Sat 2024.03.02

CricaVPR: Cross-image Correlation-aware Representation Learning for Visual Place Recognition

Feng Lu, Xiangyuan Lan, Lijun Zhang, Dongmei Jiang, Yaowei Wang, Chun Yuan

Over the past decade, most methods in visual place recognition (VPR) have used neural networks to produce feature representations. These networks typically produce a global representation of a place image using only this image itself and neglect the cross-image variations (e.g. viewpoint and illumination), which limits their robustness in challenging scenes. In this paper, we propose a robust global representation method with cross-image correlation awareness for VPR, named CricaVPR. Our method uses the self-attention mechanism to correlate multiple images within a batch. These images can be taken in the same place with different conditions or viewpoints, or even captured from different places. Therefore, our method can utilize the cross-image variations as a cue to guide the representation learning, which ensures more robust features are produced. To further facilitate the robustness, we propose a multi-scale convolution-enhanced adaptation method to adapt pre-trained visual foundation models to the VPR task, which introduces the multi-scale local information to further enhance the cross-image correlation-aware representation. Experimental results show that our method outperforms state-of-the-art methods by a large margin with significantly less training time. Our method achieves 94.5% R@1 on Pitts30k using 512-dim global features. The code is released at https://github.com/Lu-Feng/CricaVPR.

link: http://arxiv.org/abs/2402.19231v1

Trained Random Forests Completely Reveal your Dataset

Julien Ferry, Ricardo Fukasawa, Timothée Pascal, Thibaut Vidal

We introduce an optimization-based reconstruction attack capable of completely or near-completely reconstructing a dataset utilized for training a random forest. Notably, our approach relies solely on information readily available in commonly used libraries such as scikit-learn. To achieve this, we formulate the reconstruction problem as a combinatorial problem under a maximum likelihood objective. We demonstrate that this problem is NP-hard, though solvable at scale using constraint programming -- an approach rooted in constraint propagation and solution-domain reduction. Through an extensive computational investigation, we demonstrate that random forests trained without bootstrap aggregation but with feature randomization are susceptible to a complete reconstruction. This holds true even with a small number of trees. Even with bootstrap aggregation, the majority of the data can also be reconstructed. These findings underscore a critical vulnerability inherent in widely adopted ensemble methods, warranting attention and mitigation. Although the potential for such reconstruction attacks has been discussed in privacy research, our study provides clear empirical evidence of their practicability.

link: http://arxiv.org/abs/2402.19232v1

Derivative-enhanced Deep Operator Network

Yuan Qiu, Nolan Bridges, Peng Chen

Deep operator networks (DeepONets), a class of neural operators that learn mappings between function spaces, have recently been developed as surrogate models for parametric partial differential equations (PDEs). In this work we propose a derivative-enhanced deep operator network (DE-DeepONet), which leverages the derivative information to enhance the prediction accuracy, and provide a more accurate approximation of the derivatives, especially when the training data are limited. DE-DeepONet incorporates dimension reduction of input into DeepONet and includes two types of derivative labels in the loss function for training, that is, the directional derivatives of the output function with respect to the input function and the gradient of the output function with respect to the physical domain variables. We test DE-DeepONet on three different equations with increasing complexity to demonstrate its effectiveness compared to the vanilla DeepONet.

link: http://arxiv.org/abs/2402.19242v1

Let LLMs Take on the Latest Challenges! A Chinese Dynamic Question Answering Benchmark

Zhikun Xu, Yinghui Li, Ruixue Ding, Xinyu Wang, Boli Chen, Yong Jiang, Xiaodong Deng, Jianxin Ma, Hai-Tao Zheng, Wenlian Lu, Pengjun Xie, Chang Zhou, Fei Huang

How to better evaluate the capabilities of Large Language Models (LLMs) is the focal point and hot topic in current LLMs research. Previous work has noted that due to the extremely high cost of iterative updates of LLMs, they are often unable to answer the latest dynamic questions well. To promote the improvement of Chinese LLMs' ability to answer dynamic questions, in this paper, we introduce CDQA, a Chinese Dynamic QA benchmark containing question-answer pairs related to the latest news on the Chinese Internet. We obtain high-quality data through a pipeline that combines humans and models, and carefully classify the samples according to the frequency of answer changes to facilitate a more fine-grained observation of LLMs' capabilities. We have also evaluated and analyzed mainstream and advanced Chinese LLMs on CDQA. Extensive experiments and valuable insights suggest that our proposed CDQA is challenging and worthy of more further study. We believe that the benchmark we provide will become the key data resource for improving LLMs' Chinese question-answering ability in the future.

link: http://arxiv.org/abs/2402.19248v1

Feature boosting with efficient attention for scene parsing

Vivek Singh, Shailza Sharma, Fabio Cuzzolin

The complexity of scene parsing grows with the number of object and scene classes, which is higher in unrestricted open scenes. The biggest challenge is to model the spatial relation between scene elements while succeeding in identifying objects at smaller scales. This paper presents a novel feature-boosting network that gathers spatial context from multiple levels of feature extraction and computes the attention weights for each level of representation to generate the final class labels. A novel `channel attention module' is designed to compute the attention weights, ensuring that features from the relevant extraction stages are boosted while the others are attenuated. The model also learns spatial context information at low resolution to preserve the abstract spatial relationships among scene elements and reduce computation cost. Spatial attention is subsequently concatenated into a final feature set before applying feature boosting. Low-resolution spatial attention features are trained using an auxiliary task that helps learning a coarse global scene structure. The proposed model outperforms all state-of-the-art models on both the ADE20K and the Cityscapes datasets.

link: http://arxiv.org/abs/2402.19250v1

A Cognitive-Based Trajectory Prediction Approach for Autonomous Driving

Haicheng Liao, Yongkang Li, Zhenning Li, Chengyue Wang, Zhiyong Cui, Shengbo Eben Li, Chengzhong Xu

In autonomous vehicle (AV) technology, the ability to accurately predict the movements of surrounding vehicles is paramount for ensuring safety and operational efficiency. Incorporating human decision-making insights enables AVs to more effectively anticipate the potential actions of other vehicles, significantly improving prediction accuracy and responsiveness in dynamic environments. This paper introduces the Human-Like Trajectory Prediction (HLTP) model, which adopts a teacher-student knowledge distillation framework inspired by human cognitive processes. The HLTP model incorporates a sophisticated teacher-student knowledge distillation framework. The "teacher" model, equipped with an adaptive visual sector, mimics the visual processing of the human brain, particularly the functions of the occipital and temporal lobes. The "student" model focuses on real-time interaction and decision-making, drawing parallels to prefrontal and parietal cortex functions. This approach allows for dynamic adaptation to changing driving scenarios, capturing essential perceptual cues for accurate prediction. Evaluated using the Macao Connected and Autonomous Driving (MoCAD) dataset, along with the NGSIM and HighD benchmarks, HLTP

demonstrates superior performance compared to existing models, particularly in challenging environments with incomplete data. The project page is available at Github.

link: http://arxiv.org/abs/2402.19251v1

Machine learning for modular multiplication

Kristin Lauter, Cathy Yuanchen Li, Krystal Maughan, Rachel Newton, Megha Srivastava

Motivated by cryptographic applications, we investigate two machine learning approaches to modular multiplication: namely circular regression and a sequence-to-sequence transformer model. The limited success of both methods demonstrated in our results gives evidence for the hardness of tasks involving modular multiplication upon which cryptosystems are based.

link: http://arxiv.org/abs/2402.19254v1

GSM-Plus: A Comprehensive Benchmark for Evaluating the Robustness of LLMs as Mathematical Problem Solvers

Qintong Li, Leyang Cui, Xueliang Zhao, Lingpeng Kong, Wei Bi

Large language models (LLMs) have achieved impressive performance across various mathematical reasoning benchmarks. However, there are increasing debates regarding whether these models truly understand and apply mathematical knowledge or merely rely on shortcuts for mathematical reasoning. One essential and frequently occurring evidence is that when the math questions are slightly changed, LLMs can behave incorrectly. This motivates us to evaluate the robustness of LLMs' math reasoning capability by testing a wide range of question variations. We introduce the adversarial grade school math (\datasetname) dataset, an extension of GSM8K augmented with various mathematical perturbations. Our experiments on 25 LLMs and 4 prompting techniques show that while LLMs exhibit different levels of math reasoning abilities, their performances are far from robust. In particular, even for problems that have been solved in GSM8K, LLMs can make mistakes when new statements are added or the question targets are altered. We also explore whether more robust performance can be achieved by composing existing prompting methods, in which we try an iterative method that generates and verifies each intermediate thought based on its reasoning goal and calculation result. Code and data are available at \url{https://github.com/qtli/GSM-Plus}.

link: http://arxiv.org/abs/2402.19255v1

MaskFi: Unsupervised Learning of WiFi and Vision Representations for Multimodal Human Activity Recognition

Jianfei Yang, Shijie Tang, Yuecong Xu, Yunjiao Zhou, Lihua Xie

Human activity recognition (HAR) has been playing an increasingly important role in various domains such as healthcare, security monitoring, and metaverse gaming. Though numerous HAR methods based on computer vision have been developed to show prominent performance, they still suffer from poor robustness in adverse visual conditions in particular low illumination, which motivates WiFi-based HAR to serve as a good complementary modality. Existing solutions using WiFi and vision modalities rely on massive labeled data that are very cumbersome to collect. In this paper, we propose a novel unsupervised multimodal HAR solution, MaskFi, that leverages only unlabeled video and WiFi activity data for model training. We propose a new algorithm, masked WiFi-vision modeling (MI2M), that enables the model to learn cross-modal and single-modal features by predicting the masked sections in representation learning. Benefiting from our unsupervised learning procedure, the network requires only a small amount of annotated data for finetuning and can adapt to the new environment with better performance. We conduct extensive experiments on two WiFi-vision datasets collected in-house, and our method achieves human activity recognition and human identification in terms of both robustness and accuracy.

link: http://arxiv.org/abs/2402.19258v1

Masks, Signs, And Learning Rate Rewinding

Advait Gadhikar, Rebekka Burkholz

Learning Rate Rewinding (LRR) has been established as a strong variant of Iterative Magnitude Pruning (IMP) to find lottery tickets in deep overparameterized neural networks. While both iterative pruning schemes couple structure and parameter learning, understanding how LRR excels in both aspects can bring us closer to the design of more flexible deep learning algorithms that can optimize diverse sets of sparse architectures. To this end, we conduct experiments that disentangle the effect of mask learning and parameter optimization and how both benefit from overparameterization. The ability of LRR to flip parameter signs early and stay robust to sign perturbations seems to make it not only more effective in mask identification but also in optimizing diverse sets of masks, including random ones. In support of this hypothesis, we prove in a simplified single hidden neuron setting that LRR succeeds in more cases than IMP, as it can escape initially problematic sign configurations.

link: http://arxiv.org/abs/2402.19262v1

Spinal Osteophyte Detection via Robust Patch Extraction on minimally annotated X-rays

Soumya Snigdha Kundu, Yuanhan Mo, Nicharee Srikijkasemwat, Bart■omiej W. Papiez

The development and progression of arthritis is strongly associated with osteophytes, which are small and elusive bone growths. This paper presents one of the first efforts towards automated spinal osteophyte detection in spinal X-rays. A novel automated patch extraction process, called SegPatch, has been proposed based on deep learning-driven vertebrae segmentation and the enlargement of mask contours. A final patch classification accuracy of 84.5\% is secured, surpassing a baseline tiling-based patch generation technique by 9.5%. This demonstrates that even with limited annotations, SegPatch can deliver superior performance for detection of tiny structures such as osteophytes. The proposed approach has potential to assist clinicians in expediting the process of manually identifying osteophytes in spinal X-ray.

link: http://arxiv.org/abs/2402.19263v1

T3DNet: Compressing Point Cloud Models for Lightweight 3D Recognition

Zhiyuan Yang, Yunjiao Zhou, Lihua Xie, Jianfei Yang

3D point cloud has been widely used in many mobile application scenarios, including autonomous driving and 3D sensing on mobile devices. However, existing 3D point cloud models tend to be large and cumbersome, making them hard to deploy on edged devices due to their high memory requirements and non-real-time latency. There has been a lack of research on how to compress 3D point cloud models into lightweight models. In this paper, we propose a method called T3DNet (Tiny 3D Network with augmEntation and disTillation) to address this issue. We find that the tiny model after network augmentation is much easier for a teacher to distill. Instead of gradually reducing the parameters through techniques such as pruning or quantization, we pre-define a tiny model and improve its performance through auxiliary supervision from augmented networks and the original model. We evaluate our method on several public datasets, including ModelNet40, ShapeNet, and ScanObjectNN. Our method can achieve high compression rates without significant accuracy sacrifice, achieving state-of-the-art performances on three datasets against existing methods. Amazingly, our T3DNet is 58 times smaller and 54 times faster than the original model yet with only 1.4% accuracy descent on the ModelNet40 dataset.

link: http://arxiv.org/abs/2402.19264v1

Learning Logic Specifications for Policy Guidance in POMDPs: an Inductive Logic Programming Approach

Daniele Meli, Alberto Castellini, Alessandro Farinelli

Partially Observable Markov Decision Processes (POMDPs) are a powerful framework for planning under uncertainty. They allow to model state uncertainty as a belief probability distribution. Approximate solvers based on Monte Carlo sampling show great success to relax the

computational demand and perform online planning. However, scaling to complex realistic domains with many actions and long planning horizons is still a major challenge, and a key point to achieve good performance is guiding the action-selection process with domain-dependent policy heuristics which are tailored for the specific application domain. We propose to learn high-quality heuristics from POMDP traces of executions generated by any solver. We convert the belief-action pairs to a logical semantics, and exploit data- and time-efficient Inductive Logic Programming (ILP) to generate interpretable belief-based policy specifications, which are then used as online heuristics. We evaluate thoroughly our methodology on two notoriously challenging POMDP problems, involving large action spaces and long planning horizons, namely, rocksample and pocman. Considering different state-of-the-art online POMDP solvers, including POMCP, DESPOT and AdaOPS, we show that learned heuristics expressed in Answer Set Programming (ASP) yield performance superior to neural networks and similar to optimal handcrafted task-specific heuristics within lower computational time. Moreover, they well generalize to more challenging scenarios not experienced in the training phase (e.g., increasing rocks and grid size in rocksample, incrementing the size of the map and the aggressivity of ghosts in pocman).

link: http://dx.doi.org/10.1613/jair.1.15826

Robust Guidance for Unsupervised Data Selection: Capturing Perplexing Named Entities for Domain-Specific Machine Translation

Seunghyun Ji, Hagai Raja Sinulingga, Darongsae Kwon

Employing extensive datasets enables the training of multilingual machine translation models; however, these models often fail to accurately translate sentences within specialized domains. Although obtaining and translating domain-specific data incurs high costs, it is inevitable for high-quality translations. Hence, finding the most 'effective' data with an unsupervised setting becomes a practical strategy for reducing labeling costs. Recent research indicates that this effective data could be found by selecting 'properly difficult data' based on its volume. This means the data should not be excessively challenging or overly simplistic, especially if the amount of data is limited. However, we found that establishing a criterion for unsupervised data selection remains challenging, as the 'proper difficulty' might vary based on the data domain being trained on. We introduce a novel unsupervised data selection method, 'Capturing Perplexing Named Entities', which adopts the maximum inference entropy in translated named entities as a selection measure. The motivation was that named entities in domain-specific data are considered the most complex portion of the data and should be predicted with high confidence. When verified with the 'Korean-English Parallel Corpus of Specialized Domains,' our method served as a robust guidance for unsupervised data selection, in contrast to existing methods.

link: http://arxiv.org/abs/2402.19267v1

Learning Intra-view and Cross-view Geometric Knowledge for Stereo Matching

Rui Gong, Weide Liu, Zaiwang Gu, Xulei Yang, Jun Cheng

Geometric knowledge has been shown to be beneficial for the stereo matching task. However, prior attempts to integrate geometric insights into stereo matching algorithms have largely focused on geometric knowledge from single images while crucial cross-view factors such as occlusion and matching uniqueness have been overlooked. To address this gap, we propose a novel Intra-view and Cross-view Geometric knowledge learning Network (ICGNet), specifically crafted to assimilate both intra-view and cross-view geometric knowledge. ICGNet harnesses the power of interest points to serve as a channel for intra-view geometric understanding. Simultaneously, it employs the correspondences among these points to capture cross-view geometric relationships. This dual incorporation empowers the proposed ICGNet to leverage both intra-view and cross-view geometric knowledge in its learning process, substantially improving its ability to estimate disparities. Our extensive experiments demonstrate the superiority of the ICGNet over contemporary leading models.

link: http://arxiv.org/abs/2402.19270v1