Sun 2024.03.31

SG-PGM: Partial Graph Matching Network with Semantic Geometric Fusion for 3D Scene Graph Alignment and Its Downstream Tasks

Yaxu Xie, Alain Pagani, Didier Stricker

Scene graphs have been recently introduced into 3D spatial understanding as a comprehensive representation of the scene. The alignment between 3D scene graphs is the first step of many downstream tasks such as scene graph aided point cloud registration, mosaicking, overlap checking, and robot navigation. In this work, we treat 3D scene graph alignment as a partial graph-matching problem and propose to solve it with a graph neural network. We reuse the geometric features learned by a point cloud registration method and associate the clustered point-level geometric features with the node-level semantic feature via our designed feature fusion module. Partial matching is enabled by using a learnable method to select the top-k similar node pairs. Subsequent downstream tasks such as point cloud registration are achieved by running a pre-trained registration network within the matched regions. We further propose a point-matching rescoring method, that uses the node-wise alignment of the 3D scene graph to reweight the matching candidates from a pre-trained point cloud registration method. It reduces the false point correspondences estimated especially in low-overlapping cases. Experiments show that our method improves the alignment accuracy by 10~20% in low-overlap and random transformation scenarios and outperforms the existing work in multiple downstream tasks.

link: http://arxiv.org/abs/2403.19474v1

H-Consistency Guarantees for Regression

Anqi Mao, Mehryar Mohri, Yutao Zhong

We present a detailed study of H-consistency bounds for regression. We first present new theorems that generalize the tools previously given to establish H-consistency bounds. This generalization proves essential for analyzing H-consistency bounds specific to regression. Next, we prove a series of novel H-consistency bounds for surrogate loss functions of the squared loss, under the assumption of a symmetric distribution and a bounded hypothesis set. This includes positive results for the Huber loss, all ℓ_p losses, $p \geq 1$, the squared ϵ-insensitive loss, as well as a negative result for the ϵ-insensitive loss used in squared Support Vector Regression (SVR). We further leverage our analysis of H-consistency for regression and derive principled surrogate losses for adversarial regression (Section 5). This readily establishes novel algorithms for adversarial regression, for which we report favorable experimental results in Section 6.

link: http://arxiv.org/abs/2403.19480v1

Improved Genetic Algorithm Based on Greedy and Simulated Annealing Ideas for Vascular Robot Ordering Strategy

Zixi Wang, Yubo Huang, Changshuo Fan, Xin Lai, Peng Lu

This study presents a comprehensive approach for optimizing the acquisition, utilization, and maintenance of ABLVR vascular robots in healthcare settings. Medical robotics, particularly in vascular treatments, necessitates precise resource allocation and optimization due to the complex nature of robot and operator maintenance. Traditional heuristic methods, though intuitive, often fail to achieve global optimization. To address these challenges, this research introduces a novel strategy, combining mathematical modeling, a hybrid genetic algorithm, and ARIMA time series forecasting. Considering the dynamic healthcare environment, our approach includes a robust resource allocation model for robotic vessels and operators. We incorporate the unique requirements of the adaptive learning process for operators and the maintenance needs of robotic components. The hybrid genetic algorithm, integrating simulated annealing and greedy approaches, efficiently solves the optimization problem. Additionally, ARIMA time series forecasting predicts the demand for vascular robots, further enhancing the adaptability of our strategy.

Experimental results demonstrate the superiority of our approach in terms of optimization, transparency, and convergence speed from other state-of-the-art methods.

link: http://arxiv.org/abs/2403.19484v1

Evolving Assembly Code in an Adversarial Environment

Irina Maliukov, Gera Weiss, Oded Margalit, Achiya Elyasaf

In this work, we evolve assembly code for the CodeGuru competition. The competition's goal is to create a survivor -- an assembly program that runs the longest in shared memory, by resisting attacks from adversary survivors and finding their weaknesses. For evolving top-notch solvers, we specify a Backus Normal Form (BNF) for the assembly language and synthesize the code from scratch using Genetic Programming (GP). We evaluate the survivors by running CodeGuru games against human-written winning survivors. Our evolved programs found weaknesses in the programs they were trained against and utilized them. In addition, we compare our approach with a Large-Language Model, demonstrating that the latter cannot generate a survivor that can win at any competition. This work has important applications for cyber-security, as we utilize evolution to detect weaknesses in survivors. The assembly BNF is domain-independent; thus, by modifying the fitness function, it can detect code weaknesses and help fix them. Finally, the CodeGuru competition offers a novel platform for analyzing GP and code evolution in adversarial environments. To support further research in this direction, we provide a thorough qualitative analysis of the evolved survivors and the weaknesses found.

link: http://arxiv.org/abs/2403.19489v1

Jointly Training and Pruning CNNs via Learnable Agent Guidance and Alignment

Alireza Ganjdanesh, Shangqian Gao, Heng Huang

Structural model pruning is a prominent approach used for reducing the computational cost of Convolutional Neural Networks (CNNs) before their deployment on resource-constrained devices. Yet, the majority of proposed ideas require a pretrained model before pruning, which is costly to secure. In this paper, we propose a novel structural pruning approach to jointly learn the weights and structurally prune architectures of CNN models. The core element of our method is a Reinforcement Learning (RL) agent whose actions determine the pruning ratios of the CNN model's layers, and the resulting model's accuracy serves as its reward. We conduct the joint training and pruning by iteratively training the model's weights and the agent's policy, and we regularize the model's weights to align with the selected structure by the agent. The evolving model's weights result in a dynamic reward function for the agent, which prevents using prominent episodic RL methods with stationary environment assumption for our purpose. We address this challenge by designing a mechanism to model the complex changing dynamics of the reward function and provide a representation of it to the RL agent. To do so, we take a learnable embedding for each training epoch and employ a recurrent model to calculate a representation of the changing environment. We train the recurrent model and embeddings using a decoder model to reconstruct observed rewards. Such a design empowers our agent to effectively leverage episodic observations along with the environment representations to learn a proper policy to determine performant sub-networks of the CNN model. Our extensive experiments on CIFAR-10 and ImageNet using ResNets and MobileNets demonstrate the effectiveness of our method.

link: http://arxiv.org/abs/2403.19490v1

Segmentation tool for images of cracks

Andrii Kompanets, Remco Duits, Davide Leonetti, Nicky van den Berg, H. H., Snijder

Safety-critical infrastructures, such as bridges, are periodically inspected to check for existing damage, such as fatigue cracks and corrosion, and to guarantee the safe use of the infrastructure. Visual inspection is the most frequent type of general inspection, despite the fact that its detection capability is rather limited, especially for fatigue cracks. Machine learning algorithms can be used for augmenting the capability of classical visual inspection of bridge structures, however, the implementation of such an algorithm requires a massive annotated training dataset, which is

time-consuming to produce. This paper proposes a semi-automatic crack segmentation tool that eases the manual segmentation of cracks on images needed to create a training dataset for a machine learning algorithm. Also, it can be used to measure the geometry of the crack. This tool makes use of an image processing algorithm, which was initially developed for the analysis of vascular systems on retinal images. The algorithm relies on a multi-orientation wavelet transform, which is applied to the image to construct the so-called "orientation scores", i.e. a modified version of the image. Afterwards, the filtered orientation scores are used to formulate an optimal path problem that identifies the crack. The globally optimal path between manually selected crack endpoints is computed, using a state-of-the-art geometric tracking method. The pixel-wise segmentation is done afterwards using the obtained crack path. The proposed method outperforms fully automatic methods and shows potential to be an adequate alternative to the manual data annotation.

link: http://arxiv.org/abs/2403.19492v1

Regression with Multi-Expert Deferral

Anqi Mao, Mehryar Mohri, Yutao Zhong

Learning to defer with multiple experts is a framework where the learner can choose to defer the prediction to several experts. While this problem has received significant attention in classification contexts, it presents unique challenges in regression due to the infinite and continuous nature of the label space. In this work, we introduce a novel framework of regression with deferral, which involves deferring the prediction to multiple experts. We present a comprehensive analysis for both the single-stage scenario, where there is simultaneous learning of predictor and deferral functions, and the two-stage scenario, which involves a pre-trained predictor with a learned deferral function. We introduce new surrogate loss functions for both scenarios and prove that they are supported by H-consistency bounds. These bounds provide consistency guarantees that are stronger than Bayes consistency, as they are non-asymptotic and hypothesis set-specific. Our framework is versatile, applying to multiple experts, accommodating any bounded regression losses, addressing both instance-dependent and label-dependent costs, and supporting both single-stage and two-stage methods. A by-product is that our single-stage formulation includes the recent regression with abstention framework (Cheng et al., 2023) as a special case, where only a single expert, the squared loss and a label-independent cost are considered. Minimizing our proposed loss functions directly leads to novel algorithms for regression with deferral. We report the results of extensive experiments showing the effectiveness of our proposed algorithms.

link: http://arxiv.org/abs/2403.19494v1

CoherentGS: Sparse Novel View Synthesis with Coherent 3D Gaussians

Avinash Paliwal, Wei Ye, Jinhui Xiong, Dmytro Kotovenko, Rakesh Ranjan, Vikas Chandra, Nima Khademi Kalantari

The field of 3D reconstruction from images has rapidly evolved in the past few years, first with the introduction of Neural Radiance Field (NeRF) and more recently with 3D Gaussian Splatting (3DGS). The latter provides a significant edge over NeRF in terms of the training and inference speed, as well as the reconstruction quality. Although 3DGS works well for dense input images, the unstructured point-cloud like representation quickly overfits to the more challenging setup of extremely sparse input images (e.g., 3 images), creating a representation that appears as a jumble of needles from novel views. To address this issue, we propose regularized optimization and depth-based initialization. Our key idea is to introduce a structured Gaussian representation that can be controlled in 2D image space. We then constraint the Gaussians, in particular their position, and prevent them from moving independently during optimization. Specifically, we introduce single and multiview constraints through an implicit convolutional decoder and a total variation loss, respectively. With the coherency introduced to the Gaussians, we further constrain the optimization through a flow-based loss function. To support our regularized optimization, we propose an approach to initialize the Gaussians using monocular depth estimates at each input view. We demonstrate significant improvements compared to the state-of-the-art sparse-view NeRF-based approaches on a variety of scenes.

link: http://arxiv.org/abs/2403.19495v1

Surface-based parcellation and vertex-wise analysis of ultra high-resolution ex vivo 7 tesla MRI in neurodegenerative diseases

Pulkit Khandelwal, Michael Tran Duong, Constanza Fuentes, Amanda Denning, Winifred Trotman, Ranjit Ittyerah, Alejandra Bahena, Theresa Schuck, Marianna Gabrielyan, Karthik Prabhakaran, Daniel Ohm, Gabor Mizsei, John Robinson, Monica Munoz, John Detre, Edward Lee, David Irwin, Corey McMillan, M. Dylan Tisdall, Sandhitsu Das, David Wolk, Paul A. Yushkevich

Magnetic resonance imaging (MRI) is the standard modality to understand human brain structure and function in vivo (antemortem). Decades of research in human neuroimaging has led to the widespread development of methods and tools to provide automated volume-based segmentations and surface-based parcellations which help localize brain functions to specialized anatomical regions. Recently ex vivo (postmortem) imaging of the brain has opened-up avenues to study brain structure at sub-millimeter ultra high-resolution revealing details not possible to observe with in vivo MRI. Unfortunately, there has been limited methodological development in ex vivo MRI primarily due to lack of datasets and limited centers with such imaging resources. Therefore, in this work, we present one-of-its-kind dataset of 82 ex vivo T2w whole brain hemispheres MRI at 0.3 mm isotropic resolution spanning Alzheimer's disease and related dementias. We adapted and developed a fast and easy-to-use automated surface-based pipeline to parcellate, for the first time, ultra high-resolution ex vivo brain tissue at the native subject space resolution using the Desikan-Killiany-Tourville (DKT) brain atlas. This allows us to perform vertex-wise analysis in the template space and thereby link morphometry measures with pathology measurements derived from histology. We will open-source our dataset docker container, Jupyter notebooks for ready-to-use out-of-the-box set of tools and command line options to advance ex vivo MRI clinical brain imaging research on the project webpage.

link: http://arxiv.org/abs/2403.19497v1

Client-supervised Federated Learning: Towards One-model-for-all Personalization

Peng Yan, Guodong Long

Personalized Federated Learning (PerFL) is a new machine learning paradigm that delivers personalized models for diverse clients under federated learning settings. Most PerFL methods require extra learning processes on a client to adapt a globally shared model to the client-specific personalized model using its own local data. However, the model adaptation process in PerFL is still an open challenge in the stage of model deployment and test time. This work tackles the challenge by proposing a novel federated learning framework to learn only one robust global model to achieve competitive performance to those personalized models on unseen/test clients in the FL system. Specifically, we design a new Client-Supervised Federated Learning (FedCS) to unravel clients' bias on instances' latent representations so that the global model can learn both client-specific and client-agnostic knowledge. Experimental study shows that the FedCS can learn a robust FL global model for the changing data distributions of unseen/test clients. The FedCS's global model can be directly deployed to the test clients while achieving comparable performance to other personalized FL methods that require model adaptation.

link: http://arxiv.org/abs/2403.19499v1

Tensor Network-Constrained Kernel Machines as Gaussian Processes

Frederiek Wesel, Kim Batselier

Tensor Networks (TNs) have recently been used to speed up kernel machines by constraining the model weights, yielding exponential computational and storage savings. In this paper we prove that the outputs of Canonical Polyadic Decomposition (CPD) and Tensor Train (TT)-constrained kernel machines recover a Gaussian Process (GP), which we fully characterize, when placing i.i.d. priors over their parameters. We analyze the convergence of both CPD and TT-constrained models, and show how TT yields models exhibiting more GP behavior compared to CPD, for the same number of model parameters. We empirically observe this behavior in two numerical experiments where we

respectively analyze the convergence to the GP and the performance at prediction. We thereby establish a connection between TN-constrained kernel machines and GPs.

link: http://arxiv.org/abs/2403.19500v1

RELI11D: A Comprehensive Multimodal Human Motion Dataset and Method

Ming Yan, Yan Zhang, Shuqiang Cai, Shuqi Fan, Xincheng Lin, Yudi Dai, Siqi Shen, Chenglu Wen, Lan Xu, Yuexin Ma, Cheng Wang

Comprehensive capturing of human motions requires both accurate captures of complex poses and precise localization of the human within scenes. Most of the HPE datasets and methods primarily rely on RGB, LiDAR, or IMU data. However, solely using these modalities or a combination of them may not be adequate for HPE, particularly for complex and fast movements. For holistic human motion understanding, we present RELI11D, a high-quality multimodal human motion dataset involves LiDAR, IMU system, RGB camera, and Event camera. It records the motions of 10 actors performing 5 sports in 7 scenes, including 3.32 hours of synchronized LiDAR point clouds, IMU measurement data, RGB videos and Event steams. Through extensive experiments, we demonstrate that the RELI11D presents considerable challenges and opportunities as it contains many rapid and complex motions that require precise location. To address the challenge of integrating different modalities, we propose LEIR, a multimodal baseline that effectively utilizes LiDAR Point Cloud, Event stream, and RGB through our cross-attention fusion strategy. We show that LEIR exhibits promising results for rapid motions and daily motions and that utilizing the characteristics of multiple modalities can indeed improve HPE performance. Both the dataset and source code will be released publicly to the research community, fostering collaboration and enabling further exploration in this field.

link: http://arxiv.org/abs/2403.19501v1

SineNet: Learning Temporal Dynamics in Time-Dependent Partial Differential Equations

Xuan Zhang, Jacob Helwig, Yuchao Lin, Yaochen Xie, Cong Fu, Stephan Wojtowytsch, Shuiwang Ji

We consider using deep neural networks to solve time-dependent partial differential equations (PDEs), where multi-scale processing is crucial for modeling complex, time-evolving dynamics. While the U-Net architecture with skip connections is commonly used by prior studies to enable multi-scale processing, our analysis shows that the need for features to evolve across layers results in temporally misaligned features in skip connections, which limits the model's performance. To address this limitation, we propose SineNet, consisting of multiple sequentially connected U-shaped network blocks, referred to as waves. In SineNet, high-resolution features are evolved progressively through multiple stages, thereby reducing the amount of misalignment within each stage. We furthermore analyze the role of skip connections in enabling both parallel and sequential processing of multi-scale information. Our method is rigorously tested on multiple PDE datasets, including the Navier-Stokes equations and shallow water equations, showcasing the advantages of our proposed approach over conventional U-Nets with a comparable parameter budget. We further demonstrate that increasing the number of waves in SineNet while maintaining the same number of parameters leads to a monotonically improved performance. The results highlight the effectiveness of SineNet and the potential of our approach in advancing the state-of-the-art in neural PDE solver design. Our code is available as part of AIRS (https://github.com/divelab/AIRS).

link: http://arxiv.org/abs/2403.19507v1

Debiasing Cardiac Imaging with Controlled Latent Diffusion Models

Grzegorz Skorupko, Richard Osuala, Zuzanna Szafranowska, Kaisar Kushibar, Nay Aung, Steffen E Petersen, Karim Lekadir, Polyxeni Gkontra

The progress in deep learning solutions for disease diagnosis and prognosis based on cardiac magnetic resonance imaging is hindered by highly imbalanced and biased training data. To address this issue, we propose a method to alleviate imbalances inherent in datasets through the generation

of synthetic data based on sensitive attributes such as sex, age, body mass index, and health condition. We adopt ControlNet based on a denoising diffusion probabilistic model to condition on text assembled from patient metadata and cardiac geometry derived from segmentation masks using a large-cohort study, specifically, the UK Biobank. We assess our method by evaluating the realism of the generated images using established quantitative metrics. Furthermore, we conduct a downstream classification task aimed at debiasing a classifier by rectifying imbalances within underrepresented groups through synthetically generated samples. Our experiments demonstrate the effectiveness of the proposed approach in mitigating dataset imbalances, such as the scarcity of younger patients or individuals with normal BMI level suffering from heart failure. This work represents a major step towards the adoption of synthetic data for the development of fair and generalizable models for medical classification tasks. Notably, we conduct all our experiments using a single, consumer-level GPU to highlight the feasibility of our approach within resource-constrained environments. Our code is available at https://github.com/faildeny/debiasing-cardiac-mri.

link: http://arxiv.org/abs/2403.19508v1

Phonetic Segmentation of the UCLA Phonetics Lab Archive

Eleanor Chodroff, Blaž Pažon, Annie Baker, Steven Moran

Research in speech technologies and comparative linguistics depends on access to diverse and accessible speech data. The UCLA Phonetics Lab Archive is one of the earliest multilingual speech corpora, with long-form audio recordings and phonetic transcriptions for 314 languages (Ladefoged et al., 2009). Recently, 95 of these languages were time-aligned with word-level phonetic transcriptions (Li et al., 2021). Here we present VoxAngeles, a corpus of audited phonetic transcriptions and phone-level alignments of the UCLA Phonetics Lab Archive, which uses the 95-language CMU re-release as our starting point. VoxAngeles also includes word- and phone-level segmentations from the original UCLA corpus, as well as phonetic measurements of word and phone durations, vowel formants, and vowel f0. This corpus enhances the usability of the original data, particularly for quantitative phonetic typology, as demonstrated through a case study of vowel intrinsic f0. We also discuss the utility of the VoxAngeles corpus for general research and pedagogy in crosslinguistic phonetics, as well as for low-resource and multilingual speech technologies. VoxAngeles is free to download and use under a CC-BY-NC 4.0 license.

link: http://arxiv.org/abs/2403.19509v1