Sun 2024.05.05

Sparse multi-view hand-object reconstruction for unseen environments

Yik Lung Pang, Changjae Oh, Andrea Cavallaro

Recent works in hand-object reconstruction mainly focus on the single-view and dense multi-view settings. On the one hand, single-view methods can leverage learned shape priors to generalise to unseen objects but are prone to inaccuracies due to occlusions. On the other hand, dense multi-view methods are very accurate but cannot easily adapt to unseen objects without further data collection. In contrast, sparse multi-view methods can take advantage of the additional views to tackle occlusion, while keeping the computational cost low compared to dense multi-view methods. In this paper, we consider the problem of hand-object reconstruction with unseen objects in the sparse multi-view setting. Given multiple RGB images of the hand and object captured at the same time, our model SVHO combines the predictions from each view into a unified reconstruction without optimisation across views. We train our model on a synthetic hand-object dataset and evaluate directly on a real world recorded hand-object dataset with unseen objects. We show that while reconstruction of unseen hands and objects from RGB is challenging, additional views can help improve the reconstruction quality.

link: http://arxiv.org/abs/2405.01353v1

Improving Subject-Driven Image Synthesis with Subject-Agnostic Guidance

Kelvin C. K. Chan, Yang Zhao, Xuhui Jia, Ming-Hsuan Yang, Huisheng Wang

In subject-driven text-to-image synthesis, the synthesis process tends to be heavily influenced by the reference images provided by users, often overlooking crucial attributes detailed in the text prompt. In this work, we propose Subject-Agnostic Guidance (SAG), a simple yet effective solution to remedy the problem. We show that through constructing a subject-agnostic condition and applying our proposed dual classifier-free guidance, one could obtain outputs consistent with both the given subject and input text prompts. We validate the efficacy of our approach through both optimization-based and encoder-based methods. Additionally, we demonstrate its applicability in second-order customization methods, where an encoder-based model is fine-tuned with DreamBooth. Our approach is conceptually simple and requires only minimal code modifications, but leads to substantial quality improvements, as evidenced by our evaluations and user studies.

link: http://arxiv.org/abs/2405.01356v1

Dynamic Online Ensembles of Basis Expansions

Daniel Waxman, Petar M. Djuri■

Practical Bayesian learning often requires (1) online inference, (2) dynamic models, and (3) ensembling over multiple different models. Recent advances have shown how to use random feature approximations to achieve scalable, online ensembling of Gaussian processes with desirable theoretical properties and fruitful applications. One key to these methods' success is the inclusion of a random walk on the model parameters, which makes models dynamic. We show that these methods can be generalized easily to any basis expansion model and that using alternative basis expansions, such as Hilbert space Gaussian processes, often results in better performance. To simplify the process of choosing a specific basis expansion, our method's generality also allows the ensembling of several entirely different models, for example, a Gaussian process and polynomial regression. Finally, we propose a novel method to ensemble static and dynamic models together.

link: http://arxiv.org/abs/2405.01365v1

Completing the Node-Averaged Complexity Landscape of LCLs on Trees

Alkida Balliu, Sebastian Brandt, Fabian Kuhn, Dennis Olivetti, Gustav Schmid

The node-averaged complexity of a problem captures the number of rounds nodes of a graph have to spend on average to solve the problem in the LOCAL model. A challenging line of research with regards to this new complexity measure is to understand the complexity landscape of locally checkable labelings (LCLs) on families of bounded-degree graphs. Particularly interesting in this context is the family of bounded-degree trees as there, for the worst-case complexity, we know a complete characterization of the possible complexities and structures of LCL problems. A first step for the node-averaged complexity case has been achieved recently [DISC '23], where the authors in particular showed that in bounded-degree trees, there is a large complexity gap: There are no LCL problems with a deterministic node-averaged complexity between $\omega(\log^* n)$ and $n^{o(1)}$. For randomized algorithms, they even showed that the node-averaged complexity is either $O(1)$ or $n^{\Omega(1)}$. In this work we fill in the remaining gaps and give a complete description of the node-averaged complexity landscape of LCLs on bounded-degree trees. Our contributions are threefold. - On bounded-degree trees, there is no LCL with a node-averaged complexity between $\omega(1)$ and $(\log^*n)^{o(1)}$. - For any constants 00, there exists a constant c and an LCL problem with node-averaged complexity between $\Omega((\log^* n)^c)$ and $O((\log^* n)^{c+\varepsilon})$. - For any constants $0<\alpha\leq 1/2$ and $\varepsilon>0$, there exists an LCL problem with node-averaged complexity $\Theta(n^x)$ for some $x\in [\alpha, \alpha+\varepsilon]$.

link: http://arxiv.org/abs/2405.01366v1

ATOM: Attention Mixer for Efficient Dataset Distillation

Samir Khaki, Ahmad Sajedi, Kai Wang, Lucy Z. Liu, Yuri A. Lawryshyn, Konstantinos N. Plataniotis

Recent works in dataset distillation seek to minimize training expenses by generating a condensed synthetic dataset that encapsulates the information present in a larger real dataset. These approaches ultimately aim to attain test accuracy levels akin to those achieved by models trained on the entirety of the original dataset. Previous studies in feature and distribution matching have achieved significant results without incurring the costs of bi-level optimization in the distillation process. Despite their convincing efficiency, many of these methods suffer from marginal downstream performance improvements, limited distillation of contextual information, and subpar cross-architecture generalization. To address these challenges in dataset distillation, we propose the ATtentiOn Mixer (ATOM) module to efficiently distill large datasets using a mixture of channel and spatial-wise attention in the feature matching process. Spatial-wise attention helps guide the learning process based on consistent localization of classes in their respective images, allowing for distillation from a broader receptive field. Meanwhile, channel-wise attention captures the contextual information associated with the class itself, thus making the synthetic image more informative for training. By integrating both types of attention, our ATOM module demonstrates superior performance across various computer vision datasets, including CIFAR10/100 and TinyImagenet. Notably, our method significantly improves performance in scenarios with a low number of images per class, thereby enhancing its potential. Furthermore, we maintain the improvement in cross-architectures and applications such as neural architecture search.

link: http://arxiv.org/abs/2405.01373v1

Topics in the Study of the Pragmatic Functions of Phonetic Reduction in Dialog

Nigel G. Ward, Carlos A. Ortega

Reduced articulatory precision is common in speech, but for dialog its acoustic properties and pragmatic functions have been little studied. We here try to remedy this gap. This technical report contains content that was omitted from the journal article (Ward et al. 2024, submitted). Specifically, we here report 1) lessons learned about annotating for perceived reduction, 2) the finding that, unlike in read speech, the correlates of reduction in dialog include high pitch, wide pitch range, and intensity, and 3) a baseline model for predicting reduction in dialog, using simple acoustic/prosodic features, that achieves correlations with human perceptions of 0.24 for English, and 0.17 for Spanish. We also provide examples of additional possible pragmatic functions of reduction in English, and various discussion, observations and speculations

link: http://arxiv.org/abs/2405.01376v1

Verification and Refinement of Natural Language Explanations through LLM-Symbolic Theorem Proving

Xin Quan, Marco Valentino, Louise A. Dennis, André Freitas

Natural language explanations have become a proxy for evaluating explainable and multi-step Natural Language Inference (NLI) models. However, assessing the validity of explanations for NLI is challenging as it typically involves the crowd-sourcing of apposite datasets, a process that is time-consuming and prone to logical errors. To address existing limitations, this paper investigates the verification and refinement of natural language explanations through the integration of Large Language Models (LLMs) and Theorem Provers (TPs). Specifically, we present a neuro-symbolic framework, named Explanation-Refiner, that augments a TP with LLMs to generate and formalise explanatory sentences and suggest potential inference strategies for NLI. In turn, the TP is employed to provide formal guarantees on the logical validity of the explanations and to generate feedback for subsequent improvements. We demonstrate how Explanation-Refiner can be jointly used to evaluate explanatory reasoning, autoformalisation, and error correction mechanisms of state-of-the-art LLMs as well as to automatically enhance the quality of human-annotated explanations of variable complexity in different domains.

link: http://arxiv.org/abs/2405.01379v1

Invariant Risk Minimization Is A Total Variation Model

Zhao-Rong Lai, Wei-Wen Wang

Invariant risk minimization (IRM) is an arising approach to generalize invariant features to different environments in machine learning. While most related works focus on new IRM settings or new application scenarios, the mathematical essence of IRM remains to be properly explained. We verify that IRM is essentially a total variation based on L^2 norm (TV-ℓ_2) of the learning risk with respect to the classifier variable. Moreover, we propose a novel IRM framework based on the TV-ℓ_1 model. It not only expands the classes of functions that can be used as the learning risk, but also has robust performance in denoising and invariant feature preservation based on the coarea formula. We also illustrate some requirements for IRM-TV-ℓ_1 to achieve out-of-distribution generalization. Experimental results show that the proposed framework achieves competitive performance in several benchmark machine learning scenarios.

link: http://arxiv.org/abs/2405.01389v1

Learning Force Control for Legged Manipulation

Tifanny Portela, Gabriel B. Margolis, Yandong Ji, Pulkit Agrawal

Controlling contact forces during interactions is critical for locomotion and manipulation tasks. While sim-to-real reinforcement learning (RL) has succeeded in many contact-rich problems, current RL methods achieve forceful interactions implicitly without explicitly regulating forces. We propose a method for training RL policies for direct force control without requiring access to force sensing. We showcase our method on a whole-body control platform of a quadruped robot with an arm. Such force control enables us to perform gravity compensation and impedance control, unlocking compliant whole-body manipulation. The learned whole-body controller with variable compliance makes it intuitive for humans to teleoperate the robot by only commanding the manipulator, and the robot's body adjusts automatically to achieve the desired position and force. Consequently, a human teleoperator can easily demonstrate a wide variety of loco-manipulation tasks. To the best of our knowledge, we provide the first deployment of learned whole-body force control in legged manipulators, paving the way for more versatile and adaptable legged robots.

link: http://arxiv.org/abs/2405.01402v1

Unsupervised Flow Discovery from Task-oriented Dialogues

Patrícia Ferreira, Daniel Martins, Ana Alves, Catarina Silva, Hugo Gonçalo Oliveira

The design of dialogue flows is a critical but time-consuming task when developing task-oriented dialogue (TOD) systems. We propose an approach for the unsupervised discovery of flows from dialogue history, thus making the process applicable to any domain for which such an history is available. Briefly, utterances are represented in a vector space and clustered according to their semantic similarity. Clusters, which can be seen as dialogue states, are then used as the vertices of a transition graph for representing the flows visually. We present concrete examples of flows, discovered from MultiWOZ, a public TOD dataset. We further elaborate on their significance and relevance for the underlying conversations and introduce an automatic validation metric for their assessment. Experimental results demonstrate the potential of the proposed approach for extracting meaningful flows from task-oriented conversations.

link: http://arxiv.org/abs/2405.01403v1

Random Pareto front surfaces

Ben Tu, Nikolas Kantas, Robert M. Lee, Behrang Shafei

The Pareto front of a set of vectors is the subset which is comprised solely of all of the best trade-off points. By interpolating this subset, we obtain the optimal trade-off surface. In this work, we prove a very useful result which states that all Pareto front surfaces can be explicitly parametrised using polar coordinates. In particular, our polar parametrisation result tells us that we can fully characterise any Pareto front surface using the length function, which is a scalar-valued function that returns the projected length along any positive radial direction. Consequently, by exploiting this representation, we show how it is possible to generalise many useful concepts from linear algebra, probability and statistics, and decision theory to function over the space of Pareto front surfaces. Notably, we focus our attention on the stochastic setting where the Pareto front surface itself is a stochastic process. Among other things, we showcase how it is possible to define and estimate many statistical quantities of interest such as the expectation, covariance and quantile of any Pareto front surface distribution. As a motivating example, we investigate how these statistics can be used within a design of experiments setting, where the goal is to both infer and use the Pareto front surface distribution in order to make effective decisions. Besides this, we also illustrate how these Pareto front ideas can be used within the context of extreme value theory. Finally, as a numerical example, we applied some of our new methodology on a real-world air pollution data set.

link: http://arxiv.org/abs/2405.01404v1

Goal-conditioned reinforcement learning for ultrasound navigation guidance

Abdoul Aziz Amadou, Vivek Singh, Florin C. Ghesu, Young-Ho Kim, Laura Stanciulescu, Harshitha P. Sai, Puneet Sharma, Alistair Young, Ronak Rajani, Kawal Rhode

Transesophageal echocardiography (TEE) plays a pivotal role in cardiology for diagnostic and interventional procedures. However, using it effectively requires extensive training due to the intricate nature of image acquisition and interpretation. To enhance the efficiency of novice sonographers and reduce variability in scan acquisitions, we propose a novel ultrasound (US) navigation assistance method based on contrastive learning as goal-conditioned reinforcement learning (GCRL). We augment the previous framework using a novel contrastive patient batching method (CPB) and a data-augmented contrastive loss, both of which we demonstrate are essential to ensure generalization to anatomical variations across patients. The proposed framework enables navigation to both standard diagnostic as well as intricate interventional views with a single model. Our method was developed with a large dataset of 789 patients and obtained an average error of 6.56 mm in position and 9.36 degrees in angle on a testing dataset of 140 patients, which is competitive or superior to models trained on individual views. Furthermore, we quantitatively validate our method's ability to navigate to interventional views such as the Left Atrial Appendage (LAA) view used in LAA closure. Our approach holds promise in providing valuable guidance during transesophageal ultrasound examinations, contributing to the advancement of skill acquisition for cardiac ultrasound practitioners.

link: http://arxiv.org/abs/2405.01409v1

Applying Transparent Shaping for Zero Trust Architecture Implementation in AWS: A Case Study

Wenjia Wang, Seyed Masoud Sadjadi, Naphtali Rishe

This study introduces a methodology integrating Zero Trust Architecture (ZTA) principles and Transparent Shaping into an AWS-hosted Online File Manager (OFM) application, enhancing security without substantial code modifications. We evaluate our approach with the Mozilla Observatory, highlighting significant security improvements and outlining a promising direction for applying Transparent Shaping and ZTA in cloud environments.

link: http://arxiv.org/abs/2405.01412v1

MiniGPT-3D: Efficiently Aligning 3D Point Clouds with Large Language Models using 2D Priors

Yuan Tang, Xu Han, Xianzhi Li, Qiao Yu, Yixue Hao, Long Hu, Min Chen

Large 2D vision-language models (2D-LLMs) have gained significant attention by bridging Large Language Models (LLMs) with images using a simple projector. Inspired by their success, large 3D point cloud-language models (3D-LLMs) also integrate point clouds into LLMs. However, directly aligning point clouds with LLM requires expensive training costs, typically in hundreds of GPU-hours on A100, which hinders the development of 3D-LLMs. In this paper, we introduce MiniGPT-3D, an efficient and powerful 3D-LLM that achieves multiple SOTA results while training for only 27 hours on one RTX 3090. Specifically, we propose to align 3D point clouds with LLMs using 2D priors from 2D-LLMs, which can leverage the similarity between 2D and 3D visual information. We introduce a novel four-stage training strategy for modality alignment in a cascaded way, and a mixture of query experts module to adaptively aggregate features with high efficiency. Moreover, we utilize parameter-efficient fine-tuning methods LoRA and Norm fine-tuning, resulting in only 47.8M learnable parameters, which is up to 260x fewer than existing methods. Extensive experiments show that MiniGPT-3D achieves SOTA on 3D object classification and captioning tasks, with significantly cheaper training costs. Notably, MiniGPT-3D gains an 8.12 increase on GPT-4 evaluation score for the challenging object captioning task compared to ShapeLLM-13B, while the latter costs 160 total GPU-hours on 8 A800. We are the first to explore the efficient 3D-LLM, offering new insights to the community. Code and weights are available at https://github.com/TangYuan96/MiniGPT-3D.

link: http://arxiv.org/abs/2405.01413v1

GTX: A Write-Optimized Latch-free Graph Data System with Transactional Support

Libin Zhou, Yeasir Rayhan, Lu Xing, Walid. G. Aref

This paper introduces GTX a standalone main-memory write-optimized graph system that specializes in structural and graph property updates while maintaining concurrent reads and graph analytics with snapshot isolation-level transactional concurrency. Recent graph libraries target efficient concurrent read and write support while guaranteeing transactional consistency. However, their performance suffers for updates with strong temporal locality over the same vertexes and edges due to vertex-centric lock contentions. GTX introduces a new delta-chain-centric concurrency-control protocol that eliminates traditional mutually exclusive latches. GTX resolves the conflicts caused by vertex-level locking, and adapts to real-life workloads while maintaining sequential access to the graph's adjacency lists storage. This combination of features has been demonstrated to provide good performance in graph analytical queries. GTX's transactions support fast group commit, novel write-write conflict prevention, and lazy garbage collection. Based on extensive experimental and comparative studies, in addition to maintaining competitive concurrent read and analytical performance, GTX demonstrates high throughput over state-of-the-art techniques when handling concurrent transaction+analytics workloads. For write-heavy transactional workloads, GTX performs up to 11x better than the best-performing state-of-the-art systems in transaction throughput. At the same time, GTX does not sacrifice the performance of read-heavy analytical workloads, and has competitive performance similar to state-of-the-art systems.

link: http://arxiv.org/abs/2405.01418v1