resulting subspace acquired from the auto-encoding process can easily generalize to a streamlined set of target domains, making our method more efficient for practical usage. Extensive experiments show that MPA achieves state-of-the-art results on three popular datasets with an impressive average accuracy of 54.1% on DomainNet.
****
Characterization and Learning of Causal Graphs with Small Conditioning Sets
Murat Kocaoglu
Constraint-based causal discovery algorithms learn part of the causal graph structure by systematically testing conditional independences observed in the data.

These algorithms, such as the PC algorithm and its variants, rely on graphical characterizations of the so-called equivalence class of causal graphs proposed by Pearl. However, constraint-based causal discovery algorithms struggle when data is limited since conditional independence tests quickly lose their statistical power, especially when the conditioning set is large. To address this, we propose using conditional independence tests where the size of the conditioning set is upper bounded by some integer k for robust causal discovery. The existing graphical characterizations of the equivalence classes of causal graphs are not applicable when we cannot leverage all the conditional independence statements. We first define the notion of k-Markov equivalence: Two causal graphs are k-Markov equivalent if they entail the same conditional independence constraints where the conditioning set size is upper bounded by k. We propose a novel representation that allows us to graphically characterize k-Markov equivalence between two causal graphs. We propose a sound constraint-based algorithm called the k-PC algorithm for learning this equivalence class. Finally, we conduct synthetic, and semi-synthetic experiments to demonstrate that the k-PC algorithm enables more robust causal discovery in the small sample regime compared to the baseline algorithms.
****
Finite Population Regression Adjustment and Non-asymptotic Guarantees for Treatment Effect Estimation
Mehrdad Ghadiri, David Arbour, Tung Mai, Cameron Musco, Anup B. Rao
The design and analysis of randomized experiments is fundamental to many areas, from the physical and social sciences to industrial settings. Regression adjustment is a popular technique to reduce the variance of estimates obtained from experiments, by utilizing information contained in auxiliary covariates. While there is a large literature within the statistics community studying various approaches to regression adjustment and their asymptotic properties, little focus has been given to approaches in the finite population setting with non-asymptotic accuracy bounds. Further, prior work typically assumes that an entire population is exposed to an experiment, whereas practitioners often seek to minimize the number of subjects exposed to an experiment, for ethical and pragmatic reasons.In this work, we study the problems of estimating the sample mean, individual treatment effects, and average treatment effect with regression adjustment. We propose approaches that use techniques from randomized numerical linear algebra to sample a subset of the population on which to perform an experiment. We give non-asymptotic accuracy bounds for our methods and demonstrate that they compare favorably with prior approaches.
****
P-Flow: A Fast and Data-Efficient Zero-Shot TTS through Speech Prompting
Sungwon Kim, Kevin Shih, rohan badlani, Joao Felipe Santos, Evelina Bakhturina, Mikyas Desta, Rafael Valle, Sungroh Yoon, Bryan Catanzaro
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Implicit Bias of Gradient Descent for Logistic Regression at the Edge of Stability
Jingfeng Wu, Vladimir Braverman, Jason D. Lee
Recent research has observed that in machine learning optimization, gradient descent (GD) often operates at the edge of stability (EoS) [Cohen et al., 2021], where the stepsizes are set to be large, resulting in non-monotonic losses induced by the GD iterates. This paper studies the convergence and implicit bias of constant-stepsize GD for logistic regression on linearly separable data in the EoS regime. Despite the presence of local oscillations, we prove that the logistic loss can be minimized by GD with any constant stepsize over a long time scale. Furthermore, we prove that with any constant stepsize, the GD iterates tend to infinity when projected to a max-margin direction (the hard-margin SVM direction) and converge to a fixed vector that minimizes a strongly convex potential when pr

ojected to the orthogonal complement of the max-margin direction. In contrast, we also show that in the EoS regime, GD iterates may diverge catastrophically under the exponential loss, highlighting the superiority of the logistic loss. These theoretical findings are in line with numerical simulations and complement existing theories on the convergence and implicit bias of GD for logistic regression, which are only applicable when the stepsizes are sufficiently small.
****
Two Sides of One Coin: the Limits of Untuned SGD and the Power of Adaptive Methods
Junchi YANG, Xiang Li, Ilyas Fatkhullin, Niao He
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Theoretical Analysis of the Inductive Biases in Deep Convolutional Networks
Zihao Wang, Lei Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Object Reprojection Error (ORE): Camera pose benchmarks from lightweight tracking annotations
Xingyu Chen, Weiyao Wang, Hao Tang, Matt Feiszli
3D spatial understanding is highly valuable in the context of semantic modeling of environments, agents, and their relationships. Semantic modeling approaches employed on monocular video often ingest outputs from off-the-shelf SLAM/SfM pipelines, which are anecdotally observed to perform poorly or fail completely on some fraction of the videos of interest. These target videos may vary widely in complexity of scenes, activities, camera trajectory, etc. Unfortunately, such semantically-rich video data often comes with no ground-truth 3D information, and in practice it is prohibitively costly or impossible to obtain ground truth reconstructions or camera pose post-hoc. This paper proposes a novel evaluation protocol, Object Reprojection Error (ORE) to benchmark camera trajectories; ORE computes reprojection error for static objects within the video and requires only lightweight object tracklet annotations. These annotations are easy to gather on new or existing video, enabling ORE to be calculated on essentially arbitrary datasets. We show that ORE maintains high rank correlation with standard metrics based on groundtruth. Leveraging ORE, we source videos and annotations from Ego4D-EgoTracks, resulting in EgoStatic, a large-scale diverse dataset for evaluating camera trajectories in-the-wild.
****
Rethinking Bias Mitigation: Fairer Architectures Make for Fairer Face Recognition
Samuel Dooley, Rhea Sukthanker, John Dickerson, Colin White, Frank Hutter, Micah Goldblum
Face recognition systems are widely deployed in safety-critical applications, including law enforcement, yet they exhibit bias across a range of socio-demographic dimensions, such as gender and race. Conventional wisdom dictates that model biases arise from biased training data. As a consequence, previous works on bias mitigation largely focused on pre-processing the training data, adding penalties to prevent bias from effecting the model during training, or post-processing predictions to debias them, yet these approaches have shown limited success on hard problems such as face recognition. In our work, we discover that biases are actually inherent to neural network architectures themselves. Following this reframing, we conduct the first neural architecture search for fairness, jointly with a search for hyperparameters. Our search outputs a suite of models which Pareto-dominate all other high-performance architectures and existing bias mitigation methods in terms of accuracy and fairness, often by large margins, on the t

wo most widely used datasets for face identification, CelebA and VGGFace2. Furthermore, these models generalize to other datasets and sensitive attributes. We release our code, models and raw data files at https://github.com/dooleys/FR-NAS.
****
H-InDex: Visual Reinforcement Learning with Hand-Informed Representations for Dexterous Manipulation
Yanjie Ze, Yuyao Liu, Ruizhe Shi, Jiaxin Qin, Zhecheng Yuan, Jiashun Wang, Huazhe Xu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
DynGFN: Towards Bayesian Inference of Gene Regulatory Networks with GFlowNets
Lazar Atanackovic, Alexander Tong, Bo Wang, Leo J Lee, Yoshua Bengio, Jason S. Hartford
One of the grand challenges of cell biology is inferring the gene regulatory network (GRN) which describes interactions between genes and their products that control gene expression and cellular function. We can treat this as a causal discovery problem but with two non-standard challenges: (1) regulatory networks are inherently cyclic so we should not model a GRN as a directed acyclic graph (DAG), and (2) observations have significant measurement noise so for typical sample sizes, there will always be a large equivalence class of graphs that are likely given the data, and we want methods that capture this uncertainty. Existing methods either focus on challenge (1), identifying cyclic structure from dynamics, or on challenge (2) learning complex Bayesian posteriors over directed acyclic graphs, but not both. In this paper we leverage the fact that it is possible to estimate the ``velocity'' of the expression of a gene with RNA velocity techniques to develop an approach that addresses both challenges. Because we have access to velocity information, we can treat the Bayesian structure learning problem as a problem of sparse identification of a dynamical system, capturing cyclic feedback loops through time. We leverage Generative Flow Networks (GFlowNets) to estimate the posterior distribution over the combinatorial space of possible sparse dependencies. Our results indicate that our method learns posteriors that better encapsulate the distributions of cyclic structures compared to counterpart state-of-the-art Bayesian structure learning approaches.
****
RaLEs: a Benchmark for Radiology Language Evaluations
Juanma Zambrano Chaves, Nandita Bhaskhar, Maayane Attias, Jean-Benoit Delbrouck, Daniel Rubin, Andreas Loening, Curtis Langlotz, Akshay Chaudhari
The radiology report is the main form of communication between radiologists and other clinicians. Prior work in natural language processing in radiology reports has shown the value of developing methods tailored for individual tasks such as identifying reports with critical results or disease detection. Meanwhile, English and biomedical natural language understanding benchmarks such as the General Language Understanding and Evaluation as well as Biomedical Language Understanding and Reasoning Benchmark have motivated the development of models that can be easily adapted to address many tasks in those domains. Here, we characterize the radiology report as a distinct domain and introduce RaLEs, the Radiology Language Evaluations, as a benchmark for natural language understanding and generation in radiology. RaLEs is comprised of seven natural language understanding and generation evaluations including the extraction of anatomical and disease entities and their relations, procedure selection, and report summarization. We characterize the performance of models designed for the general, biomedical, clinical and radiology domains across these tasks. We find that advances in the general and biomedical domains do not necessarily translate to radiology, and that improved models from the general domain can perform comparably to smaller clinical-specific models. The limited performance of existing pre-trained models on RaLEs highlights the opportunity to improve domain-specific self-supervised models for natural language processing in radiology. We propose RaLEs as a benchmark to promo

te and track the development of such domain-specific radiology language models.
****
AutoGO: Automated Computation Graph Optimization for Neural Network Evolution
Mohammad Salameh, Keith Mills, Negar Hassanpour, Fred Han, Shuting Zhang, Wei Lu, Shangling Jui, CHUNHUA ZHOU, Fengyu Sun, Di Niu
Optimizing Deep Neural Networks (DNNs) to obtain high-quality models for efficient real-world deployment has posed multi-faceted challenges to machine learning engineers. Existing methods either search for neural architectures in heuristic design spaces or apply low-level adjustments to computation primitives to improve inference efficiency on hardware. We present Automated Graph Optimization (AutoGO), a framework to evolve neural networks in a low-level Computation Graph (CG) of primitive operations to improve both its performance and hardware friendliness. Through a tokenization scheme, AutoGO performs variable-sized segment mutations, making both primitive changes and larger-grained changes to CGs. We introduce our segmentation and mutation algorithms, efficient frequent segment mining technique, as well as a pretrained context-aware predictor to estimate the impact of segment replacements. Extensive experimental results show that AutoGO can automatically evolve several typical large convolutional networks to achieve significant task performance improvement and FLOPs reduction on a range of CV tasks, ranging from Classification, Semantic Segmentation, Human Pose Estimation, to Super Resolution, yet without introducing any newer primitive operations. We also demonstrate the lightweight deployment results of AutoGO-optimized super-resolution and denoising U-Nets on a cycle simulator for a Neural Processing Unit (NPU), achieving PSNR improvement and latency/power reduction simultaneously. Code available at https://github.com/Ascend-Research/AutoGO.
****
Where Did I Come From? Origin Attribution of AI-Generated Images
Zhenting Wang, Chen Chen, Yi Zeng, Lingjuan Lyu, Shiqing Ma
Image generation techniques have been gaining increasing attention recently, but concerns have been raised about the potential misuse and intellectual property (IP) infringement associated with image generation models. It is, therefore, necessary to analyze the origin of images by inferring if a specific image was generated by a particular model, i.e., origin attribution. Existing methods only focus on specific types of generative models and require additional procedures during the training phase or generation phase. This makes them unsuitable for pre-trained models that lack these specific operations and may impair generation quality. To address this problem, we first develop an alteration-free and model-agnostic origin attribution method via reverse-engineering on image generation models, i.e., inverting the input of a particular model for a specific image. Given a particular model, we first analyze the differences in the hardness of reverse-engineering tasks for generated samples of the given model and other images. Based on our analysis, we then propose a method that utilizes the reconstruction loss of reverse-engineering to infer the origin. Our proposed method effectively distinguishes between generated images of a specific generative model and other images, i.e., images generated by other models and real images.
****
Robust Data Pruning under Label Noise via Maximizing Re-labeling Accuracy
Dongmin Park, Seola Choi, Doyoung Kim, Hwanjun Song, Jae-Gil Lee
Data pruning, which aims to downsize a large training set into a small informative subset, is crucial for reducing the enormous computational costs of modern deep learning. Though large-scale data collections invariably contain annotation noise and numerous robust learning methods have been developed, data pruning for the noise-robust learning scenario has received little attention. With state-of-the-art Re-labeling methods that self-correct erroneous labels while training, it is challenging to identify which subset induces the most accurate re-labeling of erroneous labels in the entire training set. In this paper, we formalize the problem of data pruning with re-labeling. We first show that the likelihood of a training example being correctly re-labeled is proportional to the prediction confidence of its neighborhood in the subset. Therefore, we propose a novel data pruning algorithm, Prune4Rel, that finds a subset maximizing the total neighborh

ood confidence of all training examples, thereby maximizing the re-labeling accuracy and generalization performance. Extensive experiments on four real and one synthetic noisy datasets show that Prune4Rel outperforms the baselines with Re-labeling models by up to 9.1% as well as those with a standard model by up to 21.6%.
****
WildfireSpreadTS: A dataset of multi-modal time series for wildfire spread prediction
Sebastian Gerard, Yu Zhao, Josephine Sullivan
We present a multi-temporal, multi-modal remote-sensing dataset for predicting how active wildfires will spread at a resolution of 24 hours. The dataset consists of 13607 images across 607 fire events in the United States from January 2018 to October 2021. For each fire event, the dataset contains a full time series of daily observations, containing detected active fires and variables related to fuel, topography and weather conditions. The dataset is challenging due to: a) its inputs being multi-temporal, b) the high number of 23 multi-modal input channels, c) highly imbalanced labels and d) noisy labels, due to smoke, clouds, and inaccuracies in the active fire detection. The underlying complexity of the physical processes adds to these challenges. Compared to existing public datasets in this area, WildfireSpreadTS allows for multi-temporal modeling of spreading wildfires, due to its time series structure. Furthermore, we provide additional input modalities and a high spatial resolution of 375m for the active fire maps. We publish this dataset to encourage further research on this important task with multi-temporal, noise-resistant or generative methods, uncertainty estimation or advanced optimization techniques that deal with the high-dimensional input space.
****
Augmenting Language Models with Long-Term Memory
Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, Furu Wei
Existing large language models (LLMs) can only afford fix-sized inputs due to the input length limit, preventing them from utilizing rich long-context information from past inputs. To address this, we propose a framework, Language Models Augmented with Long-Term Memory (LongMem), which enables LLMs to memorize long history. We design a novel decoupled network architecture with the original backbone LLM frozen as a memory encoder and an adaptive residual side-network as a memory retriever and reader. Such a decoupled memory design can easily cache and update long-term past contexts for memory retrieval without suffering from memory staleness. Enhanced with memory-augmented adaptation training, LongMem can thus memorize long past context and use long-term memory for language modeling. The proposed memory retrieval module can handle unlimited-length context in its memory bank to benefit various downstream tasks. Typically, LongMem can enlarge the long-form memory to 65k tokens and thus cache many-shot extra demonstration examples as long-form memory for in-context learning. Experiments show that our method outperforms strong long-context models on ChapterBreak, a challenging long-context modeling benchmark, and achieves remarkable improvements on memory-augmented in-context learning over LLMs. The results demonstrate that the proposed method is effective in helping language models to memorize and utilize long-form contents.
****
Expressivity-Preserving GNN Simulation
Fabian Jogl, Maximilian Thiessen, Thomas Gärtner
We systematically investigate graph transformations that enable standard message passing to simulate state-of-the-art graph neural networks (GNNs) without loss of expressivity. Using these, many state-of-the-art GNNs can be implemented with message passing operations from standard libraries, eliminating many sources of implementation issues and allowing for better code optimization. We distinguish between weak and strong simulation: weak simulation achieves the same expressivity only after several message passing steps while strong simulation achieves this after every message passing step. Our contribution leads to a direct way to t

ranslate common operations of non-standard GNNs to graph transformations that al low for strong or weak simulation. Our empirical evaluation shows competitive pr edictive performance of message passing on transformed graphs for various molecu lar benchmark datasets, in several cases surpassing the original GNNs.
****
Rethinking Incentives in Recommender Systems: Are Monotone Rewards Always Benefi cial?
Fan Yao, Chuanhao Li, Karthik Abinav Sankararaman, Yiming Liao, Yan Zhu, Qifan W ang, Hongning Wang, Haifeng Xu
The past decade has witnessed the flourishing of a new profession as media conte nt creators, who rely on revenue streams from online content recommendation plat forms. The reward mechanism employed by these platforms creates a competitive en vironment among creators which affects their production choices and, consequentl y, content distribution and system welfare. It is thus crucial to design the pla tform's reward mechanism in order to steer the creators' competition towards a d esirable welfare outcome in the long run. This work makes two major contribution s in this regard: first, we uncover a fundamental limit about a class of widely adopted mechanisms, coined \emph{Merit-based Monotone Mechanisms}, by showing th at they inevitably lead to a constant fraction loss of the optimal welfare. To c ircumvent this limitation, we introduce \emph{Backward Rewarding Mechanisms} (BR Ms) and show that the competition game resultant from BRMs possesses a potential game structure. BRMs thus naturally induce strategic creators' collective behav iors towards optimizing the potential function, which can be designed to match a ny given welfare metric. In addition, the class of BRM can be parameterized so t hat it allows the platform to directly optimize welfare within the feasible mech anism space even when the welfare metric is not explicitly defined.
****
Gaussian Partial Information Decomposition: Bias Correction and Application to H igh-dimensional Data
Praveen Venkatesh, Corbett Bennett, Sam Gale, Tamina Ramirez, Greggory Heller, S everine Durand, Shawn Olsen, Stefan Mihalas
Recent advances in neuroscientific experimental techniques have enabled us to si multaneously record the activity of thousands of neurons across multiple brain r egions. This has led to a growing need for computational tools capable of analyz ing how task-relevant information is represented and communicated between severa l brain regions. Partial information decompositions (PIDs) have emerged as one s uch tool, quantifying how much unique, redundant and synergistic information two or more brain regions carry about a task-relevant message. However, computing P IDs is computationally challenging in practice, and statistical issues such as t he bias and variance of estimates remain largely unexplored. In this paper, we p ropose a new method for efficiently computing and estimating a PID definition on multivariate Gaussian distributions. We show empirically that our method satisf ies an intuitive additivity property, and recovers the ground truth in a battery of canonical examples, even at high dimensionality. We also propose and evaluat e, for the first time, a method to correct the bias in PID estimates at finite s ample sizes. Finally, we demonstrate that our Gaussian PID effectively character izes inter-areal interactions in the mouse brain, revealing higher redundancy be tween visual areas when a stimulus is behaviorally relevant.
****
Unconstrained Dynamic Regret via Sparse Coding
Zhiyu Zhang, Ashok Cutkosky, Yannis Paschalidis
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Efficient Test-Time Adaptation for Super-Resolution with Second-Order Degradatio n and Reconstruction
Zeshuai Deng, Zhuokun Chen, Shuaicheng Niu, Thomas Li, Bohan Zhuang, Mingkui Tan
Image super-resolution (SR) aims to learn a mapping from low-resolution (LR) to

high-resolution (HR) using paired HR-LR training images. Conventional SR methods typically gather the paired training data by synthesizing LR images from HR images using a predetermined degradation model, e.g., Bicubic down-sampling. However, the realistic degradation type of test images may mismatch with the training-time degradation type due to the dynamic changes of the real-world scenarios, resulting in inferior-quality SR images. To address this, existing methods attempt to estimate the degradation model and train an image-specific model, which, however, is quite time-consuming and impracticable to handle rapidly changing domain shifts. Moreover, these methods largely concentrate on the estimation of one degradation type (e.g., blur degradation), overlooking other degradation types like noise and JPEG in real-world test-time scenarios, thus limiting their practicality. To tackle these problems, we present an efficient test-time adaptation framework for SR, named SRTTA, which is able to quickly adapt SR models to test domains with different/unknown degradation types. Specifically, we design a second-order degradation scheme to construct paired data based on the degradation type of the test image, which is predicted by a pre-trained degradation classifier. Then, we adapt the SR model by implementing feature-level reconstruction learning from the initial test image to its second-order degraded counterparts, which helps the SR model generate plausible HR images. Extensive experiments are conducted on newly synthesized corrupted DIV2K datasets with 8 different degradations and several real-world datasets, demonstrating that our SRTTA framework achieves an impressive improvement over existing methods with satisfying speed. The source code is available at https://github.com/DengZeshuai/SRTTA.
****
Replicability in Reinforcement Learning
Amin Karbasi, Grigoris Velegkas, Lin Yang, Felix Zhou
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Probabilistic Invariant Learning with Randomized Linear Classifiers
Leonardo Cotta, Gal Yehuda, Assaf Schuster, Chris J. Maddison
Designing models that are both expressive and preserve known invariances of tasks is an increasingly hard problem. Existing solutions tradeoff invariance for computational or memory resources. In this work, we show how to leverage randomness and design models that are both expressive and invariant but use less resources. Inspired by randomized algorithms, our key insight is that accepting probabilistic notions of universal approximation and invariance can reduce our resource requirements. More specifically, we propose a class of binary classification models called Randomized Linear Classifiers (RLCs). We give parameter and sample size conditions in which RLCs can, with high probability, approximate any (smooth) function while preserving invariance to compact group transformations. Leveraging this result, we design three RLCs that are provably probabilistic invariant for classification tasks over sets, graphs, and spherical data. We show how these models can achieve probabilistic invariance and universality using less resources than (deterministic) neural networks and their invariant counterparts. Finally, we empirically demonstrate the benefits of this new class of models on invariant tasks where deterministic invariant neural networks are known to struggle.
****
FedNAR: Federated Optimization with Normalized Annealing Regularization
Junbo Li, Ang Li, Chong Tian, Qirong Ho, Eric Xing, Hongyi Wang
Weight decay is a standard technique to improve generalization performance in modern deep neural network optimization, and is also widely adopted in federated learning (FL) to prevent overfitting in local clients. In this paper, we first explore the choices of weight decay and identify that weight decay value appreciably influences the convergence of existing FL algorithms. While preventing overfitting is crucial, weight decay can introduce a different optimization goal towards the global objective, which is further amplified in FL due to multiple local updates and heterogeneous data distribution.To address this challenge, we develo

p {\it Federated optimization with Normalized Annealing Regularization} (FedNAR), a simple yet effective and versatile algorithmic plug-in that can be seamlessly integrated into any existing FL algorithms. Essentially, we regulate the magnitude of each update by performing co-clipping of the gradient and weight decay.We provide a comprehensive theoretical analysis of FedNAR's convergence rate and conduct extensive experiments on both vision and language datasets with different backbone federated optimization algorithms. Our experimental results consistently demonstrate that incorporating FedNAR into existing FL algorithms leads to accelerated convergence and heightened model accuracy. Moreover, FedNAR exhibits resilience in the face of various hyperparameter configurations. Specifically, FedNAR has the ability to self-adjust the weight decay when the initial specification is not optimal, while the accuracy of traditional FL algorithms would markedly decline. Our codes are released at \href{https://anonymous.4open.science/r/fednar-BE8F}{https://anonymous.4open.science/r/fednar-BE8F}.
****
How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources
Yizhong Wang, Hamish Ivison, Pradeep Dasigi, Jack Hessel, Tushar Khot, Khyathi Chandu, David Wadden, Kelsey MacMillan, Noah A. Smith, Iz Beltagy, Hannaneh Hajishirzi
In this work we explore recent advances in instruction-tuning language models on a range of open instruction-following datasets. Despite recent claims that open models can be on par with state-of-the-art proprietary models, these claims are often accompanied by limited evaluation, making it difficult to compare models across the board and determine the utility of various resources. We provide a large set of instruction-tuned models from 6.7B to 65B parameters in size, trained on 12 instruction datasets ranging from manually curated (e.g., OpenAssistant) to synthetic and distilled (e.g., Alpaca) and systematically evaluate them on their factual knowledge, reasoning, multilinguality, coding, safety, and open-ended instruction following abilities through a collection of automatic, model-based, and human-based metrics. We further introduce Tülu, our best performing instruction-tuned model suite finetuned on a combination of high-quality open resources.Our experiments show that different instruction-tuning datasets can uncover or enhance specific skills, while no single dataset (or combination) provides the best performance across all evaluations. Interestingly, we find that model and human preference-based evaluations fail to reflect differences in model capabilities exposed by benchmark-based evaluations, suggesting the need for the type of systemic evaluation performed in this work. Our evaluations show that the best model in any given evaluation reaches on average 87% of ChatGPT performance, and 73% of GPT-4 performance, suggesting that further investment in building better base models and instruction-tuning data is required to close the gap. We release our instruction-tuned models, including a fully finetuned 65B Tülu, along with our code, data, and evaluation framework to facilitate future research.
****
Trial matching: capturing variability with data-constrained spiking neural networks
Christos Sourmpis, Carl Petersen, Wulfram Gerstner, Guillaume Bellec
Simultaneous behavioral and electrophysiological recordings call for new methods to reveal the interactions between neural activity and behavior. A milestone would be an interpretable model of the co-variability of spiking activity and behavior across trials. Here, we model a mouse cortical sensory-motor pathway in a tactile detection task reported by licking with a large recurrent spiking neural network (RSNN), fitted to the recordings via gradient-based optimization. We focus specifically on the difficulty to match the trial-to-trial variability in the data. Our solution relies on optimal transport to define a distance between the distributions of generated and recorded trials. The technique is applied to artificial data and neural recordings covering six cortical areas. We find that the resulting RSNN can generate realistic cortical activity and predict jaw movements across the main modes of trial-to-trial variability. Our analysis also identifies an unexpected mode of variability in the data corresponding to task-irrelev

ant movements of the mouse.
****
Decentralized Randomly Distributed Multi-agent Multi-armed Bandit with Heterogeneous Rewards
Mengfan Xu, Diego Klabjan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
(Amplified) Banded Matrix Factorization: A unified approach to private training
Christopher A. Choquette-Choo, Arun Ganesh, Ryan McKenna, H. Brendan McMahan, John Rush, Abhradeep Guha Thakurta, Zheng Xu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Anchor Data Augmentation
Nora Schneider, Shirin Goshtasbpour, Fernando Perez-Cruz
We propose a novel algorithm for data augmentation in nonlinear over-parametrized regression. Our data augmentation algorithm borrows from the literature on causality. Contrary to the current state-of-the-art solutions that rely on modifications of Mixup algorithm, we extend the recently proposed distributionally robust Anchor regression (AR) method for data augmentation. Our Anchor Data Augmentation (ADA) uses several replicas of the modified samples in AR to provide more training examples, leading to more robust regression predictions. We apply ADA to linear and nonlinear regression problems using neural networks. ADA is competitive with state-of-the-art C-Mixup solutions.
****
The noise level in linear regression with dependent data
Ingvar Ziemann, Stephen Tu, George J. Pappas, Nikolai Matni
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SafeDICE: Offline Safe Imitation Learning with Non-Preferred Demonstrations
Youngsoo Jang, Geon-Hyeong Kim, Jongmin Lee, Sungryull Sohn, Byoungjip Kim, Honglak Lee, Moontae Lee
We consider offline safe imitation learning (IL), where the agent aims to learn the safe policy that mimics preferred behavior while avoiding non-preferred behavior from non-preferred demonstrations and unlabeled demonstrations. This problem setting corresponds to various real-world scenarios, where satisfying safety constraints is more important than maximizing the expected return. However, it is very challenging to learn the policy to avoid constraint-violating (i.e. non-preferred) behavior, as opposed to standard imitation learning which learns the policy to mimic given demonstrations. In this paper, we present a hyperparameter-free offline safe IL algorithm, SafeDICE, that learns safe policy by leveraging the non-preferred demonstrations in the space of stationary distributions. Our algorithm directly estimates the stationary distribution corrections of the policy that imitate the demonstrations excluding the non-preferred behavior. In the experiments, we demonstrate that our algorithm learns a more safe policy that satisfies the cost constraint without degrading the reward performance, compared to baseline algorithms.
****
Language Models Don't Always Say What They Think: Unfaithful Explanations in Chain-of-Thought Prompting
Miles Turpin, Julian Michael, Ethan Perez, Samuel Bowman
Large Language Models (LLMs) can achieve strong performance on many tasks by pro

ducing step-by-step reasoning before giving a final output, often referred to as chain-of-thought reasoning (CoT). It is tempting to interpret these CoT explanations as the LLM's process for solving a task. This level of transparency into LLMs' predictions would yield significant safety benefits. However, we find that CoT explanations can systematically misrepresent the true reason for a model's prediction. We demonstrate that CoT explanations can be heavily influenced by adding biasing features to model inputs—e.g., by reordering the multiple-choice options in a few-shot prompt to make the answer always "(A)"—which models systematically fail to mention in their explanations. When we bias models toward incorrect answers, they frequently generate CoT explanations rationalizing those answers. This causes accuracy to drop by as much as 36% on a suite of 13 tasks from BIG-Bench Hard, when testing with GPT-3.5 from OpenAI and Claude 1.0 from Anthropic. On a social-bias task, model explanations justify giving answers in line with stereotypes without mentioning the influence of these social biases. Our findings indicate that CoT explanations can be plausible yet misleading, which risks increasing our trust in LLMs without guaranteeing their safety. Building more transparent and explainable systems will require either improving CoT faithfulness through targeted efforts or abandoning CoT in favor of alternative methods.
****

Static and Sequential Malicious Attacks in the Context of Selective Forgetting
Chenxu Zhao, Wei Qian, Rex Ying, Mengdi Huai
With the growing demand for the right to be forgotten, there is an increasing need for machine learning models to forget sensitive data and its impact. To address this, the paradigm of selective forgetting (a.k.a machine unlearning) has been extensively studied, which aims to remove the impact of requested data from a well-trained model without retraining from scratch. Despite its significant success, limited attention has been given to the security vulnerabilities of the unlearning system concerning malicious data update requests. Motivated by this, in this paper, we explore the possibility and feasibility of malicious data update requests during the unlearning process. Specifically, we first propose a new class of malicious selective forgetting attacks, which involves a static scenario where all the malicious data update requests are provided by the adversary at once. Additionally, considering the sequential setting where the data update requests arrive sequentially, we also design a novel framework for sequential forgetting attacks, which is formulated as a stochastic optimal control problem. We also propose novel optimization algorithms that can find the effective malicious data update requests. We perform theoretical analyses for the proposed selective forgetting attacks, and extensive experimental results validate the effectiveness of our proposed selective forgetting attacks. The source code is available in the supplementary material.
****

Language-based Action Concept Spaces Improve Video Self-Supervised Learning
Kanchana Ranasinghe, Michael S Ryoo
Recent contrastive language image pre-training has led to learning highly transferable and robust image representations. However, adapting these models to video domain with minimal supervision remains an open problem. We explore a simple step in that direction, using language tied self-supervised learning to adapt an image CLIP model to the video domain. A backbone modified for temporal modeling is trained under self-distillation settings with train objectives operating in an action concept space. Feature vectors of various action concepts extracted from a language encoder using relevant textual prompts construct this space. A large language model aware of actions and their attributes generates the relevant textual prompts.We introduce two train objectives, concept distillation and concept alignment, that retain generality of original representations while enforcing relations between actions and their attributes. Our approach improves zero-shot and linear probing performance on three action recognition benchmarks.
****

TRIAGE: Characterizing and auditing training data for improved regression
Nabeel Seedat, Jonathan Crabbé, Zhaozhi Qian, Mihaela van der Schaar
Data quality is crucial for robust machine learning algorithms, with the recent

interest in data-centric AI emphasizing the importance of training data characte rization. However, current data characterization methods are largely focused on classification settings, with regression settings largely understudied. To addre ss this, we introduce TRIAGE, a novel data characterization framework tailored t o regression tasks and compatible with a broad class of regressors. TRIAGE utili zes conformal predictive distributions to provide a model-agnostic scoring metho d, the TRIAGE score. We operationalize the score to analyze individual samples' training dynamics and characterize samples as under-, over-, or well-estimated b y the model. We show that TRIAGE's characterization is consistent and highlight its utility to improve performance via data sculpting/filtering, in multiple reg ression settings. Additionally, beyond sample level, we show TRIAGE enables new approaches to dataset selection and feature acquisition. Overall, TRIAGE highlig hts the value unlocked by data characterization in real-world regression applica tions.
****
ClimateLearn: Benchmarking Machine Learning for Weather and Climate Modeling
Tung Nguyen, Jason Jewik, Hritik Bansal, Prakhar Sharma, Aditya Grover
Modeling weather and climate is an essential endeavor to understand the near- an d long-term impacts of climate change, as well as to inform technology and polic ymaking for adaptation and mitigation efforts. In recent years, there has been a surging interest in applying data-driven methods based on machine learning for solving core problems such as weather forecasting and climate downscaling. Despi te promising results, much of this progress has been impaired due to the lack of large-scale, open-source efforts for reproducibility, resulting in the use of i nconsistent or underspecified datasets, training setups, and evaluations by both domain scientists and artificial intelligence researchers. We introduce Climate Learn, an open-source PyTorch library that vastly simplifies the training and ev aluation of machine learning models for data-driven climate science. ClimateLear n consists of holistic pipelines for dataset processing (e.g., ERA5, CMIP6, PRIS M), implementing state-of-the-art deep learning models (e.g., Transformers, ResN ets), and quantitative and qualitative evaluation for standard weather and clima te modeling tasks. We supplement these functionalities with extensive documentat ion, contribution guides, and quickstart tutorials to expand access and promote community growth. We have also performed comprehensive forecasting and downscali ng experiments to showcase the capabilities and key features of our library. To our knowledge, ClimateLearn is the first large-scale, open-source effort for bri dging research in weather and climate modeling with modern machine learning syst ems. Our library is available publicly at https://github.com/aditya-grover/clima te-learn.
****
Advice Querying under Budget Constraint for Online Algorithms
Ziyad Benomar, Vianney Perchet
Several problems have been extensively studied in the learning-augmented setting , where the algorithm has access to some, possibly incorrect, predictions. Howev er, it is assumed in most works that the predictions are provided to the algorit hm as input, with no constraint on their size. In this paper, we consider algori thms with access to a limited number of predictions, that they can request at an y time during their execution. We study three classical problems in competitive analysis, the ski rental problem, the secretary problem, and the non-clairvoyant job scheduling. We address the question of when to query predictions and how to use them.
****
DIFFER:Decomposing Individual Reward for Fair Experience Replay in Multi-Agent R einforcement Learning
Xunhan Hu, Jian Zhao, Wengang Zhou, Ruili Feng, Houqiang Li
Cooperative multi-agent reinforcement learning (MARL) is a challenging task, as agents must learn complex and diverse individual strategies from a shared team r eward. However, existing methods struggle to distinguish and exploit important i ndividual experiences, as they lack an effective way to decompose the team rewar d into individual rewards. To address this challenge, we propose DIFFER, a power

ful theoretical framework for decomposing individual rewards to enable fair experience replay in MARL.By enforcing the invariance of network gradients, we establish a partial differential equation whose solution yields the underlying individual reward function. The individual TD-error can then be computed from the solved closed-form individual rewards, indicating the importance of each piece of experience in the learning task and guiding the training process. Our method elegantly achieves an equivalence to the original learning framework when individual experiences are homogeneous, while also adapting to achieve more muscular efficiency and fairness when diversity is observed.Our extensive experiments on popular benchmarks validate the effectiveness of our theory and method, demonstrating significant improvements in learning efficiency and fairness. Code is available in supplement material.
****
Quantizable Transformers: Removing Outliers by Helping Attention Heads Do Nothing
Yelysei Bondarenko, Markus Nagel, Tijmen Blankevoort
Transformer models have been widely adopted in various domains over the last years and especially large language models have advanced the field of AI significantly. Due to their size, the capability of these networks has increased tremendously, but this has come at the cost of a significant increase in necessary compute. Quantization is one of the most effective ways for reducing the computational time and memory consumption of neural networks. Many studies have shown, however, that modern transformer models tend to learn strong outliers in their activations, making them difficult to quantize. To retain acceptable performance, the existence of these outliers requires activations to be in higher-bitwidth or the use of different numeric formats, extra fine-tuning, or other workarounds. We show that strong outliers are related to very specific behavior of attention heads that try to learn a "no-op", or just a partial update of the residual. To achieve the exact zeros needed in the attention matrix for a no-update, the input to the softmax is pushed to be larger and larger during training, causing outliers in other parts of the network. Based on these observations, we propose two simple (independent) modifications to the attention mechanism - clipped softmax and gated attention. We empirically show that models pre-trained using our methods learn significantly smaller outliers while maintaining and sometimes even improving the floating-point task performance. This enables us to quantize transformers to full INT8 quantization of the activations without any additional effort. We demonstrate the effectiveness of our methods on both language models (BERT, OPT) and vision transformers.
****
Adversarial Training from Mean Field Perspective
Soichiro Kumano, Hiroshi Kera, Toshihiko Yamasaki
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
MMD-Fuse: Learning and Combining Kernels for Two-Sample Testing Without Data Splitting
Felix Biggs, Antonin Schrab, Arthur Gretton
We propose novel statistics which maximise the power of a two-sample test based on the Maximum Mean Discrepancy (MMD), byadapting over the set of kernels used in defining it.For finite sets, this reduces to combining (normalised) MMD values under each of these kernels via a weighted soft maximum.Exponential concentration bounds are proved for our proposed statistics under the null and alternative.We further show how these kernels can be chosen in a data-dependent but permutation-independent way, in a well-calibrated test, avoiding data splitting.This technique applies more broadly to general permutation-based MMD testing, and includes the use of deep kernels with features learnt using unsupervised models such as auto-encoders.We highlight the applicability of our MMD-Fuse tests on both synthetic low-dimensional and real-world high-dimensional data, and compare its pe

rformance in terms of power against current state-of-the-art kernel tests.
****
DAC-DETR: Divide the Attention Layers and Conquer
Zhengdong Hu, Yifan Sun, Jingdong Wang, Yi Yang
This paper reveals a characteristic of DEtection Transformer (DETR) that negatively impacts its training efficacy, i.e., the cross-attention and self-attention layers in DETR decoder have contrary impacts on the object queries (though both impacts are important). Specifically, we observe the cross-attention tends to gather multiple queries around the same object, while the self-attention disperses these queries far away. To improve the training efficacy, we propose a Divide-And-Conquer DETR (DAC-DETR) that divides the cross-attention out from this contrary for better conquering. During training, DAC-DETR employs an auxiliary decoder that focuses on learning the cross-attention layers. The auxiliary decoder, while sharing all the other parameters, has NO self-attention layers and employs one-to-many label assignment to improve the gathering effect. Experiments show that DAC-DETR brings remarkable improvement over popular DETRs. For example, under the 12 epochs training scheme on MS-COCO, DAC-DETR improves Deformable DETR (ResNet-50) by +3.4 AP and achieves 50.9 (ResNet-50) / 58.1 AP (Swin-Large) based on some popular methods (i.e., DINO and an IoU-related loss). Our code will be made available at https://github.com/huzhengdongcs/DAC-DETR.
****
Streaming Algorithms and Lower Bounds for Estimating Correlation Clustering Cost
Sepehr Assadi, Vihan Shah, Chen Wang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Episodic Multi-Task Learning with Heterogeneous Neural Processes
Jiayi Shen, Xiantong Zhen, Qi Wang, Marcel Worring
This paper focuses on the data-insufficiency problem in multi-task learning within an episodic training setup. Specifically, we explore the potential of heterogeneous information across tasks and meta-knowledge among episodes to effectively tackle each task with limited data. Existing meta-learning methods often fail to take advantage of crucial heterogeneous information in a single episode, while multi-task learning models neglect reusing experience from earlier episodes. To address the problem of insufficient data, we develop Heterogeneous Neural Processes (HNPs) for the episodic multi-task setup. Within the framework of hierarchical Bayes, HNPs effectively capitalize on prior experiences as meta-knowledge and capture task-relatedness among heterogeneous tasks, mitigating data-insufficiency. Meanwhile, transformer-structured inference modules are designed to enable efficient inferences toward meta-knowledge and task-relatedness. In this way, HNPs can learn more powerful functional priors for adapting to novel heterogeneous tasks in each meta-test episode. Experimental results show the superior performance of the proposed HNPs over typical baselines, and ablation studies verify the effectiveness of the designed inference modules.
****
Knowledge Distillation Performs Partial Variance Reduction
Mher Safaryan, Alexandra Peste, Dan Alistarh
Knowledge distillation is a popular approach for enhancing the performance of "student" models, with lower representational capacity, by taking advantage of more powerful "teacher" models. Despite its apparent simplicity, the underlying mechanics behind knowledge distillation (KD) are not yet fully understood. In this work, we shed new light on the inner workings of this method, by examining it from an optimization perspective. Specifically, we show that, in the context of linear and deep linear models, KD can be interpreted as a novel type of stochastic variance reduction mechanism. We provide a detailed convergence analysis of the resulting dynamics, which hold under standard assumptions for both strongly-convex and non-convex losses, showing that KD acts as a form of \emph{partial variance reduction}, which can reduce the stochastic gradient noise, but may not elim

inate it completely, depending on the properties of the ``teacher'' model. Our analysis puts further emphasis on the need for careful parametrization of KD, in particular w.r.t. the weighting of the distillation loss, and is validated empirically on both linear models and deep neural networks.
****
Jaccard Metric Losses: Optimizing the Jaccard Index with Soft Labels
Zifu Wang, Xuefei Ning, Matthew Blaschko
Intersection over Union (IoU) losses are surrogates that directly optimize the Jaccard index. Leveraging IoU losses as part of the loss function have demonstrated superior performance in semantic segmentation tasks compared to optimizing pixel-wise losses such as the cross-entropy loss alone. However, we identify a lack of flexibility in these losses to support vital training techniques like label smoothing, knowledge distillation, and semi-supervised learning, mainly due to their inability to process soft labels. To address this, we introduce Jaccard Metric Losses (JMLs), which are identical to the soft Jaccard loss in standard settings with hard labels but are fully compatible with soft labels. We apply JMLs to three prominent use cases of soft labels: label smoothing, knowledge distillation and semi-supervised learning, and demonstrate their potential to enhance model accuracy and calibration. Our experiments show consistent improvements over the cross-entropy loss across 4 semantic segmentation datasets (Cityscapes, PASCAL VOC, ADE20K, DeepGlobe Land) and 13 architectures, including classic CNNs and recent vision transformers. Remarkably, our straightforward approach significantly outperforms state-of-the-art knowledge distillation and semi-supervised learning methods. The code is available at \href{https://github.com/zifuwanggg/JDTLosses}{https://github.com/zifuwanggg/JDTLosses}.
****
Towards Stable Backdoor Purification through Feature Shift Tuning
Rui Min, Zeyu Qin, Li Shen, Minhao Cheng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Scalable 3D Captioning with Pretrained Models
Tiange Luo, Chris Rockwell, Honglak Lee, Justin Johnson
We introduce Cap3D, an automatic approach for generating descriptive text for 3D objects. This approach utilizes pretrained models from image captioning, image-text alignment, and LLM to consolidate captions from multiple views of a 3D asset, completely side-stepping the time-consuming and costly process of manual annotation. We apply Cap3D to the recently introduced large-scale 3D dataset, Objaverse, resulting in 660k 3D-text pairs. Our evaluation, conducted using 41k human annotations from the same dataset, demonstrates that Cap3D surpasses human-authored descriptions in terms of quality, cost, and speed. Through effective prompt engineering, Cap3D rivals human performance in generating geometric descriptions on 17k collected annotations from the ABO dataset. Finally, we finetune Text-to-3D models on Cap3D and human captions, and show Cap3D outperforms; and benchmark the SOTA including Point·E, Shape·E, and DreamFusion.
****
Langevin Quasi-Monte Carlo
Sifan Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
LargeST: A Benchmark Dataset for Large-Scale Traffic Forecasting
Xu Liu, Yutong Xia, Yuxuan Liang, Junfeng Hu, Yiwei Wang, LEI BAI, Chao Huang, Zhenguang Liu, Bryan Hooi, Roger Zimmermann
Road traffic forecasting plays a critical role in smart city initiatives and has experienced significant advancements thanks to the power of deep learning in ca

pturing non-linear patterns of traffic data. However, the promising results achi eved on current public datasets may not be applicable to practical scenarios due to limitations within these datasets. First, the limited sizes of them may not reflect the real-world scale of traffic networks. Second, the temporal coverage of these datasets is typically short, posing hurdles in studying long-term patte rns and acquiring sufficient samples for training deep models. Third, these data sets often lack adequate metadata for sensors, which compromises the reliability and interpretability of the data. To mitigate these limitations, we introduce t he LargeST benchmark dataset. It encompasses a total number of 8,600 sensors in California with a 5-year time coverage and includes comprehensive metadata. Usin g LargeST, we perform in-depth data analysis to extract data insights, benchmark well-known baselines in terms of their performance and efficiency, and identify challenges as well as opportunities for future research. We release the dataset s and baseline implementations at: https://github.com/liuxu77/LargeST.
****

What Can We Learn from Unlearnable Datasets?
Pedro Sandoval-Segura, Vasu Singla, Jonas Geiping, Micah Goldblum, Tom Goldstein
In an era of widespread web scraping, unlearnable dataset methods have the poten tial to protect data privacy by preventing deep neural networks from generalizin g. But in addition to a number of practical limitations that make their use unli kely, we make a number of findings that call into question their ability to safe guard data. First, it is widely believed that neural networks trained on unlearn able datasets only learn shortcuts, simpler rules that are not useful for genera lization. In contrast, we find that networks actually can learn useful features that can be reweighed for high test performance, suggesting that image protectio n is not assured. Unlearnable datasets are also believed to induce learning shor tcuts through linear separability of added perturbations. We provide a counterex ample, demonstrating that linear separability of perturbations is not a necessar y condition. To emphasize why linearly separable perturbations should not be rel ied upon, we propose an orthogonal projection attack which allows learning from unlearnable datasets published in ICML 2021 and ICLR 2023. Our proposed attack i s significantly less complex than recently proposed techniques.
****

Language Models Meet World Models: Embodied Experiences Enhance Language Models
Jiannan Xiang, Tianhua Tao, Yi Gu, Tianmin Shu, Zirui Wang, Zichao Yang , Zhitin g Hu
While large language models (LMs) have shown remarkable capabilities across nume rous tasks, they often struggle with simple reasoning and planning in physical e nvironments, such as understanding object permanence or planning household activ ities. The limitation arises from the fact that LMs are trained only on written text and miss essential embodied knowledge and skills. In this paper, we propose a new paradigm of enhancing LMs by finetuning them with world models, to gain d iverse embodied knowledge while retaining their general language capabilities. O ur approach deploys an embodied agent in a world model, particularly a simulator of the physical world (VirtualHome), and acquires a diverse set of embodied exp eriences through both goal-oriented planning and random exploration. These exper iences are then used to finetune LMs to teach diverse abilities of reasoning and acting in the physical world, e.g., planning and completing goals, object perma nence and tracking, etc. Moreover, it is desirable to preserve the generality of LMs during finetuning, which facilitates generalizing the embodied knowledge ac ross tasks rather than being tied to specific simulations. We thus further intro duce the classical elastic weight consolidation (EWC) for selective weight updat es, combined with low-rank adapters (LoRA) for training efficiency. Extensive ex periments show our approach substantially improves base LMs on 18 downstream tas ks by 64.28% on average. In particular, the small LMs (1.3B, 6B, and 13B) enhanc ed by our approach match or even outperform much larger LMs (e.g., ChatGPT).
****

Activity Grammars for Temporal Action Segmentation
Dayoung Gong, Joonseok Lee, Deunsol Jung, Suha Kwak, Minsu Cho
Sequence prediction on temporal data requires the ability to understand composit

ional structures of multi-level semantics beyond individual and contextual prope rties of parts. The task of temporal action segmentation remains challenging for the reason, aiming at translating an untrimmed activity video into a sequence o f action segments. This paper addresses the problem by introducing an effective activity grammar to guide neural predictions for temporal action segmentation. We propose a novel grammar induction algorithm, dubbed KARI, that extracts a pow erful context-free grammar from action sequence data. We also develop an efficie nt generalized parser, dubbed BEP, that transforms frame-level probability distr ibutions into a reliable sequence of actions according to the induced grammar wi th recursive rules. Our approach can be combined with any neural network for tem poral action segmentation to enhance the sequence prediction and discover its co mpositional structure. Experimental results demonstrate that our method signific antly improves temporal action segmentation in terms of both performance and int erpretability on two standard benchmarks, Breakfast and 50 Salads.
****
SANFlow: Semantic-Aware Normalizing Flow for Anomaly Detection
Daehyun Kim, Sungyong Baik, Tae Hyun Kim
Visual anomaly detection, the task of detecting abnormal characteristics in imag es, is challenging due to the rarity and unpredictability of anomalies. In order to reliably model the distribution of normality and detect anomalies, a few wor ks have attempted to exploit the density estimation ability of normalizing flow (NF). However, previous NF-based methods have relied solely on the capability of NF and forcibly transformed the distribution of all features to a single distri bution (e.g., unit normal distribution), when features can have different semant ic information and thus follow different distributions. We claim that forcibly l earning to transform such diverse distributions to a single distribution with a single network will cause the learning difficulty, limiting the capacity of a ne twork to discriminate normal and abnormal data. As such, we propose to transform the distribution of features at each location of a given image to different dis tributions. In particular, we train NF to map normal data distribution to distri butions with the same mean but different variances at each location of the given image. To enhance the discriminability, we also train NF to map abnormal data d istribution to a distribution with a mean that is different from that of normal data, where abnormal data is synthesized with data augmentation. The experimenta l results outline the effectiveness of the proposed framework in improving the d ensity modeling and thus anomaly detection performance.
****
AirDelhi: Fine-Grained Spatio-Temporal Particulate Matter Dataset From Delhi For ML based Modeling
Sachin Chauhan, Zeel Bharatkumar Patel, Sayan Ranu, Rijurekha Sen, Nipun Batra
Air pollution poses serious health concerns in developing countries, such as In dia, necessitating large-scale measurement for correlation analysis, policy reco mmendations, and informed decision-making. However, fine-grained data collection is costly. Specifically, static sensors for pollution measurement cost several thousand dollars per unit, leading to inadequate deployment and coverage. To co mplement the existing sparse static sensor network, we propose a mobile sensor n etwork utilizing lower-cost PM2.5 sensors mounted on public buses in the Delhi-N CR region of India. Through this exercise, we introduce a novel dataset AirDelhi comprising PM2.5 and PM10 measurements. This dataset is made publicly available , at https://www.cse.iitd.ac.in/pollutiondata, serving as a valuable resource fo r machine learning (ML) researchers and environmentalists. We present three key contributions with the release of this dataset. Firstly, through in-depth statis tical analysis, we demonstrate that the released dataset significantly differs f rom existing pollution datasets, highlighting its uniqueness and potential for n ew insights. Secondly, the dataset quality been validated against existing expen sive sensors. Thirdly, we conduct a benchmarking exercise (https://github.com/sa chin-iitd/DelhiPMDatasetBenchmark), evaluating state-of-the-art methods for inte rpolation, feature imputation, and forecasting on this dataset, which is the lar gest publicly available PM dataset to date. The results of the benchmarking exer cise underscore the substantial disparities in accuracy between the proposed dat

aset and other publicly available datasets. This finding highlights the complexity and richness of our dataset, emphasizing its value for advancing research in the field of air pollution.
****
On Convergence of Polynomial Approximations to the Gaussian Mixture Entropy
Caleb Dahlke, Jason Pacheco
Gaussian mixture models (GMMs) are fundamental to machine learning due to their flexibility as approximating densities. However, uncertainty quantification of GMMs remains a challenge as differential entropy lacks a closed form. This paper explores polynomial approximations, specifically Taylor and Legendre, to the GMM entropy from a theoretical and practical perspective. We provide new analysis of a widely used approach due to Huber et al.(2008) and show that the series diverges under simple conditions. Motivated by this divergence we provide a novel Taylor series that is provably convergent to the true entropy of any GMM. We demonstrate a method for selecting a center such that the series converges from below, providing a lower bound on GMM entropy. Furthermore, we demonstrate that orthogonal polynomial series result in more accurate polynomial approximations. Experimental validation supports our theoretical results while showing that our method is comparable in computation to Huber et al. We also show that in application, the use of these polynomial approximations, such as in Nonparametric Variational Inference by Gershamn et al. (2012), rely on the convergence of the methods in computing accurate approximations. This work contributes useful analysis to existing methods while introducing a novel approximation supported by firm theoretical guarantees.
****
CEIL: Generalized Contextual Imitation Learning
Jinxin Liu, Li He, Yachen Kang, Zifeng Zhuang, Donglin Wang, Huazhe Xu
In this paper, we present ContExtual Imitation Learning (CEIL), a general and broadly applicable algorithm for imitation learning (IL). Inspired by the formulation of hindsight information matching, we derive CEIL by explicitly learning a hindsight embedding function together with a contextual policy using the hindsight embeddings. To achieve the expert matching objective for IL, we advocate for optimizing a contextual variable such that it biases the contextual policy towards mimicking expert behaviors. Beyond the typical learning from demonstrations (LfD) setting, CEIL is a generalist that can be effectively applied to multiple settings including: 1) learning from observations (LfO), 2) offline IL, 3) cross-domain IL (mismatched experts), and 4) one-shot IL settings. Empirically, we evaluate CEIL on the popular MuJoCo tasks (online) and the D4RL dataset (offline). Compared to prior state-of-the-art baselines, we show that CEIL is more sample-efficient in most online IL tasks and achieves better or competitive performances in offline tasks.
****
Entropic Neural Optimal Transport via Diffusion Processes
Nikita Gushchin, Alexander Kolesov, Alexander Korotin, Dmitry P. Vetrov, Evgeny Burnaev
We propose a novel neural algorithm for the fundamental problem of computing the entropic optimal transport (EOT) plan between probability distributions which are accessible by samples. Our algorithm is based on the saddle point reformulation of the dynamic version of EOT which is known as the Schrödinger Bridge problem. In contrast to the prior methods for large-scale EOT, our algorithm is end-to-end and consists of a single learning step, has fast inference procedure, and allows handling small values of the entropy regularization coefficient which is of particular importance in some applied problems. Empirically, we show the performance of the method on several large-scale EOT tasks. The code for the ENOT solver can be found at https://github.com/ngushchin/EntropicNeuralOptimalTransport
****
On the Convergence to a Global Solution of Shuffling-Type Gradient Algorithms
Lam Nguyen, Trang H. Tran
Stochastic gradient descent (SGD) algorithm is the method of choice in many machine learning tasks thanks to its scalability and efficiency in dealing with larg

e-scale problems. In this paper, we focus on the shuffling version of SGD which matches the mainstream practical heuristics. We show the convergence to a global solution of shuffling SGD for a class of non-convex functions under over-parameterized settings. Our analysis employs more relaxed non-convex assumptions than previous literature. Nevertheless, we maintain the desired computational complexity as shuffling SGD has achieved in the general convex setting.
****
Evaluating Robustness and Uncertainty of Graph Models Under Structural Distributional Shifts
Gleb Bazhenov, Denis Kuznedelev, Andrey Malinin, Artem Babenko, Liudmila Prokhorenkova
In reliable decision-making systems based on machine learning, models have to be robust to distributional shifts or provide the uncertainty of their predictions. In node-level problems of graph learning, distributional shifts can be especially complex since the samples are interdependent. To evaluate the performance of graph models, it is important to test them on diverse and meaningful distributional shifts. However, most graph benchmarks considering distributional shifts for node-level problems focus mainly on node features, while structural properties are also essential for graph problems. In this work, we propose a general approach for inducing diverse distributional shifts based on graph structure. We use this approach to create data splits according to several structural node properties: popularity, locality, and density. In our experiments, we thoroughly evaluate the proposed distributional shifts and show that they can be quite challenging for existing graph models. We also reveal that simple models often outperform more sophisticated methods on the considered structural shifts. Finally, our experiments provide evidence that there is a trade-off between the quality of learned representations for the base classification task under structural distributional shift and the ability to separate the nodes from different distributions using these representations.
****
Learning Causal Models under Independent Changes
Sarah Mameche, David Kaltenpoth, Jilles Vreeken
In many scientific applications, we observe a system in different conditions in which its components may change, rather than in isolation. In our work, we are interested in explaining the generating process of such a multi-context system using a finite mixture of causal mechanisms. Recent work shows that this causal model is identifiable from data, but is limited to settings where the sparse mechanism shift hypothesis holds and only a subset of the causal conditionals change. As this assumption is not easily verifiable in practice, we study the more general principle that mechanism shifts are independent, which we formalize using the algorithmic notion of independence. We introduce an approach for causal discovery beyond partially directed graphs using Gaussian Process models, and give conditions under which we provably identify the correct causal model. In our experiments, we show that our method performs well in a range of synthetic settings, on realistic gene expression simulations, as well as on real-world cell signaling data.
****
Universality and Limitations of Prompt Tuning
Yihan Wang, Jatin Chauhan, Wei Wang, Cho-Jui Hsieh
Despite the demonstrated empirical efficacy of prompt tuning to adapt a pretrained language model for a new task, the theoretical underpinnings of the difference between "tuning parameters before the input" against "the tuning of model weights" are limited. We thus take one of the first steps to understand the role of soft-prompt tuning for transformer-based architectures. By considering a general purpose architecture, we analyze prompt tuning from the lens of both: universal approximation and limitations with finite-depth fixed-weight pretrained transformers for continuous-valued functions. Our universality result guarantees the existence of a strong transformer with a prompt to approximate any sequence-to-sequence function in the set of Lipschitz functions. The limitations of prompt tuning for limited-depth transformers are first proved by constructing a set of data

sets, that cannot be memorized by a prompt of any length for a given single enco der layer. We also provide a lower bound on the required number of tunable promp t parameters and compare the result with the number of parameters required for a low-rank update (based on LoRA) for a single-layer setting. We finally extend o ur analysis to multi-layer settings by providing sufficient conditions under whi ch the transformer can at best learn datasets from invertible functions only. Ou r theoretical claims are also corroborated by empirical results.
****

Evaluating Neuron Interpretation Methods of NLP Models

Yimin Fan, Fahim Dalvi, Nadir Durrani, Hassan Sajjad

Neuron interpretation offers valuable insights into how knowledge is structured within a deep neural network model. While a number of neuron interpretation meth ods have been proposed in the literature, the field lacks a comprehensive compar ison among these methods. This gap hampers progress due to the absence of standa rdized metrics and benchmarks. The commonly used evaluation metric has limitatio ns, and creating ground truth annotations for neurons is impractical. Addressing these challenges, we propose an evaluation framework based on voting theory. Ou r hypothesis posits that neurons consistently identified by different methods ca rry more significant information. We rigorously assess our framework across a di verse array of neuron interpretation methods. Notable findings include: i) despi te the theoretical differences among the methods, neuron ranking methods share o ver 60% of their rankings when identifying salient neurons, ii) the neuron inter pretation methods are most sensitive to the last layer representations, iii) Pro beless neuron ranking emerges as the most consistent method.
****

How Re-sampling Helps for Long-Tail Learning?

Jiang-Xin Shi, Tong Wei, Yuke Xiang, Yu-Feng Li

Long-tail learning has received significant attention in recent years due to the challenge it poses with extremely imbalanced datasets. In these datasets, only a few classes (known as the head classes) have an adequate number of training sa mples, while the rest of the classes (known as the tail classes) are infrequent in the training data. Re-sampling is a classical and widely used approach for ad dressing class imbalance issues. Unfortunately, recent studies claim that re-sam pling brings negligible performance improvements in modern long-tail learning ta sks. This paper aims to investigate this phenomenon systematically. Our research shows that re-sampling can considerably improve generalization when the trainin g images do not contain semantically irrelevant contexts. In other scenarios, ho wever, it can learn unexpected spurious correlations between irrelevant contexts and target labels. We design experiments on two homogeneous datasets, one conta ining irrelevant context and the other not, to confirm our findings. To prevent the learning of spurious correlations, we propose a new context shift augmentati on module that generates diverse training images for the tail class by maintaini ng a context bank extracted from the head-class images. Experiments demonstrate that our proposed module can boost the generalization and outperform other appro aches, including class-balanced re-sampling, decoupled classifier re-training, a nd data augmentation methods. The source code is available at https://www.lamda. nju.edu.cn/code_CSA.ashx.
****

FIND: A Function Description Benchmark for Evaluating Interpretability Methods

Sarah Schwettmann, Tamar Shaham, Joanna Materzynska, Neil Chowdhury, Shuang Li, Jacob Andreas, David Bau, Antonio Torralba

Labeling neural network submodules with human-legible descriptions is useful for many downstream tasks: such descriptions can surface failures, guide interventi ons, and perhaps even explain important model behaviors. To date, most mechanist ic descriptions of trained networks have involved small models, narrowly delimit ed phenomena, and large amounts of human labor. Labeling all human-interpretable sub-computations in models of increasing size and complexity will almost certai nly require tools that can generate and validate descriptions automatically. Rec ently, techniques that use learned models in-the-loop for labeling have begun to gain traction, but methods for evaluating their efficacy are limited and ad-hoc

. How should we validate and compare open-ended labeling tools? This paper intro duces FIND (Function INterpretation and Description), a benchmark suite for eval uating the building blocks of automated interpretability methods. FIND contains functions that resemble components of trained neural networks, and accompanying descriptions of the kind we seek to generate. The functions are procedurally con structed across textual and numeric domains, and involve a range of real-world c omplexities, including noise, composition, approximation, and bias. We evaluate methods that use pretrained language models (LMs) to produce code-based and natu ral language descriptions of function behavior. Additionally, we introduce a new interactive method in which an Automated Interpretability Agent (AIA) generates function descriptions. We find that an AIA, built with an off-the-shelf LM augm ented with black-box access to functions, can sometimes infer function structure —acting as a scientist by forming hypotheses, proposing experiments, and updatin g descriptions in light of new data. However, FIND also reveals that LM-based de scriptions capture global function behavior while missing local details. These r esults suggest that FIND will be useful for characterizing the performance of mo re sophisticated interpretability methods before they are applied to real-world models.
****
EgoTracks: A Long-term Egocentric Visual Object Tracking Dataset
Hao Tang, Kevin J Liang, Kristen Grauman, Matt Feiszli, Weiyao Wang
Visual object tracking is a key component to many egocentric vision problems. Ho wever, the full spectrum of challenges of egocentric tracking faced by an embodi ed AI is underrepresented in many existing datasets; these tend to focus on rela tively short, third-person videos. Egocentric video has several distinguishing c haracteristics from those commonly found in past datasets: frequent large camera motions and hand interactions with objects commonly lead to occlusions or objec ts exiting the frame, and object appearance can change rapidly due to widely dif ferent points of view, scale, or object states. Embodied tracking is also natura lly long-term, and being able to consistently (re-)associate objects to their ap pearances and disappearances over as long as a lifetime is critical. Previous da tasets under-emphasize this re-detection problem, and their "framed" nature has led to adoption of various spatiotemporal priors that we find do not necessarily generalize to egocentric video. We thus introduce EgoTracks, a new dataset for long-term egocentric visual object tracking. Sourced from the Ego4D dataset, thi s new dataset presents a significant challenge to recent state-of-the-art single -object tracking models, which we find score poorly on traditional tracking metr ics for our new dataset, compared to popular benchmarks. We further show improve ments that can be made to a STARK tracker to significantly increase its performa nce on egocentric data, resulting in a baseline model we call EgoSTARK. We publi cly release our annotations and benchmark, hoping our dataset leads to further a dvancements in tracking.
****
Brain Diffusion for Visual Exploration: Cortical Discovery using Large Scale Gen erative Models
Andrew Luo, Maggie Henderson, Leila Wehbe, Michael Tarr
A long standing goal in neuroscience has been to elucidate the functional organi zation of the brain. Within higher visual cortex, functional accounts have remai ned relatively coarse, focusing on regions of interest (ROIs) and taking the for m of selectivity for broad categories such as faces, places, bodies, food, or wo rds. Because the identification of such ROIs has typically relied on manually as sembled stimulus sets consisting of isolated objects in non-ecological contexts, exploring functional organization without robust a priori hypotheses has been c hallenging. To overcome these limitations, we introduce a data-driven approach i n which we synthesize images predicted to activate a given brain region using pa ired natural images and fMRI recordings, bypassing the need for category-specifi c stimuli. Our approach -- Brain Diffusion for Visual Exploration ("BrainDiVE") -- builds on recent generative methods by combining large-scale diffusion models with brain-guided image synthesis. Validating our method, we demonstrate the ab ility to synthesize preferred images with appropriate semantic specificity for w

ell-characterized category-selective ROIs. We then show that BrainDiVE can chara cterize differences between ROIs selective for the same high-level category. Fin ally we identify novel functional subdivisions within these ROIs, validated with behavioral data. These results advance our understanding of the fine-grained f unctional organization of human visual cortex, and provide well-specified constr aints for further examination of cortical organization using hypothesis-driven m ethods.
****
Query-based Temporal Fusion with Explicit Motion for 3D Object Detection
Jinghua Hou, Zhe Liu, dingkang liang, Zhikang Zou, Xiaoqing Ye, Xiang Bai
Effectively utilizing temporal information to improve 3D detection performance i s vital for autonomous driving vehicles. Existing methods either conduct tempora l fusion based on the dense BEV features or sparse 3D proposal features. However , the former does not pay more attention to foreground objects, leading to more computation costs and sub-optimal performance. The latter implements time-consum ing operations to generate sparse 3D proposal features, and the performance is l imited by the quality of 3D proposals. In this paper, we propose a simple and ef fective Query-based Temporal Fusion Network (QTNet). The main idea is to exploit the object queries in previous frames to enhance the representation of current object queries by the proposed Motion-guided Temporal Modeling (MTM) module, whi ch utilizes the spatial position information of object queries along the tempora l dimension to construct their relevance between adjacent frames reliably. Exper imental results show our proposed QTNet outperforms BEV-based or proposal-based manners on the nuScenes dataset. Besides, the MTM is a plug-and-play module, whi ch can be integrated into some advanced LiDAR-only or multi-modality 3D detector s and even brings new SOTA performance with negligible computation cost and late ncy on the nuScenes dataset. These experiments powerfully illustrate the superio rity and generalization of our method. The code is available at https://github.c om/AlmoonYsl/QTNet.
****
Efficient Adversarial Contrastive Learning via Robustness-Aware Coreset Selectio n
Xilie Xu, Jingfeng ZHANG, Feng Liu, Masashi Sugiyama, Mohan S. Kankanhalli
Adversarial contrastive learning (ACL) does not require expensive data annotatio ns but outputs a robust representation that withstands adversarial attacks and a lso generalizes to a wide range of downstream tasks. However, ACL needs tremendo us running time to generate the adversarial variants of all training data, which limits its scalability to large datasets. To speed up ACL, this paper proposes a robustness-aware coreset selection (RCS) method. RCS does not require label in formation and searches for an informative subset that minimizes a representation al divergence, which is the distance of the representation between natural data and their virtual adversarial variants. The vanilla solution of RCS via traversi ng all possible subsets is computationally prohibitive. Therefore, we theoretica lly transform RCS into a surrogate problem of submodular maximization, of which the greedy search is an efficient solution with an optimality guarantee for the original problem. Empirically, our comprehensive results corroborate that RCS ca n speed up ACL by a large margin without significantly hurting the robustness tr ansferability. Notably, to the best of our knowledge, we are the first to conduc t ACL efficiently on the large-scale ImageNet-1K dataset to obtain an effective robust representation via RCS. Our source code is at https://github.com/GodXuxil ie/EfficientACLvia_RCS.
****
A Finite-Sample Analysis of Payoff-Based Independent Learning in Zero-Sum Stocha stic Games
Zaiwei Chen, Kaiqing Zhang, Eric Mazumdar, Asuman Ozdaglar, Adam Wierman
In this work, we study two-player zero-sum stochastic games and develop a varian t of the smoothed best-response learning dynamics that combines independent lear ning dynamics for matrix games with the minimax value iteration for stochastic g ames. The resulting learning dynamics are payoff-based, convergent, rational, an d symmetric between the two players. Our theoretical results present to the best

of our knowledge the first last-iterate finite-sample analysis of such independent learning dynamics. To establish the results, we develop a coupled Lyapunov drift approach to capture the evolution of multiple sets of coupled and stochastic iterates, which might be of independent interest.
****
Compact Neural Volumetric Video Representations with Dynamic Codebooks
Haoyu Guo, Sida Peng, Yunzhi Yan, Linzhan Mou, Yujun Shen, Hujun Bao, Xiaowei Zhou
This paper addresses the challenge of representing high-fidelity volumetric videos with low storage cost. Some recent feature grid-based methods have shown superior performance of fast learning implicit neural representations from input 2D images. However, such explicit representations easily lead to large model sizes when modeling dynamic scenes. To solve this problem, our key idea is reducing the spatial and temporal redundancy of feature grids, which intrinsically exist due to the self-similarity of scenes. To this end, we propose a novel neural representation, named dynamic codebook, which first merges similar features for the model compression and then compensates for the potential decline in rendering quality by a set of dynamic codes. Experiments on the NHR and DyNeRF datasets demonstrate that the proposed approach achieves state-of-the-art rendering quality, while being able to achieve more storage efficiency. The source code is available at https://github.com/zju3dv/compact_vv.
****
Towards Label-free Scene Understanding by Vision Foundation Models
Runnan Chen, Youquan Liu, Lingdong Kong, Nenglun Chen, Xinge ZHU, Yuexin Ma, Tongliang Liu, Wenping Wang
Vision foundation models such as Contrastive Vision-Language Pre-training (CLIP) and Segment Anything (SAM) have demonstrated impressive zero-shot performance on image classification and segmentation tasks. However, the incorporation of CLIP and SAM for label-free scene understanding has yet to be explored. In this paper, we investigate the potential of vision foundation models in enabling networks to comprehend 2D and 3D worlds without labelled data. The primary challenge lies in effectively supervising networks under extremely noisy pseudo labels, which are generated by CLIP and further exacerbated during the propagation from the 2D to the 3D domain. To tackle these challenges, we propose a novel Cross-modality Noisy Supervision (CNS) method that leverages the strengths of CLIP and SAM to supervise 2D and 3D networks simultaneously. In particular, we introduce a prediction consistency regularization to co-train 2D and 3D networks, then further impose the networks' latent space consistency using the SAM's robust feature representation. Experiments conducted on diverse indoor and outdoor datasets demonstrate the superior performance of our method in understanding 2D and 3D open environments. Our 2D and 3D network achieves label-free semantic segmentation with 28.4\% and 33.5\% mIoU on ScanNet, improving 4.7\% and 7.9\%, respectively. For nuImages and nuScenes datasets, the performance is 22.1\% and 26.8\% with improvements of 3.5\% and 6.0\%, respectively. Code is available. (https://github.com/runnanchen/Label-Free-Scene-Understanding)
****
Sketchy: Memory-efficient Adaptive Regularization with Frequent Directions
Vladimir Feinberg, Xinyi Chen, Y. Jennifer Sun, Rohan Anil, Elad Hazan
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
SatBird: a Dataset for Bird Species Distribution Modeling using Remote Sensing and Citizen Science Data
Mélisande Teng, Amna Elmustafa, Benjamin Akera, Yoshua Bengio, Hager Radi, Hugo Larochelle, David Rolnick
Biodiversity is declining at an unprecedented rate, impacting ecosystem services necessary to ensure food, water, and human health and well-being. Understanding the distribution of species and their habitats is crucial for conservation poli

cy planning. However, traditional methods in ecology for species distribution mo
dels (SDMs) generally focus either on narrow sets of species or narrow geographi
cal areas and there remain significant knowledge gaps about the distribution of
species. A major reason for this is the limited availability of data traditional
ly used, due to the prohibitive amount of effort and expertise required for trad
itional field monitoring. The wide availability of remote sensing data and the g
rowing adoption of citizen science tools to collect species observations data at
 low cost offer an opportunity for improving biodiversity monitoring and enablin
g the modelling of complex ecosystems. We introduce a novel task for mapping bir
d species to their habitats by predicting species encounter rates from satellite
 images, and present SatBird, a satellite dataset of locations in the USA with l
abels derived from presence-absence observation data from the citizen science da
tabase eBird, considering summer (breeding) and winter seasons. We also provide
a dataset in Kenya representing low-data regimes. We additionally provide enviro
nmental data and species range maps for each location. We benchmark a set of ba
selines on our dataset, including SOTA models for remote sensing tasks. SatBird
opens up possibilities for scalably modelling properties of ecosystems worldwide
.
****
DiffComplete: Diffusion-based Generative 3D Shape Completion
Ruihang Chu, Enze Xie, Shentong Mo, Zhenguo Li, Matthias Niessner, Chi-Wing Fu,
Jiaya Jia
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Bayesian Risk-Averse Q-Learning with Streaming Observations
Yuhao Wang, Enlu Zhou
We consider a robust reinforcement learning problem, where a learning agent lear
ns from a simulated training environment. To account for the model mis-specifica
tion between this training environment and the true environment due to lack of d
ata, we adopt a formulation of Bayesian risk MDP (BRMDP) with infinite horizon,
which uses Bayesian posterior to estimate the transition model and impose a risk
 functional to account for the model uncertainty. Observations from the real env
ironment that is out of the agent's control arrive periodically and are utilized
 by the agent to update the Bayesian posterior to reduce model uncertainty. We t
heoretically demonstrate that BRMDP balances the trade-off between robustness an
d conservativeness, and we further develop a multi-stage Bayesian risk-averse Q-
learning algorithm to solve BRMDP with streaming observations from real environm
ent. The proposed algorithm learns a risk-averse yet optimal policy that depends
 on the availability of real-world observations. We provide a theoretical guaran
tee of strong convergence for the proposed algorithm.
****
On the Planning Abilities of Large Language Models - A Critical Investigation
Karthik Valmeekam, Matthew Marquez, Sarath Sreedharan, Subbarao Kambhampati
Intrigued by the claims of emergent reasoning capabilities in LLMs trained on ge
neral web corpora, in this paper, we set out to investigate their planning capab
ilities. We aim to evaluate (1) the effectiveness of LLMs in generating plans au
tonomously in commonsense planning tasks and (2) the potential of LLMs as a sour
ce of heuristic guidance for other agents (AI planners) in their planning tasks.
 We conduct a systematic study by generating a suite of instances on domains sim
ilar to the ones employed in the International Planning Competition and evaluate
 LLMs in two distinct modes: autonomous and heuristic. Our findings reveal that
LLMs’ ability to generate executable plans autonomously is rather limited, with
the best model (GPT-4) having an average success rate of ~12% across the domains
. However, the results in the heuristic mode show more promise. In the heuristic
 mode, we demonstrate that LLM-generated plans can improve the search process fo
r underlying sound planners and additionally show that external verifiers can he
lp provide feedback on the generated plans and back-prompt the LLM for better pl

an generation.
****
Is RLHF More Difficult than Standard RL? A Theoretical Perspective
Yuanhao Wang, Qinghua Liu, Chi Jin
Reinforcement learning from Human Feedback (RLHF) learns from preference signals, while standard Reinforcement Learning (RL) directly learns from reward signals. Preferences arguably contain less information than rewards, which makes preference-based RL seemingly more difficult. This paper theoretically proves that, for a wide range of preference models, we can solve preference-based RL directly using existing algorithms and techniques for reward-based RL, with small or no extra costs. Specifically, (1) for preferences that are drawn from reward-based probabilistic models, we reduce the problem to robust reward-based RL that can tolerate small errors in rewards; (2) for general arbitrary preferences where the objective is to find the von Neumann winner, we reduce the problem to multiagent reward-based RL which finds Nash equilibria for factored Markov games under a restricted set of policies. The latter case can be further reduce to adversarial MDP when preferences only depend on the final state. We instantiate all reward-based RL subroutines by concrete provable algorithms, and apply our theory to a large class of models including tabular MDPs and MDPs with generic function approximation. We further provide guarantees when K-wise comparisons are available.
****
How does GPT-2 compute greater-than?: Interpreting mathematical abilities in a pre-trained language model
Michael Hanna, Ollie Liu, Alexandre Variengien
Pre-trained language models can be surprisingly adept at tasks they were not explicitly trained on, but how they implement these capabilities is poorly understood. In this paper, we investigate the basic mathematical abilities often acquired by pre-trained language models. Concretely, we use mechanistic interpretability techniques to explain the (limited) mathematical abilities of GPT-2 small. As a case study, we examine its ability to take in sentences such as "The war lasted from the year 1732 to the year 17", and predict valid two-digit end years (years > 32). We first identify a circuit, a small subset of GPT-2 small's computational graph that computes this task's output. Then, we explain the role of each circuit component, showing that GPT-2 small's final multi-layer perceptrons boost the probability of end years greater than the start year. Finally, we find related tasks that activate our circuit. Our results suggest that GPT-2 small computes greater-than using a complex but general mechanism that activates across diverse contexts.
****
NAVI: Category-Agnostic Image Collections with High-Quality 3D Shape and Pose Annotations
Varun Jampani, Kevis-kokitsi Maninis, Andreas Engelhardt, Arjun Karpur, Karen Truong, Kyle Sargent, Stefan Popov, Andre Araujo, Ricardo Martin Brualla, Kaushal Patel, Daniel Vlasic, Vittorio Ferrari, Ameesh Makadia, Ce Liu, Yuanzhen Li, Howard Zhou
Recent advances in neural reconstruction enable high-quality 3D object reconstruction from casually captured image collections. Current techniques mostly analyze their progress on relatively simple image collections where SfM techniques can provide ground-truth (GT) camera poses. We note that SfM techniques tend to fail on in-the-wild image collections such as image search results with varying backgrounds and illuminations. To enable systematic research progress on 3D reconstruction from casual image captures, we propose `NAVI': a new dataset of category-agnostic image collections of objects with high-quality 3D scans along with per-image 2D-3D alignments providing near-perfect GT camera parameters. These 2D-3D alignments allow us to extract accurate derivative annotations such as dense pixel correspondences, depth and segmentation maps. We demonstrate the use of NAVI image collections on different problem settings and show that NAVI enables more thorough evaluations that were not possible with existing datasets. We believe NAVI is beneficial for systematic research progress on 3D reconstruction and correspondence estimation.

****
Multi-body SE(3) Equivariance for Unsupervised Rigid Segmentation and Motion Est imation
Jia-Xing Zhong, Ta-Ying Cheng, Yuhang He, Kai Lu, Kaichen Zhou, Andrew Markham, Niki Trigoni
A truly generalizable approach to rigid segmentation and motion estimation is fu ndamental to 3D understanding of articulated objects and moving scenes. In view of the closely intertwined relationship between segmentation and motion estimate s, we present an SE(3) equivariant architecture and a training strategy to tackl e this task in an unsupervised manner. Our architecture is composed of two inter connected, lightweight heads. These heads predict segmentation masks using point -level invariant features and estimate motion from SE(3) equivariant features, a ll without the need for category information. Our training strategy is unified a nd can be implemented online, which jointly optimizes the predicted segmentation and motion by leveraging the interrelationships among scene flow, segmentation mask, and rigid transformations. We conduct experiments on four datasets to demo nstrate the superiority of our method. The results show that our method excels i n both model performance and computational efficiency, with only 0.25M parameter s and 0.92G FLOPs. To the best of our knowledge, this is the first work designed for category-agnostic part-level SE(3) equivariance in dynamic point clouds.
****
Suggesting Variable Order for Cylindrical Algebraic Decomposition via Reinforcem ent Learning
Fuqi Jia, Yuhang Dong, Minghao Liu, Pei Huang, Feifei Ma, Jian Zhang
Cylindrical Algebraic Decomposition (CAD) is one of the pillar algorithms of sym bolic computation, and its worst-case complexity is double exponential to the nu mber of variables. Researchers found that variable order dramatically affects ef ficiency and proposed various heuristics. The existing learning-based methods ar e all supervised learning methods that cannot cope with diverse polynomial sets. This paper proposes two Reinforcement Learning (RL) approaches combined with Gra ph Neural Networks (GNN) for Suggesting Variable Order (SVO). One is GRL-SVO(UP) , a branching heuristic integrated with CAD. The other is GRL-SVO(NUP), a fast h euristic providing a total order directly. We generate a random dataset and coll ect a real-world dataset from SMT-LIB. The experiments show that our approaches outperform state-of-the-art learning-based heuristics and are competitive with t he best expert-based heuristics. Interestingly, our models show a strong general ization ability, working well on various datasets even if they are only trained on a 3-var random dataset. The source code and data are available at https://git hub.com/dongyuhang22/GRL-SVO.
****
GLOBER: Coherent Non-autoregressive Video Generation via GLOBal Guided Video Dec odER
Mingzhen Sun, Weining Wang, Zihan Qin, Jiahui Sun, Sihan Chen, Jing Liu
Video generation necessitates both global coherence and local realism. This work presents a novel non-autoregressive method GLOBER, which first generates global features to obtain comprehensive global guidance and then synthesizes video fra mes based on the global features to generate coherent videos. Specifically, we p ropose a video auto-encoder, where a video encoder encodes videos into global fe atures, and a video decoder, built on a diffusion model, decodes the global feat ures and synthesizes video frames in a non-autoregressive manner. To achieve max imum flexibility, our video decoder perceives temporal information through norma lized frame indexes, which enables it to synthesize arbitrary sub video clips wi th predetermined starting and ending frame indexes. Moreover, a novel adversaria l loss is introduced to improve the global coherence and local realism between t he synthesized video frames. Finally, we employ a diffusion-based video generato r to fit the global features outputted by the video encoder for video generation . Extensive experimental results demonstrate the effectiveness and efficiency of our proposed method, and new state-of-the-art results have been achieved on mul tiple benchmarks.
****

Dense and Aligned Captions (DAC) Promote Compositional Reasoning in VL Models
Sivan Doveh, Assaf Arbelle, Sivan Harary, Roei Herzig, Donghyun Kim, Paola Casca nte-Bonilla, Amit Alfassy, Rameswar Panda, Raja Giryes, Rogerio Feris, Shimon Ul lman, Leonid Karlinsky
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Latent SDEs on Homogeneous Spaces
Sebastian Zeng, Florian Graf, Roland Kwitt
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Balancing Risk and Reward: A Batched-Bandit Strategy for Automated Phased Releas e
Yufan Li, Jialiang Mao, Iavor Bojinov
Phased releases are a common strategy in the technology industry for gradually r eleasing new products or updates through a sequence of A/B tests in which the nu mber of treated units gradually grows until full deployment or deprecation. Perf orming phased releases in a principled way requires selecting the proportion of units assigned to the new release in a way that balances the risk of an adverse effect with the need to iterate and learn from the experiment rapidly. In this p aper, we formalize this problem and propose an algorithm that automatically dete rmines the release percentage at each stage in the schedule, balancing the need to control risk while maximizing ramp-up speed. Our framework models the challen ge as a constrained batched bandit problem that ensures that our pre-specified e xperimental budget is not depleted with high probability. Our proposed algorithm leverages an adaptive Bayesian approach in which the maximal number of units as signed to the treatment is determined by the posterior distribution, ensuring th at the probability of depleting the remaining budget is low. Notably, our approa ch analytically solves the ramp sizes by inverting probability bounds, eliminati ng the need for challenging rare-event Monte Carlo simulation. It only requires computing means and variances of outcome subsets, making it highly efficient and parallelizable.
****
Preconditioning Matters: Fast Global Convergence of Non-convex Matrix Factorizat ion via Scaled Gradient Descent
Xixi Jia, Hailin Wang, Jiangjun Peng, Xiangchu Feng, Deyu Meng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Deep Insights into Noisy Pseudo Labeling on Graph Data
Botao WANG, Jia Li, Yang Liu, Jiashun Cheng, Yu Rong, Wenjia Wang, Fugee Tsung
Pseudo labeling (PL) is a wide-applied strategy to enlarge the labeled dataset b y self-annotating the potential samples during the training process. Several wor ks have shown that it can improve the graph learning model performance in genera l. However, we notice that the incorrect labels can be fatal to the graph traini ng process. Inappropriate PL may result in the performance degrading, especially on graph data where the noise can propagate. Surprisingly, the corresponding er ror is seldom theoretically analyzed in the literature. In this paper, we aim to give deep insights of PL on graph learning models. We first present the error a nalysis of PL strategy by showing that the error is bounded by the confidence of PL threshold and consistency of multi-view prediction. Then, we theoretically i llustrate the effect of PL on convergence property. Based on the analysis, we pr opose a cautious pseudo labeling methodology in which we pseudo label the sample

s with highest confidence and multi-view consistency. Finally, extensive experim ents demonstrate that the proposed strategy improves graph learning process and outperforms other PL strategies on link prediction and node classification tasks .
****
Predicting a Protein's Stability under a Million Mutations
Jeffrey Ouyang-Zhang, Daniel Diaz, Adam Klivans, Philipp Kraehenbuehl
Stabilizing proteins is a foundational step in protein engineering. However, the evolutionary pressure of all extant proteins makes identifying the scarce numbe r of mutations that will improve thermodynamic stability challenging. Deep learn ing has recently emerged as a powerful tool for identifying promising mutations. Existing approaches, however, are computationally expensive, as the number of mo del inferences scales with the number of mutations queried. Our main contributio n is a simple, parallel decoding algorithm.Mutate Everything is capable of predi cting the effect of all single and double mutations in one forward pass. It is e ven versatile enough to predict higher-order mutations with minimal computationa l overhead.We build Mutate Everything on top of ESM2 and AlphaFold, neither of w hich were trained to predict thermodynamic stability.We trained on the Mega-Scal e cDNA proteolysis dataset and achieved state-of-the-art performance on single a nd higher-order mutations on S669, ProTherm, and ProteinGym datasets.Our code is available at https://github.com/jozhang97/MutateEverything.
****
Deep Gaussian Markov Random Fields for Graph-Structured Dynamical Systems
Fiona Lippert, Bart Kranstauber, Emiel van Loon, Patrick Forré
Probabilistic inference in high-dimensional state-space models is computationall y challenging. For many spatiotemporal systems, however, prior knowledge about t he dependency structure of state variables is available. We leverage this struct ure to develop a computationally efficient approach to state estimation and lear ning in graph-structured state-space models with (partially) unknown dynamics an d limited historical data. Building on recent methods that combine ideas from de ep learning with principled inference in Gaussian Markov random fields (GMRF), w e reformulate graph-structured state-space models as Deep GMRFs defined by simpl e spatial and temporal graph layers. This results in a flexible spatiotemporal p rior that can be learned efficiently from a single time sequence via variational inference. Under linear Gaussian assumptions, we retain a closed-form posterior , which can be sampled efficiently using the conjugate gradient method, scaling favourably compared to classical Kalman filter based approaches.
****
Computational Complexity of Learning Neural Networks: Smoothness and Degeneracy
Amit Daniely, Nati Srebro, Gal Vardi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
LoCoOp: Few-Shot Out-of-Distribution Detection via Prompt Learning
Atsuyuki Miyai, Qing Yu, Go Irie, Kiyoharu Aizawa
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
AdANNS: A Framework for Adaptive Semantic Search
Aniket Rege, Aditya Kusupati, Sharan Ranjit S, Alan Fan, Qingqing Cao, Sham Kaka de, Prateek Jain, Ali Farhadi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

When Do Neural Nets Outperform Boosted Trees on Tabular Data?
Duncan McElfresh, Sujay Khandagale, Jonathan Valverde, Vishak Prasad C, Ganesh Ramakrishnan, Micah Goldblum, Colin White
Tabular data is one of the most commonly used types of data in machine learning. Despite recent advances in neural nets (NNs) for tabular data, there is still an active discussion on whether or not NNs generally outperform gradient-boosted decision trees (GBDTs) on tabular data, with several recent works arguing either that GBDTs consistently outperform NNs on tabular data, or vice versa. In this work, we take a step back and question the importance of this debate. To this end, we conduct the largest tabular data analysis to date, comparing 19 algorithms across 176 datasets, and we find that the 'NN vs. GBDT' debate is overemphasized: for a surprisingly high number of datasets, either the performance difference between GBDTs and NNs is negligible, or light hyperparameter tuning on a GBDT is more important than choosing between NNs and GBDTs. Next, we analyze dozens of metafeatures to determine what \emph{properties} of a dataset make NNs or GBDTs better-suited to perform well. For example, we find that GBDTs are much better than NNs at handling skewed or heavy-tailed feature distributions and other forms of dataset irregularities. Our insights act as a guide for practitioners to determine which techniques may work best on their dataset. Finally, with the goal of accelerating tabular data research, we release the TabZilla Benchmark Suite: a collection of the 36 'hardest' of the datasets we study. Our benchmark suite, codebase, and all raw results are available at https://github.com/naszilla/tabzilla.
****
Adversarially Robust Distributed Count Tracking via Partial Differential Privacy
Zhongzheng Xiong, Xiaoyi Zhu, zengfeng Huang
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Energy-Based Cross Attention for Bayesian Context Update in Text-to-Image Diffusion Models
Geon Yeong Park, Jeongsol Kim, Beomsu Kim, Sang Wan Lee, Jong Chul Ye
Despite the remarkable performance of text-to-image diffusion models in image generation tasks, recent studies have raised the issue that generated images sometimes cannot capture the intended semantic contents of the text prompts, which phenomenon is often called semantic misalignment. To address this, here we present a novel energy-based model (EBM) framework for adaptive context control by modeling the posterior of context vectors. Specifically, we first formulate EBMs of latent image representations and text embeddings in each cross-attention layer of the denoising autoencoder. Then, we obtain the gradient of the log posterior of context vectors, which can be updated and transferred to the subsequent cross-attention layer, thereby implicitly minimizing a nested hierarchy of energy functions. Our latent EBMs further allow zero-shot compositional generation as a linear combination of cross-attention outputs from different contexts. Using extensive experiments, we demonstrate that the proposed method is highly effective in handling various image generation tasks, including multi-concept generation, text-guided image inpainting, and real and synthetic image editing. Code: https://github.com/EnergyAttention/Energy-Based-CrossAttention.
****
Optimal Algorithms for the Inhomogeneous Spiked Wigner Model
Aleksandr Pak, Justin Ko, Florent Krzakala
We study a spiked Wigner problem with an inhomogeneous noise profile. Our aim in this problem is to recover the signal passed through an inhomogeneous low-rank matrix channel. While the information-theoretic performances are well-known, we focus on the algorithmic problem. First, we derive an approximate message-passing algorithm (AMP) for the inhomogeneous problem and show that its rigorous state evolution coincides with the information-theoretic optimal Bayes fixed-point equations. Second, we deduce a simple and efficient spectral method that outperfor

ms PCA and is shown to match the information-theoretic transition.
****
Learning Repeatable Speech Embeddings Using An Intra-class Correlation Regularizer
Jianwei Zhang, Suren Jayasuriya, Visar Berisha
A good supervised embedding for a specific machine learning task is only sensitive to changes in the label of interest and is invariant to other confounding factors. We leverage the concept of repeatability from measurement theory to describe this property and propose to use the intra-class correlation coefficient (ICC) to evaluate the repeatability of embeddings. We then propose a novel regularizer, the ICC regularizer, as a complementary component for contrastive losses to guide deep neural networks to produce embeddings with higher repeatability. We use simulated data to explain why the ICC regularizer works better on minimizing the intra-class variance than the contrastive loss alone. We implement the ICC regularizer and apply it to three speech tasks: speaker verification, voice style conversion, and a clinical application for detecting dysphonic voice. The experimental results demonstrate that adding an ICC regularizer can improve the repeatability of learned embeddings compared to only using the contrastive loss; further, these embeddings lead to improved performance in these downstream tasks.
****
Momentum Provably Improves Error Feedback!
Ilyas Fatkhullin, Alexander Tyurin, Peter Richtarik
Due to the high communication overhead when training machine learning models in a distributed environment, modern algorithms invariably rely on lossy communication compression. However, when untreated, the errors caused by compression propagate, and can lead to severely unstable behavior, including exponential divergence. Almost a decade ago, Seide et al. [2014] proposed an error feedback (EF) mechanism, which we refer to as EF14, as an immensely effective heuristic for mitigating this issue. However, despite steady algorithmic and theoretical advances in the EF field in the last decade, our understanding is far from complete. In this work we address one of the most pressing issues. In particular, in the canonical nonconvex setting, all known variants of EF rely on very large batch sizes to converge, which can be prohibitive in practice. We propose a surprisingly simple fix which removes this issue both theoretically, and in practice: the application of Polyak's momentum to the latest incarnation of EF due to Richtárik et al. [2021] known as EF21. Our algorithm, for which we coin the name EF21-SGDM, improves the communication and sample complexities of previous error feedback algorithms under standard smoothness and bounded variance assumptions, and does not require any further strong assumptions such as bounded gradient dissimilarity. Moreover, we propose a double momentum version of our method that improves the complexities even further. Our proof seems to be novel even when compression is removed form the method, and as such, our proof technique is of independent interest in the study of nonconvex stochastic optimization enriched with Polyak's momentum.
****
Optimal Convergence Rate for Exact Policy Mirror Descent in Discounted Markov Decision Processes
Emmeran Johnson, Ciara Pike-Burke, Patrick Rebeschini
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Diff-Instruct: A Universal Approach for Transferring Knowledge From Pre-trained Diffusion Models
Weijian Luo, Tianyang Hu, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, Zhihua Zhang
Due to the ease of training, ability to scale, and high sample quality, diffusion models (DMs) have become the preferred option for generative modeling, with numerous pre-trained models available for a wide variety of datasets. Containing intricate information about data distributions, pre-trained DMs are valuable asse

ts for downstream applications. In this work, we consider learning from pre-trained DMs and transferring their knowledge to other generative models in a data-free fashion. Specifically, we propose a general framework called Diff-Instruct to instruct the training of arbitrary generative models as long as the generated samples are differentiable with respect to the model parameters. Our proposed Diff-Instruct is built on a rigorous mathematical foundation where the instruction process directly corresponds to minimizing a novel divergence we call Integral Kullback-Leibler (IKL) divergence. IKL is tailored for DMs by calculating the integral of the KL divergence along a diffusion process, which we show to be more robust in comparing distributions with misaligned supports. We also reveal non-trivial connections of our method to existing works such as DreamFusion \citep{poole2022dreamfusion}, and generative adversarial training. To demonstrate the effectiveness and universality of Diff-Instruct, we consider two scenarios: distilling pre-trained diffusion models and refining existing GAN models. The experiments on distilling pre-trained diffusion models show that Diff-Instruct results in state-of-the-art single-step diffusion-based models. The experiments on refining GAN models show that the Diff-Instruct can consistently improve the pre-trained generators of GAN models across various settings. Our official code is released through \url{https://github.com/pkulwj1994/diff_instruct}.
****
Characterization of Overfitting in Robust Multiclass Classification
Jingyuan Xu, Weiwei Liu
This paper considers the following question: Given the number of classes m, the number of robust accuracy queries k, and the number of test examples in the dataset n, how much can adaptive algorithms robustly overfit the test dataset? We solve this problem by equivalently giving near-matching upper and lower bounds of the robust overfitting bias in multiclass classification problems.
****
Expanding Small-Scale Datasets with Guided Imagination
Yifan Zhang, Daquan Zhou, Bryan Hooi, Kai Wang, Jiashi Feng
The power of DNNs relies heavily on the quantity and quality of training data. However, collecting and annotating data on a large scale is often expensive and time-consuming. To address this issue, we explore a new task, termed dataset expansion, aimed at expanding a ready-to-use small dataset by automatically creating new labeled samples. To this end, we present a Guided Imagination Framework (GIF) that leverages cutting-edge generative models like DALL-E2 and Stable Diffusion (SD) to "imagine" and create informative new data from the input seed data. Specifically, GIF conducts data imagination by optimizing the latent features of the seed data in the semantically meaningful space of the prior model, resulting in the creation of photo-realistic images with new content. To guide the imagination towards creating informative samples for model training, we introduce two key criteria, i.e., class-maintained information boosting and sample diversity promotion. These criteria are verified to be essential for effective dataset expansion: GIF-SD obtains 13.5% higher model accuracy on natural image datasets than unguided expansion with SD. With these essential criteria, GIF successfully expands small datasets in various scenarios, boosting model accuracy by 36.9% on average over six natural image datasets and by 13.5% on average over three medical datasets. The source code is available at https://github.com/Vanint/DatasetExpansion.
****
Parallel-mentoring for Offline Model-based Optimization
Can (Sam) Chen, Christopher Beckham, Zixuan Liu, Xue (Steve) Liu, Chris Pal
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Nominality Score Conditioned Time Series Anomaly Detection by Point/Sequential Reconstruction
Chih-Yu (Andrew) Lai, Fan-Keng Sun, Zhengqi Gao, Jeffrey H Lang, Duane Boning

Time series anomaly detection is challenging due to the complexity and variety o
f patterns that can occur. One major difficulty arises from modeling time-depend
ent relationships to find contextual anomalies while maintaining detection accur
acy for point anomalies. In this paper, we propose a framework for unsupervised
time series anomaly detection that utilizes point-based and sequence-based recon
struction models. The point-based model attempts to quantify point anomalies, an
d the sequence-based model attempts to quantify both point and contextual anomal
ies. Under the formulation that the observed time point is a two-stage deviated
value from a nominal time point, we introduce a nominality score calculated from
the ratio of a combined value of the reconstruction errors. We derive an induce
d anomaly score by further integrating the nominality score and anomaly score, t
hen theoretically prove the superiority of the induced anomaly score over the or
iginal anomaly score under certain conditions. Extensive studies conducted on se
veral public datasets show that the proposed framework outperforms most state-of
-the-art baselines for time series anomaly detection.
****
Frequency-domain MLPs are More Effective Learners in Time Series Forecasting
Kun Yi, Qi Zhang, Wei Fan, Shoujin Wang, Pengyang Wang, Hui He, Ning An, Defu Li
an, Longbing Cao, Zhendong Niu
Time series forecasting has played the key role in different industrial, includi
ng finance, traffic, energy, and healthcare domains. While existing literatures
have designed many sophisticated architectures based on RNNs, GNNs, or Transform
ers, another kind of approaches based on multi-layer perceptrons (MLPs) are prop
osed with simple structure, low complexity, and superior performance. However, m
ost MLP-based forecasting methods suffer from the point-wise mappings and inform
ation bottleneck, which largely hinders the forecasting performance. To overcome
this problem, we explore a novel direction of applying MLPs in the frequency do
main for time series forecasting. We investigate the learned patterns of frequen
cy-domain MLPs and discover their two inherent characteristic benefiting forecas
ting, (i) global view: frequency spectrum makes MLPs own a complete view for sig
nals and learn global dependencies more easily, and (ii) energy compaction: freq
uency-domain MLPs concentrate on smaller key part of frequency components with c
ompact signal energy. Then, we propose FreTS, a simple yet effective architectur
e built upon Frequency-domain MLPs for Time Series forecasting. FreTS mainly inv
olves two stages, (i) Domain Conversion, that transforms time-domain signals int
o complex numbers of frequency domain; (ii) Frequency Learning, that performs ou
r redesigned MLPs for the learning of real and imaginary part of frequency compo
nents. The above stages operated on both inter-series and intra-series scales fu
rther contribute to channel-wise and time-wise dependency learning. Extensive ex
periments on 13 real-world benchmarks (including 7 benchmarks for short-term for
ecasting and 6 benchmarks for long-term forecasting) demonstrate our consistent
superiority over state-of-the-art methods. Code is available at this repository:
https://github.com/aikunyi/FreTS.
****
Q-DM: An Efficient Low-bit Quantized Diffusion Model
Yanjing Li, Sheng Xu, Xianbin Cao, Xiao Sun, Baochang Zhang
Denoising diffusion generative models are capable of generating high-quality dat
a, but suffers from the computation-costly generation process, due to a iterativ
e noise estimation using full-precision networks. As an intuitive solution, qua
ntization can significantly reduce the computational and memory consumption by l
ow-bit parameters and operations. However, low-bit noise estimation networks in
diffusion models (DMs) remain unexplored yet and perform much worse than the ful
l-precision counterparts as observed in our experimental studies. In this paper,
we first identify that the bottlenecks of low-bit quantized DMs come from a lar
ge distribution oscillation on activations and accumulated quantization error
caused by the multi-step denoising process. To address these issues, we first de
velop a Timestep-aware Quantization (TaQ) method and a Noise-estimating Mimickin
g (NeM) scheme for low-bit quantized DMs (Q-DM) to effectively eliminate such os
cillation and accumulated error respectively, leading to well-performed low-bit
DMs. In this way, we propose an efficient Q-DM to calculate low-bit DMs by consi

dering both training and inference process in the same framework. We evaluate ou
r methods on popular DDPM and DDIM models. Extensive experimental results show t
hat our method achieves a much better performance than the prior arts. For examp
le, the 4-bit Q-DM theoretically accelerates the 1000-step DDPM by 7.8x and achi
eves a FID score of 5.17, on the unconditional CIFAR-10 dataset.
****
Beyond Exponential Graph: Communication-Efficient Topologies for Decentralized L
earning via Finite-time Convergence
Yuki Takezawa, Ryoma Sato, Han Bao, Kenta Niwa, Makoto Yamada
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Alternating Updates for Efficient Transformers
Cenk Baykal, Dylan Cutler, Nishanth Dikkala, Nikhil Ghosh, Rina Panigrahy, Xin W
ang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Interpretable Prototype-based Graph Information Bottleneck
Sangwoo Seo, Sungwon Kim, Chanyoung Park
The success of Graph Neural Networks (GNNs) has led to a need for understanding
their decision-making process and providing explanations for their predictions,
which has given rise to explainable AI (XAI) that offers transparent explanation
s for black-box models. Recently, the use of prototypes has successfully improve
d the explainability of models by learning prototypes to imply training graphs t
hat affect the prediction. However, these approaches tend to provide prototypes
with excessive information from the entire graph, leading to the exclusion of ke
y substructures or the inclusion of irrelevant substructures, which can limit bo
th the interpretability and the performance of the model in downstream tasks. In
 this work, we propose a novel framework of explainable GNNs, called interpretab
le Prototype-based Graph Information Bottleneck (PGIB) that incorporates prototy
pe learning within the information bottleneck framework to provide prototypes wi
th the key subgraph from the input graph that is important for the model predict
ion. This is the first work that incorporates prototype learning into the proces
s of identifying the key subgraphs that have a critical impact on the prediction
 performance. Extensive experiments, including qualitative analysis, demonstrate
 that PGIB outperforms state-of-the-art methods in terms of both prediction perf
ormance and explainability.
****
Self-Chained Image-Language Model for Video Localization and Question Answering
Shoubin Yu, Jaemin Cho, Prateek Yadav, Mohit Bansal
Recent studies have shown promising results on utilizing large pre-trained image
-language models for video question answering. While these image-language models
 can efficiently bootstrap the representation learning of video-language models,
 they typically concatenate uniformly sampled video frames as visual inputs with
out explicit language-aware, temporal modeling. When only a portion of a video i
nput is relevant to the language query, such uniform frame sampling can often le
ad to missing important visual cues. Although humans often find a video moment t
o focus on and rewind the moment to answer questions, training a query-aware vid
eo moment localizer often requires expensive annotations and high computational
costs. To address this issue, we propose Self-Chained Video Localization-Answeri
ng (SeViLA), a novel framework that leverages a single image-language model (BLI
P- 2) to tackle both temporal keyframe localization and question answering on vi
deos. SeViLA framework consists of two modules: Localizer and Answerer, where bo
th are parameter-efficiently fine-tuned from BLIP-2. We propose two ways of chai
ning these modules for cascaded inference and self-refinement. First, in the for

ward chain, the Localizer finds multiple language-aware keyframes in a video, wh
ich the Answerer uses to predict the answer. Second, in the reverse chain, the A
nswerer generates keyframe pseudo-labels to refine the Localizer, alleviating th
e need for expensive video moment localization annotations. Our SeViLA framework
 outperforms several strong baselines/previous works on five challenging video q
uestion answering and event prediction benchmarks, and achieves the state-of-the
-art in both fine-tuning (NExT-QA and STAR) and zero-shot (NExT-QA, STAR, How2QA
, and VLEP) settings. We show a comprehensive analysis of our framework, includi
ng the impact of Localizer, comparisons of Localizer with other temporal localiz
ation models, pre-training/self-refinement of Localizer, and varying the number
of keyframes.
****
The Tunnel Effect: Building Data Representations in Deep Neural Networks
Wojciech Masarczyk, Mateusz Ostaszewski, Ehsan Imani, Razvan Pascanu, Piotr Mi■o
■, Tomasz Trzcinski
Deep neural networks are widely known for their remarkable effectiveness across
various tasks, with the consensus that deeper networks implicitly learn more com
plex data representations. This paper shows that sufficiently deep networks trai
ned for supervised image classification split into two distinct parts that contr
ibute to the resulting data representations differently. The initial layers crea
te linearly-separable representations, while the subsequent layers, which we ref
er to as \textit{the tunnel}, compress these representations and have a minimal
impact on the overall performance. We explore the tunnel's behavior through comp
rehensive empirical studies, highlighting that it emerges early in the training
process. Its depth depends on the relation between the network's capacity and ta
sk complexity. Furthermore, we show that the tunnel degrades out-of-distribution
 generalization and discuss its implications for continual learning.
****
Restart Sampling for Improving Generative Processes
Yilun Xu, Mingyang Deng, Xiang Cheng, Yonglong Tian, Ziming Liu, Tommi Jaakkola
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Constructing Non-isotropic Gaussian Diffusion Model Using Isotropic Gaussian Dif
fusion Model for Image Editing
Xi Yu, Xiang Gu, Haozhi Liu, Jian Sun
Score-based diffusion models (SBDMs) have achieved state-of-the-art results in i
mage generation. In this paper, we propose a Non-isotropic Gaussian Diffusion Mo
del (NGDM) for image editing, which requires editing the source image while pres
erving the image regions irrelevant to the editing task. We construct NGDM by ad
ding independent Gaussian noises with different variances to different image pix
els. Instead of specifically training the NGDM, we rectify the NGDM into an isot
ropic Gaussian diffusion model with different pixels having different total forw
ard diffusion time. We propose to reverse the diffusion by designing a sampling
method that starts at different time for different pixels for denoising to gener
ate images using the pre-trained isotropic Gaussian diffusion model. Experimenta
l results show that NGDM achieves state-of-the-art performance for image editing
 tasks, considering the trade-off between the fidelity to the source image and a
lignment with the desired editing target.
****
Flocks of Stochastic Parrots: Differentially Private Prompt Learning for Large L
anguage Models
Haonan Duan, Adam Dziedzic, Nicolas Papernot, Franziska Boenisch
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****

Dataset Diffusion: Diffusion-based Synthetic Data Generation for Pixel-Level Sem antic Segmentation
Quang Nguyen, Truong Vu, Anh Tran, Khoi Nguyen
Preparing training data for deep vision models is a labor-intensive task. To add ress this, generative models have emerged as an effective solution for generatin g synthetic data. While current generative models produce image-level category l abels, we propose a novel method for generating pixel-level semantic segmentatio n labels using the text-to-image generative model Stable Diffusion (SD). By util izing the text prompts, cross-attention, and self-attention of SD, we introduce three new techniques: class-prompt appending, class-prompt cross-attention, and self-attention exponentiation. These techniques enable us to generate segmentati on maps corresponding to synthetic images. These maps serve as pseudo-labels for training semantic segmenters, eliminating the need for labor-intensive pixel-wi se annotation. To account for the imperfections in our pseudo-labels, we incorpo rate uncertainty regions into the segmentation, allowing us to disregard loss fr om those regions. We conduct evaluations on two datasets, PASCAL VOC and MSCOCO, and our approach significantly outperforms concurrent work. Our benchmarks and code will be released at https://github.com/VinAIResearch/Dataset-Diffusion.
****
ASL Citizen: A Community-Sourced Dataset for Advancing Isolated Sign Language Re cognition
Aashaka Desai, Lauren Berger, Fyodor Minakov, Nessa Milano, Chinmay Singh, Krist on Pumphrey, Richard Ladner, Hal Daumé III, Alex X Lu, Naomi Caselli, Danielle B ragg
Sign languages are used as a primary language by approximately 70 million D/deaf people world-wide. However, most communication technologies operate in spoken a nd written languages, creating inequities in access. To help tackle this problem , we release ASL Citizen, the first crowdsourced Isolated Sign Language Recognit ion (ISLR) dataset, collected with consent and containing 83,399 videos for 2,73 1 distinct signs filmed by 52 signers in a variety of environments. We propose t hat this dataset be used for sign language dictionary retrieval for American Sig n Language (ASL), where a user demonstrates a sign to their webcam to retrieve m atching signs from a dictionary. We show that training supervised machine learni ng classifiers with our dataset advances the state-of-the-art on metrics relevan t for dictionary retrieval, achieving 63\% accuracy and a recall-at-10 of 91\%, evaluated entirely on videos of users who are not present in the training or val idation sets.
****
Learning-to-Rank Meets Language: Boosting Language-Driven Ordering Alignment for Ordinal Classification
Rui Wang, Peipei Li, Huaibo Huang, Chunshui Cao, Ran He, Zhaofeng He
We present a novel language-driven ordering alignment method for ordinal classif ication. The labels in ordinal classification contain additional ordering relati ons, making them prone to overfitting when relying solely on training data. Rece nt developments in pre-trained vision-language models inspire us to leverage the rich ordinal priors in human language by converting the original task into a vi sion-language alignment task. Consequently, we propose L2RCLIP, which fully util izes the language priors from two perspectives. First, we introduce a complement ary prompt tuning technique called RankFormer, designed to enhance the ordering relation of original rank prompts. It employs token-level attention with residua l-style prompt blending in the word embedding space. Second, to further incorpor ate language priors, we revisit the approximate bound optimization of vanilla cr oss-entropy loss and restructure it within the cross-modal embedding space. Cons equently, we propose a cross-modal ordinal pairwise loss to refine the CLIP feat ure space, where texts and images maintain both semantic alignment and ordering alignment. Extensive experiments on three ordinal classification tasks, includi ng facial age estimation, historical color image (HCI) classification, and aesth etic assessment demonstrate its promising performance.
****
GAUCHE: A Library for Gaussian Processes in Chemistry

Ryan-Rhys Griffiths, Leo Klarner, Henry Moss, Aditya Ravuri, Sang Truong, Yuanqi Du, Samuel Stanton, Gary Tom, Bojana Rankovic, Arian Jamasb, Aryan Deshwal, Julius Schwartz, Austin Tripp, Gregory Kell, Simon Frieder, Anthony Bourached, Alex Chan, Jacob Moss, Chengzhi Guo, Johannes Peter Dürholt, Saudamini Chaurasia, Ji Won Park, Felix Strieth-Kalthoff, Alpha Lee, Bingqing Cheng, Alan Aspuru-Guzik, Philippe Schwaller, Jian Tang
We introduce GAUCHE, an open-source library for GAUssian processes in CHEmistry. Gaussian processes have long been a cornerstone of probabilistic machine learning, affording particular advantages for uncertainty quantification and Bayesian optimisation. Extending Gaussian processes to molecular representations, however, necessitates kernels defined over structured inputs such as graphs, strings and bit vectors. By providing such kernels in a modular, robust and easy-to-use framework, we seek to enable expert chemists and materials scientists to make use of state-of-the-art black-box optimization techniques. Motivated by scenarios frequently encountered in practice, we showcase applications for GAUCHE in molecular discovery, chemical reaction optimisation and protein design. The codebase is made available at https://github.com/leojklarner/gauche.
****
Exponentially Convergent Algorithms for Supervised Matrix Factorization
Joowon Lee, Hanbaek Lyu, Weixin Yao
Supervised matrix factorization (SMF) is a classical machine learning method that simultaneously seeks feature extraction and classification tasks, which are not necessarily a priori aligned objectives. Our goal is to use SMF to learn low-rank latent factors that offer interpretable, data-reconstructive, and class-discriminative features, addressing challenges posed by high-dimensional data. Training SMF model involves solving a nonconvex and possibly constrained optimization with at least three blocks of parameters. Known algorithms are either heuristic or provide weak convergence guarantees for special cases. In this paper, we provide a novel framework that `lifts' SMF as a low-rank matrix estimation problem in a combined factor space and propose an efficient algorithm that provably converges exponentially fast to a global minimizer of the objective with arbitrary initialization under mild assumptions. Our framework applies to a wide range of SMF-type problems for multi-class classification with auxiliary features. To showcase an application, we demonstrate that our algorithm successfully identified well-known cancer-associated gene groups for various cancers.
****
InfoCD: A Contrastive Chamfer Distance Loss for Point Cloud Completion
Fangzhou Lin, Yun Yue, Ziming Zhang, Songlin Hou, Kazunori Yamada, Vijaya Kolachalama, Venkatesh Saligrama
A point cloud is a discrete set of data points sampled from a 3D geometric surface. Chamfer distance (CD) is a popular metric and training loss to measure the distances between point clouds, but also well known to be sensitive to outliers. To address this issue, in this paper we propose InfoCD, a novel contrastive Chamfer distance loss to learn to spread the matched points for better distribution alignments between point clouds as well as accounting for a surface similarity estimator. We show that minimizing InfoCD is equivalent to maximizing a lower bound of the mutual information between the underlying geometric surfaces represented by the point clouds, leading to a regularized CD metric which is robust and computationally efficient for deep learning. We conduct comprehensive experiments for point cloud completion using InfoCD and observe significant improvements consistently over all the popular baseline networks trained with CD-based losses, leading to new state-of-the-art results on several benchmark datasets. Demo code is available at https://github.com/Zhang-VISLab/NeurIPS2023-InfoCD.
****
Differentially Private Statistical Inference through β-Divergence One Posterior Sampling
Jack E. Jewson, Sahra Ghalebikesabi, Chris C Holmes
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth

ors prior to requesting a name change in the electronic proceedings.
****
Doubly Robust Augmented Transfer for Meta-Reinforcement Learning
Yuankun Jiang, Nuowen Kan, Chenglin Li, Wenrui Dai, Junni Zou, Hongkai Xiong
Meta-reinforcement learning (Meta-RL), though enabling a fast adaptation to learn new skills by exploiting the common structure shared among different tasks, suffers performance degradation in the sparse-reward setting. Current hindsight-based sample transfer approaches can alleviate this issue by transferring relabeled trajectories from other tasks to a new task so as to provide informative experience for the target reward function, but are unfortunately constrained with the unrealistic assumption that tasks differ only in reward functions. In this paper, we propose a doubly robust augmented transfer (DRaT) approach, aiming at addressing the more general sparse reward meta-RL scenario with both dynamics mismatches and varying reward functions across tasks. Specifically, we design a doubly robust augmented estimator for efficient value-function evaluation, which tackles dynamics mismatches with the optimal importance weight of transition distributions achieved by minimizing the theoretically derived upper bound of mean squared error (MSE) between the estimated values of transferred samples and their true values in the target task. Due to its intractability, we then propose an interval-based approximation to this optimal importance weight, which is guaranteed to cover the optimum with a constrained and sample-independent upper bound on the MSE approximation error. Based on our theoretical findings, we finally develop a DRaT algorithm for transferring informative samples across tasks during the training of meta-RL. We implement DRaT on an off-policy meta-RL baseline, and empirically show that it significantly outperforms other hindsight-based approaches on various sparse-reward MuJoCo locomotion tasks with varying dynamics and reward functions.
****
Evaluating Open-QA Evaluation
Cunxiang Wang, Sirui Cheng, Qipeng Guo, Yuanhao Yue, Bowen Ding, Zhikun Xu, Yidong Wang, Xiangkun Hu, Zheng Zhang, Yue Zhang
This study focuses on the evaluation of the Open Question Answering (Open-QA) task, which can directly estimate the factuality of large language models (LLMs). Current automatic evaluation methods have shown limitations, indicating that human evaluation still remains the most reliable approach. We introduce a new task, QA Evaluation (QA-Eval) and the corresponding dataset EVOUNA, designed to assess the accuracy of AI-generated answers in relation to standard answers within Open-QA. Our evaluation of these methods utilizes human-annotated results to measure their performance. Specifically, the work investigates methods that show high correlation with human evaluations, deeming them more reliable. We also discuss the pitfalls of current methods and methods to improve LLM-based evaluators. We believe this new QA-Eval task and corresponding dataset EVOUNA will facilitate the development of more effective automatic evaluation tools and prove valuable for future research in this area. All resources are available at https://github.com/wangcunxiang/QA-Eval and it is under the Apache-2.0 License.
****
Efficiently incorporating quintuple interactions into geometric deep learning force fields
Zun Wang, Guoqing Liu, Yichi Zhou, Tong Wang, Bin Shao
Machine learning force fields (MLFFs) have instigated a groundbreaking shift in molecular dynamics (MD) simulations across a wide range of fields, such as physics, chemistry, biology, and materials science. Incorporating higher order many-body interactions can enhance the expressiveness and accuracy of models. Recent models have achieved this by explicitly including up to four-body interactions. However, five-body interactions, which have relevance in various fields, are still challenging to incorporate efficiently into MLFFs. In this work, we propose the quintuple network (QuinNet), an end-to-end graph neural network that efficiently expresses many-body interactions up to five-body interactions with \emph{ab initio} accuracy. By analyzing the topology of diverse many-body interactions, we design the model architecture to efficiently and explicitly represent these int

eractions. We evaluate QuinNet on public datasets of small molecules, such as MD 17 and its revised version, and show that it is compatible with other state-of-the-art models on these benchmarks. Moreover, QuinNet surpasses many leading models on larger and more complex molecular systems, such as MD22 and Chignolin, without increasing the computational complexity. We also use QuinNet as a force field for molecular dynamics (MD) simulations to demonstrate its accuracy and stability, and conduct an ablation study to elucidate the significance of five-body interactions. We open source our implementation at https://github.com/Zun-Wang/QuinNet.
****

Spectral Entry-wise Matrix Estimation for Low-Rank Reinforcement Learning
Stefan Stojanovic, Yassir Jedra, Alexandre Proutiere
We study matrix estimation problems arising in reinforcement learning with low-rank structure. In low-rank bandits, the matrix to be recovered specifies the expected arm rewards, and for low-rank Markov Decision Processes (MDPs), it characterizes the transition kernel of the MDP. In both cases, each entry of the matrix carries important information, and we seek estimation methods with low entry-wise prediction error. Importantly, these methods further need to accommodate for inherent correlations in the available data (e.g. for MDPs, the data consists of system trajectories). We investigate the performance of simple spectral-based matrix estimation approaches: we show that they efficiently recover the singular subspaces of the matrix and exhibit nearly-minimal entry-wise prediction error. These new results on low-rank matrix estimation make it possible to devise reinforcement learning algorithms that fully exploit the underlying low-rank structure. We provide two examples of such algorithms: a regret minimization algorithm for low-rank bandit problems, and a best policy identification algorithm for low-rank MDPs. Both algorithms yield state-of-the-art performance guarantees.
****

Function Space Bayesian Pseudocoreset for Bayesian Neural Networks
Balhae Kim, Hyungi Lee, Juho Lee
A Bayesian pseudocoreset is a compact synthetic dataset summarizing essential information of a large-scale dataset and thus can be used as a proxy dataset for scalable Bayesian inference. Typically, a Bayesian pseudocoreset is constructed by minimizing a divergence measure between the posterior conditioning on the pseudocoreset and the posterior conditioning on the full dataset. However, evaluating the divergence can be challenging, particularly for the models like deep neural networks having high-dimensional parameters. In this paper, we propose a novel Bayesian pseudocoreset construction method that operates on a function space. Unlike previous methods, which construct and match the coreset and full data posteriors in the space of model parameters (weights), our method constructs variational approximations to the coreset posterior on a function space and matches it to the full data posterior in the function space. By working directly on the function space, our method could bypass several challenges that may arise when working on a weight space, including limited scalability and multi-modality issue. Through various experiments, we demonstrate that the Bayesian pseudocoresets constructed from our method enjoys enhanced uncertainty quantification and better robustness across various model architectures.
****

One-step differentiation of iterative algorithms
Jerome Bolte, Edouard Pauwels, Samuel Vaiter
In appropriate frameworks, automatic differentiation is transparent to the user, at the cost of being a significant computational burden when the number of operations is large. For iterative algorithms, implicit differentiation alleviates this issue but requires custom implementation of Jacobian evaluation. In this paper, we study one-step differentiation, also known as Jacobian-free backpropagation, a method as easy as automatic differentiation and as performant as implicit differentiation for fast algorithms (e.g. superlinear optimization methods). We provide a complete theoretical approximation analysis with specific examples (Newton's method, gradient descent) along with its consequences in bilevel optimization. Several numerical examples illustrate the well-foundness of the one-step e

stimator.
****
Adaptive Principal Component Regression with Applications to Panel Data
Anish Agarwal, Keegan Harris, Justin Whitehouse, Steven Z. Wu
Principal component regression (PCR) is a popular technique for fixed-design error-in-variables regression, a generalization of the linear regression setting in which the observed covariates are corrupted with random noise. We provide the first time-uniform finite sample guarantees for online (regularized) PCR whenever data is collected adaptively. Since the proof techniques for PCR in the fixed design setting do not readily extend to the online setting, our results rely on adapting tools from modern martingale concentration to the error-in-variables setting. As an application of our bounds, we provide a framework for counterfactual estimation of unit-specific treatment effects in panel data settings when interventions are assigned adaptively. Our framework may be thought of as a generalization of the synthetic interventions framework where data is collected via an adaptive intervention assignment policy.
****
VisAlign: Dataset for Measuring the Alignment between AI and Humans in Visual Perception
Jiyoung Lee, Seungho Kim, Seunghyun Won, Joonseok Lee, Marzyeh Ghassemi, James Thorne, Jaeseok Choi, O-Kil Kwon, Edward Choi
AI alignment refers to models acting towards human-intended goals, preferences, or ethical principles. Analyzing the similarity between models and humans can be a proxy measure for ensuring AI safety. In this paper, we focus on the models' visual perception alignment with humans, further referred to as AI-human visual alignment. Specifically, we propose a new dataset for measuring AI-human visual alignment in terms of image classification. In order to evaluate AI-human visual alignment, a dataset should encompass samples with various scenarios and have gold human perception labels. Our dataset consists of three groups of samples, namely Must-Act (i.e., Must-Classify), Must-Abstain, and Uncertain, based on the quantity and clarity of visual information in an image and further divided into eight categories. All samples have a gold human perception label; even Uncertain (e.g., severely blurry) sample labels were obtained via crowd-sourcing. The validity of our dataset is verified by sampling theory, statistical theories related to survey design, and experts in the related fields. Using our dataset, we analyze the visual alignment and reliability of five popular visual perception models and seven abstention methods. Our code and data is available at https://github.com/jiyounglee-0523/VisAlign.
****
The Best of Both Worlds in Network Population Games: Reaching Consensus and Convergence to Equilibrium
Shuyue Hu, Harold Soh, Georgios Piliouras
Reaching consensus and convergence to equilibrium are two major challenges of multi-agent systems. Although each has attracted significant attention, relatively few studies address both challenges at the same time. This paper examines the connection between the notions of consensus and equilibrium in a multi-agent system where multiple interacting sub-populations coexist. We argue that consensus can be seen as an intricate component of intra-population stability, whereas equilibrium can be seen as encoding inter-population stability. We show that smooth fictitious play, a well-known learning model in game theory, can achieve both consensus and convergence to equilibrium in diverse multi-agent settings. Moreover, we show that the consensus formation process plays a crucial role in the seminal thorny problem of equilibrium selection in multi-agent learning.
****
L-CAD: Language-based Colorization with Any-level Descriptions using Diffusion Priors
zheng chang, Shuchen Weng, Peixuan Zhang, Yu Li, Si Li, Boxin Shi
Language-based colorization produces plausible and visually pleasing colors under the guidance of user-friendly natural language descriptions. Previous methods implicitly assume that users provide comprehensive color descriptions for most o

f the objects in the image, which leads to suboptimal performance. In this paper, we propose a unified model to perform language-based colorization with any-level descriptions. We leverage the pretrained cross-modality generative model for its robust language understanding and rich color priors to handle the inherent ambiguity of any-level descriptions. We further design modules to align with input conditions to preserve local spatial structures and prevent the ghosting effect. With the proposed novel sampling strategy, our model achieves instance-aware colorization in diverse and complex scenarios. Extensive experimental results demonstrate our advantages of effectively handling any-level descriptions and outperforming both language-based and automatic colorization methods. The code and pretrained modelsare available at: https://github.com/changzheng123/L-CAD.
****
Convolutional Neural Operators for robust and accurate learning of PDEs
Bogdan Raonic, Roberto Molinaro, Tim De Ryck, Tobias Rohner, Francesca Bartolucci, Rima Alaifari, Siddhartha Mishra, Emmanuel de Bézenac
Although very successfully used in conventional machine learning, convolution based neural network architectures -- believed to be inconsistent in function space -- have been largely ignored in the context of learning solution operators of PDEs. Here, we present novel adaptations for convolutional neural networks to demonstrate that they are indeed able to process functions as inputs and outputs. The resulting architecture, termed as convolutional neural operators (CNOs), is designed specifically to preserve its underlying continuous nature, even when implemented in a discretized form on a computer. We prove a universality theorem to show that CNOs can approximate operators arising in PDEs to desired accuracy. CNOs are tested on a novel suite of benchmarks, encompassing a diverse set of PDEs with multi-scale solutions and are observed to significantly outperform baselines, paving the way for an alternative framework for robust and accurate operator learning.
****
Neural Image Compression: Generalization, Robustness, and Spectral Biases
Kelsey Lieberman, James Diffenderfer, Charles Godfrey, Bhavya Kailkhura
Recent advances in neural image compression (NIC) have produced models that are starting to outperform classic codecs. While this has led to growing excitement about using NIC in real-world applications, the successful adoption of any machine learning system in the wild requires it to generalize (and be robust) to unseen distribution shifts at deployment. Unfortunately, current research lacks comprehensive datasets and informative tools to evaluate and understand NIC performance in real-world settings. To bridge this crucial gap, first, this paper presents a comprehensive benchmark suite to evaluate the out-of-distribution (OOD) performance of image compression methods. Specifically, we provide CLIC-C and Kodak-C by introducing 15 corruptions to the popular CLIC and Kodak benchmarks. Next, we propose spectrally-inspired inspection tools to gain deeper insight into errors introduced by image compression methods as well as their OOD performance. We then carry out a detailed performance comparison of several classic codecs and NIC variants, revealing intriguing findings that challenge our current understanding of the strengths and limitations of NIC. Finally, we corroborate our empirical findings with theoretical analysis, providing an in-depth view of the OOD performance of NIC and its dependence on the spectral properties of the data. Our benchmarks, spectral inspection tools, and findings provide a crucial bridge to the real-world adoption of NIC. We hope that our work will propel future efforts in designing robust and generalizable NIC methods. Code and data will be made available at https://github.com/klieberman/ood_nic.
****
Estimating Koopman operators with sketching to provably learn large scale dynamical systems
Giacomo Meanti, Antoine Chatalic, Vladimir Kostic, Pietro Novelli, Massimiliano Pontil, Lorenzo Rosasco
The theory of Koopman operators allows to deploy non-parametric machine learning algorithms to predict and analyze complex dynamical systems.Estimators such as principal component regression (PCR) or reduced rank regression (RRR) in kernel

spaces can be shown to provably learn Koopman operators from finite empirical ob servations of the system's time evolution. Scaling these approaches to very long trajectories is a challenge and requires introducing suitable approximations to make computations feasible. In this paper, we boost the efficiency of different kernel-based Koopman operator estimators using random projections (sketching).W e derive, implement and test the new ``sketched'' estimators with extensive expe riments on synthetic and large-scale molecular dynamics datasets. Further, we es tablish non asymptotic error bounds giving a sharp characterization of the trade -offs between statistical learning rates and computational efficiency.Our empiri cal and theoretical analysis shows that the proposed estimators provide a sound and efficient way to learn large scale dynamical systems.In particular our exper iments indicate that the proposed estimators retain the same accuracy of PCR or RRR, while being much faster.
****

Self-Adaptive Motion Tracking against On-body Displacement of Flexible Sensors
Chengxu Zuo, Fang Jiawei, Shihui Guo, Yipeng Qin
Flexible sensors are promising for ubiquitous sensing of human status due to the ir flexibility and easy integration as wearable systems. However, on-body displa cement of sensors is inevitable since the device cannot be firmly worn at a fixe d position across different sessions. This displacement issue causes complicated patterns and significant challenges to subsequent machine learning algorithms. Our work proposes a novel self-adaptive motion tracking network to address this challenge. Our network consists of three novel components: i) a light-weight lea rnable Affine Transformation layer whose parameters can be tuned to efficiently adapt to unknown displacements; ii) a Fourier-encoded LSTM network for better pa ttern identification; iii) a novel sequence discrepancy loss equipped with auxil iary regressors for unsupervised tuning of Affine Transformation parameters.
****

Counterfactual Conservative Q Learning for Offline Multi-agent Reinforcement Lea rning
Jianzhun Shao, Yun Qu, Chen Chen, Hongchang Zhang, Xiangyang Ji
Offline multi-agent reinforcement learning is challenging due to the coupling ef fect of both distribution shift issue common in offline setting and the high dim ension issue common in multi-agent setting, making the action out-of-distributio n (OOD) and value overestimation phenomenon excessively severe. To mitigate this problem, we propose a novel multi-agent offline RL algorithm, named CounterFact ual Conservative Q-Learning (CFCQL) to conduct conservative value estimation. Ra ther than regarding all the agents as a high dimensional single one and directly applying single agent conservative methods to it, CFCQL calculates conservative regularization for each agent separately in a counterfactual way and then linea rly combines them to realize an overall conservative value estimation. We prove that it still enjoys the underestimation property and the performance guarantee as those single agent conservative methods do, but the induced regularization an d safe policy improvement bound are independent of the agent number, which is th erefore theoretically superior to the direct treatment referred to above, especi ally when the agent number is large. We further conduct experiments on four envi ronments including both discrete and continuous action settings on both existing and our man-made datasets, demonstrating that CFCQL outperforms existing method s on most datasets and even with a remarkable margin on some of them.
****

Black-Box Differential Privacy for Interactive ML
Haim Kaplan, Yishay Mansour, Shay Moran, Kobbi Nissim, Uri Stemmer
In this work we revisit an interactive variant of joint differential privacy, re cently introduced by Naor et al. [2023], and generalize it towards handling onli ne processes in which existing privacy definitions seem too restrictive. We stud y basic properties of this definition and demonstrate that it satisfies (suitabl e variants) of group privacy, composition, and post processing.In order to demon strate the advantages of this privacy definition compared to traditional forms o f differential privacy,we consider the basic setting of online classification. W e show that any (possibly non-private) learning rule can be effectively transfor

med to a private learning rule with only a polynomial overhead in the mistake bo
und. This demonstrates a stark difference with traditional forms of differential
 privacy, such as the one studied by Golowich and Livni [2021], where only a do
uble exponential overhead in the mistake bound is known (via an information theo
retic upper bound).
****
Mnemosyne: Learning to Train Transformers with Transformers
Deepali Jain, Krzysztof M Choromanski, Kumar Avinava Dubey, Sumeet Singh, Vikas
Sindhwani, Tingnan Zhang, Jie Tan
In this work, we propose a new class of learnable optimizers, called Mnemosyne.
It is based on the novel spatio-temporal low-rank implicit attention Transformer
s that can learn to train entire neural network architectures, including other T
ransformers, without any task-specific optimizer tuning. We show that Mnemosyne:
 (a) outperforms popular LSTM optimizers (also with new feature engineering to m
itigate catastrophic forgetting of LSTMs), (b) can successfully train Transforme
rs while using simple meta-training strategies that require minimal computationa
l resources, (c) matches accuracy-wise SOTA hand-designed optimizers with carefu
lly tuned hyper-parameters (often producing top performing models). Furthermore,
 Mnemosyne provides space complexity comparable to that of its hand-designed fir
st-order counterparts, which allows it to scale to training larger sets of param
eters. We conduct an extensive empirical evaluation of Mnemosyne on: (a) fine-tu
ning a wide range of Vision Transformers (ViTs) from medium-size architectures t
o massive ViT-Hs (36 layers, 16 heads), (b) pre-training BERT models and (c) sof
t prompt-tuning large 11B+ T5XXL models. We complement our results with a compre
hensive theoretical analysis of the compact associative memory used by Mnemosyne
 which we believe was never done before.
****
M^{2}Hub: Unlocking the Potential of Machine Learning for Materials Discovery
Yuanqi Du, Yingheng Wang, Yining Huang, Jianan Canal Li, Yanqiao Zhu, Tian Xie,
Chenru Duan, John Gregoire, Carla P. Gomes
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
PoET: A generative model of protein families as sequences-of-sequences
Timothy Truong Jr, Tristan Bepler
Generative protein language models are a natural way to design new proteins with
 desired functions. However, current models are either difficult to direct to pr
oduce a protein from a specific family of interest, or must be trained on a larg
e multiple sequence alignment (MSA) from the specific family of interest, making
 them unable to benefit from transfer learning across families. To address this,
 we propose Protein Evolutionary Transformer (PoET), an autoregressive generativ
e model of whole protein families that learns to generate sets of related protei
ns as sequences-of-sequences across tens of millions of natural protein sequence
 clusters. PoET can be used as a retrieval-augmented language model to generate
and score arbitrary modifications conditioned on any protein family of interest,
 and can extrapolate from short context lengths to generalize well even for smal
l families. This is enabled by a unique Transformer layer; we model tokens seque
ntially within sequences while attending between sequences order invariantly, al
lowing PoET to scale to context lengths beyond those used during training. In ex
tensive experiments on deep mutational scanning datasets, we show that PoET outp
erforms existing protein language models and evolutionary sequence models for va
riant function prediction across proteins of all MSA depths. We also demonstrate
 PoET's ability to controllably generate new protein sequences.
****
BQ-NCO: Bisimulation Quotienting for Efficient Neural Combinatorial Optimization
Darko Drakulic, Sofia Michel, Florian Mai, Arnaud Sors, Jean-Marc Andreoli
Despite the success of neural-based combinatorial optimization methods for end-t
o-end heuristic learning, out-of-distribution generalization remains a challenge

. In this paper, we present a novel formulation of Combinatorial Optimization Problems (COPs) as Markov Decision Processes (MDPs) that effectively leverages common symmetries of COPs to improve out-of-distribution robustness. Starting from a direct MDP formulation of a constructive method, we introduce a generic way to reduce the state space, based on Bisimulation Quotienting (BQ) in MDPs. Then, for COPs with a recursive nature, we specialize the bisimulation and show how the reduced state exploits the symmetries of these problems and facilitates MDP solving. Our approach is principled and we prove that an optimal policy for the proposed BQ-MDP actually solves the associated COPs. We illustrate our approach on five classical problems: the Euclidean and Asymmetric Traveling Salesman, Capacitated Vehicle Routing, Orienteering and Knapsack Problems. Furthermore, for each problem, we introduce a simple attention-based policy network for the BQ-MDPs, which we train by imitation of (near) optimal solutions of small instances from a single distribution. We obtain new state-of-the-art results for the five COPs on both synthetic and realistic benchmarks. Notably, in contrast to most existing neural approaches, our learned policies show excellent generalization performance to much larger instances than seen during training, without any additional search procedure. Our code is available at: link.
****
Turbulence in Focus: Benchmarking Scaling Behavior of 3D Volumetric Super-Resolution with BLASTNet 2.0 Data
Wai Tong Chung, Bassem Akoush, Pushan Sharma, Alex Tamkin, Ki Sung Jung, Jacqueline Chen, Jack Guo, Davy Brouzet, Mohsen Talei, Bruno Savard, Alexei Poludnenko, Matthias Ihme
Analysis of compressible turbulent flows is essential for applications related to propulsion, energy generation, and the environment. Here, we present BLASTNet 2.0, a 2.2 TB network-of-datasets containing 744 full-domain samples from 34 high-fidelity direct numerical simulations, which addresses the current limited availability of 3D high-fidelity reacting and non-reacting compressible turbulent flow simulation data. With this data, we benchmark a total of 49 variations of five deep learning approaches for 3D super-resolution - which can be applied for improving scientific imaging, simulations, turbulence models, as well as in computer vision applications. We perform neural scaling analysis on these models to examine the performance of different machine learning (ML) approaches, including two scientific ML techniques. We demonstrate that (i) predictive performance can scale with model size and cost, (ii) architecture matters significantly, especially for smaller models, and (iii) the benefits of physics-based losses can persist with increasing model size. The outcomes of this benchmark study are anticipated to offer insights that can aid the design of 3D super-resolution models, especially for turbulence models, while this data is expected to foster ML methods for a broad range of flow physics applications. This data is publicly available with download links and browsing tools consolidated at https://blastnet.github.io.
****
Neural Functional Transformers
Allan Zhou, Kaien Yang, Yiding Jiang, Kaylee Burns, Winnie Xu, Samuel Sokota, J. Zico Kolter, Chelsea Finn
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
LinkerNet: Fragment Poses and Linker Co-Design with 3D Equivariant Diffusion
Jiaqi Guan, Xingang Peng, PeiQi Jiang, Yunan Luo, Jian Peng, Jianzhu Ma
Targeted protein degradation techniques, such as PROteolysis TArgeting Chimeras (PROTACs), have emerged as powerful tools for selectively removing disease-causing proteins. One challenging problem in this field is designing a linker to connect different molecular fragments to form a stable drug-candidate molecule. Existing models for linker design assume that the relative positions of the fragments are known, which may not be the case in real scenarios. In this work, we addre

ss a more general problem where the poses of the fragments are unknown in 3D space. We develop a 3D equivariant diffusion model that jointly learns the generative process of both fragment poses and the 3D structure of the linker. By viewing fragments as rigid bodies, we design a fragment pose prediction module inspired by the Newton-Euler equations in rigid body mechanics. Empirical studies on ZINC and PROTAC-DB datasets demonstrate that our model can generate chemically valid, synthetically-accessible, and low-energy molecules under both unconstrained and constrained generation settings.
****
One Risk to Rule Them All: A Risk-Sensitive Perspective on Model-Based Offline Reinforcement Learning
Marc Rigter, Bruno Lacerda, Nick Hawes
Offline reinforcement learning (RL) is suitable for safety-critical domains where online exploration is not feasible. In such domains, decision-making should take into consideration the risk of catastrophic outcomes. In other words, decision-making should be risk-averse. An additional challenge of offline RL is avoiding distributional shift, i.e. ensuring that state-action pairs visited by the policy remain near those in the dataset. Previous offline RL algorithms that consider risk combine offline RL techniques (to avoid distributional shift), with risk-sensitive RL algorithms (to achieve risk-aversion). In this work, we propose risk-aversion as a mechanism to jointly address both of these issues. We propose a model-based approach, and use an ensemble of models to estimate epistemic uncertainty, in addition to aleatoric uncertainty. We train a policy that is risk-averse, and avoids high uncertainty actions. Risk-aversion to epistemic uncertainty prevents distributional shift, as areas not covered by the dataset have high epistemic uncertainty. Risk-aversion to aleatoric uncertainty discourages actions that are risky due to environment stochasticity. Thus, by considering epistemic uncertainty via a model ensemble and introducing risk-aversion, our algorithm (1R2R) avoids distributional shift in addition to achieving risk-aversion to aleatoric risk. Our experiments show that 1R2R achieves strong performance on deterministic benchmarks, and outperforms existing approaches for risk-sensitive objectives in stochastic domains.
****
Monarch Mixer: A Simple Sub-Quadratic GEMM-Based Architecture
Dan Fu, Simran Arora, Jessica Grogan, Isys Johnson, Evan Sabri Eyuboglu, Armin Thomas, Benjamin Spector, Michael Poli, Atri Rudra, Christopher Ré
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Bypassing spike sorting: Density-based decoding using spike localization from dense multielectrode probes
Yizi Zhang, Tianxiao He, Julien Boussard, Charles Windolf, Olivier Winter, Eric Trautmann, Noam Roth, Hailey Barrell, Mark Churchland, Nicholas A Steinmetz, Erdem Varol, Cole Hurwitz, Liam Paninski
Neural decoding and its applications to brain computer interfaces (BCI) are essential for understanding the association between neural activity and behavior. A prerequisite for many decoding approaches is spike sorting, the assignment of action potentials (spikes) to individual neurons. Current spike sorting algorithms, however, can be inaccurate and do not properly model uncertainty of spike assignments, therefore discarding information that could potentially improve decoding performance. Recent advances in high-density probes (e.g., Neuropixels) and computational methods now allow for extracting a rich set of spike features from unsorted data; these features can in turn be used to directly decode behavioral correlates. To this end, we propose a spike sorting-free decoding method that directly models the distribution of extracted spike features using a mixture of Gaussians (MoG) encoding the uncertainty of spike assignments, without aiming to solve the spike clustering problem explicitly. We allow the mixing proportion of the MoG to change over time in response to the behavior and develop variational i

nference methods to fit the resulting model and to perform decoding. We benchmar k our method with an extensive suite of recordings from different animals and pr obe geometries, demonstrating that our proposed decoder can consistently outperf orm current methods based on thresholding (i.e. multi-unit activity) and spike s orting. Open source code is available at https://github.com/yzhang511/density_de coding.

****

SA-Solver: Stochastic Adams Solver for Fast Sampling of Diffusion Models

Shuchen Xue, Mingyang Yi, Weijian Luo, Shifeng Zhang, Jiacheng Sun, Zhenguo Li, Zhi-Ming Ma

Diffusion Probabilistic Models (DPMs) have achieved considerable success in gene ration tasks. As sampling from DPMs is equivalent to solving diffusion SDE or OD E which is time-consuming, numerous fast sampling methods built upon improved di fferential equation solvers are proposed. The majority of such techniques consid er solving the diffusion ODE due to its superior efficiency. However, stochastic sampling could offer additional advantages in generating diverse and high-quali ty data. In this work, we engage in a comprehensive analysis of stochastic sampl ing from two aspects: variance-controlled diffusion SDE and linear multi-step SD E solver. Based on our analysis, we propose SA-Solver, which is an improved effi cient stochastic Adams method for solving diffusion SDE to generate data with hi gh quality. Our experiments show that SA-Solver achieves: 1) improved or compara ble performance compared with the existing state-of-the-art (SOTA) sampling meth ods for few-step sampling; 2) SOTA FID on substantial benchmark datasets under a suitable number of function evaluations (NFEs).

****

Social Motion Prediction with Cognitive Hierarchies

Wentao Zhu, Jason Qin, Yuke Lou, Hang Ye, Xiaoxuan Ma, Hai Ci, Yizhou Wang

Humans exhibit a remarkable capacity for anticipating the actions of others and planning their own actions accordingly. In this study, we strive to replicate th is ability by addressing the social motion prediction problem. We introduce a ne w benchmark, a novel formulation, and a cognition-inspired framework. We present Wusi, a 3D multi-person motion dataset under the context of team sports, which features intense and strategic human interactions and diverse pose distributions . By reformulating the problem from a multi-agent reinforcement learning perspec tive, we incorporate behavioral cloning and generative adversarial imitation lea rning to boost learning efficiency and generalization. Furthermore, we take into account the cognitive aspects of the human social action planning process and d evelop a cognitive hierarchy framework to predict strategic human social interac tions. We conduct comprehensive experiments to validate the effectiveness of our proposed dataset and approach.

****

Unbounded Differentially Private Quantile and Maximum Estimation

David Durfee

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.

****

How to Turn Your Knowledge Graph Embeddings into Generative Models

Lorenzo Loconte, Nicola Di Mauro, Robert Peharz, Antonio Vergari

Some of the most successful knowledge graph embedding (KGE) models for link pred iction – CP, RESCAL, TuckER, ComplEx – can be interpreted as energy-based models . Under this perspective they are not amenable for exact maximum-likelihood esti mation (MLE), sampling and struggle to integrate logical constraints. This work re-interprets the score functions of these KGEs as circuits – constrained comput ational graphs allowing efficient marginalisation. Then, we design two recipes t o obtain efficient generative circuit models by either restricting their activat ions to be non-negative or squaring their outputs. Our interpretation comes with little or no loss of performance for link prediction, while the circuits framew ork unlocks exact learning by MLE, efficient sampling of new triples, and guaran

tee that logical constraints are satisfied by design. Furthermore, our models scale more gracefully than the original KGEs on graphs with millions of entities.
****
Fed-GraB: Federated Long-tailed Learning with Self-Adjusting Gradient Balancer
Zikai Xiao, Zihan Chen, Songshang Liu, Hualiang Wang, YANG FENG, Jin Hao, Joey Tianyi Zhou, Jian Wu, Howard Yang, Zuozhu Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
CityRefer: Geography-aware 3D Visual Grounding Dataset on City-scale Point Cloud Data
Taiki Miyanishi, Fumiya Kitamori, Shuhei Kurita, Jungdae Lee, Motoaki Kawanabe, Nakamasa Inoue
City-scale 3D point cloud is a promising way to express detailed and complicated outdoor structures. It encompasses both the appearance and geometry features of segmented city components, including cars, streets, and buildings that can be utilized for attractive applications such as user-interactive navigation of autonomous vehicles and drones. However, compared to the extensive text annotations available for images and indoor scenes, the scarcity of text annotations for outdoor scenes poses a significant challenge for achieving these applications. To tackle this problem, we introduce the CityRefer dataset for city-level visual grounding. The dataset consists of 35k natural language descriptions of 3D objects appearing in SensatUrban city scenes and 5k landmarks labels synchronizing with OpenStreetMap. To ensure the quality and accuracy of the dataset, all descriptions and labels in the CityRefer dataset are manually verified. We also have developed a baseline system that can learn encoded language descriptions, 3D object instances, and geographical information about the city's landmarks to perform visual grounding on the CityRefer dataset. To the best of our knowledge, the CityRefer dataset is the largest city-level visual grounding dataset for localizing specific 3D objects.
****
GenImage: A Million-Scale Benchmark for Detecting AI-Generated Image
Mingjian Zhu, Hanting Chen, Qiangyu YAN, Xudong Huang, Guanyu Lin, Wei Li, Zhijun Tu, Hailin Hu, Jie Hu, Yunhe Wang
The extraordinary ability of generative models to generate photographic images has intensified concerns about the spread of disinformation, thereby leading to the demand for detectors capable of distinguishing between AI-generated fake images and real images. However, the lack of large datasets containing images from the most advanced image generators poses an obstacle to the development of such detectors. In this paper, we introduce the GenImage dataset, which has the following advantages: 1) Plenty of Images, including over one million pairs of AI-generated fake images and collected real images. 2) Rich Image Content, encompassing a broad range of image classes. 3) State-of-the-art Generators, synthesizing images with advanced diffusion models and GANs. The aforementioned advantages allow the detectors trained on GenImage to undergo a thorough evaluation and demonstrate strong applicability to diverse images. We conduct a comprehensive analysis of the dataset and propose two tasks for evaluating the detection method in resembling real-world scenarios. The cross-generator image classification task measures the performance of a detector trained on one generator when tested on the others. The degraded image classification task assesses the capability of the detectors in handling degraded images such as low-resolution, blurred, and compressed images. With the GenImage dataset, researchers can effectively expedite the development and evaluation of superior AI-generated image detectors in comparison to prevailing methodologies.
****
On Differentially Private Sampling from Gaussian and Product Distributions
Badih Ghazi, Xiao Hu, Ravi Kumar, Pasin Manurangsi
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

MedSat: A Public Health Dataset for England Featuring Medical Prescriptions and Satellite Imagery
Sanja Scepanovic, Ivica Obadic, Sagar Joglekar, Laura GIUSTARINI, Cristiano Nattero, Daniele Quercia, Xiaoxiang Zhu
As extreme weather events become more frequent, understanding their impact on human health becomes increasingly crucial. However, the utilization of Earth Observation to effectively analyze the environmental context in relation to health remains limited. This limitation is primarily due to the lack of fine-grained spatial and temporal data in public and population health studies, hindering a comprehensive understanding of health outcomes. Additionally, obtaining appropriate environmental indices across different geographical levels and timeframes poses a challenge. For the years 2019 (pre-COVID) and 2020 (COVID), we collected spatio-temporal indicators for all Lower Layer Super Output Areas in England. These indicators included: i) 111 sociodemographic features linked to health in existing literature, ii) 43 environmental point features (e.g., greenery and air pollution levels), iii) 4 seasonal composite satellite images each with 11 bands, and iv) prescription prevalence associated with five medical conditions (depression, anxiety, diabetes, hypertension, and asthma), opioids and total prescriptions. We combined these indicators into a single MedSat dataset, the availability of which presents an opportunity for the machine learning community to develop new techniques specific to public health. These techniques would address challenges such as handling large and complex data volumes, performing effective feature engineering on environmental and sociodemographic factors, capturing spatial and temporal dependencies in the models, addressing imbalanced data distributions, developing novel computer vision methods for health modeling based on satellite imagery, ensuring model explainability, and achieving generalization beyond the specific geographical region.
****

Anytime-Competitive Reinforcement Learning with Policy Prior
Jianyi Yang, Pengfei Li, Tongxin Li, Adam Wierman, Shaolei Ren
This paper studies the problem of Anytime-Competitive Markov Decision Process (A-CMDP). Existing works on Constrained Markov Decision Processes (CMDPs) aim to optimize the expected reward while constraining the expected cost over random dynamics, but the cost in a specific episode can still be unsatisfactorily high. In contrast, the goal of A-CMDP is to optimize the expected reward while guaranteeing a bounded cost in each round of any episode against a policy prior. We propose a new algorithm, called Anytime-Competitive Reinforcement Learning (ACRL), which provably guarantees the anytime cost constraints. The regret analysis shows the policy asymptotically matches the optimal reward achievable under the anytime competitive constraints. Experiments on the application of carbon-intelligent computing verify the reward performance and cost constraint guarantee of ACRL.
****

Metis: Understanding and Enhancing In-Network Regular Expressions
Zhengxin Zhang, Yucheng Huang, Guanglin Duan, Qing Li, Dan Zhao, Yong Jiang, Lianbo Ma, Xi Xiao, Hengyang Xu
Regular expressions (REs) offer one-shot solutions for many networking tasks, e.g., network intrusion detection. However, REs purely rely on expert knowledge and cannot utilize labeled data for better accuracy. Today, neural networks (NNs) have shown superior accuracy and flexibility, thanks to their ability to learn from rich labeled data. Nevertheless, NNs are often incompetent in cold-start scenarios and too complex for deployment on network devices. In this paper, we propose Metis, a general framework that converts REs to network device affordable models for superior accuracy and throughput by taking advantage of REs' expert knowledge and NNs' learning ability. In Metis, we convert REs to byte-level recurrent neural networks (BRNNs) without training. The BRNNs preserve expert knowledge from REs and offer adequate accuracy in cold-start scenarios. When rich labeled

data is available, the performance of BRNNs can be improved by training. Furthe rmore, we design a semi-supervised knowledge distillation to transform the BRNNs into pooling soft random forests (PSRFs) that can be deployed on network device s. To the best of our knowledge, this is the first method to employ model infere nce as an alternative to RE matching in network scenarios. We collect network tr affic data on our campus for three weeks and evaluate Metis on them. Experimenta l results show that Metis is more accurate than original REs and other baselines , achieving superior throughput when deployed on network devices.
****
Adaptive Test-Time Personalization for Federated Learning
Wenxuan Bao, Tianxin Wei, Haohan Wang, Jingrui He
Personalized federated learning algorithms have shown promising results in adapt ing models to various distribution shifts. However, most of these methods requir e labeled data on testing clients for personalization, which is usually unavaila ble in real-world scenarios. In this paper, we introduce a novel setting called test-time personalized federated learning (TTPFL), where clients locally adapt a global model in an unsupervised way without relying on any labeled data during test-time. While traditional test-time adaptation (TTA) can be used in this scen ario, most of them inherently assume training data come from a single domain, wh ile they come from multiple clients (source domains) with different distribution s. Overlooking these domain interrelationships can result in suboptimal generali zation. Moreover, most TTA algorithms are designed for a specific kind of distri bution shift and lack the flexibility to handle multiple kinds of distribution s hifts in FL. In this paper, we find that this lack of flexibility partially resu lts from their pre-defining which modules to adapt in the model. To tackle this challenge, we propose a novel algorithm called ATP to adaptively learns the adap tation rates for each module in the model from distribution shifts among source domains. Theoretical analysis proves the strong generalization of ATP. Extensive experiments demonstrate its superiority in handling various distribution shifts including label shift, image corruptions, and domain shift, outperforming exist ing TTA methods across multiple datasets and model architectures. Our code is av ailable at https://github.com/baowenxuan/ATP.
****
Context-lumpable stochastic bandits
Chung-Wei Lee, Qinghua Liu, Yasin Abbasi Yadkori, Chi Jin, Tor Lattimore, Csaba Szepesvari
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****
Text Alignment Is An Efficient Unified Model for Massive NLP Tasks
Yuheng Zha, Yichi Yang, Ruichen Li, Zhiting Hu
Large language models (LLMs), typically designed as a function of next-word pred iction, have excelled across extensive NLP tasks. Despite the generality, next-w ord prediction is often not an efficient formulation for many of the tasks, dema nding an extreme scale of model parameters (10s or 100s of billions) and sometim es yielding suboptimal performance.In practice, it is often desirable to build m ore efficient models---despite being less versatile, they still apply to a subst antial subset of problems, delivering on par or even superior performance with m uch smaller model sizes.In this paper, we propose text alignment as an efficient unified model for a wide range of crucial tasks involving text entailment, simi larity, question answering (and answerability), factual consistency, and so fort h. Given a pair of texts, the model measures the degree of alignment between the ir information. We instantiate an alignment model through lightweight finetuning of RoBERTa (355M parameters) using 5.9M examples from 28 datasets. Despite its compact size, extensive experiments show the model's efficiency and strong perfo rmance: (1) On over 20 datasets of aforementioned diverse tasks, the model match es or surpasses FLAN-T5 models that have around 2x or 10x more parameters; the s ingle unified model also outperforms task-specific models finetuned on individua

l datasets; (2) When applied to evaluate factual consistency of language generation on 23 datasets, our model improves over various baselines, including the much larger GPT-3.5 (ChatGPT) and sometimes even GPT-4; (3) The lightweight model can also serve as an add-on component for LLMs such as GPT-3.5 in question answering tasks, improving the average exact match (EM) score by 17.94 and F1 score by 15.05 through identifying unanswerable questions.
****
Learning from Active Human Involvement through Proxy Value Propagation
Zhenghao (Mark) Peng, Wenjie Mo, Chenda Duan, Quanyi Li, Bolei Zhou
Learning from active human involvement enables the human subject to actively intervene and demonstrate to the AI agent during training. The interaction and corrective feedback from human brings safety and AI alignment to the learning process. In this work, we propose a new reward-free active human involvement method called Proxy Value Propagation for policy optimization. Our key insight is that a proxy value function can be designed to express human intents, wherein state- action pairs in the human demonstration are labeled with high values, while those agents' actions that are intervened receive low values. Through the TD-learning framework, labeled values of demonstrated state-action pairs are further propagated to other unlabeled data generated from agents' exploration. The proxy value function thus induces a policy that faithfully emulates human behaviors. Human-in-the-loop experiments show the generality and efficiency of our method. With minimal modification to existing reinforcement learning algorithms, our method can learn to solve continuous and discrete control tasks with various human control devices, including the challenging task of driving in Grand Theft Auto V. Demo video and code are available at: https://metadriverse.github.io/pvp.
****
PrObeD: Proactive Object Detection Wrapper
Vishal Asnani, Abhinav Kumar, Suya You, Xiaoming Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Waypoint Transformer: Reinforcement Learning via Supervised Learning with Intermediate Targets
Anirudhan Badrinath, Yannis Flet-Berliac, Allen Nie, Emma Brunskill
Despite the recent advancements in offline reinforcement learning via supervised learning (RvS) and the success of the decision transformer (DT) architecture in various domains, DTs have fallen short in several challenging benchmarks. The root cause of this underperformance lies in their inability to seamlessly connect segments of suboptimal trajectories. To overcome this limitation, we present a novel approach to enhance RvS methods by integrating intermediate targets. We introduce the Waypoint Transformer (WT), using an architecture that builds upon the DT framework and conditioned on automatically-generated waypoints. The results show a significant increase in the final return compared to existing RvS methods, with performance on par or greater than existing state-of-the-art temporal difference learning-based methods. Additionally, the performance and stability improvements are largest in the most challenging environments and data configurations, including AntMaze Large Play/Diverse and Kitchen Mixed/Partial.
****
Should Under-parameterized Student Networks Copy or Average Teacher Weights?
Berfin Simsek, Amire Bendjeddou, Wulfram Gerstner, Johanni Brea
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
A Dataset for Analyzing Streaming Media Performance over HTTP/3 Browsers
Sapna Chaudhary, Mukulika Maity, Sandip Chakraborty, Naval Shukla
HTTP/3 is a new application layer protocol supported by most browsers. It uses Q

UIC as an underlying transport protocol. QUIC provides multiple benefits, like faster connection establishment, reduced latency, and improved connection migration. Hence, most popular browsers like Chrome/Chromium, Microsoft Edge, Apple Safari, and Mozilla Firefox have started supporting it. In this paper, we present an HTTP/3-supported browser dataset collection tool named H3B. It collects the application and network-level logs during YouTube streaming. We consider YouTube, as it the most popular video streaming application supporting QUIC. Using this tool, we collected a dataset of over 5936 YouTube sessions covering 5464 hours of streaming over 5 different geographical locations and 5 different bandwidth patterns. We believe our tool and as well as the dataset could be used in multiple applications such as a better configuration of application/transport protocols based on the network conditions, intelligent integration of network and application, predicting YouTube's QoE etc. We analyze the dataset and observe that during an HTTP/3 streaming not all requests are served by HTTP/3. Instead whenever the network condition is not favorable the browser chooses to fallback, and the application requests are transmitted using HTTP/2 over the old-standing transport protocol TCP. We observe that such switching of protocols impacts the performance of video streaming applications.
****
Learning from Rich Semantics and Coarse Locations for Long-tailed Object Detection
Lingchen Meng, Xiyang Dai, Jianwei Yang, Dongdong Chen, Yinpeng Chen, Mengchen Liu, Yi-Ling Chen, Zuxuan Wu, Lu Yuan, Yu-Gang Jiang
Long-tailed object detection (LTOD) aims to handle the extreme data imbalance in real-world datasets, where many tail classes have scarce instances. One popular strategy is to explore extra data with image-level labels, yet it produces limited results due to (1) semantic ambiguity---an image-level label only captures a salient part of the image, ignoring the remaining rich semantics within the image; and (2) location sensitivity---the label highly depends on the locations and crops of the original image, which may change after data transformations like random cropping.To remedy this, we propose RichSem, a simple but effective method, which is robust to learn rich semantics from coarse locations without the need of accurate bounding boxes. RichSem leverages rich semantics from images, which are then served as additional ``soft supervision'' for training detectors. Specifically, we add a semantic branch to our detector to learn these soft semantics and enhance feature representations for long-tailed object detection. The semantic branch is only used for training and is removed during inference. RichSem achieves consistent improvements on both overall and rare-category of LVIS under different backbones and detectors. Our method achieves state-of-the-art performance without requiring complex training and testing procedures. Moreover, we show the effectiveness of our method on other long-tailed datasets with additional experiments.
****
StoryBench: A Multifaceted Benchmark for Continuous Story Visualization
Emanuele Bugliarello, H. Hernan Moraldo, Ruben Villegas, Mohammad Babaeizadeh, Mohammad Taghi Saffar, Han Zhang, Dumitru Erhan, Vittorio Ferrari, Pieter-Jan Kindermans, Paul Voigtlaender
Generating video stories from text prompts is a complex task. In addition to having high visual quality, videos need to realistically adhere to a sequence of text prompts whilst being consistent throughout the frames. Creating a benchmark for video generation requires data annotated over time, which contrasts with the single caption used often in video datasets. To fill this gap, we collect comprehensive human annotations on three existing datasets, and introduce StoryBench: a new, challenging multi-task benchmark to reliably evaluate forthcoming text-to-video models. Our benchmark includes three video generation tasks of increasing difficulty: action execution, where the next action must be generated starting from a conditioning video; story continuation, where a sequence of actions must be executed starting from a conditioning video; and story generation, where a video must be generated from only text prompts. We evaluate small yet strong text-to-video baselines, and show the benefits of training on story-like data algorit

hmically generated from existing video captions. Finally, we establish guidelines for human evaluation of video stories, and reaffirm the need of better automatic metrics for video generation. StoryBench aims at encouraging future research efforts in this exciting new area.
****
DiffInfinite: Large Mask-Image Synthesis via Parallel Random Patch Diffusion in Histopathology
Marco Aversa, Gabriel Nobis, Miriam Hägele, Kai Standvoss, Mihaela Chirica, Roderick Murray-Smith, Ahmed M. Alaa, Lukas Ruff, Daniela Ivanova, Wojciech Samek, Frederick Klauschen, Bruno Sanguinetti, Luis Oala
We present DiffInfinite, a hierarchical diffusion model that generates arbitrarily large histological images while preserving long-range correlation structural information. Our approach first generates synthetic segmentation masks, subsequently used as conditions for the high-fidelity generative diffusion process. The proposed sampling method can be scaled up to any desired image size while only requiring small patches for fast training. Moreover, it can be parallelized more efficiently than previous large-content generation methods while avoiding tiling artifacts. The training leverages classifier-free guidance to augment a small, sparsely annotated dataset with unlabelled data. Our method alleviates unique challenges in histopathological imaging practice: large-scale information, costly manual annotation, and protective data handling. The biological plausibility of DiffInfinite data is evaluated in a survey by ten experienced pathologists as well as a downstream classification and segmentation task. Samples from the model score strongly on anti-copying metrics which is relevant for the protection of patient data.
****
Benchmarking Foundation Models with Language-Model-as-an-Examiner
Yushi Bai, Jiahao Ying, Yixin Cao, Xin Lv, Yuze He, Xiaozhi Wang, Jifan Yu, Kaisheng Zeng, Yijia Xiao, Haozhe Lyu, Jiayin Zhang, Juanzi Li, Lei Hou
Numerous benchmarks have been established to assess the performance of foundation models on open-ended question answering, which serves as a comprehensive test of a model's ability to understand and generate language in a manner similar to humans.Most of these works focus on proposing new datasets, however, we see two main issues within previous benchmarking pipelines, namely testing leakage and evaluation automation. In this paper, we propose a novel benchmarking framework, Language-Model-as-an-Examiner, where the LM serves as a knowledgeable examiner that formulates questions based on its knowledge and evaluates responses in a reference-free manner. Our framework allows for effortless extensibility as various LMs can be adopted as the examiner, and the questions can be constantly updated given more diverse trigger topics. For a more comprehensive and equitable evaluation, we devise three strategies: (1) We instruct the LM examiner to generate questions across a multitude of domains to probe for a broad acquisition, and raise follow-up questions to engage in a more in-depth assessment. (2) Upon evaluation, the examiner combines both scoring and ranking measurements, providing a reliable result as it aligns closely with human annotations. (3) We additionally propose a decentralized Peer-examination method to address the biases in a single examiner. Our data and benchmarking results are available at: http://lmexam.xlore.cn.
****
Granger Components Analysis: Unsupervised learning of latent temporal dependencies
Jacek Dmochowski
A new technique for unsupervised learning of time series data based on the notion of Granger causality is presented. The technique learns pairs of projections of a multivariate data set such that the resulting components -- "driving" and "driven" -- maximize the strength of the Granger causality between the latent time series (how strongly the past of the driving signal predicts the present of the driven signal). A coordinate descent algorithm that learns pairs of coefficient vectors in an alternating fashion is developed and shown to blindly identify the underlying sources (up to scale) on simulated vector autoregressive (VAR) data

. The technique is tested on scalp electroencephalography (EEG) data from a motor imagery experiment where the resulting components lateralize with the side of the cued hand, and also on functional magnetic resonance imaging (fMRI) data, where the recovered components express previously reported resting-state networks.
****

Monte Carlo Tree Search with Boltzmann Exploration

Michael Painter, Mohamed Baioumy, Nick Hawes, Bruno Lacerda

Monte-Carlo Tree Search (MCTS) methods, such as Upper Confidence Bound applied to Trees (UCT), are instrumental to automated planning techniques. However, UCT can be slow to explore an optimal action when it initially appears inferior to other actions. Maximum ENtropy Tree-Search (MENTS) incorporates the maximum entropy principle into an MCTS approach, utilising Boltzmann policies to sample actions, naturally encouraging more exploration. In this paper, we highlight a major limitation of MENTS: optimal actions for the maximum entropy objective do not necessarily correspond to optimal actions for the original objective. We introduce two algorithms, Boltzmann Tree Search (BTS) and Decaying ENtropy Tree-Search (DENTS), that address these limitations and preserve the benefits of Boltzmann policies, such as allowing actions to be sampled faster by using the Alias method. Our empirical analysis shows that our algorithms show consistent high performance across several benchmark domains, including the game of Go.
****

False Discovery Proportion control for aggregated Knockoffs

Alexandre Blain, Bertrand Thirion, Olivier Grisel, Pierre Neuvial

Controlled variable selection is an important analytical step in various scientific fields, such as brain imaging or genomics. In these high-dimensional data settings, considering too many variables leads to poor models and high costs, hence the need for statistical guarantees on false positives. Knockoffs are a popular statistical tool for conditional variable selection in high dimension. However, they control for the expected proportion of false discoveries (FDR) and not the actual proportion of false discoveries (FDP). We present a new method, KOPI, that controls the proportion of false discoveries for Knockoff-based inference. The proposed method also relies on a new type of aggregation to address the undesirable randomness associated with classical Knockoff inference. We demonstrate FDP control and substantial power gains over existing Knockoff-based methods in various simulation settings and achieve good sensitivity/specificity tradeoffs on brain imaging data.
****

Interpretability at Scale: Identifying Causal Mechanisms in Alpaca

Zhengxuan Wu, Atticus Geiger, Thomas Icard, Christopher Potts, Noah Goodman

Obtaining human-interpretable explanations of large, general-purpose language models is an urgent goal for AI safety. However, it is just as important that our interpretability methods are faithful to the causal dynamics underlying model behavior and able to robustly generalize to unseen inputs. Distributed Alignment Search (DAS) is a powerful gradient descent method grounded in a theory of causal abstraction that uncovered perfect alignments between interpretable symbolic algorithms and small deep learning models fine-tuned for specific tasks. In the present paper, we scale DAS significantly by replacing the remaining brute-force search steps with learned parameters -- an approach we call Boundless DAS. This enables us to efficiently search for interpretable causal structure in large language models while they follow instructions. We apply Boundless DAS to the Alpaca model (7B parameters), which, off the shelf, solves a simple numerical reasoning problem. With Boundless DAS, we discover that Alpaca does this by implementing a causal model with two interpretable boolean variables. Furthermore, we find that the alignment of neural representations with these variables is robust to changes in inputs and instructions. These findings mark a first step toward deeply understanding the inner-workings of our largest and most widely deployed language models.
****

Large Language Models Are Semi-Parametric Reinforcement Learning Agents

Danyang Zhang, Lu Chen, Situo Zhang, Hongshen Xu, Zihan Zhao, Kai Yu

Inspired by the insights in cognitive science with respect to human memory and reasoning mechanism, a novel evolvable LLM-based (Large Language Model) agent framework is proposed as Rememberer. By equipping the LLM with a long-term experience memory, Rememberer is capable of exploiting the experiences from the past episodes even for different task goals, which excels an LLM-based agent with fixed exemplars or equipped with a transient working memory. We further introduce Reinforcement Learning with Experience Memory (RLEM) to update the memory. Thus, the whole system can learn from the experiences of both success and failure, and evolve its capability without fine-tuning the parameters of the LLM. In this way, the proposed Rememberer constitutes a semi-parametric RL agent. Extensive experiments are conducted on two RL task sets to evaluate the proposed framework. The average results with different initialization and training sets exceed the prior SOTA by 4% and 2% for the success rate on two task sets and demonstrate the superiority and robustness of Rememberer.
****
BIOT: Biosignal Transformer for Cross-data Learning in the Wild
Chaoqi Yang, M Westover, Jimeng Sun
Biological signals, such as electroencephalograms (EEG), play a crucial role in numerous clinical applications, exhibiting diverse data formats and quality profiles. Current deep learning models for biosignals (based on CNN, RNN, and Transformers) are typically specialized for specific datasets and clinical settings, limiting their broader applicability. This paper explores the development of a flexible biosignal encoder architecture that can enable pre-training on multiple datasets and fine-tuned on downstream biosignal tasks with different formats.To overcome the unique challenges associated with biosignals of various formats, such as mismatched channels, variable sample lengths, and prevalent missing val- ues, we propose Biosignal Transformer (BIOT). The proposed BIOT model can enable cross-data learning with mismatched channels, variable lengths, and missing values by tokenizing different biosignals into unified "sentences" structure. Specifically, we tokenize each channel separately into fixed-length segments containing local signal features and then rearrange the segments to form a long "sentence". Channel embeddings and relative position embeddings are added to each segment (viewed as "token") to preserve spatio-temporal features.The BIOT model is versatile and applicable to various biosignal learning settings across different datasets, including joint pre-training for larger models. Comprehensive evaluations on EEG, electrocardiogram (ECG), and human activity sensory signals demonstrate that BIOT outperforms robust baselines in common settings and facilitates learning across multiple datasets with different formats. Using CHB-MIT seizure detection task as an example, our vanilla BIOT model shows 3% improvement over baselines in balanced accuracy, and the pre-trained BIOT models (optimized from other data sources) can further bring up to 4% improvements. Our repository is public at https://github.com/ycq091044/BIOT.
****
Interactive Multi-fidelity Learning for Cost-effective Adaptation of Language Model with Sparse Human Supervision
Jiaxin Zhang, Zhuohang Li, Kamalika Das, Sricharan Kumar
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
OceanBench: The Sea Surface Height Edition
J. Emmanuel Johnson, Quentin Febvre, Anastasiia Gorbunova, Sam Metref, Maxime Ballarotta, Julien Le Sommer, ronan fablet
The ocean is a crucial component of the Earth's system. It profoundly influences human activities and plays a critical role in climate regulation. Our understanding has significantly improved over the last decades with the advent of satellite remote sensing data, allowing us to capture essential sea surface quantities over the globe, e.g., sea surface height (SSH). Despite their ever-increasing abundance, ocean satellite data presents challenges for information extraction due

to their sparsity and irregular sampling, signal complexity, and noise. Machine learning (ML) techniques have demonstrated their capabilities in dealing with large-scale, complex signals. Therefore we see an opportunity for these ML models to harness the full extent of the information contained in ocean satellite data. However, data representation and relevant evaluation metrics can be the defining factors when determining the success of applied ML. The processing steps from the raw observation data to a ML-ready state and from model outputs to interpretable quantities require domain expertise, which can be a significant barrier to entry for ML researchers. In addition, imposing fixed processing steps, like committing to specific variables, regions, and geometries, will narrow the scope of ML models and their potential impact on real-world applications. OceanBench is a unifying framework that provides standardized processing steps that comply with domain-expert standards. It is designed with a flexible and pedagogical abstraction: it a) provides plug-and-play data and pre-configured pipelines for ML researchers to benchmark their models w.r.t. ML and domain-related baselines and b) provides a transparent and configurable framework for researchers to customize and extend the pipeline for their tasks. In this work, we demonstrate the OceanBench framework through a first edition dedicated to SSH interpolation challenges. We provide datasets and ML-ready benchmarking pipelines for the long-standing problem of interpolating observations from simulated ocean satellite data, multi-modal and multi-sensor fusion issues, and transfer-learning to real ocean satellite observations. The OceanBench framework is available at https://github.com/jejjohnson/oceanbench and the dataset registry is available at https://github.com/quentinf00/oceanbench-data-registry.
****
Amortized Reparametrization: Efficient and Scalable Variational Inference for Latent SDEs
Kevin Course, Prasanth Nair
We consider the problem of inferring latent stochastic differential equations (SDEs) with a time and memory cost that scales independently with the amount of data, the total length of the time series, and the stiffness of the approximate differential equations. This is in stark contrast to typical methods for inferring latent differential equations which, despite their constant memory cost, have a time complexity that is heavily dependent on the stiffness of the approximate differential equation. We achieve this computational advancement by removing the need to solve differential equations when approximating gradients using a novel amortization strategy coupled with a recently derived reparametrization of expectations under linear SDEs. We show that, in practice, this allows us to achieve similar performance to methods based on adjoint sensitivities with more than an order of magnitude fewer evaluations of the model in training.
****
Boundary Guided Learning-Free Semantic Control with Diffusion Models
Ye Zhu, Yu Wu, Zhiwei Deng, Olga Russakovsky, Yan Yan
Applying pre-trained generative denoising diffusion models (DDMs) for downstream tasks such as image semantic editing usually requires either fine-tuning DDMs or learning auxiliary editing networks in the existing literature. In this work, we present our BoundaryDiffusion method for efficient, effective and light-weight semantic control with frozen pre-trained DDMs, without learning any extra networks. As one of the first learning-free diffusion editing works, we start by seeking a more comprehensive understanding of the intermediate high-dimensional latent spaces by theoretically and empirically analyzing their probabilistic and geometric behaviors in the Markov chain. We then propose to further explore the critical step in the denoising trajectory that characterizes the convergence of a pre-trained DDM and introduce an automatic search method. Last but not least, in contrast to the conventional understanding that DDMs have relatively poor semantic behaviors (in generic latent spaces), we prove that the critical latent space we found already forms semantic subspace boundaries at the generic level in unconditional DDMs, which allows us to do controllable manipulation by guiding the denoising trajectory towards the targeted boundary via a single-step operation. We conduct extensive experiments on multiple DPMs architectures (DDPM, iDDPM) a

nd datasets (CelebA, CelebA-HQ, LSUN-church, LSUN-bedroom, AFHQ-dog) with differ
ent resolutions (64, 256), achieving superior or state-of-the-art performance in
 various task scenarios (image semantic editing, text-based editing, uncondition
al semantic control) to demonstrate the effectiveness.
****
Kiki or Bouba? Sound Symbolism in Vision-and-Language Models
Morris Alper, Hadar Averbuch-Elor
Although the mapping between sound and meaning in human language is assumed to b
e largely arbitrary, research in cognitive science has shown that there are non-
trivial correlations between particular sounds and meanings across languages and
 demographic groups, a phenomenon known as sound symbolism. Among the many dimen
sions of meaning, sound symbolism is particularly salient and well-demonstrated
with regards to cross-modal associations between language and the visual domain.
 In this work, we address the question of whether sound symbolism is reflected i
n vision-and-language models such as CLIP and Stable Diffusion. Using zero-shot
knowledge probing to investigate the inherent knowledge of these models, we find
 strong evidence that they do show this pattern, paralleling the well-known kiki
-bouba effect in psycholinguistics. Our work provides a novel method for demonst
rating sound symbolism and understanding its nature using computational tools. O
ur code will be made publicly available.
****
MoCa: Measuring Human-Language Model Alignment on Causal and Moral Judgment Task
s
Allen Nie, Yuhui Zhang, Atharva Shailesh Amdekar, Chris Piech, Tatsunori B. Hash
imoto, Tobias Gerstenberg
Human commonsense understanding of the physical and social world is organized ar
ound intuitive theories. These theories support making causal and moral judgment
s. When something bad happens, we naturally ask: who did what, and why? A rich l
iterature in cognitive science has studied people's causal and moral intuitions.
 This work has revealed a number of factors that systematically influence people
's judgments, such as the violation of norms and whether the harm is avoidable o
r inevitable. We collected a dataset of stories from 24 cognitive science papers
 and developed a system to annotate each story with the factors they investigate
d. Using this dataset, we test whether large language models (LLMs) make causal
and moral judgments about text-based scenarios that align with those of human pa
rticipants. On the aggregate level, alignment has improved with more recent LLMs
. However, using statistical analyses, we find that LLMs weigh the different fac
tors quite differently from human participants. These results show how curated,
challenge datasets combined with insights from cognitive science can help us go
beyond comparisons based merely on aggregate metrics: we uncover LLMs implicit t
endencies and show to what extent these align with human intuitions.
****
Collapsed Inference for Bayesian Deep Learning
Zhe Zeng, Guy Van den Broeck
Bayesian neural networks (BNNs) provide a formalism to quantify and calibrate un
certainty in deep learning. Current inference approaches for BNNs often resort t
o few-sample estimation for scalability, which can harm predictive performance,
while its alternatives tend to be computationally prohibitively expensive. We ta
ckle this challenge by revealing a previously unseen connection between inferenc
e on BNNs and volume computation problems. With this observation, we introduce a
 novel collapsed inference scheme that performs Bayesian model averaging using c
ollapsed samples. It improves over a Monte-Carlo sample by limiting sampling to
a subset of the network weights while pairing it with some closed-form condition
al distribution over the rest. A collapsed sample represents uncountably many mo
dels drawn from the approximate posterior and thus yields higher sample efficien
cy. Further, we show that the marginalization of a collapsed sample can be solve
d analytically and efficiently despite the non-linearity of neural networks by l
everaging existing volume computation solvers. Our proposed use of collapsed sam
ples achieves a balance between scalability and accuracy. On various regression
and classification tasks, our collapsed Bayesian deep learning approach demonstr

ates significant improvements over existing methods and sets a new state of the art in terms of uncertainty estimation as well as predictive performance.
****

Contextual Stochastic Bilevel Optimization
Yifan Hu, Jie Wang, Yao Xie, Andreas Krause, Daniel Kuhn
We introduce contextual stochastic bilevel optimization (CSBO) -- a stochastic bilevel optimization framework with the lower-level problem minimizing an expectation conditioned on some contextual information and the upper-level decision variable. This framework extends classical stochastic bilevel optimization when the lower-level decision maker responds optimally not only to the decision of the upper-level decision maker but also to some side information and when there are multiple or even infinite many followers. It captures important applications such as meta-learning, personalized federated learning, end-to-end learning, and Wasserstein distributionally robust optimization with side information (WDRO-SI). Due to the presence of contextual information, existing single-loop methods for classical stochastic bilevel optimization are unable to converge. To overcome this challenge, we introduce an efficient double-loop gradient method based on the Multilevel Monte-Carlo (MLMC) technique and establish its sample and computational complexities. When specialized to stochastic nonconvex optimization, our method matches existing lower bounds. For meta-learning, the complexity of our method does not depend on the number of tasks. Numerical experiments further validate our theoretical results.
****

Learning Invariant Molecular Representation in Latent Discrete Space
Xiang Zhuang, Qiang Zhang, Keyan Ding, Yatao Bian, Xiao Wang, Jingsong Lv, Hongyang Chen, Huajun Chen
Molecular representation learning lays the foundation for drug discovery. However, existing methods suffer from poor out-of-distribution (OOD) generalization, particularly when data for training and testing originate from different environments. To address this issue, we propose a new framework for learning molecular representations that exhibit invariance and robustness against distribution shifts. Specifically, we propose a strategy called ``first-encoding-then-separation'' to identify invariant molecule features in the latent space, which deviates from conventional practices. Prior to the separation step, we introduce a residual vector quantization module that mitigates the over-fitting to training data distributions while preserving the expressivity of encoders. Furthermore, we design a task-agnostic self-supervised learning objective to encourage precise invariance identification, which enables our method widely applicable to a variety of tasks, such as regression and multi-label classification. Extensive experiments on 18 real-world molecular datasets demonstrate that our model achieves stronger generalization against state-of-the-art baselines in the presence of various distribution shifts. Our code is available at https://github.com/HICAI-ZJU/iMoLD.
****

Accelerating Motion Planning via Optimal Transport
An T. Le, Georgia Chalvatzaki, Armin Biess, Jan R. Peters
Motion planning is still an open problem for many disciplines, e.g., robotics, autonomous driving, due to their need for high computational resources that hinder real-time, efficient decision-making. A class of methods striving to provide smooth solutions is gradient-based trajectory optimization. However, those methods usually suffer from bad local minima, while for many settings, they may be inapplicable due to the absence of easy-to-access gradients of the optimization objectives. In response to these issues, we introduce Motion Planning via Optimal Transport (MPOT)---a \textit{gradient-free} method that optimizes a batch of smooth trajectories over highly nonlinear costs, even for high-dimensional tasks, while imposing smoothness through a Gaussian Process dynamics prior via the planning-as-inference perspective. To facilitate batch trajectory optimization, we introduce an original zero-order and highly-parallelizable update rule----the Sinkhorn Step, which uses the regular polytope family for its search directions. Each regular polytope, centered on trajectory waypoints, serves as a local cost-probing neighborhood, acting as a \textit{trust region} where the Sinkhorn Step ``tr

ansports'' local waypoints toward low-cost regions. We theoretically show that S
inkhorn Step guides the optimizing parameters toward local minima regions of non
-convex objective functions. We then show the efficiency of MPOT in a range of p
roblems from low-dimensional point-mass navigation to high-dimensional whole-bod
y robot motion planning, evincing its superiority compared to popular motion pla
nners, paving the way for new applications of optimal transport in motion planni
ng.
****
M5HisDoc: A Large-scale Multi-style Chinese Historical Document Analysis Benchma
rk
Yongxin Shi, Chongyu Liu, Dezhi Peng, Cheng Jian, Jiarong Huang, Lianwen Jin
Recognizing and organizing text in correct reading order plays a crucial role in
 historical document analysis and preservation. While existing methods have show
n promising performance, they often struggle with challenges such as diverse lay
outs, low image quality, style variations, and distortions. This is primarily du
e to the lack of consideration for these issues in the current benchmarks, which
 hinders the development and evaluation of historical document analysis and reco
gnition (HDAR) methods in complex real-world scenarios. To address this gap, thi
s paper introduces a complex multi-style Chinese historical document analysis be
nchmark, named M5HisDoc. The M5 indicates five properties of style, ie., Multipl
e layouts, Multiple document types, Multiple calligraphy styles, Multiple backgr
ounds, and Multiple challenges. The M5HisDoc dataset consists of two subsets, M5
HisDoc-R (Regular) and M5HisDoc-H (Hard). The M5HisDoc-R subset comprises 4,000
historical document images. To ensure high-quality annotations, we meticulously
perform manual annotation and triple-checking. To replicate real-world condition
s for historical document analysis applications, we incorporate image rotation,
distortion, and resolution reduction into M5HisDoc-R subset to form a new challe
nging subset named M5HisDoc-H, which contains the same number of images as M5His
Doc-R. The dataset exhibits diverse styles, significant scale variations, dense
texts, and an extensive character set. We conduct benchmarking experiments on fi
ve tasks: text line detection, text line recognition, character detection, chara
cter recognition, and reading order prediction. We also conduct cross-validation
 with other benchmarks. Experimental results demonstrate that the M5HisDoc datas
et can offer new challenges and great opportunities for future research in this
field, thereby providing deep insights into the solution for HDAR. The dataset i
s available at https://github.com/HCIILAB/M5HisDoc.
****
CWCL: Cross-Modal Transfer with Continuously Weighted Contrastive Loss
Rakshith Sharma Srinivasa, Jaejin Cho, Chouchang Yang, Yashas Malur Saidutta, Ch
ing-Hua Lee, Yilin Shen, Hongxia Jin
This paper considers contrastive training for cross-modal 0-shot transfer wherei
n a pre-trained model in one modality is used for representation learning in ano
ther domain using pairwise data. The learnt models in the latter domain can then
 be used for a diverse set of tasks in a 0-shot way, similar to Contrastive Lang
uage-Image Pre-training (CLIP) and Locked-image Tuning (LiT) that have recently
gained considerable attention. Classical contrastive training employs sets of po
sitive and negative examples to align similar and repel dissimilar training data
 samples. However, similarity amongst training examples has a more continuous na
ture, thus calling for a more `non-binary' treatment. To address this, we propos
e a new contrastive loss function called Continuously Weighted Contrastive Loss
(CWCL) that employs a continuous measure of similarity. With CWCL, we seek to tr
ansfer the structure of the embedding space from one modality to another. Owing
to the continuous nature of similarity in the proposed loss function, these mode
ls outperform existing methods for 0-shot transfer across multiple models, datas
ets and modalities. By using publicly available datasets, we achieve 5-8% (absol
ute) improvement over previous state-of-the-art methods in 0-shot image classifi
cation and 20-30% (absolute) improvement in 0-shot speech-to-intent classificati
on and keyword classification.
****
Decompose a Task into Generalizable Subtasks in Multi-Agent Reinforcement Learni

ng
Zikang Tian, Ruizhi Chen, Xing Hu, Ling Li, Rui Zhang, Fan Wu, Shaohui Peng, Jiaming Guo, Zidong Du, Qi Guo, Yunji Chen
In recent years, Multi-Agent Reinforcement Learning (MARL) techniques have made significant strides in achieving high asymptotic performance in single task. However, there has been limited exploration of model transferability across tasks. Training a model from scratch for each task can be time-consuming and expensive, especially for large-scale Multi-Agent Systems. Therefore, it is crucial to develop methods for generalizing the model across tasks. Considering that there exist task-independent subtasks across MARL tasks, a model that can decompose such subtasks from the source task could generalize to target tasks. However, ensuring true task-independence of subtasks poses a challenge. In this paper, we propose to \textbf{d}ecompose a \textbf{t}ask in\textbf{to} a series of \textbf{g}eneralizable \textbf{s}ubtasks (DT2GS), a novel framework that addresses this challenge by utilizing a scalable subtask encoder and an adaptive subtask semantic module. We show that these components endow subtasks with two properties critical for task-independence: avoiding overfitting to the source task and maintaining consistent yet scalable semantics across tasks. Empirical results demonstrate that DT2GS possesses sound zero-shot generalization capability across tasks, exhibits sufficient transferability, and outperforms existing methods in both multi-task and single-task problems.
****
The Equivalence of Dynamic and Strategic Stability under Regularized Learning in Games
Victor Boone, Panayotis Mertikopoulos
In this paper, we examine the long-run behavior of regularized, no-regret learning in finite N-player games. A well-known result in the field states that the empirical frequencies of play under no-regret learning converge to the game’s set of coarse correlated equilibria; however, our understanding of how the players' actual strategies evolve over time is much more limited – and, in many cases, non-existent. This issue is exacerbated further by a series of recent results showing that only strict Nash equilibria are stable and attracting under regularized learning, thus making the relation between learning and pointwise solution concepts particularly elusive. In lieu of this, we take a more general approach and instead seek to characterize the setwise rationality properties of the players' day-to-day trajectory of play. To do so, we focus on one of the most stringent criteria of setwise strategic stability, namely that any unilateral deviation from the set in question incurs a cost to the deviator – a property known as closedness under better replies (club). In so doing, we obtain a remarkable equivalence between strategic and dynamic stability: a product of pure strategies is closed under better replies if and only if its span is stable and attracting under regularized learning. In addition, we estimate the rate of convergence to such sets, and we show that methods based on entropic regularization (like the exponential weights algorithm) converge at a geometric rate, while projection-based methods converge within a finite number of iterations, even with bandit, payoff-based feedback.
****
HubRouter: Learning Global Routing via Hub Generation and Pin-hub Connection
Xingbo Du, Chonghua Wang, Ruizhe Zhong, Junchi Yan
Global Routing (GR) is a core yet time-consuming task in VLSI systems. It recently attracted efforts from the machine learning community, especially generative models, but they suffer from the non-connectivity of generated routes. We argue that the inherent non-connectivity can harm the advantage of its one-shot generation and has to be post-processed by traditional approaches. Thus, we propose a novel definition, called hub, which represents the key point in the route. Equipped with hubs, global routing is transferred from a pin-pin connection problem to a hub-pin connection problem. Specifically, to generate definitely-connected routes, this paper proposes a two-phase learning scheme named HubRouter, which includes 1) hub-generation phase: A condition-guided hub generator using deep generative models; 2) pin-hub-connection phase: An RSMT construction module that con

nects the hubs and pins using an actor-critic model. In the first phase, we incorporate typical generative models into a multi-task learning framework to perform hub generation and address the impact of sensitive noise points with stripe mask learning. During the second phase, HubRouter employs an actor-critic model to finish the routing, which is efficient and has very slight errors. Experiments on simulated and real-world global routing benchmarks are performed to show our approach's efficiency, particularly HubRouter outperforms the state-of-the-art generative global routing methods in wirelength, overflow, and running time. Moreover, HubRouter also shows strength in other applications, such as RSMT construction and interactive path replanning.
****
L_2-Uniform Stability of Randomized Learning Algorithms: Sharper Generalization Bounds and Confidence Boosting
Xiaotong Yuan, Ping Li
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Neural Sampling in Hierarchical Exponential-family Energy-based Models
Xingsi Dong, Si Wu
Bayesian brain theory suggests that the brain employs generative models to understand the external world. The sampling-based perspective posits that the brain infers the posterior distribution through samples of stochastic neuronal responses. Additionally, the brain continually updates its generative model to approach the true distribution of the external world. In this study, we introduce the Hierarchical Exponential-family Energy-based (HEE) model, which captures the dynamics of inference and learning. In the HEE model, we decompose the partition function into individual layers and leverage a group of neurons with shorter time constants to sample the gradient of the decomposed normalization term. This allows our model to estimate the partition function and perform inference simultaneously, circumventing the negative phase encountered in conventional energy-based models (EBMs). As a result, the learning process is localized both in time and space, and the model is easy to converge. To match the brain's rapid computation, we demonstrate that neural adaptation can serve as a momentum term, significantly accelerating the inference process. On natural image datasets, our model exhibits representations akin to those observed in the biological visual system. Furthermore, for the machine learning community, our model can generate observations through joint or marginal generation. We show that marginal generation outperforms joint generation and achieves performance on par with other EBMs.
****
Block Coordinate Plug-and-Play Methods for Blind Inverse Problems
Weijie Gan, shirin shoushtari, Yuyang Hu, Jiaming Liu, Hongyu An, Ulugbek Kamilov
Plug-and-play (PnP) prior is a well-known class of methods for solving imaging inverse problems by computing fixed-points of operators combining physical measurement models and learned image denoisers. While PnP methods have been extensively used for image recovery with known measurement operators, there is little work on PnP for solving blind inverse problems. We address this gap by presenting a new block-coordinate PnP (BC-PnP) method that efficiently solves this joint estimation problem by introducing learned denoisers as priors on both the unknown image and the unknown measurement operator. We present a new convergence theory for BC-PnP compatible with blind inverse problems by considering nonconvex data-fidelity terms and expansive denoisers. Our theory analyzes the convergence of BC-PnP to a stationary point of an implicit function associated with an approximate minimum mean-squared error (MMSE) denoiser. We numerically validate our method on two blind inverse problems: automatic coil sensitivity estimation in magnetic resonance imaging (MRI) and blind image deblurring. Our results show that BC-PnP provides an efficient and principled framework for using denoisers as PnP priors for jointly estimating measurement operators and images.

****
PreDiff: Precipitation Nowcasting with Latent Diffusion Models
Zhihan Gao, Xingjian Shi, Boran Han, Hao Wang, Xiaoyong Jin, Danielle Maddix, Yi Zhu, Mu Li, Yuyang (Bernie) Wang
Earth system forecasting has traditionally relied on complex physical models that are computationally expensive and require significant domain expertise.In the past decade, the unprecedented increase in spatiotemporal Earth observation data has enabled data-driven forecasting models using deep learning techniques.These models have shown promise for diverse Earth system forecasting tasks but either struggle with handling uncertainty or neglect domain-specific prior knowledge, resulting in averaging possible futures to blurred forecasts or generating physically implausible predictions.To address these limitations, we propose a two-stage pipeline for probabilistic spatiotemporal forecasting: 1) We develop PreDiff, a conditional latent diffusion model capable of probabilistic forecasts. 2) We incorporate an explicit knowledge alignment mechanism to align forecasts with domain-specific physical constraints. This is achieved by estimating the deviation from imposed constraints at each denoising step and adjusting the transition distribution accordingly.We conduct empirical studies on two datasets: N-body MNIST, a synthetic dataset with chaotic behavior, and SEVIR, a real-world precipitation nowcasting dataset. Specifically, we impose the law of conservation of energy in N-body MNIST and anticipated precipitation intensity in SEVIR. Experiments demonstrate the effectiveness of PreDiff in handling uncertainty, incorporating domain-specific prior knowledge, and generating forecasts that exhibit high operational utility.
****
All Points Matter: Entropy-Regularized Distribution Alignment for Weakly-supervised 3D Segmentation
Liyao Tang, Zhe Chen, Shanshan Zhao, Chaoyue Wang, Dacheng Tao
Pseudo-labels are widely employed in weakly supervised 3D segmentation tasks where only sparse ground-truth labels are available for learning.Existing methods often rely on empirical label selection strategies, such as confidence thresholding, to generate beneficial pseudo-labels for model training.This approach may, however, hinder the comprehensive exploitation of unlabeled data points.We hypothesize that this selective usage arises from the noise in pseudo-labels generated on unlabeled data. The noise in pseudo-labels may result in significant discrepancies between pseudo-labels and model predictions, thus confusing and affecting the model training greatly.To address this issue, we propose a novel learning strategy to regularize the generated pseudo-labels and effectively narrow the gaps between pseudo-labels and model predictions.More specifically, our method introduces an Entropy Regularization loss and a Distribution Alignment loss for weakly supervised learning in 3D segmentation tasks, resulting in an ERDA learning strategy.Interestingly, by using KL distance to formulate the distribution alignment loss, it reduces to a deceptively simple cross-entropy-based loss which optimizes both the pseudo-label generation network and the 3D segmentation network simultaneously.Despite the simplicity, our method promisingly improves the performance.We validate the effectiveness through extensive experiments on various baselines and large-scale datasets.Results show that ERDA effectively enables the effective usage of all unlabeled data points for learning and achieves state-of-the-art performance under different settings.Remarkably, our method can outperform fully-supervised baselines using only 1\% of true annotations.Code and model will be made publicly available at https://github.com/LiyaoTang/ERDA.
****
SimMMDG: A Simple and Effective Framework for Multi-modal Domain Generalization
Hao Dong, Ismail Nejjar, Han Sun, Eleni Chatzi, Olga Fink
In real-world scenarios, achieving domain generalization (DG) presents significant challenges as models are required to generalize to unknown target distributions. Generalizing to unseen multi-modal distributions poses even greater difficulties due to the distinct properties exhibited by different modalities. To overcome the challenges of achieving domain generalization in multi-modal scenarios, we propose SimMMDG, a simple yet effective multi-modal DG framework. We argue tha

t mapping features from different modalities into the same embedding space impedes model generalization. To address this, we propose splitting the features within each modality into modality-specific and modality-shared components. We employ supervised contrastive learning on the modality-shared features to ensure they possess joint properties and impose distance constraints on modality-specific features to promote diversity. In addition, we introduce a cross-modal translation module to regularize the learned features, which can also be used for missing-modality generalization. We demonstrate that our framework is theoretically well-supported and achieves strong performance in multi-modal DG on the EPIC-Kitchens dataset and the novel Human-Animal-Cartoon (HAC) dataset introduced in this paper. Our source code and HAC dataset are available at https://github.com/donghao51/SimMMDG.
****

Bounding training data reconstruction in DP-SGD

Jamie Hayes, Borja Balle, Saeed Mahloujifar

Differentially private training offers a protection which is usually interpreted as a guarantee against membership inference attacks. By proxy, this guarantee extends to other threats like reconstruction attacks attempting to extract complete training examples. Recent works provide evidence that if one does not need to protect against membership attacks but instead only wants to protect against a training data reconstruction, then utility of private models can be improved because less noise is required to protect against these more ambitious attacks. We investigate this question further in the context of DP-SGD, a standard algorithm for private deep learning, and provide an upper bound on the success of any reconstruction attack against DP-SGD together with an attack that empirically matches the predictions of our bound. Together, these two results open the door to fine-grained investigations on how to set the privacy parameters of DP-SGD in practice to protect against reconstruction attacks. Finally, we use our methods to demonstrate that different settings of the DP-SGD parameters leading to same DP guarantees can results in significantly different success rates for reconstruction, indicating that the DP guarantee alone might not be a good proxy for controlling the protection against reconstruction attacks.
****

T2I-CompBench: A Comprehensive Benchmark for Open-world Compositional Text-to-image Generation

Kaiyi Huang, Kaiyue Sun, Enze Xie, Zhenguo Li, Xihui Liu

Despite the stunning ability to generate high-quality images by recent text-to-image models, current approaches often struggle to effectively compose objects with different attributes and relationships into a complex and coherent scene. We propose T2I-CompBench, a comprehensive benchmark for open-world compositional text-to-image generation, consisting of 6,000 compositional text prompts from 3 categories (attribute binding, object relationships, and complex compositions) and 6 sub-categories (color binding, shape binding, texture binding, spatial relationships, non-spatial relationships, and complex compositions). We further propose several evaluation metrics specifically designed to evaluate compositional text-to-image generation and explore the potential and limitations of multimodal LLMs for evaluation. We introduce a new approach, Generative mOdel finetuning with Reward-driven Sample selection (GORS), to boost the compositional text-to-image generation abilities of pretrained text-to-image models. Extensive experiments and evaluations are conducted to benchmark previous methods on T2I-CompBench, and to validate the effectiveness of our proposed evaluation metrics and GORS approach. Project page is available at https://karine-h.github.io/T2I-CompBench/.
****

Neural Processes with Stability

Huafeng Liu, Liping Jing, Jian Yu

Unlike traditional statistical models depending on hand-specified priors, neural processes (NPs) have recently emerged as a class of powerful neural statistical models that combine the strengths of neural networks and stochastic processes. NPs can define a flexible class of stochastic processes well suited for highly non-trivial functions by encoding contextual knowledge into the function space. H

owever, noisy context points introduce challenges to the algorithmic stability that small changes in training data may significantly change the models and yield lower generalization performance. In this paper, we provide theoretical guidelines for deriving stable solutions with high generalization by introducing the notion of algorithmic stability into NPs, which can be flexible to work with various NPs and achieves less biased approximation with theoretical guarantees. To illustrate the superiority of the proposed model, we perform experiments on both synthetic and real-world data, and the results demonstrate that our approach not only helps to achieve more accurate performance but also improves model robustness.
****
Multi-Agent Learning with Heterogeneous Linear Contextual Bandits
Anh Do, Thanh Nguyen-Tang, Raman Arora
As trained intelligent systems become increasingly pervasive, multiagent learning has emerged as a popular framework for studying complex interactions between autonomous agents. Yet, a formal understanding of how and when learners in heterogeneous environments benefit from sharing their respective experiences is far from complete. In this paper, we seek answers to these questions in the context of linear contextual bandits. We present a novel distributed learning algorithm based on the upper confidence bound (UCB) algorithm, which we refer to as H-LINUCB, wherein agents cooperatively minimize the group regret under the coordination of a central server. In the setting where the level of heterogeneity or dissimilarity across the environments is known to the agents, we show that H-LINUCB is provably optimal in regimes where the tasks are highly similar or highly dissimilar.
****
A polar prediction model for learning to represent visual transformations
Pierre-Étienne Fiquet, Eero Simoncelli
All organisms make temporal predictions, and their evolutionary fitness level depends on the accuracy of these predictions. In the context of visual perception, the motions of both the observer and objects in the scene structure the dynamics of sensory signals, allowing for partial prediction of future signals based on past ones. Here, we propose a self-supervised representation-learning framework that extracts and exploits the regularities of natural videos to compute accurate predictions. We motivate the polar architecture by appealing to the Fourier shift theorem and its group-theoretic generalization, and we optimize its parameters on next-frame prediction. Through controlled experiments, we demonstrate that this approach can discover the representation of simple transformation groups acting in data. When trained on natural video datasets, our framework achieves better prediction performance than traditional motion compensation and rivals conventional deep networks, while maintaining interpretability and speed. Furthermore, the polar computations can be restructured into components resembling normalized simple and direction-selective complex cell models of primate V1 neurons. Thus, polar prediction offers a principled framework for understanding how the visual system represents sensory inputs in a form that simplifies temporal prediction.
****
MEGABYTE: Predicting Million-byte Sequences with Multiscale Transformers
LILI YU, Daniel Simig, Colin Flaherty, Armen Aghajanyan, Luke Zettlemoyer, Mike Lewis
Autoregressive transformers are spectacular models for short sequences but scale poorly to long sequences such as high-resolution images, podcasts, code, or books. We proposed Megabyte, a multi-scale decoder architecture that enables end-to-end differentiable modeling of sequences of over one million bytes. Megabyte segments sequences into patches and uses a local submodel within patches and a global model between patches. This enables sub-quadratic self-attention, much larger feedforward layers for the same compute, and improved parallelism during decoding---unlocking better performance at reduced cost for both training and generation. Extensive experiments show that Megabyte allows byte-level models to perform competitively with subword models on long context language modeling, achieve

state-of-the-art density estimation on ImageNet, and model audio from raw files. Together, these results establish the viability of tokenization-free autoregressive sequence modeling at scale.
****
CQM: Curriculum Reinforcement Learning with a Quantized World Model
Seungjae Lee, Daesol Cho, Jonghae Park, H. Jin Kim
Recent curriculum Reinforcement Learning (RL) has shown notable progress in solving complex tasks by proposing sequences of surrogate tasks. However, the previous approaches often face challenges when they generate curriculum goals in a high-dimensional space. Thus, they usually rely on manually specified goal spaces. To alleviate this limitation and improve the scalability of the curriculum, we propose a novel curriculum method that automatically defines the semantic goal space which contains vital information for the curriculum process, and suggests curriculum goals over it. To define the semantic goal space, our method discretizes continuous observations via vector quantized-variational autoencoders (VQ-VAE) and restores the temporal relations between the discretized observations by a graph. Concurrently, ours suggests uncertainty and temporal distance-aware curriculum goals that converges to the final goals over the automatically composed goal space. We demonstrate that the proposed method allows efficient explorations in an uninformed environment with raw goal examples only. Also, ours outperforms the state-of-the-art curriculum RL methods on data efficiency and performance, in various goal-reaching tasks even with ego-centric visual inputs.
****
Debiasing Conditional Stochastic Optimization
Lie He, Shiva Kasiviswanathan
In this paper, we study the conditional stochastic optimization (CSO) problem which covers a variety of applications including portfolio selection, reinforcement learning, robust learning, causal inference, etc. The sample-averaged gradient of the CSO objective is biased due to its nested structure, and therefore requires a high sample complexity for convergence. We introduce a general stochastic extrapolation technique that effectively reduces the bias. We show that for nonconvex smooth objectives, combining this extrapolation with variance reduction techniques can achieve a significantly better sample complexity than the existing bounds. Additionally, we develop new algorithms for the finite-sum variant of the CSO problem that also significantly improve upon existing results. Finally, we believe that our debiasing technique has the potential to be a useful tool for addressing similar challenges in other stochastic optimization problems.
****
Cascading Bandits: Optimizing Recommendation Frequency in Delayed Feedback Environments
Dairui Wang, Junyu Cao, Yan Zhang, Wei Qi
Delayed feedback is a critical problem in dynamic recommender systems. In practice, the feedback result often depends on the frequency of recommendation. Most existing online learning literature fails to consider optimization of the recommendation frequency, and regards the reward from each successfully recommended message to be equal. In this paper, we consider a novel cascading bandits setting, where individual messages from a selected list are sent to a user periodically. Whenever a user does not like a message, she may abandon the system with a probability positively correlated with the recommendation frequency. A learning agent needs to learn both the underlying message attraction probabilities and users' abandonment probabilities through the randomly delayed feedback. We first show a dynamic programming solution to finding the optimal message sequence in deterministic scenarios, in which the reward is allowed to vary with different messages. Then we propose a polynomial time UCB-based offline learning algorithm, and discuss its performance by characterizing its regret bound. For the online setting, we propose a learning algorithm which allows adaptive content for a given user. Numerical experiment on AmEx dataset confirms the effectiveness of our algorithms.
****
CoPriv: Network/Protocol Co-Optimization for Communication-Efficient Private Inf

erence
Wenxuan Zeng, Meng Li, Haichuan Yang, Wen-jie Lu, Runsheng Wang, Ru Huang
Deep neural network (DNN) inference based on secure 2-party computation (2PC) can offer cryptographically-secure privacy protection but suffers from orders of magnitude latency overhead due to enormous communication. Previous works heavily rely on a proxy metric of ReLU counts to approximate the communication overhead and focus on reducing the ReLUs to improve the communication efficiency. However, we observe these works achieve limited communication reduction for state-of-the-art (SOTA) 2PC protocols due to the ignorance of other linear and non-linear operations, which now contribute to the majority of communication. In this work, we present CoPriv, a framework that jointly optimizes the 2PC inference protocol and the DNN architecture. CoPriv features a new 2PC protocol for convolution based on Winograd transformation and develops DNN-aware optimization to significantly reduce the inference communication. CoPriv further develops a 2PC-aware network optimization algorithm that is compatible with the proposed protocol and simultaneously reduces the communication for all the linear and non-linear operations. We compare CoPriv with the SOTA 2PC protocol, CrypTFlow2, and demonstrate 2.1× communication reduction for both ResNet-18 and ResNet-32 on CIFAR-100. We also compare CoPriv with SOTA network optimization methods, including SNL, MetaPruning, etc. CoPriv achieves 9.98× and 3.88× online and total communication reduction with a higher accuracy compare to SNL, respectively. CoPriv also achieves 3.87× online communication reduction with more than 3% higher accuracy compared to MetaPruning.
****
Generalized equivalences between subsampling and ridge regularization
Pratik Patil, Jin-Hong Du
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
D4Explainer: In-distribution Explanations of Graph Neural Network via Discrete Denoising Diffusion
Jialin Chen, Shirley Wu, Abhijit Gupta, Rex Ying
The widespread deployment of Graph Neural Networks (GNNs) sparks significant interest in their explainability, which plays a vital role in model auditing and ensuring trustworthy graph learning. The objective of GNN explainability is to discern the underlying graph structures that have the most significant impact on model predictions. Ensuring that explanations generated are reliable necessitates consideration of the in-distribution property, particularly due to the vulnerability of GNNs to out-of-distribution data. Unfortunately, prevailing explainability methods tend to constrain the generated explanations to the structure of the original graph, thereby downplaying the significance of the in-distribution property and resulting in explanations that lack reliability.To address these challenges, we propose D4Explainer, a novel approach that provides in-distribution GNN explanations for both counterfactual and model-level explanation scenarios. The proposed D4Explainer incorporates generative graph distribution learning into the optimization objective, which accomplishes two goals: 1) generate a collection of diverse counterfactual graphs that conform to the in-distribution property for a given instance, and 2) identify the most discriminative graph patterns that contribute to a specific class prediction, thus serving as model-level explanations. It is worth mentioning that D4Explainer is the first unified framework that combines both counterfactual and model-level explanations.Empirical evaluations conducted on synthetic and real-world datasets provide compelling evidence of the state-of-the-art performance achieved by D4Explainer in terms of explanation accuracy, faithfulness, diversity, and robustness.
****
Core-sets for Fair and Diverse Data Summarization
Sepideh Mahabadi, Stojan Trajanovski
Requests for name changes in the electronic proceedings will be accepted with no

questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Energy-Efficient Scheduling with Predictions
Eric Balkanski, Noemie Perivier, Clifford Stein, Hao-Ting Wei
An important goal of modern scheduling systems is to efficiently manage power usage. In energy-efficient scheduling, the operating system controls the speed at which a machine is processing jobs with the dual objective of minimizing energy consumption and optimizing the quality of service cost of the resulting schedule. Since machine-learned predictions about future requests can often be learned from historical data, a recent line of work on learning-augmented algorithms aims to achieve improved performance guarantees by leveraging predictions. In particular, for energy-efficient scheduling, Bamas et. al. [NeurIPS '20] and Antoniadis et. al. [SWAT '22] designed algorithms with predictions for the energy minimization with deadlines problem and achieved an improved competitive ratio when the prediction error is small while also maintaining worst-case bounds even when the prediction error is arbitrarily large.In this paper, we consider a general setting for energy-efficient scheduling and provide a flexible learning-augmented algorithmic framework that takes as input an offline and an online algorithm for the desired energy-efficient scheduling problem. We show that, when the prediction error is small, this framework gives improved competitive ratios for many different energy-efficient scheduling problems, including energy minimization with deadlines, while also maintaining a bounded competitive ratio regardless of the prediction error. Finally, we empirically demonstrate that this framework achieves an improved performance on real and synthetic datasets.
****

Diversify Your Vision Datasets with Automatic Diffusion-based Augmentation
Lisa Dunlap, Alyssa Umino, Han Zhang, Jiezhi Yang, Joseph E. Gonzalez, Trevor Darrell
Many fine-grained classification tasks, like rare animal identification, have limited training data and consequently classifiers trained on these datasets often fail to generalize to variations in the domain like changes in weather or location. As such, we explore how natural language descriptions of the domains seen in training data can be used with large vision models trained on diverse pretraining datasets to generate useful variations of the training data. We introduce ALIA (Automated Language-guided Image Augmentation), a method which utilizes large vision and language models to automatically generate natural language descriptions of a dataset's domains and augment the training data via language-guided image editing. To maintain data integrity, a model trained on the original dataset filters out minimal image edits and those which corrupt class-relevant information. The resulting dataset is visually consistent with the original training data and offers significantly enhanced diversity. We show that ALIA is able to surpasses traditional data augmentation and text-to-image generated data on fine-grained classification tasks, including cases of domain generalization and contextual bias. Code is available at https://github.com/lisadunlap/ALIA.
****

DISCS: A Benchmark for Discrete Sampling
Katayoon Goshvadi, Haoran Sun, Xingchao Liu, Azade Nova, Ruqi Zhang, Will Grathwohl, Dale Schuurmans, Hanjun Dai
Sampling in discrete spaces, with critical applications in simulation and optimization, has recently been boosted by significant advances in gradient-based approaches that exploit modern accelerators like GPUs. However, two key challenges are hindering further advancement in research on discrete sampling. First, since there is no consensus on experimental settings and evaluation setups, the empirical results in different research papers are often not comparable. Second, implementing samplers and target distributions often requires a nontrivial amount of effort in terms of calibration and parallelism. To tackle these challenges, we propose DISCS (DISCrete Sampling), a tailored package and benchmark that supports unified and efficient experiment implementation and evaluations for discrete sa

mpling in three types of tasks: sampling from classical graphical models and ene
rgy based generative models, and sampling for solving combinatorial optimization
. Throughout the comprehensive evaluations in DISCS, we gained new insights into
 scalability, design principles for proposal distributions, and lessons for adap
tive sampling design. DISCS efficiently implements representative discrete sampl
ers in existing research works as baselines and offers a simple interface that r
esearchers can conveniently add new discrete samplers and directly compare their
 performance with the benchmark result in a calibrated setup.
****
Improving Robustness with Adaptive Weight Decay
Mohammad Amin Ghiasi, Ali Shafahi, Reza Ardekani
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Leveraging Pre-trained Large Language Models to Construct and Utilize World Mode
ls for Model-based Task Planning
Lin Guan, Karthik Valmeekam, Sarath Sreedharan, Subbarao Kambhampati
There is a growing interest in applying pre-trained large language models (LLMs)
 to planning problems. However, methods that use LLMs directly as planners are c
urrently impractical due to several factors, including limited correctness of pl
ans, strong reliance on feedback from interactions with simulators or even the a
ctual environment, and the inefficiency in utilizing human feedback. In this wor
k, we introduce a novel alternative paradigm that constructs an explicit world (
domain) model in planning domain definition language (PDDL) and then uses it to
plan with sound domain-independent planners. To address the fact that LLMs may n
ot generate a fully functional PDDL model initially, we employ LLMs as an interf
ace between PDDL and sources of corrective feedback, such as PDDL validators and
 humans. For users who lack a background in PDDL, we show that LLMs can translat
e PDDL into natural language and effectively encode corrective feedback back to
the underlying domain model. Our framework not only enjoys the correctness guara
ntee offered by the external planners but also reduces human involvement by allo
wing users to correct domain models at the beginning, rather than inspecting and
 correcting (through interactive prompting) every generated plan as in previous
work. On two IPC domains and a Household domain that is more complicated than co
mmonly used benchmarks such as ALFWorld, we demonstrate that GPT-4 can be levera
ged to produce high-quality PDDL models for over 40 actions, and the corrected P
DDL models are then used to successfully solve 48 challenging planning tasks. Re
sources, including the source code, are released at: https://guansuns.github.io/
pages/llm-dm.
****
Described Object Detection: Liberating Object Detection with Flexible Expression
s
Chi Xie, Zhao Zhang, Yixuan Wu, Feng Zhu, Rui Zhao, Shuang Liang
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Learning Cuts via Enumeration Oracles
Daniel Thuerck, Boro Sofranac, Marc E Pfetsch, Sebastian Pokutta
Cutting-planes are one of the most important building blocks for solving large-s
cale integer programming (IP) problems to (near) optimality. The majority of cut
ting plane approaches rely on explicit rules to derive valid inequalities that c
an separate the target point from the feasible set. Local cuts, on the other han
d, seek to directly derive the facets of the underlying polyhedron and use them
as cutting planes. However, current approaches rely on solving Linear Programmin
g (LP) problems in order to derive such a hyperplane. In this paper, we present
a novel generic approach for learning the facets of the underlying polyhedron by

accessing it implicitly via an enumeration oracle in a reduced dimension. This is achieved by embedding the oracle in a variant of the Frank-Wolfe algorithm which is capable of generating strong cutting planes, effectively turning the enumeration oracle into a separation oracle. We demonstrate the effectiveness of our approach with a case study targeting the multidimensional knapsack problem (MKP).
****
Learning Descriptive Image Captioning via Semipermeable Maximum Likelihood Estimation
Zihao Yue, Anwen Hu, Liang Zhang, Qin Jin
Image captioning aims to describe visual content in natural language. As 'a picture is worth a thousand words', there could be various correct descriptions for an image. However, with maximum likelihood estimation as the training objective, the captioning model is penalized whenever its prediction mismatches with the label. For instance, when the model predicts a word expressing richer semantics than the label, it will be penalized and optimized to prefer more concise expressions, referred to as conciseness optimization. In contrast, predictions that are more concise than labels lead to richness optimization. Such conflicting optimization directions could eventually result in the model generating general descriptions. In this work, we introduce Semipermeable MaxImum Likelihood Estimation (SMILE), which allows richness optimization while blocking conciseness optimization, thus encouraging the model to generate longer captions with more details. Extensive experiments on two mainstream image captioning datasets MSCOCO and Flickr30K demonstrate that SMILE significantly enhances the descriptiveness of generated captions. We further provide in-depth investigations to facilitate a better understanding of how SMILE works.
****
Alternating Gradient Descent and Mixture-of-Experts for Integrated Multimodal Perception
Hassan Akbari, Dan Kondratyuk, Yin Cui, Rachel Hornung, Huisheng Wang, Hartwig Adam
We present Integrated Multimodal Perception (IMP), a simple and scalable multimodal multi-task training and modeling approach. IMP integrates multimodal inputs including image, video, text, and audio into a single Transformer encoder with minimal modality-specific components. IMP makes use of a novel design that combines Alternating Gradient Descent (AGD) and Mixture-of-Experts (MoE) for efficient model & task scaling. We conduct extensive empirical studies and reveal the following key insights: 1) performing gradient descent updates by alternating on diverse modalities, loss functions, and tasks, with varying input resolutions, efficiently improves the model.   2) sparsification with MoE on a single modality-agnostic encoder substantially improves the performance, outperforming dense models that use modality-specific encoders or additional fusion layers and greatly mitigating the conflicts between modalities. IMP achieves competitive performance on a wide range of downstream tasks including video classification, image classification, image-text, and video-text retrieval. Most notably, we train a sparse IMP-MoE-L focusing on video tasks that achieves new state-of-the-art in zero-shot video classification: 77.0% on Kinetics-400, 76.8% on Kinetics-600, and 68.3% on Kinetics-700, improving the previous state-of-the-art by +5%, +6.7%, and +5.8%, respectively, while using only 15% of their total training computational cost.
****
The RefinedWeb Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data Only
Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Hamza Alobeidli, Alessandro Cappelli, Baptiste Pannier, Ebtesam Almazrouei, Julien Launay
Large language models are commonly trained on a mixture of filtered web data and curated ``high-quality'' corpora, such as social media conversations, books, or technical papers. This curation process is believed to be necessary to produce performant models with broad zero-shot generalization abilities. However, as larger models requiring pretraining on trillions of tokens are considered, it is un

clear how scalable is curation, and whether we will run out of unique high-quali
ty data soon. At variance with previous beliefs, we show that properly filtered
 and deduplicated web data alone can lead to powerful models; even significantly
 outperforming models trained on The Pile. Despite extensive filtering, the high
-quality data we extract from the web is still plentiful, and we are able to obt
ain five trillion tokens from CommonCrawl. We publicly release an extract of 500
 billion tokens from our RefinedWeb dataset, and 1.3/7.5B parameters language mo
dels trained on it.
****
Self-Correcting Bayesian Optimization through Bayesian Active Learning
Carl Hvarfner, Erik Hellsten, Frank Hutter, Luigi Nardi
Gaussian processes are the model of choice in Bayesian optimization and active l
earning. Yet, they are highly dependent on cleverly chosen hyperparameters to re
ach their full potential, and little effort is devoted to finding good hyperpara
meters in the literature. We demonstrate the impact of selecting good hyperparam
eters for GPs and present two acquisition functions that explicitly prioritize h
yperparameter learning. Statistical distance-based Active Learning (SAL) conside
rs the average disagreement between samples from the posterior, as measured by a
 statistical distance. SAL outperforms the state-of-the-art in Bayesian active l
earning on several test functions. We then introduce Self-Correcting Bayesian Op
timization (SCoreBO), which extends SAL to perform Bayesian optimization and act
ive learning simultaneously. SCoreBO learns the model hyperparameters at improve
d rates compared to vanilla BO, while outperforming the latest Bayesian optimiza
tion methods on traditional benchmarks. Moreover, we demonstrate the importance
of self-correction on atypical Bayesian optimization tasks.
****
SE(3) Equivariant Augmented Coupling Flows
Laurence Midgley, Vincent Stimper, Javier Antorán, Emile Mathieu, Bernhard Schöl
kopf, José Miguel Hernández-Lobato
Coupling normalizing flows allow for fast sampling and density evaluation, makin
g them the tool of choice for probabilistic modeling of physical systems. Howeve
r, the standard coupling architecture precludes endowing flows that operate on t
he Cartesian coordinates of atoms with the SE(3) and permutation invariances of
physical systems. This work proposes a coupling flow that preserves SE(3) and pe
rmutation equivariance by performing coordinate splits along additional augmente
d dimensions. At each layer, the flow maps atoms' positions into learned SE(3) i
nvariant bases, where we apply standard flow transformations, such as monotonic
rational-quadratic splines, before returning to the original basis.Crucially, ou
r flow preserves fast sampling and density evaluation, and may be used to produc
e unbiased estimates of expectations with respect to the target distribution via
 importance sampling.When trained on the DW4, LJ13, and QM9-positional datasets,
 our flow is competitive with equivariant continuous normalizing flows and diffu
sion models, while allowing sampling more than an order of magnitude faster.More
over, to the best of our knowledge, we are the first to learn the full Boltzmann
 distribution of alanine dipeptide by only modeling the Cartesian positions of i
ts atoms.Lastly, we demonstrate that our flow can be trained to approximately sa
mple from the Boltzmann distribution of the DW4 and LJ13 particle systems using
only their energy functions.
****
Bridging the Domain Gap: Self-Supervised 3D Scene Understanding with Foundation
Models
Zhimin Chen, Longlong Jing, Yingwei Li, Bing Li
Foundation models have achieved remarkable results in 2D and language tasks like
 image segmentation, object detection, and visual-language understanding. Howeve
r, their potential to enrich 3D scene representation learning is largely untappe
d due to the existence of the domain gap. In this work, we propose an innovative
 methodology called Bridge3D to address this gap by pre-training 3D models using
 features, semantic masks, and captions sourced from foundation models. Specific
ally, our method employs semantic masks from foundation models to guide the mask
ing and reconstruction process for the masked autoencoder, enabling more focused

attention on foreground representations. Moreover, we bridge the 3D-text gap at the scene level using image captioning foundation models, thereby facilitating scene-level knowledge distillation. We further extend this bridging effort by introducing an innovative object-level knowledge distillation method that harnesses highly accurate object-level masks and semantic text data from foundation models. Our methodology significantly surpasses the performance of existing state-of-the-art methods in 3D object detection and semantic segmentation tasks. For instance, on the ScanNet dataset, Bridge3D improves the baseline by a notable margin of 6.3%. Code will be available at: https://github.com/Zhimin-C/Bridge3D
****

Imagine That! Abstract-to-Intricate Text-to-Image Synthesis with Scene Graph Hallucination Diffusion
Shengqiong Wu, Hao Fei, Hanwang Zhang, Tat-Seng Chua
In this work, we investigate the task of text-to-image (T2I) synthesis under the abstract-to-intricate setting, i.e., generating intricate visual content from simple abstract text prompts. Inspired by human imagination intuition, we propose a novel scene-graph hallucination (SGH) mechanism for effective abstract-to-intricate T2I synthesis. SGH carries out scene hallucination by expanding the initial scene graph (SG) of the input prompt with more feasible specific scene structures, in which the structured semantic representation of SG ensures high controllability of the intrinsic scene imagination. To approach the T2I synthesis, we deliberately build an SG-based hallucination diffusion system. First, we implement the SGH module based on the discrete diffusion technique, which evolves the SG structure by iteratively adding new scene elements. Then, we utilize another continuous-state diffusion model as the T2I synthesizer, where the overt image-generating process is navigated by the underlying semantic scene structure induced from the SGH module. On the benchmark COCO dataset, our system outperforms the existing best-performing T2I model by a significant margin, especially improving on the abstract-to-intricate T2I generation. Further in-depth analyses reveal how our methods advance.
****

A unified framework for information-theoretic generalization bounds
Yifeng Chu, Maxim Raginsky
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Diverse Shape Completion via Style Modulated Generative Adversarial Networks
Wesley Khademi, Fuxin Li
Shape completion aims to recover the full 3D geometry of an object from a partial observation. This problem is inherently multi-modal since there can be many ways to plausibly complete the missing regions of a shape. Such diversity would be indicative of the underlying uncertainty of the shape and could be preferable for downstream tasks such as planning. In this paper, we propose a novel conditional generative adversarial network that can produce many diverse plausible completions of a partially observed point cloud. To enable our network to produce multiple completions for the same partial input, we introduce stochasticity into our network via style modulation. By extracting style codes from complete shapes during training, and learning a distribution over them, our style codes can explicitly carry shape category information leading to better completions. We further introduce diversity penalties and discriminators at multiple scales to prevent conditional mode collapse and to train without the need for multiple ground truth completions for each partial input. Evaluations across several synthetic and real datasets demonstrate that our method achieves significant improvements in respecting the partial observations while obtaining greater diversity in completions.
****

On the Role of Randomization in Adversarially Robust Classification
Lucas Gnecco Heredia, Muni Sreenivas Pydi, Laurent Meunier, Benjamin Negrevergne

, Yann Chevaleyre
Deep neural networks are known to be vulnerable to small adversarial perturbations in test data. To defend against adversarial attacks, probabilistic classifiers have been proposed as an alternative to deterministic ones. However, literature has conflicting findings on the effectiveness of probabilistic classifiers in comparison to deterministic ones. In this paper, we clarify the role of randomization in building adversarially robust classifiers.Given a base hypothesis set of deterministic classifiers, we show the conditions under which a randomized ensemble outperforms the hypothesis set in adversarial risk, extending previous results.Additionally, we show that for any probabilistic binary classifier (including randomized ensembles), there exists a deterministic classifier that outperforms it. Finally, we give an explicit description of the deterministic hypothesis set that contains such a deterministic classifier for many types of commonly used probabilistic classifiers, i.e. randomized ensembles and parametric/input noise injection.
****
Controlling Text-to-Image Diffusion by Orthogonal Finetuning
Zeju Qiu, Weiyang Liu, Haiwen Feng, Yuxuan Xue, Yao Feng, Zhen Liu, Dan Zhang, Adrian Weller, Bernhard Schölkopf
Large text-to-image diffusion models have impressive capabilities in generating photorealistic images from text prompts. How to effectively guide or control these powerful models to perform different downstream tasks becomes an important open problem. To tackle this challenge, we introduce a principled finetuning method -- Orthogonal Finetuning (OFT), for adapting text-to-image diffusion models to downstream tasks. Unlike existing methods, OFT can provably preserve hyperspherical energy which characterizes the pairwise neuron relationship on the unit hypersphere. We find that this property is crucial for preserving the semantic generation ability of text-to-image diffusion models. To improve finetuning stability, we further propose Constrained Orthogonal Finetuning (COFT) which imposes an additional radius constraint to the hypersphere. Specifically, we consider two important finetuning text-to-image tasks: subject-driven generation where the goal is to generate subject-specific images given a few images of a subject and a text prompt, and controllable generation where the goal is to enable the model to take in additional control signals. We empirically show that our OFT framework outperforms existing methods in generation quality and convergence speed.
****
NCDL: A Framework for Deep Learning on non-Cartesian Lattices
Joshua Horacsek, Usman Alim
The use of non-Cartesian grids is a niche but important topic in sub-fields of the numerical sciences such as simulation and scientific visualization. However, non-Cartesian approaches are virtually unexplored in machine learning. This is likely due to the difficulties in the representation of data on non-Cartesian domains and the lack of support for standard machine learning operations on non-Cartesian data. This paper proposes a new data structure called the lattice tensor which generalizes traditional tensor spatio-temporal operations to lattice tensors, enabling the use of standard machine learning algorithms on non-Cartesian data. However, data need not reside on a non-Cartesian structure, we use non-Dyadic downsampling schemes to bring Cartesian data into a non-Cartesian space for further processing.   We introduce a software library that implements the lattice tensor container (with some common machine learning operations), and   demonstrate its effectiveness. Our method provides a general framework for machine learning on non-Cartesian domains, addressing the challenges mentioned above and filling a gap in the current literature.
****
Human-in-the-Loop Optimization for Deep Stimulus Encoding in Visual Prostheses
Jacob Granley, Tristan Fauvel, Matthew Chalk, Michael Beyeler
Neuroprostheses show potential in restoring lost sensory function and enhancing human capabilities, but the sensations produced by current devices often seem unnatural or distorted. Exact placement of implants and differences in individual perception lead to significant variations in stimulus response, making personali

zed stimulus optimization a key challenge. Bayesian optimization could be usedto optimize patient-specific stimulation parameters with limited noisy observations, but is not feasible for high-dimensional stimuli. Alternatively, deep learning models can optimize stimulus encoding strategies, but typically assume perfect knowledge of patient-specific variations. Here we propose a novel, practically feasible approach that overcomes both of these fundamental limitations. First, a deep encoder network is trained to produce optimal stimuli for any individual patient by inverting a forward model mapping electrical stimuli to visual percepts. Second, a preferential Bayesian optimization strategy utilizes this encoder to learn the optimal patient-specific parameters for a new patient, using a minimal number of pairwise comparisons between candidate stimuli. We demonstrate the viability of this approach on a novel, state-of-the-art visual prosthesis model. Our approach quickly learns a personalized stimulus encoder and leads to dramatic improvements in the quality of restored vision, outperforming existing encoding strategies. Further, this approach is robust to noisy patient feedback and misspecifications in the underlying forward model. Overall, our results suggest that combining the strengths of deep learning and Bayesian optimization could significantly improve the perceptual experience of patients fitted with visual prostheses and may prove a viable solution for a range of neuroprosthetic technologies

****

CAST: Cross-Attention in Space and Time for Video Action Recognition

Dongho Lee, Jongseo Lee, Jinwoo Choi

Recognizing human actions in videos requires spatial and temporal understanding. Most existing action recognition models lack a balanced spatio-temporal understanding of videos. In this work, we propose a novel two-stream architecture, called Cross-Attention in Space and Time (CAST), that achieves a balanced spatio-temporal understanding of videos using only RGB input. Our proposed bottleneck cross-attention mechanism enables the spatial and temporal expert models to exchange information and make synergistic predictions, leading to improved performance. We validate the proposed method with extensive experiments on public benchmarks with different characteristics: EPIC-Kitchens-100, Something-Something-V2, and Kinetics-400. Our method consistently shows favorable performance across these datasets, while the performance of existing methods fluctuates depending on the dataset characteristics. The code is available at https://github.com/KHU-VLL/CAST.

****

Faster Differentially Private Convex Optimization via Second-Order Methods

Arun Ganesh, Mahdi Haghifam, Thomas Steinke, Abhradeep Guha Thakurta

Differentially private (stochastic) gradient descent is the workhorse of DP private machine learning in both the convex and non-convex settings. Without privacy constraints, second-order methods, like Newton's method, converge faster than first-order methods like gradient descent. In this work, we investigate the prospect of using the second-order information from the loss function to accelerate DP convex optimization. We first develop a private variant of the regularized cubic Newton method of Nesterov and Polyak, and show that for the class of strongly convex loss functions, our algorithm has quadratic convergence and achieves the optimal excess loss. We then design a practical second-order DP algorithm for the unconstrained logistic regression problem. We theoretically and empirically study the performance of our algorithm. Empirical results show our algorithm consistently achieves the best excess loss compared to other baselines and is 10-40x faster than DP-GD/DP-SGD for challenging datasets.

****

Auditing for Human Expertise

Rohan Alur, Loren Laine, Darrick Li, Manish Raghavan, Devavrat Shah, Dennis Shung

High-stakes prediction tasks (e.g., patient diagnosis) are often handled by trained human experts. A common source of concern about automation in these settings is that experts may exercise intuition that is difficult to model and/or have access to information (e.g., conversations with a patient) that is simply unavailable to a would-be algorithm. This raises a natural question whether human exper

ts add value which could not be captured by an algorithmic predictor.We develop a statistical framework under which we can pose this question as a natural hypothesis test. Indeed, as our framework highlights, detecting human expertise is more subtle than simply comparing the accuracy of expert predictions to those made by a particular learning algorithm. Instead, we propose a simple procedure which tests whether expert predictions are statistically independent from the outcomes of interest after conditioning on the available inputs (‘features’). A rejection of our test thus suggests that human experts may add value to any algorithm trained on the available data, and has direct implications for whether human-AI ‘complementarity’ is achievable in a given prediction task.We highlight the utility of our procedure using admissions data collected from the emergency department of a large academic hospital system, where we show that physicians’ admit/discharge decisions for patients with acute gastrointestinal bleeding (AGIB) appear to be incorporating information that is not available to a standard algorithmic screening tool. This is despite the fact that the screening tool is arguably more accurate than physicians’ discretionary decisions, highlighting that – even absent normative concerns about accountability or interpretability – accuracy is insufficient to justify algorithmic automation.
****
Smooth, exact rotational symmetrization for deep learning on point clouds
Sergey Pozdnyakov, Michele Ceriotti
Point clouds are versatile representations of 3D objects and have found widespread application in science and engineering. Many successful deep-learning models have been proposed that use them as input. The domain of chemical and materials modeling is especially challenging because exact compliance with physical constraints is highly desirable for a model to be usable in practice. These constraints include smoothness and invariance with respect to translations, rotations, and permutations of identical atoms. If these requirements are not rigorously fulfilled, atomistic simulations might lead to absurd outcomes even if the model has excellent accuracy. Consequently, dedicated architectures, which achieve invariance by restricting their design space, have been developed. General-purpose point-cloud models are more varied but often disregard rotational symmetry. We propose a general symmetrization method that adds rotational equivariance to any given model while preserving all the other requirements.Our approach simplifies the development of better atomic-scale machine-learning schemes by relaxing the constraints on the design space and making it possible to incorporate ideas that proved effective in other domains.We demonstrate this idea by introducing the Point Edge Transformer (PET) architecture, which is not intrinsically equivariant but achieves state-of-the-art performance on several benchmark datasets of molecules and solids. A-posteriori application of our general protocol makes PET exactly equivariant, with minimal changes to its accuracy.
****
Functional Equivalence and Path Connectivity of Reducible Hyperbolic Tangent Networks
Matthew Farrugia-Roberts
Understanding the learning process of artificial neural networks requires clarifying the structure of the parameter space within which learning takes place. A neural network parameter's functional equivalence class is the set of parameters implementing the same input--output function. For many architectures, almost all parameters have a simple and well-documented functional equivalence class. However, there is also a vanishing minority of reducible parameters, with richer functional equivalence classes caused by redundancies among the network's units.In this paper, we give an algorithmic characterisation of unit redundancies and reducible functional equivalence classes for a single-hidden-layer hyperbolic tangent architecture. We show that such functional equivalence classes are piecewise-linear path-connected sets, and that for parameters with a majority of redundant units, the sets have a diameter of at most 7 linear segments.
****
On Robust Streaming for Learning with Experts: Algorithms and Lower Bounds
David Woodruff, Fred Zhang, Samson Zhou

Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Stochastic Optimal Control for Collective Variable Free Sampling of Molecular Transition Paths
Lars Holdijk, Yuanqi Du, Ferry Hooft, Priyank Jaini, Berend Ensing, Max Welling
We consider the problem of sampling transition paths between two given metastable states of a molecular system, eg. a folded and unfolded protein or products and reactants of a chemical reaction. Due to the existence of high energy barriers separating the states, these transition paths are unlikely to be sampled with standard Molecular Dynamics (MD) simulation. Traditional methods to augment MD with a bias potential to increase the probability of the transition rely on a dimensionality reduction step based on Collective Variables (CVs). Unfortunately, selecting appropriate CVs requires chemical intuition and traditional methods are therefore not always applicable to larger systems. Additionally, when incorrect CVs are used, the bias potential might not be minimal and bias the system along dimensions irrelevant to the transition. Showing a formal relation between the problem of sampling molecular transition paths, the Schrodinger bridge problem and stochastic optimal control with neural network policies, we propose a machine learning method for sampling said transitions. Unlike previous non-machine learning approaches our method, named PIPS, does not depend on CVs. We show that our method successful generates low energy transitions for Alanine Dipeptide as well as the larger Polyproline and Chignolin proteins.
****
RangePerception: Taming LiDAR Range View for Efficient and Accurate 3D Object Detection
Yeqi BAI, Ben Fei, Youquan Liu, Tao MA, Yuenan Hou, Botian Shi, Yikang LI
LiDAR-based 3D detection methods currently use bird's-eye view (BEV) or range view (RV) as their primary basis. The former relies on voxelization and 3D convolutions, resulting in inefficient training and inference processes. Conversely, RV-based methods demonstrate higher efficiency due to their compactness and compatibility with 2D convolutions, but their performance still trails behind that of BEV-based methods. To eliminate this performance gap while preserving the efficiency of RV-based methods, this study presents an efficient and accurate RV-based 3D object detection framework termed RangePerception. Through meticulous analysis, this study identifies two critical challenges impeding the performance of existing RV-based methods: 1) there exists a natural domain gap between the 3D world coordinate used in output and 2D range image coordinate used in input, generating difficulty in information extraction from range images; 2) native range images suffer from vision corruption issue, affecting the detection accuracy of the objects located on the margins of the range images. To address the key challenges above, we propose two novel algorithms named Range Aware Kernel (RAK) and Vision Restoration Module (VRM), which facilitate information flow from range image representation and world-coordinate 3D detection results. With the help of RAK and VRM, our RangePerception achieves 3.25/4.18 higher averaged L1/L2 AP compared to previous state-of-the-art RV-based method RangeDet, on Waymo Open Dataset. For the first time as an RV-based 3D detection method, RangePerception achieves slightly superior averaged AP compared with the well-known BEV-based method CenterPoint and the inference speed of RangePerception is 1.3 times as fast as CenterPoint.
****
One-for-All: Bridge the Gap Between Heterogeneous Architectures in Knowledge Distillation
Zhiwei Hao, Jianyuan Guo, Kai Han, Yehui Tang, Han Hu, Yunhe Wang, Chang Xu
Knowledge distillation (KD) has proven to be a highly effective approach for enhancing model performance through a teacher-student training scheme. However, most existing distillation methods are designed under the assumption that the teacher and student models belong to the same model family, particularly the hint-bas

ed approaches. By using centered kernel alignment (CKA) to compare the learned features between heterogeneous teacher and student models, we observe significant feature divergence. This divergence illustrates the ineffectiveness of previous hint-based methods in cross-architecture distillation. To tackle the challenge in distilling heterogeneous models, we propose a simple yet effective one-for-all KD framework called OFA-KD, which significantly improves the distillation performance between heterogeneous architectures. Specifically, we project intermediate features into an aligned latent space such as the logits space, where architecture-specific information is discarded. Additionally, we introduce an adaptive target enhancement scheme to prevent the student from being disturbed by irrelevant information. Extensive experiments with various architectures, including CNN, Transformer, and MLP, demonstrate the superiority of our OFA-KD framework in enabling distillation between heterogeneous architectures. Specifically, when equipped with our OFA-KD, the student models achieve notable performance improvements, with a maximum gain of 8.0% on the CIFAR-100 dataset and 0.7% on the ImageNet-1K dataset. PyTorch code and checkpoints can be found at https://github.com/Hao840/OFAKD.
****
The Graph Pencil Method: Mapping Subgraph Densities to Stochastic Block Models
Lee Gunderson, Gecia Bravo-Hermsdorff, Peter Orbanz
In this work, we describe a method that determines an exact map from a finite set of subgraph densities to the parameters of a stochastic block model (SBM) matching these densities. Given a number K of blocks, the subgraph densities of a finite number of stars and bistars uniquely determines a single element of the class of all degree-separated stochastic block models with K blocks. Our method makes it possible to translate estimates of these subgraph densities into model parameters, and hence to use subgraph densities directly for inference. The computational overhead is negligible; computing the translation map is polynomial in K, but independent of the graph size once the subgraph densities are given.
****
Cross-links Matter for Link Prediction: Rethinking the Debiased GNN from a Data Perspective
Zihan Luo, Hong Huang, Jianxun Lian, Xiran Song, Xing Xie, Hai Jin
Recently, the bias-related issues in GNN-based link prediction have raised widely spread concerns. In this paper, we emphasize the bias on links across different node clusters, which we call cross-links, after considering its significance in both easing information cocoons and preserving graph connectivity. Instead of following the objective-oriented mechanism in prior works with compromised utility, we empirically find that existing GNN models face severe data bias between internal-links (links within the same cluster) and cross-links, and this inspires us to rethink the bias issue on cross-links from a data perspective. Specifically, we design a simple yet effective twin-structure framework, which can be easily applied to most of GNNs to mitigate the bias as well as boost their utility in an end-to-end manner. The basic idea is to generate debiased node embeddings as demonstrations, and fuse them into the embeddings of original GNNs. In particular, we learn debiased node embeddings with the help of augmented supervision signals, and a novel dynamic training strategy is designed to effectively fuse debiased node embeddings with the original node embeddings. Experiments on three datasets with six common GNNs show that our framework can not only alleviate the bias between internal-links and cross-links, but also boost the overall accuracy. Comparisons with other state-of-the-art methods also verify the superiority of our method.
****
On the Robustness of Removal-Based Feature Attributions
Chris Lin, Ian Covert, Su-In Lee
To explain predictions made by complex machine learning models, many feature attribution methods have been developed that assign importance scores to input features. Some recent work challenges the robustness of these methods by showing that they are sensitive to input and model perturbations, while other work addresses this issue by proposing robust attribution methods. However, previous work on

attribution robustness has focused primarily on gradient-based feature attributions, whereas the robustness of removal-based attribution methods is not currently well understood. To bridge this gap, we theoretically characterize the robustness properties of removal-based feature attributions. Specifically, we provide a unified analysis of such methods and derive upper bounds for the difference between intact and perturbed attributions, under settings of both input and model perturbations. Our empirical results on synthetic and real-world data validate our theoretical results and demonstrate their practical implications, including the ability to increase attribution robustness by improving the model’s Lipschitz regularity.
****
Sample-efficient Multi-objective Molecular Optimization with GFlowNets
Yiheng Zhu, Jialu Wu, Chaowen Hu, Jiahuan Yan, kim hsieh, Tingjun Hou, Jian Wu
Many crucial scientific problems involve designing novel molecules with desired properties, which can be formulated as a black-box optimization problem over the discrete chemical space. In practice, multiple conflicting objectives and costly evaluations (e.g., wet-lab experiments) make the diversity of candidates paramount. Computational methods have achieved initial success but still struggle with considering diversity in both objective and search space. To fill this gap, we propose a multi-objective Bayesian optimization (MOBO) algorithm leveraging the hypernetwork-based GFlowNets (HN-GFN) as an acquisition function optimizer, with the purpose of sampling a diverse batch of candidate molecular graphs from an approximate Pareto front. Using a single preference-conditioned hypernetwork, HN-GFN learns to explore various trade-offs between objectives. We further propose a hindsight-like off-policy strategy to share high-performing molecules among different preferences in order to speed up learning for HN-GFN. We empirically illustrate that HN-GFN has adequate capacity to generalize over preferences. Moreover, experiments in various real-world MOBO settings demonstrate that our framework predominantly outperforms existing methods in terms of candidate quality and sample efficiency. The code is available at https://github.com/violet-sto/HN-GFN.
****
DeepSimHO: Stable Pose Estimation for Hand-Object Interaction via Physics Simulation
Rong Wang, Wei Mao, Hongdong Li
This paper addresses the task of 3D pose estimation for a hand interacting with an object from a single image observation. When modeling hand-object interaction, previous works mainly exploit proximity cues, while overlooking the dynamical nature that the hand must stably grasp the object to counteract gravity and thus preventing the object from slipping or falling. These works fail to leverage dynamical constraints in the estimation and consequently often produce unstable results. Meanwhile, refining unstable configurations with physics-based reasoning remains challenging, both by the complexity of contact dynamics and by the lack of effective and efficient physics inference in the data-driven learning framework. To address both issues, we present DeepSimHO: a novel deep-learning pipeline that combines forward physics simulation and backward gradient approximation with a neural network. Specifically, for an initial hand-object pose estimated by a base network, we forward it to a physics simulator to evaluate its stability. However, due to non-smooth contact geometry and penetration, existing differentiable simulators can not provide reliable state gradient. To remedy this, we further introduce a deep network to learn the stability evaluation process from the simulator, while smoothly approximating its gradient and thus enabling effective back-propagation. Extensive experiments show that our method noticeably improves the stability of the estimation and achieves superior efficiency over test-time optimization. The code is available at https://github.com/rongakowang/DeepSimHO.
****
Towards Data-Algorithm Dependent Generalization: a Case Study on Overparameterized Linear Regression
Jing Xu, Jiaye Teng, Yang Yuan, Andrew Yao

One of the major open problems in machine learning is to characterize generaliza
tion in the overparameterized regime, where most traditional generalization boun
ds become inconsistent even for overparameterized linear regression. In many sce
narios, this failure can be attributed to obscuring the crucial interplay betwee
n the training algorithm and the underlying data distribution. This paper demons
trate that the generalization behavior of overparameterized model should be anal
yzed in a both data-relevant and algorithm-relevant manner. To make a formal cha
racterization, We introduce a notion called data-algorithm compatibility, which
considers the generalization behavior of the entire data-dependent training traj
ectory, instead of traditional last-iterate analysis. We validate our claim by
studying the setting of solving overparameterized linear regression with gradien
t descent. Specifically, we perform a data-dependent trajectory analysis and der
ive a sufficient condition for compatibility in such a setting. Our theoretical
results demonstrate that if we take early stopping iterates into consideration,
generalization can hold with significantly weaker restrictions on the problem in
stance than the previous last-iterate analysis.
****
Global Structure-Aware Diffusion Process for Low-light Image Enhancement
Jinhui HOU, Zhiyu Zhu, Junhui Hou, Hui LIU, Huanqiang Zeng, Hui Yuan
This paper studies a diffusion-based framework to address the low-light image en
hancement problem. To harness the capabilities of diffusion models, we delve int
o this intricate process and advocate for the regularization of its inherent ODE
-trajectory. To be specific, inspired by the recent research that low curvature
ODE-trajectory results in a stable and effective diffusion process, we formulate
 a curvature regularization term anchored in the intrinsic non-local structures
of image data, i.e., global structure-aware regularization, which gradually faci
litates the preservation of complicated details and the augmentation of contrast
 during the diffusion process. This incorporation mitigates the adverse effects
of noise and artifacts resulting from the diffusion process, leading to a more p
recise and flexible enhancement. To additionally promote learning in challenging
 regions, we introduce an uncertainty-guided regularization technique, which wis
ely relaxes constraints on the most extreme regions of the image. Experimental e
valuations reveal that the proposed diffusion-based framework, complemented by r
ank-informed regularization, attains distinguished performance in low-light enha
ncement. The outcomes indicate substantial advancements in image quality, noise
suppression, and contrast amplification in comparison with state-of-the-art meth
ods. We believe this innovative approach will stimulate further exploration and
advancement in low-light image processing, with potential implications for other
 applications of diffusion models. The code is publicly available at https://git
hub.com/jinnh/GSAD.
****
FedGCN: Convergence-Communication Tradeoffs in Federated Training of Graph Convo
lutional Networks
Yuhang Yao, Weizhao Jin, Srivatsan Ravi, Carlee Joe-Wong
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
SegRefiner: Towards Model-Agnostic Segmentation Refinement with Discrete Diffusi
on Process
Mengyu Wang, Henghui Ding, Jun Hao Liew, Jiajun Liu, Yao Zhao, Yunchao Wei
In this paper, we explore a principal way to enhance the quality of object masks
 produced by different segmentation models. We propose a model-agnostic solution
 called SegRefiner, which offers a novel perspective on this problem by interpre
ting segmentation refinement as a data generation process. As a result, the refi
nement process can be smoothly implemented through a series of denoising diffusi
on steps. Specifically, SegRefiner takes coarse masks as inputs and refines them
 using a discrete diffusion process. By predicting the label and corresponding s
tates-transition probabilities for each pixel, SegRefiner progressively refines

the noisy masks in a conditional denoising manner. To assess the effectiveness of SegRefiner, we conduct comprehensive experiments on various segmentation tasks, including semantic segmentation, instance segmentation, and dichotomous image segmentation. The results demonstrate the superiority of our SegRefiner from multiple aspects. Firstly, it consistently improves both the segmentation metrics and boundary metrics across different types of coarse masks. Secondly, it outperforms previous model-agnostic refinement methods by a significant margin. Lastly, it exhibits a strong capability to capture extremely fine details when refining high-resolution images. The source code and trained models are available at SegRefiner.git
****

On Masked Pre-training and the Marginal Likelihood
Pablo Moreno-Muñoz, Pol Garcia Recasens, Søren Hauberg
Masked pre-training removes random input dimensions and learns a model that can predict the missing values. Empirical results indicate that this intuitive form of self-supervised learning yields models that generalize very well to new domains. A theoretical understanding is, however, lacking. This paper shows that masked pre-training with a suitable cumulative scoring function corresponds to maximizing the model's marginal likelihood, which is de facto the Bayesian model selection measure of generalization. Beyond shedding light on the success of masked pre-training, this insight also suggests that Bayesian models can be trained with appropriately designed self-supervision. Empirically, we confirm the developed theory and explore the main learning principles of masked pre-training in large language models.
****

AQuA: A Benchmarking Tool for Label Quality Assessment
Mononito Goswami, Vedant Sanil, Arjun Choudhry, Arvind Srinivasan, Chalisa Udompanyawit, Artur Dubrawski
Machine learning (ML) models are only as good as the data they are trained on. But recent studies have found datasets widely used to train and evaluate ML models, e.g. ImageNet, to have pervasive labeling errors. Erroneous labels on the train set hurt ML models' ability to generalize, and they impact evaluation and model selection using the test set. Consequently, learning in the presence of labeling errors is an active area of research, yet this field lacks a comprehensive benchmark to evaluate these methods. Most of these methods are evaluated on a few computer vision datasets with significant variance in the experimental protocols. With such a large pool of methods and inconsistent evaluation, it is also unclear how ML practitioners can choose the right models to assess label quality in their data. To this end, we propose a benchmarking environment AQuA to rigorously evaluate methods that enable machine learning in the presence of label noise. We also introduce a design space to delineate concrete design choices of label error detection models. We hope that our proposed design space and benchmark enable practitioners to choose the right tools to improve their label quality and that our benchmark enables objective and rigorous evaluation of machine learning tools facing mislabeled data.
****

Smoothed Analysis of Sequential Probability Assignment
Alankrita Bhatt, Nika Haghtalab, Abhishek Shetty
We initiate the study of smoothed analysis for the sequential probability assignment problem with contexts. We study information-theoretically optimal minmax rates as well as a framework for algorithmic reduction involving the maximum likelihood estimator oracle. Our approach establishes a general-purpose reduction from minimax rates for sequential probability assignment for smoothed adversaries to minimax rates for transductive learning. This leads to optimal (logarithmic) fast rates for parametric classes and classes with finite VC dimension. On the algorithmic front, we develop an algorithm that efficiently taps into the MLE oracle, for general classes of functions. We show that under general conditions this algorithmic approach yields sublinear regret.
****

Invariant Learning via Probability of Sufficient and Necessary Causes

Mengyue Yang, Zhen Fang, Yonggang Zhang, Yali Du, Furui Liu, Jean-Francois Ton, Jianhong Wang, Jun Wang

Out-of-distribution (OOD) generalization is indispensable for learning models in the wild, where testing distribution typically unknown and different from the training. Recent methods derived from causality have shown great potential in achieving OOD generalization. However, existing methods mainly focus on the invariance property of causes, while largely overlooking the property of sufficiency and necessity conditions. Namely, a necessary but insufficient cause (feature) is invariant to distribution shift, yet it may not have required accuracy. By contrast, a sufficient yet unnecessary cause (feature) tends to fit specific data well but may have a risk of adapting to a new domain. To capture the information of sufficient and necessary causes, we employ a classical concept, the probability of sufficiency and necessary causes (PNS), which indicates the probability of whether one is the necessary and sufficient cause. To associate PNS with OOD generalization, we propose PNS risk and formulate an algorithm to learn representation with a high PNS value. We theoretically analyze and prove the generalizability of the PNS risk. Experiments on both synthetic and real-world benchmarks demonstrate the effectiveness of the proposed method. The detailed implementation can be found at the GitHub repository: https://github.com/ymy4323460/CaSN.

****

DPOK: Reinforcement Learning for Fine-tuning Text-to-Image Diffusion Models

Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, Kimin Lee

Learning from human feedback has been shown to improve text-to-image models. These techniques first learn a reward function that captures what humans care about in the task and then improve the models based on the learned reward function. Even though relatively simple approaches (e.g., rejection sampling based on reward scores) have been investigated, fine-tuning text-to-image models with the reward function remains challenging. In this work, we propose using online reinforcement learning (RL) to fine-tune text-to-image models. We focus on diffusion models, defining the fine-tuning task as an RL problem, and updating the pre-trained text-to-image diffusion models using policy gradient to maximize the feedback-trained reward. Our approach, coined DPOK, integrates policy optimization with KL regularization. We conduct an analysis of KL regularization for both RL fine-tuning and supervised fine-tuning. In our experiments, we show that DPOK is generally superior to supervised fine-tuning with respect to both image-text alignment and image quality. Our code is available at https://github.com/google-research/google-research/tree/master/dpok.

****

Online POMDP Planning with Anytime Deterministic Guarantees

Moran Barenboim, Vadim Indelman

Autonomous agents operating in real-world scenarios frequently encounter uncertainty and make decisions based on incomplete information. Planning under uncertainty can be mathematically formalized using partially observable Markov decision processes (POMDPs). However, finding an optimal plan for POMDPs can be computationally expensive and is feasible only for small tasks. In recent years, approximate algorithms, such as tree search and sample-based methodologies, have emerged as state-of-the-art POMDP solvers for larger problems. Despite their effectiveness, these algorithms offer only probabilistic and often asymptotic guarantees toward the optimal solution due to their dependence on sampling. To address these limitations, we derive a deterministic relationship between a simplified solution that iseasier to obtain and the theoretically optimal one. First, we derive bounds for selecting a subset of the observations to branch from while computing a complete belief at each posterior node. Then, since a complete belief update may be computationally demanding, we extend the bounds to support reduction of both the state and the observation spaces. We demonstrate how our guarantees can be integrated with existing state-of-the-art solvers that sample a subset of states and observations. As a result, the returned solution holds deterministic bounds relative to the optimal policy. Lastly, we substantiate our findings with supporting experimental results.

****
The Curious Price of Distributional Robustness in Reinforcement Learning with a Generative Model
Laixi Shi, Gen Li, Yuting Wei, Yuxin Chen, Matthieu Geist, Yuejie Chi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Strategic Apple Tasting
Keegan Harris, Chara Podimata, Steven Z. Wu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
GAIA: Delving into Gradient-based Attribution Abnormality for Out-of-distribution Detection
Jinggang Chen, Junjie Li, Xiaoyang Qu, Jianzong Wang, Jiguang Wan, Jing Xiao
Detecting out-of-distribution (OOD) examples is crucial to guarantee the reliability and safety of deep neural networks in real-world settings. In this paper, we offer an innovative perspective on quantifying the disparities between in-distribution (ID) and OOD data---analyzing the uncertainty that arises when models attempt to explain their predictive decisions. This perspective is motivated by our observation that gradient-based attribution methods encounter challenges in assigning feature importance to OOD data, thereby yielding divergent explanation patterns. Consequently, we investigate how attribution gradients lead to uncertain explanation outcomes and introduce two forms of abnormalities for OOD detection: the zero-deflation abnormality and the channel-wise average abnormality. We then propose GAIA, a simple and effective approach that incorporates Gradient Abnormality Inspection and Aggregation. The effectiveness of GAIA is validated on both commonly utilized (CIFAR) and large-scale (ImageNet-1k) benchmarks. Specifically, GAIA reduces the average FPR95 by 23.10% on CIFAR10 and by 45.41% on CIFAR100 compared to advanced post-hoc methods.
****
Context-guided Embedding Adaptation for Effective Topic Modeling in Low-Resource Regimes
Yishi Xu, Jianqiao Sun, Yudi Su, Xinyang Liu, Zhibin Duan, Bo Chen, Mingyuan Zhou
Embedding-based neural topic models have turned out to be a superior option for low-resourced topic modeling. However, current approaches consider static word embeddings learnt from source tasks as general knowledge that can be transferred directly to the target task, discounting the dynamically changing nature of word meanings in different contexts, thus typically leading to sub-optimal results when adapting to new tasks with unfamiliar contexts. To settle this issue, we provide an effective method that centers on adaptively generating semantically tailored word embeddings for each task by fully exploiting contextual information. Specifically, we first condense the contextual syntactic dependencies of words into a semantic graph for each task, which is then modeled by a Variational Graph Auto-Encoder to produce task-specific word representations. On this basis, we further impose a learnable Gaussian mixture prior on the latent space of words to efficiently learn topic representations from a clustering perspective, which contributes to diverse topic discovery and fast adaptation to novel tasks. We have conducted a wealth of quantitative and qualitative experiments, and the results show that our approach comprehensively outperforms established topic models.
****
Discovering General Reinforcement Learning Algorithms with Adversarial Environment Design
Matthew T Jackson, Minqi Jiang, Jack Parker-Holder, Risto Vuorio, Chris Lu, Greg Farquhar, Shimon Whiteson, Jakob Foerster

The past decade has seen vast progress in deep reinforcement learning (RL) on the back of algorithms manually designed by human researchers. Recently, it has been shown that it is possible to meta-learn update rules, with the hope of discovering algorithms that can perform well on a wide range of RL tasks. Despite impressive initial results from algorithms such as Learned Policy Gradient (LPG), there remains a generalization gap when these algorithms are applied to unseen environments. In this work, we examine how characteristics of the meta-training distribution impact the generalization performance of these algorithms. Motivated by this analysis and building on ideas from Unsupervised Environment Design (UED), we propose a novel approach for automatically generating curricula to maximize the regret of a meta-learned optimizer, in addition to a novel approximation of regret, which we name algorithmic regret (AR). The result is our method, General RL Optimizers Obtained Via Environment Design (GROOVE). In a series of experiments, we show that GROOVE achieves superior generalization to LPG, and evaluate AR against baseline metrics from UED, identifying it as a critical component of environment design in this setting. We believe this approach is a step towards the discovery of truly general RL algorithms, capable of solving a wide range of real-world environments.
****
A Riemannian Exponential Augmented Lagrangian Method for Computing the Projection Robust Wasserstein Distance
Bo Jiang, Ya-Feng Liu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Context Shift Reduction for Offline Meta-Reinforcement Learning
Yunkai Gao, Rui Zhang, Jiaming Guo, Fan Wu, Qi Yi, Shaohui Peng, Siming Lan, Ruizhi Chen, Zidong Du, Xing Hu, Qi Guo, Ling Li, Yunji Chen
Offline meta-reinforcement learning (OMRL) utilizes pre-collected offline datasets to enhance the agent's generalization ability on unseen tasks. However, the context shift problem arises due to the distribution discrepancy between the contexts used for training (from the behavior policy) and testing (from the exploration policy). The context shift problem leads to incorrect task inference and further deteriorates the generalization ability of the meta-policy. Existing OMRL methods either overlook this problem or attempt to mitigate it with additional information. In this paper, we propose a novel approach called Context Shift Reduction for OMRL (CSRO) to address the context shift problem with only offline datasets. The key insight of CSRO is to minimize the influence of policy in context during both the meta-training and meta-test phases. During meta-training, we design a max-min mutual information representation learning mechanism to diminish the impact of the behavior policy on task representation. In the meta-test phase, we introduce the non-prior context collection strategy to reduce the effect of the exploration policy. Experimental results demonstrate that CSRO significantly reduces the context shift and improves the generalization ability, surpassing previous methods across various challenging domains.
****
Towards Data-Agnostic Pruning At Initialization: What Makes a Good Sparse Mask?
Hoang Pham, The Anh Ta, Shiwei Liu, Lichuan Xiang, Dung Le, Hongkai Wen, Long Tran-Thanh
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
New Bounds for Hyperparameter Tuning of Regression Problems Across Instances
Maria-Florina F. Balcan, Anh Nguyen, Dravyansh Sharma
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Jailbroken: How Does LLM Safety Training Fail?
Alexander Wei, Nika Haghtalab, Jacob Steinhardt
Large language models trained for safety and harmlessness remain susceptible to adversarial misuse, as evidenced by the prevalence of “jailbreak” attacks on early releases of ChatGPT that elicit undesired behavior. Going beyond recognition of the issue, we investigate why such attacks succeed and how they can be created. We hypothesize two failure modes of safety training: competing objectives and mismatched generalization. Competing objectives arise when a model’s capabilities and safety goals conflict, while mismatched generalization occurs when safety training fails to generalize to a domain for which capabilities exist. We use these failure modes to guide jailbreak design and then evaluate state-of-the-art models, including OpenAI’s GPT-4 and Anthropic’s Claude v1.3, against both existing and newly designed attacks. We find that vulnerabilities persist despite the extensive red-teaming and safety-training efforts behind these models. Notably, new attacks utilizing our failure modes succeed on every prompt in a collection of unsafe requests from the models’ red-teaming evaluation sets and outperform existing ad hoc jailbreaks. Our analysis emphasizes the need for safety-capability parity—that safety mechanisms should be as sophisticated as the underlying model—and argues against the idea that scaling alone can resolve these safety failure modes.
****
Conditional Mutual Information for Disentangled Representations in Reinforcement Learning
Mhairi Dunion, Trevor McInroe, Kevin Luck, Josiah Hanna, Stefano Albrecht
Reinforcement Learning (RL) environments can produce training data with spurious correlations between features due to the amount of training data or its limited feature coverage. This can lead to RL agents encoding these misleading correlations in their latent representation, preventing the agent from generalising if the correlation changes within the environment or when deployed in the real world. Disentangled representations can improve robustness, but existing disentanglement techniques that minimise mutual information between features require independent features, thus they cannot disentangle correlated features. We propose an auxiliary task for RL algorithms that learns a disentangled representation of high-dimensional observations with correlated features by minimising the conditional mutual information between features in the representation. We demonstrate experimentally, using continuous control tasks, that our approach improves generalisation under correlation shifts, as well as improving the training performance of RL algorithms in the presence of correlated features.
****
Comparing Causal Frameworks: Potential Outcomes, Structural Models, Graphs, and Abstractions
Duligur Ibeling, Thomas Icard
The aim of this paper is to make clear and precise the relationship between the Rubin causal model (RCM) and structural causal model (SCM) frameworks for causal inference. Adopting a neutral logical perspective, and drawing on previous work, we show what is required for an RCM to be representable by an SCM. A key result then shows that every RCM---including those that violate algebraic principles implied by the SCM framework---emerges as an abstraction of some representable RCM. Finally, we illustrate the power of this ameliorative perspective by pinpointing an important role for SCM principles in classic applications of RCMs; conversely, we offer a characterization of the algebraic constraints implied by a graph, helping to substantiate further comparisons between the two frameworks.
****
LogSpecT: Feasible Graph Learning Model from Stationary Signals with Recovery Guarantees
Shangyuan LIU, Linglingzhi Zhu, Anthony Man-Cho So
Graph learning from signals is a core task in graph signal processing (GSP). A s

ignificant subclass of graph signals called the stationary graph signals that br
oadens the concept of stationarity of data defined on regular domains to signals
 on graphs is gaining increasing popularity in the GSP community. The most commo
nly used model to learn graphs from these stationary signals is SpecT, which for
ms the foundation for nearly all the subsequent, more advanced models. Despite i
ts strengths, the practical formulation of the model, known as rSpecT, has been
identified to be susceptible to the choice of hyperparameters. More critically,
it may suffer from infeasibility as an optimization problem. In this paper, we i
ntroduce the first condition that ensures the infeasibility of rSpecT and design
 a novel model called LogSpecT, along with its practical formulation rLogSpecT t
o overcome this issue. Contrary to rSpecT, our novel practical model rLogSpecT i
s always feasible. Furthermore, we provide recovery guarantees of rLogSpecT from
 modern optimization tools related to epi-convergence, which could be of indepen
dent interest and significant for various learning problems. To demonstrate the
practical advantages of rLogSpecT, a highly efficient algorithm based on the lin
earized alternating direction method of multipliers (L-ADMM) that allows closed-
form solutions for each subproblem is proposed with convergence guarantees. Exte
nsive numerical results on both synthetic and real networks not only corroborate
 the stability of our proposed methods, but also highlight their comparable and
even superior performance than existing models.
****
Depth-discriminative Metric Learning for Monocular 3D Object Detection
Wonhyeok Choi, Mingyu Shin, Sunghoon Im
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
PLANNER: Generating Diversified Paragraph via Latent Language Diffusion Model
Yizhe Zhang, Jiatao Gu, Zhuofeng Wu, Shuangfei Zhai, Joshua Susskind, Navdeep Ja
itly
Autoregressive models for text sometimes generate repetitive and low-quality out
put because errors accumulate during the steps of generation. This issue is ofte
n attributed to exposure bias -- the difference between how a model is trained,
and how it is used during inference. Denoising diffusion models provide an alter
native approach in which a model can revisit and revise its output. However, the
y can be computationally expensive and prior efforts on text have led to models
that produce less fluent output compared to autoregressive models, especially fo
r longer text and paragraphs. In this paper, we propose PLANNER, a model that co
mbines latent semantic diffusion with autoregressive generation, to generate flu
ent text while exercising global control over paragraphs. The model achieves thi
s by combining an autoregressive "decoding" module with a "planning" module tha
t uses latent diffusion to generate semantic paragraph embeddings in a coarse-to
-fine manner. The proposed method is evaluated on various conditional generation
 tasks, and results on semantic generation, text completion and summarization sh
ow its effectiveness in generating high-quality long-form text in an efficient m
anner.
****
Generalized Bayesian Inference for Scientific Simulators via Amortized Cost Esti
mation
Richard Gao, Michael Deistler, Jakob H Macke
Simulation-based inference (SBI) enables amortized Bayesian inference for simula
tors with implicit likelihoods. But when we are primarily interested in the qual
ity of predictive simulations, or when the model cannot exactly reproduce the ob
served data (i.e., is misspecified), targeting the Bayesian posterior may be ove
rly restrictive. Generalized Bayesian Inference (GBI) aims to robustify inferenc
e for (misspecified) simulator models, replacing the likelihood-function with a
cost function that evaluates the goodness of parameters relative to data. Howeve
r, GBI methods generally require running multiple simulations to estimate the co
st function at each parameter value during inference, making the approach comput

ationally infeasible for even moderately complex simulators. Here, we propose am ortized cost estimation (ACE) for GBI to address this challenge: We train a neur al network to approximate the cost function, which we define as the expected dis tance between simulations produced by a parameter and observed data. The trained network can then be used with MCMC to infer GBI posteriors for any observation without running additional simulations. We show that, on several benchmark tasks , ACE accurately predicts cost and provides predictive simulations that are clos er to synthetic observations than other SBI methods, especially for misspecified simulators. Finally, we apply ACE to infer parameters of the Hodgkin-Huxley mod el given real intracellular recordings from the Allen Cell Types Database. ACE i dentifies better data-matching parameters while being an order of magnitude more simulation-efficient than a standard SBI method. In summary, ACE combines the s trengths of SBI methods and GBI to perform robust and simulation-amortized infer ence for scientific simulators.
****

Harnessing the power of choices in decision tree learning
Guy Blanc, Jane Lange, Chirag Pabbaraju, Colin Sullivan, Li-Yang Tan, Mo Tiwari
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Mitigating Over-smoothing in Transformers via Regularized Nonlocal Functionals
Tam Nguyen, Tan Nguyen, Richard Baraniuk
Transformers have achieved remarkable success in a wide range of natural languag e processing and computer vision applications. However, the representation capac ity of a deep transformer model is degraded due to the over-smoothing issue in w hich the token representations become identical when the model's depth grows. In this work, we show that self-attention layers in transformers minimize a functi onal which promotes smoothness, thereby causing token uniformity. We then propos e a novel regularizer that penalizes the norm of the difference between the smoo th output tokens from self-attention and the input tokens to preserve the fideli ty of the tokens. Minimizing the resulting regularized energy functional, we der ive the Neural Transformer with a Regularized Nonlocal Functional (NeuTRENO), a novel class of transformer models that can mitigate the over-smoothing issue. We empirically demonstrate the advantages of NeuTRENO over the baseline transforme rs and state-of-the-art methods in reducing the over-smoothing of token represen tations on various practical tasks, including object classification, image segme ntation, and language modeling.
****

On the Role of Noise in the Sample Complexity of Learning Recurrent Neural Netwo rks: Exponential Gaps for Long Sequences
Alireza F. Pour, Hassan Ashtiani
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Provably Bounding Neural Network Preimages
Suhas Kotha, Christopher Brix, J. Zico Kolter, Krishnamurthy Dvijotham, Huan Zha ng
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-auth ors prior to requesting a name change in the electronic proceedings.
****

Scaling Riemannian Diffusion Models
Aaron Lou, Minkai Xu, Adam Farris, Stefano Ermon
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues.

Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****

Decision Stacks: Flexible Reinforcement Learning via Modular Generative Models
Siyan Zhao, Aditya Grover
Reinforcement learning presents an attractive paradigm to reason about several distinct aspects of sequential decision making, such as specifying complex goals, planning future observations and actions, and critiquing their utilities. However, the combined integration of these capabilities poses competing algorithmic challenges in retaining maximal expressivity while allowing for flexibility in modeling choices for efficient learning and inference. We present Decision Stacks, a generative framework that decomposes goal-conditioned policy agents into 3 generative modules. These modules simulate the temporal evolution of observations, rewards, and actions via independent generative models that can be learned in parallel via teacher forcing. Our framework guarantees both expressivity and flexibility in designing individual modules to account for key factors such as architectural bias, optimization objective and dynamics, transferrability across domains, and inference speed. Our empirical results demonstrate the effectiveness of Decision Stacks for offline policy optimization for several MDP and POMDP environments, outperforming existing methods and enabling flexible generative decision making.
****

Conformal Prediction for Uncertainty-Aware Planning with Diffusion Dynamics Model
Jiankai Sun, Yiqi Jiang, Jianing Qiu, Parth Nobel, Mykel J Kochenderfer, Mac Schwager
Robotic applications often involve working in environments that are uncertain, dynamic, and partially observable. Recently, diffusion models have been proposed for learning trajectory prediction models trained from expert demonstrations, which can be used for planning in robot tasks. Such models have demonstrated a strong ability to overcome challenges such as multi-modal action distributions, high-dimensional output spaces, and training instability. It is crucial to quantify the uncertainty of these dynamics models when using them for planning. In this paper, we quantify the uncertainty of diffusion dynamics models using Conformal Prediction (CP). Given a finite number of exchangeable expert trajectory examples (called the “calibration set”), we use CP to obtain a set in the trajectory space (called the “coverage region”) that is guaranteed to contain the output of the diffusion model with a user-defined probability (called the “coverage level”). In PlanCP, inspired by concepts from conformal prediction, we modify the loss function for training the diffusion model to include a quantile term to encourage more robust performance across the variety of training examples. At test time, we then calibrate PlanCP with a conformal prediction process to obtain coverage sets for the trajectory prediction with guaranteed coverage level. We evaluate our algorithm on various planning tasks and model-based offline reinforcement learning tasks and show that it reduces the uncertainty of the learned trajectory prediction model. As a by-product, our algorithm PlanCP outperforms prior algorithms on existing offline RL benchmarks and challenging continuous planning tasks. Our method can be combined with most model-based planning approaches to produce uncertainty estimates of the closed-loop system.
****

Max-Sliced Mutual Information
Dor Tsur, Ziv Goldfeld, Kristjan Greenewald
Quantifying dependence between high-dimensional random variables is central to statistical learning and inference. Two classical methods are canonical correlation analysis (CCA), which identifies maximally correlated projected versions of the original variables, and Shannon's mutual information, which is a universal dependence measure that also captures high-order dependencies. However, CCA only accounts for linear dependence, which may be insufficient for certain applications, while mutual information is often infeasible to compute/estimate in high dimensions. This work proposes a middle ground in the form of a scalable information

-theoretic generalization of CCA, termed max-sliced mutual information (mSMI). mSMI equals the maximal mutual information between low-dimensional projections of the high-dimensional variables, which reduces back to CCA in the Gaussian case. It enjoys the best of both worlds: capturing intricate dependencies in the data while being amenable to fast computation and scalable estimation from samples. We show that mSMI retains favorable structural properties of Shannon's mutual information, like variational forms and identification of independence. We then study statistical estimation of mSMI, propose an efficiently computable neural estimator, and couple it with formal non-asymptotic error bounds. We present experiments that demonstrate the utility of mSMI for several tasks, encompassing independence testing, multi-view representation learning, algorithmic fairness, and generative modeling. We observe that mSMI consistently outperforms competing methods with little-to-no computational overhead.
****
Neural Data Transformer 2: Multi-context Pretraining for Neural Spiking Activity
Joel Ye, Jennifer Collinger, Leila Wehbe, Robert Gaunt
The neural population spiking activity recorded by intracortical brain-computer interfaces (iBCIs) contain rich structure. Current models of such spiking activity are largely prepared for individual experimental contexts, restricting data volume to that collectable within a single session and limiting the effectiveness of deep neural networks (DNNs). The purported challenge in aggregating neural spiking data is the pervasiveness of context-dependent shifts in the neural data distributions. However, large scale unsupervised pretraining by nature spans heterogeneous data, and has proven to be a fundamental recipe for successful representation learning across deep learning. We thus develop Neural Data Transformer 2 (NDT2), a spatiotemporal Transformer for neural spiking activity, and demonstrate that pretraining can leverage motor BCI datasets that span sessions, subjects, and experimental tasks. NDT2 enables rapid adaptation to novel contexts in downstream decoding tasks and opens the path to deployment of pretrained DNNs for iBCI control. Code: https://github.com/joel99/contextgeneralbci
****
Data Quality in Imitation Learning
Suneel Belkhale, Yuchen Cui, Dorsa Sadigh
In supervised learning, the question of data quality and curation has been sidelined in recent years in favor of increasingly more powerful and expressive models that can ingest internet-scale data. However, in offline learning for robotics, we simply lack internet scale data, and so high quality datasets are a necessity. This is especially true in imitation learning (IL), a sample efficient paradigm for robot learning using expert demonstrations. Policies learned through IL suffer from state distribution shift at test time due to compounding errors in action prediction, which leads to unseen states that the policy cannot recover from.Instead of designing new algorithms to address distribution shift, an alternative perspective is to develop new ways of assessing and curating datasets. There is growing evidence that the same IL algorithms can have substantially different performance across different datasets. This calls for a formalism for defining metrics of "data quality" that can further be leveraged for data curation.In this work, we take the first step toward formalizing data quality for imitation learning through the lens of distribution shift: a high quality dataset encourages the policy to stay in distribution at test time. We propose two fundamental properties that are necessary for a high quality datasets: i) action divergence: the mismatch between the expert and learned policy at certain states; and ii) transition diversity: the noise present in the system for a given state and action. We investigate the combined effect of these two key properties in imitation learning theoretically, and we empirically analyze models trained on a variety of different data sources. We show that state diversity is not always beneficial, and we demonstrate how action divergence and transition diversity interact in practice.
****
Align Your Prompts: Test-Time Prompting with Distribution Alignment for Zero-Shot Generalization

Jameel Abdul Samadh, Mohammad Hanan Gani, Noor Hussein, Muhammad Uzair Khattak, Muhammad Muzammal Naseer, Fahad Shahbaz Khan, Salman H. Khan
The promising zero-shot generalization of vision-language models such as CLIP has led to their adoption using prompt learning for numerous downstream tasks. Previous works have shown test-time prompt tuning using entropy minimization to adapt text prompts for unseen domains. While effective, this overlooks the key cause for performance degradation to unseen domains -- distribution shift. In this work, we explicitly handle this problem by aligning the out-of-distribution (OOD) test sample statistics to those of the source data using prompt tuning. We use a single test sample to adapt multi-modal prompts at test time by minimizing the feature distribution shift to bridge the gap in the test domain. Evaluating against the domain generalization benchmark, our method improves zero-shot top-1 accuracy beyond existing prompt-learning techniques, with a 3.08% improvement over the baseline MaPLe. In cross-dataset generalization with unseen categories across 10 datasets, our method improves consistently across all datasets compared to the existing state-of-the-art. Our source code and models are available at https://jameelhassan.github.io/promptalign
****
SLM: A Smoothed First-Order Lagrangian Method for Structured Constrained Nonconvex Optimization
Songtao Lu
Functional constrained optimization (FCO) has emerged as a powerful tool for solving various machine learning problems. However, with the rapid increase in applications of neural networks in recent years, it has become apparent that both the objective and constraints often involve nonconvex functions, which poses significant challenges in obtaining high-quality solutions. In this work, we focus on a class of nonconvex FCO problems with nonconvex constraints, where the two optimization variables are nonlinearly coupled in the inequality constraint. Leveraging the primal-dual optimization framework, we propose a smoothed first-order Lagrangian method (SLM) for solving this class of problems. We establish the theoretical convergence guarantees of SLM to the Karush-Kuhn-Tucker (KKT) solutions through quantifying dual error bounds. By establishing connections between this structured FCO and equilibrium-constrained nonconvex problems (also known as bilevel optimization), we apply the proposed SLM to tackle bilevel optimization oriented problems where the lower-level problem is nonconvex. Numerical results obtained from both toy examples and hyper-data cleaning problems demonstrate the superiority of SLM compared to benchmark methods.
****
Auslan-Daily: Australian Sign Language Translation for Daily Communication and News
Xin Shen, Shaozu Yuan, Hongwei Sheng, Heming Du, Xin Yu
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Red Teaming Deep Neural Networks with Feature Synthesis Tools
Stephen Casper, Tong Bu, Yuxiao Li, Jiawei Li, Kevin Zhang, Kaivalya Hariharan, Dylan Hadfield-Menell
Interpretable AI tools are often motivated by the goal of understanding model behavior in out-of-distribution (OOD) contexts. Despite the attention this area of study receives, there are comparatively few cases where these tools have identified previously unknown bugs in models. We argue that this is due, in part, to a common feature of many interpretability methods: they analyze model behavior by using a particular dataset. This only allows for the study of the model in the context of features that the user can sample in advance. To address this, a growing body of research involves interpreting models using feature synthesis methods that do not depend on a dataset. In this paper, we benchmark the usefulness of interpretability tools for model debugging. Our key insight is that we can implant human-interpretable trojans into models and then evaluate these tools based

on whether they can help humans discover them. This is analogous to finding OOD bugs, except the ground truth is known, allowing us to know when a user's interpretation is correct. We make four contributions. (1) We propose trojan discovery as an evaluation task for interpretability tools and introduce a benchmark with 12 trojans of 3 different types. (2) We demonstrate the difficulty of this benchmark with a preliminary evaluation of 16 state-of-the-art feature attribution/saliency tools. Even under ideal conditions, given direct access to data with the trojan trigger, these methods still often fail to identify bugs. (3) We evaluate 7 feature-synthesis methods on our benchmark. (4) We introduce and evaluate 2 new variants of the best-performing method from the previous evaluation.
****

Rehearsal Learning for Avoiding Undesired Future

Tian Qin, Tian-Zuo Wang, Zhi-Hua Zhou

Machine learning (ML) models have been widely used to make predictions. Instead of a predictive statement about future outcomes, in many situations we want to pursue a decision: what can we do to avoid the undesired future if an ML model predicts so? In this paper, we present a rehearsal learning framework, in which decisions that can persuasively avoid the happening of undesired outcomes can be found and recommended. Based on the influence relation, we characterize the generative process of variables with structural rehearsal models, consisting of a probabilistic graphical model called rehearsal graphs and structural equations, and find actionable decisions that can alter the outcome by reasoning under a Bayesian framework. Moreover, we present a probably approximately correct bound to quantify the associated risk of a decision. Experiments validate the effectiveness of the proposed rehearsal learning framework and the informativeness of the bound.
****

Understanding How Consistency Works in Federated Learning via Stage-wise Relaxed Initialization

Yan Sun, Li Shen, Dacheng Tao

Federated learning (FL) is a distributed paradigm that coordinates massive local clients to collaboratively train a global model via stage-wise local training processes on the heterogeneous dataset. Previous works have implicitly studied that FL suffers from the "client-drift" problem, which is caused by the inconsistent optimum across local clients. However, till now it still lacks solid theoretical analysis to explain the impact of this local inconsistency. To alleviate the negative impact of the "client drift" and explore its substance in FL, in this paper, we first design an efficient FL algorithm FedInit, which allows employing the personalized relaxed initialization state at the beginning of each local training stage. Specifically, FedInit initializes the local state by moving away from the current global state towards the reverse direction of the latest local state. This relaxed initialization helps to revise the local divergence and enhance the local consistency level. Moreover, to further understand how inconsistency disrupts performance in FL, we introduce the excess risk analysis and study the divergence term to investigate the test error of the proposed FedInit method. Our studies show that on the non-convex objectives, optimization error is not sensitive to this local inconsistency, while it mainly affects the generalization error bound in FedInit. Extensive experiments are conducted to validate this conclusion. Our proposed FedInit could achieve state-of-the-art (SOTA) results compared to several advanced benchmarks without any additional costs. Meanwhile, stage-wise relaxed initialization could also be incorporated into the current advanced algorithms to achieve higher performance in the FL paradigm.
****

Errors-in-variables Fr\'echet Regression with Low-rank Covariate Approximation

Dogyoon Song, Kyunghee Han

Fr\'echet regression has emerged as a promising approach for regression analysis involving non-Euclidean response variables. However, its practical applicability has been hindered by its reliance on ideal scenarios with abundant and noiseless covariate data. In this paper, we present a novel estimation method that tackles these limitations by leveraging the low-rank structure inherent in the covar

iate matrix. Our proposed framework combines the concepts of global Fr\'echet regression and principal component regression, aiming to improve the efficiency and accuracy of the regression estimator. By incorporating the low-rank structure, our method enables more effective modeling and estimation, particularly in high-dimensional and errors-in-variables regression settings. We provide a theoretical analysis of the proposed estimator's large-sample properties, including a comprehensive rate analysis of bias, variance, and additional variations due to measurement errors. Furthermore, our numerical experiments provide empirical evidence that supports the theoretical findings, demonstrating the superior performance of our approach. Overall, this work introduces a promising framework for regression analysis of non-Euclidean variables, effectively addressing the challenges associated with limited and noisy covariate data, with potential applications in diverse fields.
****
Coupled Reconstruction of Cortical Surfaces by Diffeomorphic Mesh Deformation
Hao Zheng, Hongming Li, Yong Fan
Accurate reconstruction of cortical surfaces from brain magnetic resonance images (MRIs) remains a challenging task due to the notorious partial volume effect in brain MRIs and the cerebral cortex's thin and highly folded patterns. Although many promising deep learning-based cortical surface reconstruction methods have been developed, they typically fail to model the interdependence between inner (white matter) and outer (pial) cortical surfaces, which can help generate cortical surfaces with spherical topology. To robustly reconstruct the cortical surfaces with topological correctness, we develop a new deep learning framework to jointly reconstruct the inner, outer, and their in-between (midthickness) surfaces and estimate cortical thickness directly from 3D MRIs. Our method first estimates the midthickness surface and then learns three diffeomorphic flows jointly to optimize the midthickness surface and deform it inward and outward to the inner and outer cortical surfaces respectively, regularized by topological correctness. Our method also outputs a cortex thickness value for each surface vertex, estimated from its diffeomorphic deformation trajectory. Our method has been evaluated on two large-scale neuroimaging datasets, including ADNI and OASIS, achieving state-of-the-art cortical surface reconstruction performance in terms of accuracy, surface regularity, and computation efficiency.
****
Active representation learning for general task space with applications in robotics
Yifang Chen, Yingbing Huang, Simon S. Du, Kevin G. Jamieson, Guanya Shi
Requests for name changes in the electronic proceedings will be accepted with no questions asked. However name changes may cause bibliographic tracking issues. Authors are asked to consider this carefully and discuss it with their co-authors prior to requesting a name change in the electronic proceedings.
****
Model and Feature Diversity for Bayesian Neural Networks in Mutual Learning
Van Cuong Pham, Cuong Nguyen, Trung Le, Dinh Phung, Gustavo Carneiro, Thanh-Toan Do
Bayesian Neural Networks (BNNs) offer probability distributions for model parameters, enabling uncertainty quantification in predictions. However, they often underperform compared to deterministic neural networks. Utilizing mutual learning can effectively enhance the performance of peer BNNs. In this paper, we propose a novel approach to improve BNNs performance through deep mutual learning. The proposed approaches aim to increase diversity in both network parameter distributions and feature distributions, promoting peer networks to acquire distinct features that capture different characteristics of the input, which enhances the effectiveness of mutual learning. Experimental results demonstrate significant improvements in the classification accuracy, negative log-likelihood, and expected calibration error when compared to traditional mutual learning for BNNs.
****
Fair Graph Distillation
Qizhang Feng, Zhimeng (Stephen) Jiang, Ruiquan Li, Yicheng Wang, Na Zou, Jiang B

ian, Xia Hu
As graph neural networks (GNNs) struggle with large-scale graphs due to high com
putational demands, data distillation for graph data promises to alleviate this
issue by distilling a large real graph into a smaller distilled graph while main
taining comparable prediction performance for GNNs trained on both graphs. Howev
er, we observe that GNNs trained on distilled graphs may exhibit more severe gro
up fairness problems than those trained on real graphs. Motivated by this observ
ation, we propose \textit{fair graph distillation} (\Algnameabbr), an approach f
or generating small distilled \textit{fair and informative} graphs based on the
graph distillation method. The challenge lies in the deficiency of sensitive att
ributes for nodes in the distilled graph, making most debiasing methods (e.g., r
egularization and adversarial debiasing) intractable for distilled graphs. We de
velop a simple yet effective bias metric, called coherence, for distilled graphs
. Based on the proposed coherence metric, we introduce a framework for fair grap
h distillation using a bi-level optimization algorithm. Extensive experiments de
monstrate that the proposed algorithm can achieve better prediction performance-
fairness trade-offs across various datasets and GNN architectures.
****
Optimal testing using combined test statistics across independent studies
Lasse Vuursteen, Botond Szabo, Aad van der Vaart, Harry van Zanten
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
Regret-Optimal Model-Free Reinforcement Learning for Discounted MDPs with Short
Burn-In Time
Xiang Ji, Gen Li
A crucial problem in reinforcement learning is learning the optimal policy. We s
tudy this in tabular infinite-horizon discounted Markov decision processes under
 the online setting. The existing algorithms either fail to achieve regret optim
ality or have to incur a high memory and computational cost. In addition, existi
ng optimal algorithms all require a long burn-in time in order to achieve optima
l sample efficiency, i.e., their optimality is not guaranteed unless sample size
 surpasses a high threshold. We address both open problems by introducing a mode
l-free algorithm that employs variance reduction and a novel technique that swit
ches the execution policy in a slow-yet-adaptive manner. This is the first regre
t-optimal model-free algorithm in the discounted setting, with the additional be
nefit of a low burn-in time.
****
Convolutional State Space Models for Long-Range Spatiotemporal Modeling
Jimmy Smith, Shalini De Mello, Jan Kautz, Scott Linderman, Wonmin Byeon
Requests for name changes in the electronic proceedings will be accepted with no
 questions asked. However name changes may cause bibliographic tracking issues.
  Authors are asked to consider this carefully and discuss it with their co-auth
ors prior to requesting a name change in the electronic proceedings.
****
CRoSS: Diffusion Model Makes Controllable, Robust and Secure Image Steganography
Jiwen Yu, Xuanyu Zhang, Youmin Xu, Jian Zhang
Current image steganography techniques are mainly focused on cover-based methods
, which commonly have the risk of leaking secret images and poor robustness agai
nst degraded container images. Inspired by recent developments in diffusion mode
ls, we discovered that two properties of diffusion models, the ability to achiev
e translation between two images without training, and robustness to noisy data,
 can be used to improve security and natural robustness in image steganography t
asks. For the choice of diffusion model, we selected Stable Diffusion, a type of
 conditional diffusion model, and fully utilized the latest tools from open-sour
ce communities, such as LoRAs and ControlNets, to improve the controllability an
d diversity of container images. In summary, we propose a novel image steganogra
phy framework, named Controllable, Robust and Secure Image Steganography (CRoSS)

, which has significant advantages in controllability, robustness, and security compared to cover-based image steganography methods. These benefits are obtained without additional training. To our knowledge, this is the first work to introduce diffusion models to the field of image steganography. In the experimental section, we conducted detailed experiments to demonstrate the advantages of our proposed CRoSS framework in controllability, robustness, and security.
****

American Stories: A Large-Scale Structured Text Dataset of Historical U.S. Newspapers

Melissa Dell, Jacob Carlson, Tom Bryan, Emily Silcock, Abhishek Arora, Zejiang Shen, Luca D'Amico-Wong, Quan Le, Pablo Querubin, Leander Heldring

Existing full text datasets of U.S. public domain newspapers do not recognize the often complex layouts of newspaper scans, and as a result the digitized content scrambles texts from articles, headlines, captions, advertisements, and other layout regions. OCR quality can also be low. This study develops a novel, deep learning pipeline for extracting full article texts from newspaper images and applies it to the nearly 20 million scans in Library of Congress's public domain Chronicling America collection. The pipeline includes layout detection, legibility classification, custom OCR, and association of article texts spanning multiple bounding boxes. To achieve high scalability, it is built with efficient architectures designed for mobile phones. The resulting American Stories dataset provides high quality data that could be used for pre-training a large language model to achieve better understanding of historical English and historical world knowledge. The dataset could also be added to the external database of a retrieval-augmented language model to make historical information - ranging from interpretations of political events to minutiae about the lives of people's ancestors - more widely accessible. Furthermore, structured article texts facilitate using transformer-based methods for popular social science applications like topic classification, detection of reproduced content, and news story clustering. Finally, American Stories provides a massive silver quality dataset for innovating multimodal layout analysis models and other multimodal applications.
****