Wed 2024.04.03

Unknown Prompt, the only Lacuna: Unveiling CLIP's Potential for Open Domain Generalization

Mainak Singha, Ankit Jha, Shirsha Bose, Ashwin Nair, Moloud Abdar, Biplab Banerjee

We delve into Open Domain Generalization (ODG), marked by domain and category shifts between training's labeled source and testing's unlabeled target domains. Existing solutions to ODG face limitations due to constrained generalizations of traditional CNN backbones and errors in detecting target open samples in the absence of prior knowledge. Addressing these pitfalls, we introduce ODG-CLIP, harnessing the semantic prowess of the vision-language model, CLIP. Our framework brings forth three primary innovations: Firstly, distinct from prevailing paradigms, we conceptualize ODG as a multi-class classification challenge encompassing both known and novel categories. Central to our approach is modeling a unique prompt tailored for detecting unknown class samples, and to train this, we employ a readily accessible stable diffusion model, elegantly generating proxy images for the open class. Secondly, aiming for domain-tailored classification (prompt) weights while ensuring a balance of precision and simplicity, we devise a novel visual stylecentric prompt learning mechanism. Finally, we infuse images with class-discriminative knowledge derived from the prompt space to augment the fidelity of CLIP's visual embeddings. We introduce a novel objective to safeguard the continuity of this infused semantic intel across domains, especially for the shared classes. Through rigorous testing on diverse datasets, covering closed and open-set DG contexts, ODG-CLIP demonstrates clear supremacy, consistently outpacing peers with performance boosts between 8%-16%. Code will be available at https://github.com/mainaksingha01/ODG-CLIP.

link: http://arxiv.org/abs/2404.00710v1

Survey of Computerized Adaptive Testing: A Machine Learning Perspective

Qi Liu, Yan Zhuang, Haoyang Bi, Zhenya Huang, Weizhe Huang, Jiatong Li, Junhao Yu, Zirui Liu, Zirui Hu, Yuting Hong, Zachary A. Pardos, Haiping Ma, Mengxiao Zhu, Shijin Wang, Enhong Chen

Computerized Adaptive Testing (CAT) provides an efficient and tailored method for assessing the proficiency of examinees, by dynamically adjusting test questions based on their performance. Widely adopted across diverse fields like education, healthcare, sports, and sociology, CAT has revolutionized testing practices. While traditional methods rely on psychometrics and statistics, the increasing complexity of large-scale testing has spurred the integration of machine learning techniques. This paper aims to provide a machine learning-focused survey on CAT, presenting a fresh perspective on this adaptive testing method. By examining the test question selection algorithm at the heart of CAT's adaptivity, we shed light on its functionality. Furthermore, we delve into cognitive diagnosis models, question bank construction, and test control within CAT, exploring how machine learning can optimize these components. Through an analysis of current methods, strengths, limitations, and challenges, we strive to develop robust, fair, and efficient CAT systems. By bridging psychometric-driven CAT research with machine learning, this survey advocates for a more inclusive and interdisciplinary approach to the future of adaptive testing.

link: http://arxiv.org/abs/2404.00712v1

Neural Radiance Field-based Visual Rendering: A Comprehensive Review

Mingyuan Yao, Yukang Huo, Yang Ran, Qingbin Tian, Ruifeng Wang, Haihua Wang

In recent years, Neural Radiance Fields (NeRF) has made remarkable progress in the field of computer vision and graphics, providing strong technical support for solving key tasks including 3D scene understanding, new perspective synthesis, human body reconstruction, robotics, and so on, the attention of academics to this research result is growing. As a revolutionary neural implicit field representation, NeRF has caused a continuous research boom in the academic community. Therefore, the purpose of this review is to provide an in-depth analysis of the research literature on NeRF within the past two years, to provide a comprehensive academic perspective for budding

researchers. In this paper, the core architecture of NeRF is first elaborated in detail, followed by a discussion of various improvement strategies for NeRF, and case studies of NeRF in diverse application scenarios, demonstrating its practical utility in different domains. In terms of datasets and evaluation metrics, This paper details the key resources needed for NeRF model training. Finally, this paper provides a prospective discussion on the future development trends and potential challenges of NeRF, aiming to provide research inspiration for researchers in the field and to promote the further development of related technologies.

link: http://arxiv.org/abs/2404.00714v1

End-to-End Autonomous Driving through V2X Cooperation

Haibao Yu, Wenxian Yang, Jiaru Zhong, Zhenwei Yang, Siqi Fan, Ping Luo, Zaiqing Nie

Cooperatively utilizing both ego-vehicle and infrastructure sensor data via V2X communication has emerged as a promising approach for advanced autonomous driving. However, current research mainly focuses on improving individual modules, rather than taking end-to-end learning to optimize final planning performance, resulting in underutilized data potential. In this paper, we introduce UniV2X, a pioneering cooperative autonomous driving framework that seamlessly integrates all key driving modules across diverse views into a unified network. We propose a sparse-dense hybrid data transmission and fusion mechanism for effective vehicle-infrastructure cooperation, offering three advantages: 1) Effective for simultaneously enhancing agent perception, online mapping, and occupancy prediction, ultimately improving planning performance. 2) Transmission-friendly for practical and limited communication conditions. 3) Reliable data fusion with interpretability of this hybrid data. We implement UniV2X, as well as reproducing several benchmark methods, on the challenging DAIR-V2X, the real-world cooperative driving dataset. Experimental results demonstrate the effectiveness of UniV2X in significantly enhancing planning performance, as well as all intermediate output performance. Code is at https://github.com/AIR-THU/UniV2X.

link: http://arxiv.org/abs/2404.00717v1

DRCT: Saving Image Super-resolution away from Information Bottleneck

Chih-Chung Hsu, Chia-Ming Lee, Yi-Shiuan Chou

In recent years, Vision Transformer-based applications to low-level vision tasks have achieved widespread success. Unlike CNN-based models, Transformers are more adept at capturing long-range dependencies, enabling the reconstruction of images utilizing information from non-local areas. In the domain of super-resolution, Swin-transformer-based approaches have become mainstream due to their capacity to capture global spatial information and their shifting-window attention mechanism that facilitates the interchange of information between different windows. Many researchers have enhanced image quality and network efficiency by expanding the receptive field or designing complex networks, yielding commendable results. However, we observed that spatial information tends to diminish during the forward propagation process due to increased depth, leading to a loss of spatial information and, consequently, limiting the model's potential. To address this, we propose the Dense-residual-connected Transformer (DRCT), aimed at mitigating the loss of spatial information through dense-residual connections between layers, thereby unleashing the model's potential and enhancing performance. Experiment results indicate that our approach is not only straightforward but also achieves remarkable efficiency, surpassing state-of-the-art methods and performing commendably at NTIRE2024.

link: http://arxiv.org/abs/2404.00722v2

Absolute-Unified Multi-Class Anomaly Detection via Class-Agnostic Distribution Alignment

Jia Guo, Shuai Lu, Weihang Zhang, Huiqi Li

Conventional unsupervised anomaly detection (UAD) methods build separate models for each object category. Recent studies have proposed to train a unified model for multiple classes, namely model-unified UAD. However, such methods still implement the unified model separately on each class during inference with respective anomaly decision thresholds, which hinders their application

when the image categories are entirely unavailable. In this work, we present a simple yet powerful method to address multi-class anomaly detection without any class information, namely \textit{absolute-unified} UAD. We target the crux of prior works in this challenging setting: different objects have mismatched anomaly score distributions. We propose Class-Agnostic Distribution Alignment (CADA) to align the mismatched score distribution of each implicit class without knowing class information, which enables unified anomaly detection for all classes and samples. The essence of CADA is to predict each class's score distribution of normal samples given any image, normal or anomalous, of this class. As a general component, CADA can activate the potential of nearly all UAD methods under absolute-unified setting. Our approach is extensively evaluated under the proposed setting on two popular UAD benchmark datasets, MVTec AD and VisA, where we exceed previous state-of-the-art by a large margin.

link: http://arxiv.org/abs/2404.00724v1

The Larger the Better? Improved LLM Code-Generation via Budget Reallocation

Michael Hassid, Tal Remez, Jonas Gehring, Roy Schwartz, Yossi Adi

It is a common belief that large language models (LLMs) are better than smaller-sized ones. However, larger models also require significantly more time and compute during inference. This begs the question: what happens when both models operate under the same budget? (e.g., compute, run-time). To address this question, we analyze code generation LLMs of various sizes and make comparisons such as running a 70B model once vs. generating five outputs from a 13B model and selecting one. Our findings reveal that, in a standard unit-test setup, the repeated use of smaller models can yield consistent improvements, with gains of up to 15% across five tasks. On the other hand, in scenarios where unit-tests are unavailable, a ranking-based selection of candidates from the smaller model falls short of the performance of a single output from larger ones. Our results highlight the potential of using smaller models instead of larger ones, and the importance of studying approaches for ranking LLM outputs.

link: http://arxiv.org/abs/2404.00725v1

MugenNet: A Novel Combined Convolution Neural Network and Transformer Network with its Application for Colonic Polyp Image Segmentation

Chen Peng, Zhiqin Qian, Kunyu Wang, Qi Luo, Zhuming Bi, Wenjun Zhang

Biomedical image segmentation is a very important part in disease diagnosis. The term "colonic polyps" refers to polypoid lesions that occur on the surface of the colonic mucosa within the intestinal lumen. In clinical practice, early detection of polyps is conducted through colonoscopy examinations and biomedical image processing. Therefore, the accurate polyp image segmentation is of great significance in colonoscopy examinations. Convolutional Neural Network (CNN) is a common automatic segmentation method, but its main disadvantage is the long training time. Transformer utilizes a self-attention mechanism, which essentially assigns different importance weights to each piece of information, thus achieving high computational efficiency during segmentation. However, a potential drawback is the risk of information loss. In the study reported in this paper, based on the well-known hybridization principle, we proposed a method to combine CNN and Transformer to retain the strengths of both, and we applied this method to build a system called MugenNet for colonic polyp image segmentation. We conducted a comprehensive experiment to compare MugenNet with other CNN models on five publicly available datasets. The ablation experiment on MugentNet was conducted as well. The experimental results show that MugenNet achieves significantly higher processing speed and accuracy compared with CNN alone. The generalized implication with our work is a method to optimally combine two complimentary methods of machine learning.

link: http://arxiv.org/abs/2404.00726v1

A Controlled Reevaluation of Coreference Resolution Models

Ian Porada, Xiyuan Zou, Jackie Chi Kit Cheung

All state-of-the-art coreference resolution (CR) models involve finetuning a pretrained language model. Whether the superior performance of one CR model over another is due to the choice of language model or other factors, such as the task-specific architecture, is difficult or impossible to determine due to lack of a standardized experimental setup. To resolve this ambiguity, we systematically evaluate five CR models and control for certain design decisions including the pretrained language model used by each. When controlling for language model size, encoder-based CR models outperform more recent decoder-based models in terms of both accuracy and inference speed. Surprisingly, among encoder-based CR models, more recent models are not always more accurate, and the oldest CR model that we test generalizes the best to out-of-domain textual genres. We conclude that controlling for the choice of language model reduces most, but not all, of the increase in F1 score reported in the past five years.

link: http://arxiv.org/abs/2404.00727v1

Nonparametric End-to-End Probabilistic Forecasting of Distributed Generation Outputs Considering Missing Data Imputation

Minghui Chen, Zichao Meng, Yanping Liu, Longbo Luo, Ye Guo, Kang Wang

In this paper, we introduce a nonparametric end-to-end method for probabilistic forecasting of distributed renewable generation outputs while including missing data imputation. Firstly, we employ a nonparametric probabilistic forecast model utilizing the long short-term memory (LSTM) network to model the probability distributions of distributed renewable generations' outputs. Secondly, we design an end-to-end training process that includes missing data imputation through iterative imputation and iterative loss-based training procedures. This two-step modeling approach effectively combines the strengths of the nonparametric method with the end-to-end approach. Consequently, our approach demonstrates exceptional capabilities in probabilistic forecasting for the outputs of distributed renewable generations while effectively handling missing values. Simulation results confirm the superior performance of our approach compared to existing alternatives.

link: http://arxiv.org/abs/2404.00729v1

Smooth Information Gathering in Two-Player Noncooperative Games

Fernando Palafox, Jesse Milzman, Dong Ho Lee, Ryan Park, David Fridovich-Keil

We present a mathematical framework for modeling two-player noncooperative games in which one player (the defender) is uncertain of the costs of the game and the second player's (the attacker's) intention but can preemptively allocate information-gathering resources to reduce this uncertainty. We obtain the defender's decisions by solving a two-stage problem. In Stage 1, the defender allocates information-gathering resources, and in Stage 2, the information-gathering resources output a signal that informs the defender about the costs of the game and the attacker's intent, and then both players play a noncooperative game. We provide a gradient-based algorithm to solve the two-stage game and apply this framework to a tower-defense game which can be interpreted as a variant of a Colonel Blotto game with smooth payoff functions and uncertainty over battlefield valuations. Finally, we analyze how optimal decisions shift with changes in information-gathering allocations and perturbations in the cost functions.

link: http://arxiv.org/abs/2404.00733v1

Opera Graeca Adnotata: Building a 34M+ Token Multilayer Corpus for Ancient Greek

Giuseppe G. A. Celano

In this article, the beta version 0.1.0 of Opera Graeca Adnotata (OGA), the largest open-access multilayer corpus for Ancient Greek (AG) is presented. OGA consists of 1,687 literary works and 34M+ tokens coming from the PerseusDL and OpenGreekAndLatin GitHub repositories, which host AG texts ranging from about 800 BCE to about 250 CE. The texts have been enriched with seven annotation layers: (i) tokenization layer; (ii) sentence segmentation layer; (iii) lemmatization layer; (iv) morphological layer; (v) dependency layer; (vi) dependency function layer; (vii) Canonical Text Services (CTS) citation layer. The creation of each layer is described by highlighting the main

technical and annotation-related issues encountered. Tokenization, sentence segmentation, and CTS citation are performed by rule-based algorithms, while morphosyntactic annotation is the output of the COMBO parser trained on the data of the Ancient Greek Dependency Treebank. For the sake of scalability and reusability, the corpus is released in the standoff formats PAULA XML and its offspring LAULA XML.

link: http://arxiv.org/abs/2404.00739v1

Rethinking Interactive Image Segmentation with Low Latency, High Quality, and Diverse Prompts

Qin Liu, Jaemin Cho, Mohit Bansal, Marc Niethammer

The goal of interactive image segmentation is to delineate specific regions within an image via visual or language prompts. Low-latency and high-quality interactive segmentation with diverse prompts remain challenging for existing specialist and generalist models. Specialist models, with their limited prompts and task-specific designs, experience high latency because the image must be recomputed every time the prompt is updated, due to the joint encoding of image and visual prompts. Generalist models, exemplified by the Segment Anything Model (SAM), have recently excelled in prompt diversity and efficiency, lifting image segmentation to the foundation model era. However, for high-quality segmentations, SAM still lags behind state-of-the-art specialist models despite SAM being trained with x100 more segmentation masks. In this work, we delve deep into the architectural differences between the two types of models. We observe that dense representation and fusion of visual prompts are the key design choices contributing to the high segmentation quality of specialist models. In light of this, we reintroduce this dense design into the generalist models, to facilitate the development of generalist models with high segmentation quality. To densely represent diverse visual prompts, we propose to use a dense map to capture five types: clicks, boxes, polygons, scribbles, and masks. Thus, we propose SegNext, a next-generation interactive segmentation approach offering low latency, high quality, and diverse prompt support. Our method outperforms current state-of-the-art methods on HQSeg-44K and DAVIS, both quantitatively and qualitatively.

link: http://arxiv.org/abs/2404.00741v1

Adapting to Length Shift: FlexiLength Network for Trajectory Prediction

Yi Xu, Yun Fu

Trajectory prediction plays an important role in various applications, including autonomous driving, robotics, and scene understanding. Existing approaches mainly focus on developing compact neural networks to increase prediction precision on public datasets, typically employing a standardized input duration. However, a notable issue arises when these models are evaluated with varying observation lengths, leading to a significant performance drop, a phenomenon we term the Observation Length Shift. To address this issue, we introduce a general and effective framework, the FlexiLength Network (FLN), to enhance the robustness of existing trajectory prediction techniques against varying observation periods. Specifically, FLN integrates trajectory data with diverse observation lengths, incorporates FlexiLength Calibration (FLC) to acquire temporal invariant representations, and employs FlexiLength Adaptation (FLA) to further refine these representations for more accurate future trajectory predictions. Comprehensive experiments on multiple datasets, ie, ETH/UCY, nuScenes, and Argoverse 1, demonstrate the effectiveness and flexibility of our proposed FLN framework.

link: http://arxiv.org/abs/2404.00742v1

Mining Weighted Sequential Patterns in Incremental Uncertain Databases

Kashob Kumar Roy, Md Hasibul Haque Moon, Md Mahmudur Rahman, Chowdhury Farhan Ahmed, Carson Kai-Sang Leung

Due to the rapid development of science and technology, the importance of imprecise, noisy, and uncertain data is increasing at an exponential rate. Thus, mining patterns in uncertain databases have drawn the attention of researchers. Moreover, frequent sequences of items from these

databases need to be discovered for meaningful knowledge with great impact. In many real cases, weights of items and patterns are introduced to find interesting sequences as a measure of importance. Hence, a constraint of weight needs to be handled while mining sequential patterns. Besides, due to the dynamic nature of databases, mining important information has become more challenging. Instead of mining patterns from scratch after each increment, incremental mining algorithms utilize previously mined information to update the result immediately. Several algorithms exist to mine frequent patterns and weighted sequences from incremental databases. However, these algorithms are confined to mine the precise ones. Therefore, we have developed an algorithm to mine frequent sequences in an uncertain database in this work. Furthermore, we have proposed two new techniques for mining when the database is incremental. Extensive experiments have been conducted for performance evaluation. The analysis showed the efficiency of our proposed framework.

link: http://arxiv.org/abs/2404.00746v1