Tue 2024.04.23

Approximate Algorithms For k-Sparse Wasserstein Barycenter With Outliers

Qingyuan Yang, Hu Ding

Wasserstein Barycenter (WB) is one of the most fundamental optimization problems in optimal transportation. Given a set of distributions, the goal of WB is to find a new distribution that minimizes the average Wasserstein distance to them. The problem becomes even harder if we restrict the solution to be ``k-sparse''. In this paper, we study the k-sparse WB problem in the presence of outliers, which is a more practical setting since real-world data often contains noise. Existing WB algorithms cannot be directly extended to handle the case with outliers, and thus it is urgently needed to develop some novel ideas. First, we investigate the relation between k-sparse WB with outliers and the clustering (with outliers) problems. In particular, we propose a clustering based LP method that yields constant approximation factor for the k-sparse WB with outliers problem. Further, we utilize the coreset technique to achieve the $(1+\epsilon)$-approximation factor for any $\epsilon>0$, if the dimensionality is not high. Finally, we conduct the experiments for our proposed algorithms and illustrate their efficiencies in practice.

link: http://arxiv.org/abs/2404.13401v1

Intrusion Detection at Scale with the Assistance of a Command-line Language Model

Jiongliang Lin, Yiwen Guo, Hao Chen

Intrusion detection is a long standing and crucial problem in security. A system capable of detecting intrusions automatically is on great demand in enterprise security solutions. Existing solutions rely heavily on hand-crafted rules designed by security operators, which suffer from high false negative rates and poor generalization ability to new, zero-day attacks at scale. AI and machine learning offer promising solutions to address the issues, by inspecting abnormal user behaviors intelligently and automatically from data. However, existing learning-based intrusion detection systems in the literature are mostly designed for small data, and they lack the ability to leverage the power of big data in cloud environments. In this paper, we target at this problem and introduce an intrusion detection system which incorporates large-scale pre-training, so as to train a large language model based on tens of millions of command lines for AI-based intrusion detection. Experiments performed on 30 million training samples and 10 million test samples verify the effectiveness of our solution.

link: http://arxiv.org/abs/2404.13402v1

Converter: Enhancing Interoperability in Research Data Management

Sefika Efeoglu, Zongxiong Chen, Sonja Schimmler, Bianca Wentzel

Research Data Management (RDM) is essential in handling and organizing data in the research field. The Berlin Open Science Platform (BOP) serves as a case study that exemplifies the significance of standardization within the Berlin University Alliance (BUA), employing different vocabularies when publishing their data, resulting in data heterogeneity. The meta portals of the NFDI4Cat and the NFDI4DataScience project serve as additional case studies in the context of the NFDI initiative. To establish consistency among the harvested repositories in the respective systems, this study focuses on developing a novel component, namely the converter, that breaks barriers between data collection and various schemas. With the minor modification of the existing Piveau framework, the development of the converter, contributes to enhanced data accessibility, streamlined collaboration, and improved interoperability within the research community.

link: http://arxiv.org/abs/2404.13406v1

AMMUNet: Multi-Scale Attention Map Merging for Remote Sensing Image Segmentation

Yang Yang, Shunyi Zheng

The advancement of deep learning has driven notable progress in remote sensing semantic segmentation. Attention mechanisms, while enabling global modeling and utilizing contextual information, face challenges of high computational costs and require window-based operations that weaken capturing long-range dependencies, hindering their effectiveness for remote sensing image processing. In this letter, we propose AMMUNet, a UNet-based framework that employs multi-scale attention map merging, comprising two key innovations: the granular multi-head self-attention (GMSA) module and the attention map merging mechanism (AMMM). GMSA efficiently acquires global information while substantially mitigating computational costs in contrast to global multi-head self-attention mechanism. This is accomplished through the strategic utilization of dimension correspondence to align granularity and the reduction of relative position bias parameters, thereby optimizing computational efficiency. The proposed AMMM effectively combines multi-scale attention maps into a unified representation using a fixed mask template, enabling the modeling of global attention mechanism. Experimental evaluations highlight the superior performance of our approach, achieving remarkable mean intersection over union (mIoU) scores of 75.48\% on the challenging Vaihingen dataset and an exceptional 77.90\% on the Potsdam dataset, demonstrating the superiority of our method in precise remote sensing semantic segmentation. Codes are available at https://github.com/interpretty/AMMUNet.

link: http://arxiv.org/abs/2404.13408v1

Urgent Edge Computing

Patrizio Dazzi, Luca Ferrucci, Marco Danelutto, Konstantinos Tserpes, Antonis Makris, Theodoros Theodoropoulos, Jacopo Massa, Emanuele Carlini, Matteo Mordacchini

This position paper introduces Urgent Edge Computing (UEC) as a paradigm shift addressing the evolving demands of time-sensitive applications in distributed edge environments, in time-critical scenarios. With a focus on ultra-low latency, availability, resource management, decentralization, self-organization, and robust security, UEC aims to facilitate operations in critical scenarios such as disaster response, environmental monitoring, and smart city management. This paper outlines and discusses the key requirements, challenges, and enablers along with a conceptual architecture. The paper also outlines the potential applications of Urgent Edge Computing

link: http://dx.doi.org/10.1145/3659994.3660315

Efficient and Concise Explanations for Object Detection with Gaussian-Class Activation Mapping Explainer

Quoc Khanh Nguyen, Truong Thanh Hung Nguyen, Vo Thanh Khang Nguyen, Van Binh Truong, Tuong Phan, Hung Cao

To address the challenges of providing quick and plausible explanations in Explainable AI (XAI) for object detection models, we introduce the Gaussian Class Activation Mapping Explainer (G-CAME). Our method efficiently generates concise saliency maps by utilizing activation maps from selected layers and applying a Gaussian kernel to emphasize critical image regions for the predicted object. Compared with other Region-based approaches, G-CAME significantly reduces explanation time to 0.5 seconds without compromising the quality. Our evaluation of G-CAME, using Faster-RCNN and YOLOX on the MS-COCO 2017 dataset, demonstrates its ability to offer highly plausible and faithful explanations, especially in reducing the bias on tiny object detection.

link: http://arxiv.org/abs/2404.13417v1

NeurCADRecon: Neural Representation for Reconstructing CAD Surfaces by Enforcing Zero Gaussian Curvature

Qiujie Dong, Rui Xu, Pengfei Wang, Shuangmin Chen, Shiqing Xin, Xiaohong Jia, Wenping Wang, Changhe Tu

Despite recent advances in reconstructing an organic model with the neural signed distance function (SDF), the high-fidelity reconstruction of a CAD model directly from low-quality unoriented point clouds remains a significant challenge. In this paper, we address this challenge based on the prior observation that the surface of a CAD model is generally composed of piecewise surface patches, each approximately developable even around the feature line. Our approach, named NeurCADRecon, is self-supervised, and its loss includes a developability term to encourage the Gaussian curvature toward 0 while ensuring fidelity to the input points. Noticing that the Gaussian curvature is non-zero at tip points, we introduce a double-trough curve to tolerate the existence of these tip points. Furthermore, we develop a dynamic sampling strategy to deal with situations where the given points are incomplete or too sparse. Since our resulting neural SDFs can clearly manifest sharp feature points/lines, one can easily extract the feature-aligned triangle mesh from the SDF and then decompose it into smooth surface patches, greatly reducing the difficulty of recovering the parametric CAD design. A comprehensive comparison with existing state-of-the-art methods shows the significant advantage of our approach in reconstructing faithful CAD shapes.

link: http://arxiv.org/abs/2404.13420v1

MultiConfederated Learning: Inclusive Non-IID Data handling with Decentralized Federated Learning

Michael Duchesne, Kaiwen Zhang, Chamseddine Talhi

Federated Learning (FL) has emerged as a prominent privacy-preserving technique for enabling use cases like confidential clinical machine learning. FL operates by aggregating models trained by remote devices which owns the data. Thus, FL enables the training of powerful global models using crowd-sourced data from a large number of learners, without compromising their privacy. However, the aggregating server is a single point of failure when generating the global model. Moreover, the performance of the model suffers when the data is not independent and identically distributed (non-IID data) on all remote devices. This leads to vastly different models being aggregated, which can reduce the performance by as much as 50% in certain scenarios. In this paper, we seek to address the aforementioned issues while retaining the benefits of FL. We propose MultiConfederated Learning: a decentralized FL framework which is designed to handle non-IID data. Unlike traditional FL, MultiConfederated Learning will maintain multiple models in parallel (instead of a single global model) to help with convergence when the data is non-IID. With the help of transfer learning, learners can converge to fewer models. In order to increase adaptability, learners are allowed to choose which updates to aggregate from their peers.

link: http://dx.doi.org/10.1145/3605098.3636000

PIPER: Primitive-Informed Preference-based Hierarchical Reinforcement Learning via Hindsight Relabeling

Utsav Singh, Wesley A. Suttle, Brian M. Sadler, Vinay P. Namboodiri, Amrit Singh Bedi

In this work, we introduce PIPER: Primitive-Informed Preference-based Hierarchical reinforcement learning via Hindsight Relabeling, a novel approach that leverages preference-based learning to learn a reward model, and subsequently uses this reward model to relabel higher-level replay buffers. Since this reward is unaffected by lower primitive behavior, our relabeling-based approach is able to mitigate non-stationarity, which is common in existing hierarchical approaches, and demonstrates impressive performance across a range of challenging sparse-reward tasks. Since obtaining human feedback is typically impractical, we propose to replace the human-in-the-loop approach with our primitive-in-the-loop approach, which generates feedback using sparse rewards provided by the environment. Moreover, in order to prevent infeasible subgoal prediction and avoid degenerate solutions, we propose primitive-informed regularization that conditions higher-level policies to generate feasible subgoals for lower-level policies. We perform extensive experiments to show that PIPER mitigates non-stationarity in hierarchical reinforcement learning and achieves greater than 50$\%$ success rates in challenging, sparse-reward robotic environments, where most other baselines fail to achieve any significant progress.

link: http://arxiv.org/abs/2404.13423v1

AdvLoRA: Adversarial Low-Rank Adaptation of Vision-Language Models

Yuheng Ji, Yue Liu, Zhicheng Zhang, Zhao Zhang, Yuting Zhao, Gang Zhou, Xingwei Zhang, Xinwang Liu, Xiaolong Zheng

Vision-Language Models (VLMs) are a significant technique for Artificial General Intelligence (AGI). With the fast growth of AGI, the security problem become one of the most important challenges for VLMs. In this paper, through extensive experiments, we demonstrate the vulnerability of the conventional adaptation methods for VLMs, which may bring significant security risks. In addition, as the size of the VLMs increases, performing conventional adversarial adaptation techniques on VLMs results in high computational costs. To solve these problems, we propose a parameter-efficient \underline{Adv}ersarial adaptation method named \underline{AdvLoRA} by \underline{Lo}w-\underline{R}ank \underline{A}daptation. At first, we investigate and reveal the intrinsic low-rank property during the adversarial adaptation for VLMs. Different from LoRA, we improve the efficiency and robustness of adversarial adaptation by designing a novel reparameterizing method based on parameter clustering and parameter alignment. In addition, an adaptive parameter update strategy is proposed to further improve the robustness. By these settings, our proposed AdvLoRA alleviates the model security and high resource waste problems. Extensive experiments demonstrate the effectiveness and efficiency of the AdvLoRA.

link: http://arxiv.org/abs/2404.13425v1

Text-dependent Speaker Verification (TdSV) Challenge 2024: Challenge Evaluation Plan

Zeinali Hossein, Lee Kong Aik, Alam Jahangir, Burget Lukas

This document outlines the Text-dependent Speaker Verification (TdSV) Challenge 2024, which centers on analyzing and exploring novel approaches for text-dependent speaker verification. The primary goal of this challenge is to motive participants to develop single yet competitive systems, conduct thorough analyses, and explore innovative concepts such as multi-task learning, self-supervised learning, few-shot learning, and others, for text-dependent speaker verification.

link: http://arxiv.org/abs/2404.13428v1

Nested-TNT: Hierarchical Vision Transformers with Multi-Scale Feature Processing

Yuang Liu, Zhiheng Qiu, Xiaokai Qin

Transformer has been applied in the field of computer vision due to its excellent performance in natural language processing, surpassing traditional convolutional neural networks and achieving new state-of-the-art. ViT divides an image into several local patches, known as "visual sentences". However, the information contained in the image is vast and complex, and focusing only on the features at the "visual sentence" level is not enough. The features between local patches should also be taken into consideration. In order to achieve further improvement, the TNT model is proposed, whose algorithm further divides the image into smaller patches, namely "visual words," achieving more accurate results. The core of Transformer is the Multi-Head Attention mechanism, and traditional attention mechanisms ignore interactions across different attention heads. In order to reduce redundancy and improve utilization, we introduce the nested algorithm and apply the Nested-TNT to image classification tasks. The experiment confirms that the proposed model has achieved better classification performance over ViT and TNT, exceeding 2.25%, 1.1% on dataset CIFAR10 and 2.78%, 0.25% on dataset FLOWERS102 respectively.

link: http://arxiv.org/abs/2404.13434v1

High-fidelity Endoscopic Image Synthesis by Utilizing Depth-guided Neural Surfaces

Baoru Huang, Yida Wang, Anh Nguyen, Daniel Elson, Francisco Vasconcelos, Danail Stoyanov

In surgical oncology, screening colonoscopy plays a pivotal role in providing diagnostic assistance, such as biopsy, and facilitating surgical navigation, particularly in polyp detection.
Computer-assisted endoscopic surgery has recently gained attention and amalgamated various 3D computer vision techniques, including camera localization, depth estimation, surface reconstruction,

etc. Neural Radiance Fields (NeRFs) and Neural Implicit Surfaces (NeuS) have emerged as promising methodologies for deriving accurate 3D surface models from sets of registered images, addressing the limitations of existing colon reconstruction approaches stemming from constrained camera movement. However, the inadequate tissue texture representation and confused scale problem in monocular colonoscopic image reconstruction still impede the progress of the final rendering results. In this paper, we introduce a novel method for colon section reconstruction by leveraging NeuS applied to endoscopic images, supplemented by a single frame of depth map. Notably, we pioneered the exploration of utilizing only one frame depth map in photorealistic reconstruction and neural rendering applications while this single depth map can be easily obtainable from other monocular depth estimation networks with an object scale. Through rigorous experimentation and validation on phantom imagery, our approach demonstrates exceptional accuracy in completely rendering colon sections, even capturing unseen portions of the surface. This breakthrough opens avenues for achieving stable and consistently scaled reconstructions, promising enhanced quality in cancer screening procedures and treatment interventions.

link: http://arxiv.org/abs/2404.13437v1

Fine-Grained Named Entities for Corona News

Sefika Efeoglu, Adrian Paschke

Information resources such as newspapers have produced unstructured text data in various languages related to the corona outbreak since December 2019. Analyzing these unstructured texts is time-consuming without representing them in a structured format; therefore, representing them in a structured format is crucial. An information extraction pipeline with essential tasks -- named entity tagging and relation extraction -- to accomplish this goal might be applied to these texts. This study proposes a data annotation pipeline to generate training data from corona news articles, including generic and domain-specific entities. Named entity recognition models are trained on this annotated corpus and then evaluated on test sentences manually annotated by domain experts evaluating the performance of a trained model. The code base and demonstration are available at https://github.com/sefeoglu/coronanews-ner.git.

link: http://arxiv.org/abs/2404.13439v1

FisheyeDetNet: Object Detection on Fisheye Surround View Camera Systems for Automated Driving

Ganesh Sistu, Senthil Yogamani

Object detection is a mature problem in autonomous driving with pedestrian detection being one of the first deployed algorithms. It has been comprehensively studied in the literature. However, object detection is relatively less explored for fisheye cameras used for surround-view near field sensing. The standard bounding box representation fails in fisheye cameras due to heavy radial distortion, particularly in the periphery. To mitigate this, we explore extending the standard object detection output representation of bounding box. We design rotated bounding boxes, ellipse, generic polygon as polar arc/angle representations and define an instance segmentation mIOU metric to analyze these representations. The proposed model FisheyeDetNet with polygon outperforms others and achieves a mAP score of 49.5 % on Valeo fisheye surround-view dataset for automated driving applications. This dataset has 60K images captured from 4 surround-view cameras across Europe, North America and Asia. To the best of our knowledge, this is the first detailed study on object detection on fisheye cameras for autonomous driving scenarios.

link: http://arxiv.org/abs/2404.13443v1