Wed 2024.03.27

EgoExoLearn: A Dataset for Bridging Asynchronous Ego- and Exo-centric View of Procedural Activities in Real World

Yifei Huang, Guo Chen, Jilan Xu, Mingfang Zhang, Lijin Yang, Baoqi Pei, Hongjie Zhang, Lu Dong, Yali Wang, Limin Wang, Yu Qiao

Being able to map the activities of others into one's own point of view is one fundamental human skill even from a very early age. Taking a step toward understanding this human ability, we introduce EgoExoLearn, a large-scale dataset that emulates the human demonstration following process, in which individuals record egocentric videos as they execute tasks guided by demonstration videos. Focusing on the potential applications in daily assistance and professional support, EgoExoLearn contains egocentric and demonstration video data spanning 120 hours captured in daily life scenarios and specialized laboratories. Along with the videos we record high-quality gaze data and provide detailed multimodal annotations, formulating a playground for modeling the human ability to bridge asynchronous procedural actions from different viewpoints. To this end, we present benchmarks such as cross-view association, cross-view action planning, and cross-view referenced skill assessment, along with detailed analysis. We expect EgoExoLearn can serve as an important resource for bridging the actions across views, thus paving the way for creating AI agents capable of seamlessly learning by observing humans in the real world. Code and data can be found at: https://github.com/OpenGVLab/EgoExoLearn

link: http://arxiv.org/abs/2403.16182v1

Improving Scene Graph Generation with Relation Words' Debiasing in Vision-Language Models

Yuxuan Wang, Xiaoyuan Liu

Scene Graph Generation (SGG) provides basic language representation of visual scenes, requiring models to grasp complex and diverse semantics between various objects. However, this complexity and diversity in SGG also leads to underrepresentation, where part of test triplets are rare or even unseen during training, resulting in imprecise predictions. To tackle this, we propose using the SGG models with pretrained vision-language models (VLMs) to enhance representation. However, due to the gap between the pretraining and SGG, directly ensembling the pretrained VLMs leads to severe biases across relation words. Thus, we introduce LM Estimation to approximate the words' distribution underlies in the pretraining language sets, and then use the distribution for debiasing. After that, we ensemble VLMs with SGG models to enhance representation. Considering that each model may represent better at different samples, we use a certainty-aware indicator to score each sample and dynamically adjust the ensemble weights. Our method effectively addresses the words biases, enhances SGG's representation, and achieve markable performance enhancements. It is training-free and integrates well with existing SGG models.

link: http://arxiv.org/abs/2403.16184v1

ALoRA: Allocating Low-Rank Adaptation for Fine-tuning Large Language Models

Zequan Liu, Jiawen Lyn, Wei Zhu, Xing Tian, Yvette Graham

Parameter-efficient fine-tuning (PEFT) is widely studied for its effectiveness and efficiency in the era of large language models. Low-rank adaptation (LoRA) has demonstrated commendable performance as a popular and representative method. However, it is implemented with a fixed intrinsic rank that might not be the ideal setting for the downstream tasks. Recognizing the need for more flexible downstream task adaptation, we extend the methodology of LoRA to an innovative approach we call allocating low-rank adaptation (ALoRA) that enables dynamic adjustments to the intrinsic rank during the adaptation process. First, we propose a novel method, AB-LoRA, that can effectively estimate the importance score of each LoRA rank. Second, guided by AB-LoRA, we gradually prune abundant and negatively impacting LoRA ranks and allocate the pruned LoRA budgets to important Transformer modules needing higher ranks. We have conducted experiments on various tasks, and the experimental results demonstrate that our ALoRA method can outperform the recent baselines with comparable tunable parameters.

link: http://arxiv.org/abs/2403.16187v1

Cross-domain Multi-modal Few-shot Object Detection via Rich Text

Zeyu Shangguan, Daniel Seita, Mohammad Rostami

Cross-modal feature extraction and integration have led to steady performance improvements in few-shot learning tasks due to generating richer features. However, existing multi-modal object detection (MM-OD) methods degrade when facing significant domain-shift and are sample insufficient. We hypothesize that rich text information could more effectively help the model to build a knowledge relationship between the vision instance and its language description and can help mitigate domain shift. Specifically, we study the Cross-Domain few-shot generalization of MM-OD (CDMM-FSOD) and propose a meta-learning based multi-modal few-shot object detection method that utilizes rich text semantic information as an auxiliary modality to achieve domain adaptation in the context of FSOD. Our proposed network contains (i) a multi-modal feature aggregation module that aligns the vision and language support feature embeddings and (ii) a rich text semantic rectify module that utilizes bidirectional text feature generation to reinforce multi-modal feature alignment and thus to enhance the model's language understanding capability. We evaluate our model on common standard cross-domain object detection datasets and demonstrate that our approach considerably outperforms existing FSOD methods.

link: http://arxiv.org/abs/2403.16188v1

Interference Management for Integrated Sensing and Communication Systems: A Survey

Yangyang Niu, Zhiqing Wei, Lin Wang, Huici Wu, Zhiyong Feng

Emerging applications such as autonomous driving and Internet of things (IoT) services put forward the demand for simutaneous sensing and communication functions in the same system. Integrated sensing and communication (ISAC) has the potential to meet the demands of ubiquitous communication and high-precision sensing due to the advantages of spectrum and hardware resource sharing, as well as the mutual enhancement of sensing and communication. However, ISAC system faces severe interference requiring effective interference suppression, avoidance, and exploitation techniques. This article provides a comprehensive survey on the interference management techniques in ISAC systems, involving network architecture, system design, signal processing, and resource allocation. We first review the channel modeling and performance metrics of the ISAC system. Then, the methods for managing self-interference (SI), mutual interference (MI), and clutter in a single base station (BS) system are summarized, including interference suppression, interference avoidance and interference exploitation methods. Furthermore, cooperative interference management methods are studied to address the cross-link interference (CLI) in a coordinated multipoint ISAC (CoMP-ISAC) system. Finally, future trends are revealed. This article may provide a reference for the study of interference management in ISAC systems.

link: http://arxiv.org/abs/2403.16189v1

Logic-based Explanations for Linear Support Vector Classifiers with Reject Option

Francisco Mateus Rocha Filho, Thiago Alves Rocha, Reginaldo Pereira Fernandes Ribeiro, Ajalmar Rêgo da Rocha Neto

Support Vector Classifier (SVC) is a well-known Machine Learning (ML) model for linear classification problems. It can be used in conjunction with a reject option strategy to reject instances that are hard to correctly classify and delegate them to a specialist. This further increases the confidence of the model. Given this, obtaining an explanation of the cause of rejection is important to not blindly trust the obtained results. While most of the related work has developed means to give such explanations for machine learning models, to the best of our knowledge none have done so for when reject option is present. We propose a logic-based approach with formal guarantees on the correctness and minimality of explanations for linear SVCs with reject option. We evaluate our approach by comparing it to Anchors, which is a heuristic algorithm for generating explanations. Obtained results show that our proposed method gives shorter explanations with reduced time cost.

link: http://dx.doi.org/10.1007/978-3-031-45368-7_10

Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery

Siddharth Tourani, Ahmed Alwheibi, Arif Mahmood, Muhammad Haris Khan

Unsupervised landmarks discovery (ULD) for an object category is a challenging computer vision problem. In pursuit of developing a robust ULD framework, we explore the potential of a recent paradigm of self-supervised learning algorithms, known as diffusion models. Some recent works have shown that these models implicitly contain important correspondence cues. Towards harnessing the potential of diffusion models for the ULD task, we make the following core contributions. First, we propose a ZeroShot ULD baseline based on simple clustering of random pixel locations with nearest neighbour matching. It delivers better results than existing ULD methods. Second, motivated by the ZeroShot performance, we develop a ULD algorithm based on diffusion features using self-training and clustering which also outperforms prior methods by notable margins. Third, we introduce a new proxy task based on generating latent pose codes and also propose a two-stage clustering mechanism to facilitate effective pseudo-labeling, resulting in a significant performance improvement. Overall, our approach consistently outperforms state-of-the-art methods on four challenging benchmarks AFLW, MAFL, CatHeads and LS3D by significant margins.

link: http://arxiv.org/abs/2403.16194v1

Diffusion Model is a Good Pose Estimator from 3D RF-Vision

Junqiao Fan, Jianfei Yang, Yuecong Xu, Lihua Xie

Human pose estimation (HPE) from Radio Frequency vision (RF-vision) performs human sensing using RF signals that penetrate obstacles without revealing privacy (e.g., facial information). Recently, mmWave radar has emerged as a promising RF-vision sensor, providing radar point clouds by processing RF signals. However, the mmWave radar has a limited resolution with severe noise, leading to inaccurate and inconsistent human pose estimation. This work proposes mmDiff, a novel diffusion-based pose estimator tailored for noisy radar data. Our approach aims to provide reliable guidance as conditions to diffusion models. Two key challenges are addressed by mmDiff: (1) miss-detection of parts of human bodies, which is addressed by a module that isolates feature extraction from different body parts, and (2) signal inconsistency due to environmental interference, which is tackled by incorporating prior knowledge of body structure and motion. Several modules are designed to achieve these goals, whose features work as the conditions for the subsequent diffusion model, eliminating the miss-detection and instability of HPE based on RF-vision. Extensive experiments demonstrate that mmDiff outperforms existing methods significantly, achieving state-of-the-art performances on public datasets.

link: http://arxiv.org/abs/2403.16198v1

From Discrete to Continuous: Deep Fair Clustering With Transferable Representations

Xiang Zhang

We consider the problem of deep fair clustering, which partitions data into clusters via the representations extracted by deep neural networks while hiding sensitive data attributes. To achieve fairness, existing methods present a variety of fairness-related objective functions based on the group fairness criterion. However, these works typically assume that the sensitive attributes are discrete and do not work for continuous sensitive variables, such as the proportion of the female population in an area. Besides, the potential of the representations learned from clustering tasks to improve performance on other tasks is ignored by existing works. In light of these limitations, we propose a flexible deep fair clustering method that can handle discrete and continuous sensitive attributes simultaneously. Specifically, we design an information bottleneck style objective function to learn fair and clustering-friendly representations. Furthermore, we explore for the first time the transferability of the extracted representations to other downstream tasks. Unlike existing works, we impose fairness at the representation level, which could guarantee fairness for the transferred task regardless of clustering results. To verify the effectiveness of the proposed method, we perform extensive experiments on datasets with discrete and continuous sensitive attributes, demonstrating the advantage of our method in comparison with state-of-the-art methods.

link: http://arxiv.org/abs/2403.16201v1

FH-SSTNet: Forehead Creases based User Verification using Spatio-Spatial Temporal Network

Geetanjali Sharma, Gaurav Jaswal, Aditya Nigam, Raghavendra Ramachandra

Biometric authentication, which utilizes contactless features, such as forehead patterns, has become increasingly important for identity verification and access management. The proposed method is based on learning a 3D spatio-spatial temporal convolution to create detailed pictures of forehead patterns. We introduce a new CNN model called the Forehead Spatio-Spatial Temporal Network (FH-SSTNet), which utilizes a 3D CNN architecture with triplet loss to capture distinguishing features. We enhance the model's discrimination capability using Arcloss in the network's head. Experimentation on the Forehead Creases version 1 (FH-V1) dataset, containing 247 unique subjects, demonstrates the superior performance of FH-SSTNet compared to existing methods and pre-trained CNNs like ResNet50, especially for forehead-based user verification. The results demonstrate the superior performance of FH-SSTNet for forehead-based user verification, confirming its effectiveness in identity authentication.

link: http://arxiv.org/abs/2403.16202v1

SQL-Encoder: Improving NL2SQL In-Context Learning Through a Context-Aware Encoder

Mohammadreza Pourreza, Davood Rafiei, Yuxi Feng, Raymond Li, Zhenan Fan, Weiwei Zhang

Detecting structural similarity between queries is essential for selecting examples in in-context learning models. However, assessing structural similarity based solely on the natural language expressions of queries, without considering SQL queries, presents a significant challenge. This paper explores the significance of this similarity metric and proposes a model for accurately estimating it. To achieve this, we leverage a dataset comprising 170k question pairs, meticulously curated to train a similarity prediction model. Our comprehensive evaluation demonstrates that the proposed model adeptly captures the structural similarity between questions, as evidenced by improvements in Kendall-Tau distance and precision@k metrics. Notably, our model outperforms strong competitive embedding models from OpenAI and Cohere. Furthermore, compared to these competitive models, our proposed encoder enhances the downstream performance of NL2SQL models in 1-shot in-context learning scenarios by 1-2\% for GPT-3.5-turbo, 4-8\% for CodeLlama-7B, and 2-3\% for CodeLlama-13B.

link: http://arxiv.org/abs/2403.16204v1

Blur2Blur: Blur Conversion for Unsupervised Image Deblurring on Unknown Domains

Bang-Dang Pham, Phong Tran, Anh Tran, Cuong Pham, Rang Nguyen, Minh Hoai

This paper presents an innovative framework designed to train an image deblurring algorithm tailored to a specific camera device. This algorithm works by transforming a blurry input image, which is challenging to deblur, into another blurry image that is more amenable to deblurring. The transformation process, from one blurry state to another, leverages unpaired data consisting of sharp and blurry images captured by the target camera device. Learning this blur-to-blur transformation is inherently simpler than direct blur-to-sharp conversion, as it primarily involves modifying blur patterns rather than the intricate task of reconstructing fine image details. The efficacy of the proposed approach has been demonstrated through comprehensive experiments on various benchmarks, where it significantly outperforms state-of-the-art methods both quantitatively and qualitatively. Our code and data are available at https://zero1778.github.io/blur2blur/

link: http://arxiv.org/abs/2403.16205v1

Rumor Detection with a novel graph neural network approach

Tianrui Liu, Qi Cai, Changxin Xu, Bo Hong, Fanghao Ni, Yuxin Qiao, Tsungwei Yang

The wide spread of rumors on social media has caused a negative impact on people's daily life, leading to potential panic, fear, and mental health problems for the public. How to debunk rumors as early as possible remains a challenging problem. Existing studies mainly leverage information propagation structure to detect rumors, while very few works focus on correlation among users that they may coordinate to spread rumors in order to gain large popularity. In this paper, we propose a new detection model, that jointly learns both the representations of user correlation and information propagation to detect rumors on social media. Specifically, we leverage graph neural networks to learn the representations of user correlation from a bipartite graph that describes the correlations between users and source tweets, and the representations of information propagation with a tree structure. Then we combine the learned representations from these two modules to classify the rumors. Since malicious users intend to subvert our model after deployment, we further develop a greedy attack scheme to analyze the cost of three adversarial attacks: graph attack, comment attack, and joint attack. Evaluation results on two public datasets illustrate that the proposed MODEL outperforms the state-of-the-art rumor detection models. We also demonstrate our method performs well for early rumor detection. Moreover, the proposed detection method is more robust to adversarial attacks compared to the best existing method. Importantly, we show that it requires a high cost for attackers to subvert user correlation pattern, demonstrating the importance of considering user correlation for rumor detection.

link: http://arxiv.org/abs/2403.16206v2

Skull-to-Face: Anatomy-Guided 3D Facial Reconstruction and Editing

Yongqing Liang, Congyi Zhang, Junli Zhao, Wenping Wang, Xin Li

Deducing the 3D face from a skull is an essential but challenging task in forensic science and archaeology. Existing methods for automated facial reconstruction yield inaccurate results, suffering from the non-determinative nature of the problem that a skull with a sparse set of tissue depth cannot fully determine the skinned face. Additionally, their texture-less results require further post-processing stages to achieve a photo-realistic appearance. This paper proposes an end-to-end 3D face reconstruction and exploration tool, providing textured 3D faces for reference. With the help of state-of-the-art text-to-image diffusion models and image-based facial reconstruction techniques, we generate an initial reference 3D face, whose biological profile aligns with the given skull. We then adapt these initial faces to meet the statistical expectations of extruded anatomical landmarks on the skull through an optimization process. The joint statistical distribution of tissue depths is learned on a small set of anatomical landmarks on the skull. To support further adjustment, we propose an efficient face adaptation tool to assist users in tuning tissue depths, either globally or at local regions, while observing plausible visual feedback. Experiments conducted on a real skull-face dataset demonstrated the effectiveness of our proposed pipeline in terms of reconstruction accuracy, diversity, and stability.

link: http://arxiv.org/abs/2403.16207v1

Convergence analysis of OT-Flow for sample generation

Yang Jing, Lei Li

Deep generative models aim to learn the underlying distribution of data and generate new ones. Despite the diversity of generative models and their high-quality generation performance in practice, most of them lack rigorous theoretical convergence proofs. In this work, we aim to establish some convergence results for OT-Flow, one of the deep generative models. First, by reformulating the framework of OT-Flow model, we establish the Γ-convergence of the formulation of OT-flow to the corresponding optimal transport (OT) problem as the regularization term parameter α goes to infinity. Second, since the loss function will be approximated by

Monte Carlo method in training, we established the convergence between the discrete loss function and the continuous one when the sample number N goes to infinity as well. Meanwhile, the approximation capability of the neural network provides an upper bound for the discrete loss function of the minimizers. The proofs in both aspects provide convincing assurances for OT-Flow.

link: http://arxiv.org/abs/2403.16208v1