Fri 2024.03.22

Parallel Gaussian process with kernel approximation in CUDA

Davide Carminati

This paper introduces a parallel implementation in CUDA/C++ of the Gaussian process with a decomposed kernel. This recent formulation, introduced by Joukov and Kuli\'c (2022), is characterized by an approximated -- but much smaller -- matrix to be inverted compared to plain Gaussian process. However, it exhibits a limitation when dealing with higher-dimensional samples which degrades execution times. The solution presented in this paper relies on parallelizing the computation of the predictive posterior statistics on a GPU using CUDA and its libraries. The CPU code and GPU code are then benchmarked on different CPU-GPU configurations to show the benefits of the parallel implementation on GPU over the CPU.

link: http://arxiv.org/abs/2403.12797v1

Investigating Text Shortening Strategy in BERT: Truncation vs Summarization

Mirza Alim Mutasodirin, Radityo Eko Prasojo

The parallelism of Transformer-based models comes at the cost of their input max-length. Some studies proposed methods to overcome this limitation, but none of them reported the effectiveness of summarization as an alternative. In this study, we investigate the performance of document truncation and summarization in text classification tasks. Each of the two was investigated with several variations. This study also investigated how close their performances are to the performance of full-text. We used a dataset of summarization tasks based on Indonesian news articles (IndoSum) to do classification tests. This study shows how the summaries outperform the majority of truncation method variations and lose to only one. The best strategy obtained in this study is taking the head of the document. The second is extractive summarization. This study explains what happened to the result, leading to further research in order to exploit the potential of document summarization as a shortening alternative. The code and data used in this work are publicly available in https://github.com/mirzaalimm/TruncationVsSummarization.

link: http://dx.doi.org/10.1109/ICACSIS53237.2021.9631364

Learning Neural Volumetric Pose Features for Camera Localization

Jingyu Lin, Jiaqi Gu, Bojian Wu, Lubin Fan, Renjie Chen, Ligang Liu, Jieping Ye

We introduce a novel neural volumetric pose feature, termed PoseMap, designed to enhance camera localization by encapsulating the information between images and the associated camera poses. Our framework leverages an Absolute Pose Regression (APR) architecture, together with an augmented NeRF module. This integration not only facilitates the generation of novel views to enrich the training dataset but also enables the learning of effective pose features. Additionally, we extend our architecture for self-supervised online alignment, allowing our method to be used and fine-tuned for unlabelled images within a unified framework. Experiments demonstrate that our method achieves 14.28% and 20.51% performance gain on average in indoor and outdoor benchmark scenes, outperforming existing APR methods with state-of-the-art accuracy.

link: http://arxiv.org/abs/2403.12800v1

RelationVLM: Making Large Vision-Language Models Understand Visual Relations

Zhipeng Huang, Zhizheng Zhang, Zheng-Jun Zha, Yan Lu, Baining Guo

The development of Large Vision-Language Models (LVLMs) is striving to catch up with the success of Large Language Models (LLMs), yet it faces more challenges to be resolved. Very recent works enable LVLMs to localize object-level visual contents and ground text to them. Nonetheless, current LVLMs still struggle to precisely understand visual relations due to the lack of relevant data. In this work, we present RelationVLM, a large vision-language model capable of

comprehending various levels and types of relations whether across multiple images or within a video. Specifically, we devise a multi-stage relation-aware training scheme and a series of corresponding data configuration strategies to bestow RelationVLM with the capabilities of understanding semantic relations, temporal associations and geometric transforms. Extensive case studies and quantitative evaluations show RelationVLM has strong capability in understanding such relations and emerges impressive in-context capability of reasoning from few-shot examples by comparison. This work fosters the advancements of LVLMs by enabling them to support a wider range of downstream applications toward artificial general intelligence.

link: http://arxiv.org/abs/2403.12801v1

DreamDA: Generative Data Augmentation with Diffusion Models

Yunxiang Fu, Chaoqi Chen, Yu Qiao, Yizhou Yu

The acquisition of large-scale, high-quality data is a resource-intensive and time-consuming endeavor. Compared to conventional Data Augmentation (DA) techniques (e.g. cropping and rotation), exploiting prevailing diffusion models for data generation has received scant attention in classification tasks. Existing generative DA methods either inadequately bridge the domain gap between real-world and synthesized images, or inherently suffer from a lack of diversity. To solve these issues, this paper proposes a new classification-oriented framework DreamDA, which enables data synthesis and label generation by way of diffusion models. DreamDA generates diverse samples that adhere to the original data distribution by considering training images in the original data as seeds and perturbing their reverse diffusion process. In addition, since the labels of the generated data may not align with the labels of their corresponding seed images, we introduce a self-training paradigm for generating pseudo labels and training classifiers using the synthesized data. Extensive experiments across four tasks and five datasets demonstrate consistent improvements over strong baselines, revealing the efficacy of DreamDA in synthesizing high-quality and diverse images with accurate labels. Our code will be available at https://github.com/yunxiangfu2001/DreamDA.

link: http://arxiv.org/abs/2403.12803v1

Contextual Moral Value Alignment Through Context-Based Aggregation

Pierre Dognin, Jesus Rios, Ronny Luss, Inkit Padhi, Matthew D Riemer, Miao Liu, Prasanna Sattigeri, Manish Nagireddy, Kush R. Varshney, Djallel Bouneffouf

Developing value-aligned AI agents is a complex undertaking and an ongoing challenge in the field of AI. Specifically within the domain of Large Language Models (LLMs), the capability to consolidate multiple independently trained dialogue agents, each aligned with a distinct moral value, into a unified system that can adapt to and be aligned with multiple moral values is of paramount importance. In this paper, we propose a system that does contextual moral value alignment based on contextual aggregation. Here, aggregation is defined as the process of integrating a subset of LLM responses that are best suited to respond to a user input, taking into account features extracted from the user's input. The proposed system shows better results in term of alignment to human value compared to the state of the art.

link: http://arxiv.org/abs/2403.12805v1

VisualCritic: Making LMMs Perceive Visual Quality Like Humans

Zhipeng Huang, Zhizheng Zhang, Yiting Lu, Zheng-Jun Zha, Zhibo Chen, Baining Guo

At present, large multimodal models (LMMs) have exhibited impressive generalization capabilities in understanding and generating visual signals. However, they currently still lack sufficient capability to perceive low-level visual quality akin to human perception. Can LMMs achieve this and show the same degree of generalization in this regard? If so, not only could the versatility of LMMs be further enhanced, but also the challenge of poor cross-dataset performance in the field of visual quality assessment could be addressed. In this paper, we explore this question and provide the answer "Yes!". As the result of this initial exploration, we present VisualCritic, the first LMM for broad-spectrum image subjective quality assessment. VisualCritic can be used across diverse data

right out of box, without any requirements of dataset-specific adaptation operations like conventional specialist models. As an instruction-following LMM, VisualCritic enables new capabilities of (1) quantitatively measuring the perceptual quality of given images in terms of their Mean Opinion Score (MOS), noisiness, colorfulness, sharpness, and other numerical indicators, (2) qualitatively evaluating visual quality and providing explainable descriptions, (3) discerning whether a given image is AI-generated or photographic. Extensive experiments demonstrate the efficacy of VisualCritic by comparing it with other open-source LMMs and conventional specialist models over both AI-generated and photographic images.

link: http://arxiv.org/abs/2403.12806v1

Comparing Explanation Faithfulness between Multilingual and Monolingual Fine-tuned Language Models

Zhixue Zhao, Nikolaos Aletras

In many real natural language processing application scenarios, practitioners not only aim to maximize predictive performance but also seek faithful explanations for the model predictions. Rationales and importance distribution given by feature attribution methods (FAs) provide insights into how different parts of the input contribute to a prediction. Previous studies have explored how different factors affect faithfulness, mainly in the context of monolingual English models. On the other hand, the differences in FA faithfulness between multilingual and monolingual models have yet to be explored. Our extensive experiments, covering five languages and five popular FAs, show that FA faithfulness varies between multilingual and monolingual models. We find that the larger the multilingual model, the less faithful the FAs are compared to its counterpart monolingual models.Our further analysis shows that the faithfulness disparity is potentially driven by the differences between model tokenizers. Our code is available: https://github.com/casszhao/multilingual-faith.

link: http://arxiv.org/abs/2403.12809v1

Re-identification from histopathology images

Jonathan Ganz, Jonas Ammeling, Samir Jabari, Katharina Breininger, Marc Aubreville

In numerous studies, deep learning algorithms have proven their potential for the analysis of histopathology images, for example, for revealing the subtypes of tumors or the primary origin of metastases. These models require large datasets for training, which must be anonymized to prevent possible patient identity leaks. This study demonstrates that even relatively simple deep learning algorithms can re-identify patients in large histopathology datasets with substantial accuracy. We evaluated our algorithms on two TCIA datasets including lung squamous cell carcinoma (LSCC) and lung adenocarcinoma (LUAD). We also demonstrate the algorithm's performance on an in-house dataset of meningioma tissue. We predicted the source patient of a slide with F1 scores of 50.16 % and 52.30 % on the LSCC and LUAD datasets, respectively, and with 62.31 % on our meningioma dataset. Based on our findings, we formulated a risk assessment scheme to estimate the risk to the patient's privacy prior to publication.

link: http://arxiv.org/abs/2403.12816v1

Dynamic Survival Analysis for Early Event Prediction

Hugo Yèche, Manuel Burger, Dinara Veshchezerova, Gunnar Rätsch

This study advances Early Event Prediction (EEP) in healthcare through Dynamic Survival Analysis (DSA), offering a novel approach by integrating risk localization into alarm policies to enhance clinical event metrics. By adapting and evaluating DSA models against traditional EEP benchmarks, our research demonstrates their ability to match EEP models on a time-step level and significantly improve event-level metrics through a new alarm prioritization scheme (up to 11% AuPRC difference). This approach represents a significant step forward in predictive healthcare, providing a more nuanced and actionable framework for early event prediction and management.

link: http://arxiv.org/abs/2403.12818v1

A Physics-embedded Deep Learning Framework for Cloth Simulation

Zhiwei Zhao

Delicate cloth simulations have long been desired in computer graphics. Various methods were proposed to improve engaged force interactions, collision handling, and numerical integrations. Deep learning has the potential to achieve fast and real-time simulation, but common neural network structures often demand many parameters to capture cloth dynamics. This paper proposes a physics-embedded learning framework that directly encodes physical features of cloth simulation. The convolutional neural network is used to represent spatial correlations of the mass-spring system, after which three branches are designed to learn linear, nonlinear, and time derivate features of cloth physics. The framework can also integrate with other external forces and collision handling through either traditional simulators or sub neural networks. The model is tested across different cloth animation cases, without training with new data. Agreement with baselines and predictive realism successfully validate its generalization ability. Inference efficiency of the proposed model also defeats traditional physics simulation. This framework is also designed to easily integrate with other visual refinement techniques like wrinkle carving, which leaves significant chances to incorporate prevailing macing learning techniques in 3D cloth amination.

link: http://arxiv.org/abs/2403.12820v1

FlowerFormer: Empowering Neural Architecture Encoding using a Flow-aware Graph Transformer

Dongyeong Hwang, Hyunju Kim, Sunwoo Kim, Kijung Shin

The success of a specific neural network architecture is closely tied to the dataset and task it tackles; there is no one-size-fits-all solution. Thus, considerable efforts have been made to quickly and accurately estimate the performances of neural architectures, without full training or evaluation, for given tasks and datasets. Neural architecture encoding has played a crucial role in the estimation, and graphbased methods, which treat an architecture as a graph, have shown prominent performance. For enhanced representation learning of neural architectures, we introduce FlowerFormer, a powerful graph transformer that incorporates the information flows within a neural architecture. FlowerFormer consists of two key components: (a) bidirectional asynchronous message passing, inspired by the flows; (b) global attention built on flow-based masking. Our extensive experiments demonstrate the superiority of FlowerFormer over existing neural encoding methods, and its effectiveness extends beyond computer vision models to include graph neural networks and auto speech recognition models. Our code is available at http://github.com/y0ngjaenius/CVPR2024_FLOWERFormer.

link: http://arxiv.org/abs/2403.12821v2

Answer Set Programming for Flexible Payroll Management

Benjamin Callewaert, Joost Vennekens

Payroll management is a critical business task that is subject to a large number of rules, which vary widely between companies, sectors, and countries. Moreover, the rules are often complex and change regularly. Therefore, payroll management systems must be flexible in design. In this paper, we suggest an approach based on a flexible Answer Set Programming (ASP) model and an easy-to-read tabular representation based on the Decision Model and Notation (DMN) standard. It allows HR consultants to represent complex rules without the need for a software engineer, and to ultimately design payroll systems for a variety of different scenarios. We show how the multi-shot solving capabilities of the clingo ASP system can be used to reach the performance that is necessary to handle real-world instances.

link: http://arxiv.org/abs/2403.12823v1

Has Approximate Machine Unlearning been evaluated properly? From Auditing to Side Effects

Cheng-Long Wang, Qi Li, Zihang Xiang, Di Wang

The growing concerns surrounding data privacy and security have underscored the critical necessity for machine unlearning--aimed at fully removing data lineage from machine learning models. MLaaS providers expect this to be their ultimate safeguard for regulatory compliance. Despite its critical importance, the pace at which privacy communities have been developing and implementing strong methods to verify the effectiveness of machine unlearning has been disappointingly slow, with this vital area often receiving insufficient focus. This paper seeks to address this shortfall by introducing well-defined and effective metrics for black-box unlearning auditing tasks. We transform the auditing challenge into a question of non-membership inference and develop efficient metrics for auditing. By relying exclusively on the original and unlearned models--eliminating the need to train additional shadow models--our approach simplifies the evaluation of unlearning at the individual data point level. Utilizing these metrics, we conduct an in-depth analysis of current approximate machine unlearning algorithms, identifying three key directions where these approaches fall short: utility, resilience, and equity. Our aim is that this work will greatly improve our understanding of approximate machine unlearning methods, taking a significant stride towards converting the theoretical right to data erasure into a auditable reality.

link: http://arxiv.org/abs/2403.12830v1

Embarrassingly Simple Scribble Supervision for 3D Medical Segmentation

Karol Gotkowski, Carsten Lüth, Paul F. Jäger, Sebastian Ziegler, Lars Krämer, Stefan Denner, Shuhan Xiao, Nico Disch, Klaus H. Maier-Hein, Fabian Isensee

Traditionally, segmentation algorithms require dense annotations for training, demanding significant annotation efforts, particularly within the 3D medical imaging field. Scribble-supervised learning emerges as a possible solution to this challenge, promising a reduction in annotation efforts when creating large-scale datasets. Recently, a plethora of methods for optimized learning from scribbles have been proposed, but have so far failed to position scribble annotation as a beneficial alternative. We relate this shortcoming to two major issues: 1) the complex nature of many methods which deeply ties them to the underlying segmentation model, thus preventing a migration to more powerful state-of-the-art models as the field progresses and 2) the lack of a systematic evaluation to validate consistent performance across the broader medical domain, resulting in a lack of trust when applying these methods to new segmentation problems. To address these issues, we propose a comprehensive scribble supervision benchmark consisting of seven datasets covering a diverse set of anatomies and pathologies imaged with varying modalities. We furthermore propose the systematic use of partial losses, i.e. losses that are only computed on annotated voxels. Contrary to most existing methods, these losses can be seamlessly integrated into state-of-the-art segmentation methods, enabling them to learn from scribble annotations while preserving their original loss formulations. Our evaluation using nnU-Net reveals that while most existing methods suffer from a lack of generalization, the proposed approach consistently delivers state-of-the-art performance. Thanks to its simplicity, our approach presents an embarrassingly simple yet effective solution to the challenges of scribble supervision. Source code as well as our extensive scribble benchmarking suite will be made publicly available upon publication.

link: http://arxiv.org/abs/2403.12834v1