Mon 2024.03.18

Open Stamped Parts Dataset

Sara Antiles, Sachin S. Talathi

We present the Open Stamped Parts Dataset (OSPD), featuring synthetic and real images of stamped metal sheets for auto manufacturing. The real part images, captured from 7 cameras, consist of 7,980 unlabeled images and 1,680 labeled images. In addition, we have compiled a defect dataset by overlaying synthetically generated masks on 10% of the holes. The synthetic dataset replicates the real manufacturing environment in terms of lighting and part placement relative to the cameras. The synthetic data includes 7,980 training images, 1,680 validation images and 1,680 test images, each with bounding box and segmentation mask annotations around all holes. 10% of the holes in the synthetic data mimic defects generated in the real image dataset. We trained a hole-detection model on the synthetic-OSPD, achieving a modified recall score of 67.2% and a precision of 94.4% . We anticipate researchers in the auto manufacturing and broader machine learning and computer vision communities using OSPD to advance the state of the art in defect detection of stamped holes in the metalsheet stamping process. The dataset is available for download at: https://tinyurl.com/hm6xatd7

link: http://arxiv.org/abs/2403.10369v1

An Energy-Efficient Ensemble Approach for Mitigating Data Incompleteness in IoT Applications

Yousef AlShehri, Lakshmish Ramaswamy

Machine Learning (ML) is becoming increasingly important for IoT-based applications. However, the dynamic and ad-hoc nature of many IoT ecosystems poses unique challenges to the efficacy of ML algorithms. One such challenge is data incompleteness, which is manifested as missing sensor readings. Many factors, including sensor failures and/or network disruption, can cause data incompleteness. Furthermore, most IoT systems are severely power-constrained. It is important that we build IoT-based ML systems that are robust against data incompleteness while simultaneously being energy efficient. This paper presents an empirical study of SECOE - a recent technique for alleviating data incompleteness in IoT - with respect to its energy bottlenecks. Towards addressing the energy bottlenecks of SECOE, we propose ENAMLE - a proactive, energy-aware technique for mitigating the impact of concurrent missing data. ENAMLE is unique in the sense that it builds an energy-aware ensemble of sub-models, each trained with a subset of sensors chosen carefully based on their correlations. Furthermore, at inference time, ENAMLE adaptively alters the number of the ensemble of models based on the amount of missing data rate and the energy-accuracy trade-off. ENAMLE's design includes several novel mechanisms for minimizing energy consumption while maintaining accuracy. We present extensive experimental studies on two distinct datasets that demonstrate the energy efficiency of ENAMLE and its ability to alleviate sensor failures.

link: http://arxiv.org/abs/2403.10371v1

Towards a general framework for improving the performance of classifiers using XAI methods

Andrea Apicella, Salvatore Giugliano, Francesco Isgrò, Roberto Prevete

Modern Artificial Intelligence (AI) systems, especially Deep Learning (DL) models, poses challenges in understanding their inner workings by AI researchers. eXplainable Artificial Intelligence (XAI) inspects internal mechanisms of AI models providing explanations about their decisions. While current XAI research predominantly concentrates on explaining AI systems, there is a growing interest in using XAI techniques to automatically improve the performance of AI systems themselves. This paper proposes a general framework for automatically improving the performance of pre-trained DL classifiers using XAI methods, avoiding the computational overhead associated with retraining complex models from scratch. In particular, we outline the possibility of

two different learning strategies for implementing this architecture, which we will call auto-encoder-based and encoder-decoder-based, and discuss their key aspects.

link: http://arxiv.org/abs/2403.10373v1

Overcoming Distribution Shifts in Plug-and-Play Methods with Test-Time Training

Edward P. Chandler, Shirin Shoushtari, Jiaming Liu, M. Salman Asif, Ulugbek S. Kamilov

Plug-and-Play Priors (PnP) is a well-known class of methods for solving inverse problems in computational imaging. PnP methods combine physical forward models with learned prior models specified as image denoisers. A common issue with the learned models is that of a performance drop when there is a distribution shift between the training and testing data. Test-time training (TTT) was recently proposed as a general strategy for improving the performance of learned models when training and testing data come from different distributions. In this paper, we propose PnP-TTT as a new method for overcoming distribution shifts in PnP. PnP-TTT uses deep equilibrium learning (DEQ) for optimizing a self-supervised loss at the fixed points of PnP iterations. PnP-TTT can be directly applied on a single test sample to improve the generalization of PnP. We show through simulations that given a sufficient number of measurements, PnP-TTT enables the use of image priors trained on natural images for image reconstruction in magnetic resonance imaging (MRI).

link: http://dx.doi.org/10.1109/CAMSAP58249.2023.10403502

PASTA: Towards Flexible and Efficient HDR Imaging Via Progressively Aggregated Spatio-Temporal Aligment

Xiaoning Liu, Ao Li, Zongwei Wu, Yapeng Du, Le Zhang, Yulun Zhang, Radu Timofte, Ce Zhu

Leveraging Transformer attention has led to great advancements in HDR deghosting. However, the intricate nature of self-attention introduces practical challenges, as existing state-of-the-art methods often demand high-end GPUs or exhibit slow inference speeds, especially for high-resolution images like 2K. Striking an optimal balance between performance and latency remains a critical concern. In response, this work presents PASTA, a novel Progressively Aggregated Spatio-Temporal Alignment framework for HDR deghosting. Our approach achieves effectiveness and efficiency by harnessing hierarchical representation during feature distanglement. Through the utilization of diverse granularities within the hierarchical structure, our method substantially boosts computational speed and optimizes the HDR imaging workflow. In addition, we explore within-scale feature modeling with local and global attention, gradually merging and refining them in a coarse-to-fine fashion. Experimental results showcase PASTA's superiority over current SOTA methods in both visual quality and performance metrics, accompanied by a substantial 3-fold (x3) increase in inference speed.

link: http://arxiv.org/abs/2403.10376v1

EXAMS-V: A Multi-Discipline Multilingual Multimodal Exam Benchmark for Evaluating Vision Language Models

Rocktim Jyoti Das, Simeon Emilov Hristov, Haonan Li, Dimitar Iliyanov Dimitrov, Ivan Koychev, Preslav Nakov

We introduce EXAMS-V, a new challenging multi-discipline multimodal multilingual exam benchmark for evaluating vision language models. It consists of 20,932 multiple-choice questions across 20 school disciplines covering natural science, social science, and other miscellaneous studies, e.g., religion, fine arts, business, etc. EXAMS-V includes a variety of multimodal features such as text, images, tables, figures, diagrams, maps, scientific symbols, and equations. The questions come in 11 languages from 7 language families. Unlike existing benchmarks, EXAMS-V is uniquely curated by gathering school exam questions from various countries, with a variety of education systems. This distinctive approach calls for intricate reasoning across diverse languages and relies on region-specific knowledge. Solving the problems in the dataset requires advanced perception and joint reasoning over the text and the visual content of the image. Our evaluation results demonstrate that this is a challenging dataset, which is difficult even for advanced vision-text models such as GPT-4V and Gemini; this underscores the inherent complexity of the dataset and

its significance as a future benchmark.

link: http://arxiv.org/abs/2403.10378v1

Regret Minimization via Saddle Point Optimization

Johannes Kirschner, Seyed Alireza Bakhtiari, Kushagra Chandak, Volodymyr Tkachuk, Csaba Szepesvári

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 (E2D) algorithm. 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.

link: http://arxiv.org/abs/2403.10379v1

BirdSet: A Multi-Task Benchmark for Classification in Avian Bioacoustics

Lukas Rauch, Raphael Schwinger, Moritz Wirth, René Heinrich, Jonas Lange, Stefan Kahl, Bernhard Sick, Sven Tomforde, Christoph Scholz

Deep learning (DL) models have emerged as a powerful tool in avian bioacoustics to diagnose environmental health and biodiversity. However, inconsistencies in research pose notable challenges hindering progress in this domain. Reliable DL models need to analyze bird calls flexibly across various species and environments to fully harness the potential of bioacoustics in a cost-effective passive acoustic monitoring scenario. Data fragmentation and opacity across studies complicate a comprehensive evaluation of general model performance. To overcome these challenges, we present the BirdSet benchmark, a unified framework consolidating research efforts with a holistic approach for classifying bird vocalizations in avian bioacoustics. BirdSet harmonizes open-source bird recordings into a curated dataset collection. This unified approach provides an in-depth understanding of model performance and identifies potential shortcomings across different tasks. By establishing baseline results of current models, BirdSet aims to facilitate comparability, guide subsequent data collection, and increase accessibility for newcomers to avian bioacoustics.

link: http://arxiv.org/abs/2403.10380v1

Monotonic Representation of Numeric Properties in Language Models

Benjamin Heinzerling, Kentaro Inui

Language models (LMs) can express factual knowledge involving numeric properties such as Karl Popper was born in 1902. However, how this information is encoded in the model's internal representations is not understood well. Here, we introduce a simple method for finding and editing representations of numeric properties such as an entity's birth year. Empirically, we find low-dimensional subspaces that encode numeric properties monotonically, in an interpretable and editable fashion. When editing representations along directions in these subspaces, LM output changes accordingly. For example, by patching activations along a "birthyear" direction we can make the LM express an increasingly late birthyear: Karl Popper was born in 1929, Karl Popper was born in 1957, Karl Popper was born in 1968. Property-encoding directions exist across several numeric properties in all models under consideration, suggesting the possibility that monotonic representation of numeric properties consistently emerges during LM pretraining. Code: https://github.com/bheinzerling/numeric-property-repr

link: http://arxiv.org/abs/2403.10381v1

Coordination in Noncooperative Multiplayer Matrix Games via Reduced Rank Correlated Equilibria

Jaehan Im, Yue Yu, David Fridovich-Keil, Ufuk Topcu

Coordination in multiplayer games enables players to avoid the lose-lose outcome that often arises at Nash equilibria. However, designing a coordination mechanism typically requires the consideration of the joint actions of all players, which becomes intractable in large-scale games. We develop a novel coordination mechanism, termed reduced rank correlated equilibria, which reduces the number of joint actions to be considered and thereby mitigates computational complexity. The idea is to approximate the set of all joint actions with the actions used in a set of pre-computed Nash equilibria via a convex hull operation. In a game with n players and each player having m actions, the proposed mechanism reduces the number of joint actions considered from O(m^n) to O(mn). We demonstrate the application of the proposed mechanism to an air traffic queue management problem. Compared with the correlated equilibrium-a popular benchmark coordination mechanism-the proposed approach is capable of solving a queue management problem involving four thousand times more joint actions. In the meantime, it yields a solution that shows a 58.5% to 99.5% improvement in the fairness indicator and a 1.8% to 50.4% reduction in average delay cost compared to the Nash solution, which does not involve coordination.

link: http://arxiv.org/abs/2403.10384v1

Evaluating Perceptual Distances by Fitting Binomial Distributions to Two-Alternative Forced Choice Data

Alexander Hepburn, Raul Santos-Rodriguez, Javier Portilla

The two-alternative forced choice (2AFC) experimental setup is popular in the visual perception literature, where practitioners aim to understand how human observers perceive distances within triplets that consist of a reference image and two distorted versions of that image. In the past, this had been conducted in controlled environments, with a tournament-style algorithm dictating which images are shown to each participant to rank the distorted images. Recently, crowd-sourced perceptual datasets have emerged, with no images shared between triplets, making ranking impossible. Evaluating perceptual distances using this data is non-trivial, relying on reducing the collection of judgements on a triplet to a binary decision -- which is suboptimal and prone to misleading conclusions. Instead, we statistically model the underlying decision-making process during 2AFC experiments using a binomial distribution. We use maximum likelihood estimation to fit a distribution to the perceptual judgements, conditioned on the perceptual distance to test and impose consistency and smoothness between our empirical estimates of the density. This way, we can evaluate a different number of judgements per triplet, and can calculate metrics such as likelihoods of judgements according to a set of distances -- key ingredients that neural network counterparts lack.

link: http://arxiv.org/abs/2403.10390v1

CDMAD: Class-Distribution-Mismatch-Aware Debiasing for Class-Imbalanced Semi-Supervised Learning

Hyuck Lee, Heeyoung Kim

Pseudo-label-based semi-supervised learning (SSL) algorithms trained on a class-imbalanced set face two cascading challenges: 1) Classifiers tend to be biased towards majority classes, and 2) Biased pseudo-labels are used for training. It is difficult to appropriately re-balance the classifiers in SSL because the class distribution of an unlabeled set is often unknown and could be mismatched with that of a labeled set. We propose a novel class-imbalanced SSL algorithm called class-distribution-mismatch-aware debiasing (CDMAD). For each iteration of training, CDMAD first assesses the classifier's biased degree towards each class by calculating the logits on an image without any patterns (e.g., solid color image), which can be considered irrelevant to the training set. CDMAD then refines biased pseudo-labels of the base SSL algorithm by ensuring the classifier's neutrality. CDMAD uses these refined pseudo-labels during the training of the base SSL algorithm

to improve the quality of the representations. In the test phase, CDMAD similarly refines biased class predictions on test samples. CDMAD can be seen as an extension of post-hoc logit adjustment to address a challenge of incorporating the unknown class distribution of the unlabeled set for re-balancing the biased classifier under class distribution mismatch. CDMAD ensures Fisher consistency for the balanced error. Extensive experiments verify the effectiveness of CDMAD.

link: http://arxiv.org/abs/2403.10391v1

Isotropic3D: Image-to-3D Generation Based on a Single CLIP Embedding

Pengkun Liu, Yikai Wang, Fuchun Sun, Jiafang Li, Hang Xiao, Hongxiang Xue, Xinzhou Wang

Encouraged by the growing availability of pre-trained 2D diffusion models, image-to-3D generation by leveraging Score Distillation Sampling (SDS) is making remarkable progress. Most existing methods combine novel-view lifting from 2D diffusion models which usually take the reference image as a condition while applying hard L2 image supervision at the reference view. Yet heavily adhering to the image is prone to corrupting the inductive knowledge of the 2D diffusion model leading to flat or distorted 3D generation frequently. In this work, we reexamine image-to-3D in a novel perspective and present Isotropic3D, an image-to-3D generation pipeline that takes only an image CLIP embedding as input. Isotropic3D allows the optimization to be isotropic w.r.t. the azimuth angle by solely resting on the SDS loss. The core of our framework lies in a two-stage diffusion model fine-tuning. Firstly, we fine-tune a text-to-3D diffusion model by substituting its text encoder with an image encoder, by which the model preliminarily acquires image-to-image capabilities. Secondly, we perform fine-tuning using our Explicit Multi-view Attention (EMA) which combines noisy multi-view images with the noise-free reference image as an explicit condition. CLIP embedding is sent to the diffusion model throughout the whole process while reference images are discarded once after fine-tuning. As a result, with a single image CLIP embedding, Isotropic3D is capable of generating multi-view mutually consistent images and also a 3D model with more symmetrical and neat content, well-proportioned geometry, rich colored texture, and less distortion compared with existing image-to-3D methods while still preserving the similarity to the reference image to a large extent. The project page is available at https://isotropic3d.github.io/. The code and models are available at https://github.com/pkunliu/Isotropic3D.

link: http://arxiv.org/abs/2403.10395v1

SculptDiff: Learning Robotic Clay Sculpting from Humans with Goal Conditioned Diffusion Policy

Alison Bartsch, Arvind Car, Charlotte Avra, Amir Barati Farimani

Manipulating deformable objects remains a challenge within robotics due to the difficulties of state estimation, long-horizon planning, and predicting how the object will deform given an interaction. These challenges are the most pronounced with 3D deformable objects. We propose SculptDiff, a goal-conditioned diffusion-based imitation learning framework that works with point cloud state observations to directly learn clay sculpting policies for a variety of target shapes. To the best of our knowledge this is the first real-world method that successfully learns manipulation policies for 3D deformable objects. For sculpting videos and access to our dataset and hardware CAD models, see the project website: https://sites.google.com/andrew.cmu.edu/imitation-sculpting/home

link: http://arxiv.org/abs/2403.10401v1

Energy Correction Model in the Feature Space for Out-of-Distribution Detection

Marc Lafon, Clément Rambour, Nicolas Thome

In this work, we study the out-of-distribution (OOD) detection problem through the use of the feature space of a pre-trained deep classifier. We show that learning the density of in-distribution (ID) features with an energy-based models (EBM) leads to competitive detection results. However, we found that the non-mixing of MCMC sampling during the EBM's training undermines its detection performance. To overcome this an energy-based correction of a mixture of class-conditional Gaussian distributions. We obtains favorable results when compared to a strong baseline like the KNN detector on the CIFAR-10/CIFAR-100 OOD detection benchmarks.

link: http://arxiv.org/abs/2403.10403v1