Tue 2024.04.09

HawkDrive: A Transformer-driven Visual Perception System for Autonomous Driving in Night Scene

Ziang Guo, Stepan Perminov, Mikhail Konenkov, Dzmitry Tsetserukou

Many established vision perception systems for autonomous driving scenarios ignore the influence of light conditions, one of the key elements for driving safety. To address this problem, we present HawkDrive, a novel perception system with hardware and software solutions. Hardware that utilizes stereo vision perception, which has been demonstrated to be a more reliable way of estimating depth information than monocular vision, is partnered with the edge computing device Nvidia Jetson Xavier AGX. Our software for low light enhancement, depth estimation, and semantic segmentation tasks, is a transformer-based neural network. Our software stack, which enables fast inference and noise reduction, is packaged into system modules in Robot Operating System 2 (ROS2). Our experimental results have shown that the proposed end-to-end system is effective in improving the depth estimation and semantic segmentation performance. Our dataset and codes will be released at https://github.com/ZionGo6/HawkDrive.

link: http://arxiv.org/abs/2404.04653v1

Music Recommendation Based on Facial Emotion Recognition

Rajesh B, Keerthana V, Narayana Darapaneni, Anwesh Reddy P

Introduction: Music provides an incredible avenue for individuals to express their thoughts and emotions, while also serving as a delightful mode of entertainment for enthusiasts and music lovers. Objectives: This paper presents a comprehensive approach to enhancing the user experience through the integration of emotion recognition, music recommendation, and explainable AI using GRAD-CAM. Methods: The proposed methodology utilizes a ResNet50 model trained on the Facial Expression Recognition (FER) dataset, consisting of real images of individuals expressing various emotions. Results: The system achieves an accuracy of 82% in emotion classification. By leveraging GRAD-CAM, the model provides explanations for its predictions, allowing users to understand the reasoning behind the system's recommendations. The model is trained on both FER and real user datasets, which include labelled facial expressions, and real images of individuals expressing various emotions. The training process involves pre-processing the input images, extracting features through convolutional layers, reasoning with dense layers, and generating emotion predictions through the output layer Conclusion: The proposed methodology, leveraging the Resnet50 model with ROI-based analysis and explainable AI techniques, offers a robust and interpretable solution for facial emotion detection paper.

link: http://arxiv.org/abs/2404.04654v1

Binary Classifier Optimization for Large Language Model Alignment

Seungjae Jung, Gunsoo Han, Daniel Wontae Nam, Kyoung-Woon On

Aligning Large Language Models (LLMs) to human preferences through preference optimization has been crucial but labor-intensive, necessitating for each prompt a comparison of both a chosen and a rejected text completion by evaluators. Recently, Kahneman-Tversky Optimization (KTO) has demonstrated that LLMs can be aligned using merely binary "thumbs-up" or "thumbs-down" signals on each prompt-completion pair. In this paper, we present theoretical foundations to explain the successful alignment achieved through these binary signals. Our analysis uncovers a new perspective: optimizing a binary classifier, whose logit is a reward, implicitly induces minimizing the Direct Preference Optimization (DPO) loss. In the process of this discovery, we identified two techniques for effective alignment: reward shift and underlying distribution matching. Consequently, we propose a new algorithm, \textit{Binary Classifier Optimization}, that integrates the techniques. We validate our methodology in two settings: first, on a paired preference dataset, where our method performs on par with DPO and KTO; and second, on binary signal datasets simulating real-world conditions with divergent underlying distributions between thumbs-up and thumbs-down

data. Our model consistently demonstrates effective and robust alignment across two base LLMs and three different binary signal datasets, showcasing the strength of our approach to learning from binary feedback.

link: http://arxiv.org/abs/2404.04656v1

Multilingual Pretraining and Instruction Tuning Improve Cross-Lingual Knowledge Alignment, But Only Shallowly

Changjiang Gao, Hongda Hu, Peng Hu, Jiajun Chen, Jixing Li, Shujian Huang

Despite their strong ability to retrieve knowledge in English, current large language models show imbalance abilities in different languages. Two approaches are proposed to address this, i.e., multilingual pretraining and multilingual instruction tuning. However, whether and how do such methods contribute to the cross-lingual knowledge alignment inside the models is unknown. In this paper, we propose CLiKA, a systematic framework to assess the cross-lingual knowledge alignment of LLMs in the Performance, Consistency and Conductivity levels, and explored the effect of multilingual pretraining and instruction tuning on the degree of alignment. Results show that: while both multilingual pretraining and instruction tuning are beneficial for cross-lingual knowledge alignment, the training strategy needs to be carefully designed. Namely, continued pretraining improves the alignment of the target language at the cost of other languages, while mixed pretraining affect other languages less. Also, the overall cross-lingual knowledge alignment, especially in the conductivity level, is unsatisfactory for all tested LLMs, and neither multilingual pretraining nor instruction tuning can substantially improve the cross-lingual knowledge conductivity.

link: http://arxiv.org/abs/2404.04659v1

Transform then Explore: a Simple and Effective Technique for Exploratory Combinatorial Optimization with Reinforcement Learning

Tianle Pu, Changjun Fan, Mutian Shen, Yizhou Lu, Li Zeng, Zohar Nussinov, Chao Chen, Zhong Liu

Many complex problems encountered in both production and daily life can be conceptualized as combinatorial optimization problems (COPs) over graphs. Recent years, reinforcement learning (RL) based models have emerged as a promising direction, which treat the COPs solving as a heuristic learning problem. However, current finite-horizon-MDP based RL models have inherent limitations. They are not allowed to explore adquately for improving solutions at test time, which may be necessary given the complexity of NP-hard optimization tasks. Some recent attempts solve this issue by focusing on reward design and state feature engineering, which are tedious and ad-hoc. In this work, we instead propose a much simpler but more effective technique, named gauge transformation (GT). The technique is originated from physics, but is very effective in enabling RL agents to explore to continuously improve the solutions during test. Morever, GT is very simple, which can be implemented with less than 10 lines of Python codes, and can be applied to a vast majority of RL models. Experimentally, we show that traditional RL models with GT technique produce the state-of-the-art performances on the MaxCut problem. Furthermore, since GT is independent of any RL models, it can be seamlessly integrated into various RL frameworks, paving the way of these models for more effective explorations in the solving of general COPs.

link: http://arxiv.org/abs/2404.04661v1

Learning Minimal NAP Specifications for Neural Network Verification

Chuqin Geng, Zhaoyue Wang, Haolin Ye, Saifei Liao, Xujie Si

Specifications play a crucial role in neural network verification. They define the precise input regions we aim to verify, typically represented as L-infinity norm balls. While recent research suggests using neural activation patterns (NAPs) as specifications for verifying unseen test set data, it focuses on computing the most refined NAPs, often limited to very small regions in the input space. In this paper, we study the following problem: Given a neural network, find a minimal (coarsest) NAP that is sufficient for formal verification of the network's robustness. Finding the

minimal NAP specification not only expands verifiable bounds but also provides insights into which neurons contribute to the model's robustness. To address this problem, we propose several exact and approximate approaches. Our exact approaches leverage the verification tool to find minimal NAP specifications in either a deterministic or statistical manner. Whereas the approximate methods efficiently estimate minimal NAPs using adversarial examples and local gradients, without making calls to the verification tool. This allows us to inspect potential causal links between neurons and the robustness of state-of-the-art neural networks, a task for which existing verification frameworks fail to scale. Our experimental results suggest that minimal NAP specifications require much smaller fractions of neurons compared to the most refined NAP specifications, yet they can significantly expand the verifiable boundaries to several orders of magnitude larger.

link: http://arxiv.org/abs/2404.04662v1

Focused Active Learning for Histopathological Image Classification

Arne Schmidt, Pablo Morales-Álvarez, Lee A. D. Cooper, Lee A. Newberg, Andinet Enquobahrie, Aggelos K. Katsaggelos, Rafael Molina

Active Learning (AL) has the potential to solve a major problem of digital pathology: the efficient acquisition of labeled data for machine learning algorithms. However, existing AL methods often struggle in realistic settings with artifacts, ambiguities, and class imbalances, as commonly seen in the medical field. The lack of precise uncertainty estimations leads to the acquisition of images with a low informative value. To address these challenges, we propose Focused Active Learning (FocAL), which combines a Bayesian Neural Network with Out-of-Distribution detection to estimate different uncertainties for the acquisition function. Specifically, the weighted epistemic uncertainty accounts for the class imbalance, aleatoric uncertainty for ambiguous images, and an OoD score for artifacts. We perform extensive experiments to validate our method on MNIST and the real-world Panda dataset for the classification of prostate cancer. The results confirm that other AL methods are 'distracted' by ambiguities and artifacts which harm the performance. FocAL effectively focuses on the most informative images, avoiding ambiguities and artifacts during acquisition. For both experiments, FocAL outperforms existing AL approaches, reaching a Cohen's kappa of 0.764 with only 0.69% of the labeled Panda data.

link: http://arxiv.org/abs/2404.04663v1

Adaptive Intra-Class Variation Contrastive Learning for Unsupervised Person Re-Identification

Lingzhi Liu, Haiyang Zhang, Chengwei Tang, Tiantian Zhang

The memory dictionary-based contrastive learning method has achieved remarkable results in the field of unsupervised person Re-ID. However, The method of updating memory based on all samples does not fully utilize the hardest sample to improve the generalization ability of the model, and the method based on hardest sample mining will inevitably introduce false-positive samples that are incorrectly clustered in the early stages of the model. Clustering-based methods usually discard a significant number of outliers, leading to the loss of valuable information. In order to address the issues mentioned before, we propose an adaptive intra-class variation contrastive learning algorithm for unsupervised Re-ID, called AdaInCV. And the algorithm quantitatively evaluates the learning ability of the model for each class by considering the intra-class variations after clustering, which helps in selecting appropriate samples during the training process of the model. To be more specific, two new strategies are proposed: Adaptive Sample Mining (AdaSaM) and Adaptive Outlier Filter (AdaOF). The first one gradually creates more reliable clusters to dynamically refine the memory, while the second can identify and filter out valuable outliers as negative samples.

link: http://arxiv.org/abs/2404.04665v1

Domain Generalisation via Imprecise Learning

Anurag Singh, Siu Lun Chau, Shahine Bouabid, Krikamol Muandet

Out-of-distribution (OOD) generalisation is challenging because it involves not only learning from empirical data, but also deciding among various notions of generalisation, e.g., optimising the average-case risk, worst-case risk, or interpolations thereof. While this choice should in principle be made by the model operator like medical doctors, this information might not always be available at training time. The institutional separation between machine learners and model operators leads to arbitrary commitments to specific generalisation strategies by machine learners due to these deployment uncertainties. We introduce the Imprecise Domain Generalisation framework to mitigate this, featuring an imprecise risk optimisation that allows learners to stay imprecise by optimising against a continuous spectrum of generalisation strategies during training, and a model framework that allows operators to specify their generalisation preference at deployment. Supported by both theoretical and empirical evidence, our work showcases the benefits of integrating imprecision into domain generalisation.

link: http://arxiv.org/abs/2404.04669v1

Neural-ABC: Neural Parametric Models for Articulated Body with Clothes

Honghu Chen, Yuxin Yao, Juyong Zhang

In this paper, we introduce Neural-ABC, a novel parametric model based on neural implicit functions that can represent clothed human bodies with disentangled latent spaces for identity, clothing, shape, and pose. Traditional mesh-based representations struggle to represent articulated bodies with clothes due to the diversity of human body shapes and clothing styles, as well as the complexity of poses. Our proposed model provides a unified framework for parametric modeling, which can represent the identity, clothing, shape and pose of the clothed human body. Our proposed approach utilizes the power of neural implicit functions as the underlying representation and integrates well-designed structures to meet the necessary requirements. Specifically, we represent the underlying body as a signed distance function and clothing as an unsigned distance function, and they can be uniformly represented as unsigned distance fields. Different types of clothing do not require predefined topological structures or classifications, and can follow changes in the underlying body to fit the body. Additionally, we construct poses using a controllable articulated structure. The model is trained on both open and newly constructed datasets, and our decoupling strategy is carefully designed to ensure optimal performance. Our model excels at disentangling clothing and identity in different shape and poses while preserving the style of the clothing. We demonstrate that Neural-ABC fits new observations of different types of clothing. Compared to other state-of-the-art parametric models, Neural-ABC demonstrates powerful advantages in the reconstruction of clothed human bodies, as evidenced by fitting raw scans, depth maps and images. We show that the attributes of the fitted results can be further edited by adjusting their identities, clothing, shape and pose codes.

link: http://dx.doi.org/10.1109/tvcg.2024.3364814

Order-Based Pre-training Strategies for Procedural Text Understanding

Abhilash Nandy, Yash Kulkarni, Pawan Goyal, Niloy Ganguly

In this paper, we propose sequence-based pretraining methods to enhance procedural understanding in natural language processing. Procedural text, containing sequential instructions to accomplish a task, is difficult to understand due to the changing attributes of entities in the context. We focus on recipes, which are commonly represented as ordered instructions, and use this order as a supervision signal. Our work is one of the first to compare several 'order as-supervision' transformer pre-training methods, including Permutation Classification, Embedding Regression, and Skip-Clip, and shows that these methods give improved results compared to the baselines and SoTA LLMs on two downstream Entity-Tracking datasets: NPN-Cooking dataset in recipe domain and ProPara dataset in open domain. Our proposed methods address the non-trivial Entity Tracking Task that requires prediction of entity states across procedure steps, which requires understanding the order of steps. These methods show an improvement over the best baseline by 1.6% and 7-9% on NPN-Cooking and ProPara Datasets respectively across metrics.

link: http://arxiv.org/abs/2404.04676v1

Salient Sparse Visual Odometry With Pose-Only Supervision

Siyu Chen, Kangcheng Liu, Chen Wang, Shenghai Yuan, Jianfei Yang, Lihua Xie

Visual Odometry (VO) is vital for the navigation of autonomous systems, providing accurate position and orientation estimates at reasonable costs. While traditional VO methods excel in some conditions, they struggle with challenges like variable lighting and motion blur. Deep learning-based VO, though more adaptable, can face generalization problems in new environments. Addressing these drawbacks, this paper presents a novel hybrid visual odometry (VO) framework that leverages pose-only supervision, offering a balanced solution between robustness and the need for extensive labeling. We propose two cost-effective and innovative designs: a self-supervised homographic pre-training for enhancing optical flow learning from pose-only labels and a random patch-based salient point detection strategy for more accurate optical flow patch extraction. These designs eliminate the need for dense optical flow labels for training and significantly improve the generalization capability of the system in diverse and challenging environments. Our pose-only supervised method achieves competitive performance on standard datasets and greater robustness and generalization ability in extreme and unseen scenarios, even compared to dense optical flow-supervised state-of-the-art methods.

link: http://dx.doi.org/10.1109/LRA.2024.3384757

Automatic Gradient Estimation for Calibrating Crowd Models with Discrete Decision Making

Philipp Andelfinger, Justin N. Kreikemeyer

Recently proposed gradient estimators enable gradient descent over stochastic programs with discrete jumps in the response surface, which are not covered by automatic differentiation (AD) alone. Although these estimators' capability to guide a swift local search has been shown for certain problems, their applicability to models relevant to real-world applications remains largely unexplored. As the gradients governing the choice in candidate solutions are calculated from sampled simulation trajectories, the optimization procedure bears similarities to metaheuristics such as particle swarm optimization, which puts the focus on the different methods' calibration progress per function evaluation. Here, we consider the calibration of force-based crowd evacuation models based on the popular Social Force model augmented by discrete decision making. After studying the ability of an AD-based estimator for branching programs to capture the simulation's rugged response surface, calibration problems are tackled using gradient descent and two metaheuristics. As our main insights, we find 1) that the estimation's fidelity benefits from disregarding jumps of large magnitude inherent to the Social Force model, and 2) that the common problem of calibration by adjusting a simulation input distribution obviates the need for AD across the Social Force calculations, allowing gradient descent to excel.

link: http://arxiv.org/abs/2404.04678v1

Compositional Conservatism: A Transductive Approach in Offline Reinforcement Learning

Yeda Song, Dongwook Lee, Gunhee Kim

Offline reinforcement learning (RL) is a compelling framework for learning optimal policies from past experiences without additional interaction with the environment. Nevertheless, offline RL inevitably faces the problem of distributional shifts, where the states and actions encountered during policy execution may not be in the training dataset distribution. A common solution involves incorporating conservatism into the policy or the value function to safeguard against uncertainties and unknowns. In this work, we focus on achieving the same objectives of conservatism but from a different perspective. We propose COmpositional COnservatism with Anchor-seeking (COCOA) for offline RL, an approach that pursues conservatism in a compositional manner on top of the transductive reparameterization (Netanyahu et al., 2023), which decomposes the input variable (the state in our case) into an anchor and its difference from the original input. Our COCOA seeks both in-distribution anchors and differences by utilizing the learned reverse dynamics model,

encouraging conservatism in the compositional input space for the policy or value function. Such compositional conservatism is independent of and agnostic to the prevalent behavioral conservatism in offline RL. We apply COCOA to four state-of-the-art offline RL algorithms and evaluate them on the D4RL benchmark, where COCOA generally improves the performance of each algorithm. The code is available at https://github.com/runamu/compositional-conservatism.

link: http://arxiv.org/abs/2404.04682v1

Predictive Modeling for Breast Cancer Classification in the Context of Bangladeshi Patients: A Supervised Machine Learning Approach with Explainable AI

Taminul Islam, Md. Alif Sheakh, Mst. Sazia Tahosin, Most. Hasna Hena, Shopnil Akash, Yousef A. Bin Jardan, Gezahign Fentahun Wondmie, Hiba-Allah Nafidi, Mohammed Bourhia

Breast cancer has rapidly increased in prevalence in recent years, making it one of the leading causes of mortality worldwide. Among all cancers, it is by far the most common. Diagnosing this illness manually requires significant time and expertise. Since detecting breast cancer is a time-consuming process, preventing its further spread can be aided by creating machine-based forecasts. Machine learning and Explainable AI are crucial in classification as they not only provide accurate predictions but also offer insights into how the model arrives at its decisions, aiding in the understanding and trustworthiness of the classification results. In this study, we evaluate and compare the classification accuracy, precision, recall, and F-1 scores of five different machine learning methods using a primary dataset (500 patients from Dhaka Medical College Hospital). Five different supervised machine learning techniques, including decision tree, random forest, logistic regression, naive bayes, and XGBoost, have been used to achieve optimal results on our dataset. Additionally, this study applied SHAP analysis to the XGBoost model to interpret the model's predictions and understand the impact of each feature on the model's output. We compared the accuracy with which several algorithms classified the data, as well as contrasted with other literature in this field. After final evaluation, this study found that XGBoost achieved the best model accuracy, which is 97%.

link: http://arxiv.org/abs/2404.04686v1