Thu 2024.05.09

RepVGG-GELAN: Enhanced GELAN with VGG-STYLE ConvNets for Brain Tumour Detection

Thennarasi Balakrishnan, Sandeep Singh Sengar

Object detection algorithms particularly those based on YOLO have demonstrated remarkable efficiency in balancing speed and accuracy. However, their application in brain tumour detection remains underexplored. This study proposes RepVGG-GELAN, a novel YOLO architecture enhanced with RepVGG, a reparameterized convolutional approach for object detection tasks particularly focusing on brain tumour detection within medical images. RepVGG-GELAN leverages the RepVGG architecture to improve both speed and accuracy in detecting brain tumours. Integrating RepVGG into the YOLO framework aims to achieve a balance between computational efficiency and detection performance. This study includes a spatial pyramid pooling-based Generalized Efficient Layer Aggregation Network (GELAN) architecture which further enhances the capability of RepVGG. Experimental evaluation conducted on a brain tumour dataset demonstrates the effectiveness of RepVGG-GELAN surpassing existing RCS-YOLO in terms of precision and speed. Specifically, RepVGG-GELAN achieves an increased precision of 4.91% and an increased AP50 of 2.54% over the latest existing approach while operating at 240.7 GFLOPs. The proposed RepVGG-GELAN with GELAN architecture presents promising results establishing itself as a state-of-the-art solution for accurate and efficient brain tumour detection in medical images. The implementation code is publicly available at https://github.com/ThensiB/RepVGG-GELAN.

link: http://arxiv.org/abs/2405.03541v1

Optimizing Hand Region Detection in MediaPipe Holistic Full-Body Pose Estimation to Improve Accuracy and Avoid Downstream Errors

Amit Moryossef

This paper addresses a critical flaw in MediaPipe Holistic's hand Region of Interest (ROI) prediction, which struggles with non-ideal hand orientations, affecting sign language recognition accuracy. We propose a data-driven approach to enhance ROI estimation, leveraging an enriched feature set including additional hand keypoints and the z-dimension. Our results demonstrate better estimates, with higher Intersection-over-Union compared to the current method. Our code and optimizations are available at https://github.com/sign-language-processing/mediapipe-hand-crop-fix.

link: http://arxiv.org/abs/2405.03545v1

CCDM: Continuous Conditional Diffusion Models for Image Generation

Xin Ding, Yongwei Wang, Kao Zhang, Z. Jane Wang

Continuous Conditional Generative Modeling (CCGM) aims to estimate the distribution of high-dimensional data, typically images, conditioned on scalar continuous variables known as regression labels. While Continuous conditional Generative Adversarial Networks (CcGANs) were initially designed for this task, their adversarial training mechanism remains vulnerable to extremely sparse or imbalanced data, resulting in suboptimal outcomes. To enhance the quality of generated images, a promising alternative is to replace CcGANs with Conditional Diffusion Models (CDMs), renowned for their stable training process and ability to produce more realistic images. However, existing CDMs encounter challenges when applied to CCGM tasks due to several limitations such as inadequate U-Net architectures and deficient model fitting mechanisms for handling regression labels. In this paper, we introduce Continuous Conditional Diffusion Models (CCDMs), the first CDM designed specifically for the CCGM task. CCDMs address the limitations of existing CDMs by introducing specially designed conditional diffusion processes, a modified denoising U-Net with a custom-made conditioning mechanism, a novel hard vicinal loss for model fitting, and an efficient conditional sampling procedure. With comprehensive experiments on four datasets with varying resolutions ranging from 64x64 to 192x192, we demonstrate the superiority of the proposed CCDM

over state-of-the-art CCGM models, establishing new benchmarks in CCGM. Extensive ablation studies validate the model design and implementation configuration of the proposed CCDM. Our code is publicly available at https://github.com/UBCDingXin/CCDM.

link: http://arxiv.org/abs/2405.03546v1

Position Paper: Leveraging Foundational Models for Black-Box Optimization: Benefits, Challenges, and Future Directions

Xingyou Song, Yingtao Tian, Robert Tjarko Lange, Chansoo Lee, Yujin Tang, Yutian Chen

Undeniably, Large Language Models (LLMs) have stirred an extraordinary wave of innovation in the machine learning research domain, resulting in substantial impact across diverse fields such as reinforcement learning, robotics, and computer vision. Their incorporation has been rapid and transformative, marking a significant paradigm shift in the field of machine learning research. However, the field of experimental design, grounded on black-box optimization, has been much less affected by such a paradigm shift, even though integrating LLMs with optimization presents a unique landscape ripe for exploration. In this position paper, we frame the field of black-box optimization around sequence-based foundation models and organize their relationship with previous literature. We discuss the most promising ways foundational language models can revolutionize optimization, which include harnessing the vast wealth of information encapsulated in free-form text to enrich task comprehension, utilizing highly flexible sequence models such as Transformers to engineer superior optimization strategies, and enhancing performance prediction over previously unseen search spaces.

link: http://arxiv.org/abs/2405.03547v1

MAmmoTH2: Scaling Instructions from the Web

Xiang Yue, Tuney Zheng, Ge Zhang, Wenhu Chen

Instruction tuning improves the reasoning abilities of large language models (LLMs), with data quality and scalability being the crucial factors. Most instruction tuning data come from human crowd-sourcing or GPT-4 distillation. We propose a paradigm to efficiently harvest 10 million naturally existing instruction data from the pre-training web corpus to enhance LLM reasoning. Our approach involves (1) recalling relevant documents, (2) extracting instruction-response pairs, and (3) refining the extracted pairs using open-source LLMs. Fine-tuning base LLMs on this dataset, we build MAmmoTH2 models, which significantly boost performance on reasoning benchmarks. Notably, MAmmoTH2-7B's (Mistral) performance increases from 11% to 34% on MATH and from 36% to 67% on GSM8K without training on any in-domain data. Further training MAmmoTH2 on public instruction tuning datasets yields MAmmoTH2-Plus, achieving state-of-the-art performance on several reasoning and chatbot benchmarks. Our work demonstrates how to harvest large-scale, high-quality instruction data without costly human annotation or GPT-4 distillation, providing a new paradigm for building better instruction tuning data.

link: http://arxiv.org/abs/2405.03548v1

Bridging discrete and continuous state spaces: Exploring the Ehrenfest process in time-continuous diffusion models

Ludwig Winkler, Lorenz Richter, Manfred Opper

Generative modeling via stochastic processes has led to remarkable empirical results as well as to recent advances in their theoretical understanding. In principle, both space and time of the processes can be discrete or continuous. In this work, we study time-continuous Markov jump processes on discrete state spaces and investigate their correspondence to state-continuous diffusion processes given by SDEs. In particular, we revisit the $\textit{Ehrenfest process}$, which converges to an Ornstein-Uhlenbeck process in the infinite state space limit. Likewise, we can show that the time-reversal of the Ehrenfest process converges to the time-reversed Ornstein-Uhlenbeck process. This observation bridges discrete and continuous state spaces and allows to carry over methods from one to the respective other setting. Additionally, we suggest an algorithm for training the time-reversal of Markov jump processes which relies on conditional expectations and can thus

be directly related to denoising score matching. We demonstrate our methods in multiple convincing numerical experiments.

link: http://arxiv.org/abs/2405.03549v1

AlphaMath Almost Zero: process Supervision without process

Guoxin Chen, Minpeng Liao, Chengxi Li, Kai Fan

Recent advancements in large language models (LLMs) have substantially enhanced their mathematical reasoning abilities. However, these models still struggle with complex problems that require multiple reasoning steps, frequently leading to logical or numerical errors. While numerical mistakes can largely be addressed by integrating a code interpreter, identifying logical errors within intermediate steps is more challenging. Moreover, manually annotating these steps for training is not only expensive but also demands specialized expertise. In this study, we introduce an innovative approach that eliminates the need for manual annotation by leveraging the Monte Carlo Tree Search (MCTS) framework to generate both the process supervision and evaluation signals automatically. Essentially, when a LLM is well pre-trained, only the mathematical questions and their final answers are required to generate our training data, without requiring the solutions. We proceed to train a step-level value model designed to improve the LLM's inference process in mathematical domains. Our experiments indicate that using automatically generated solutions by LLMs enhanced with MCTS significantly improves the model's proficiency in dealing with intricate mathematical reasoning tasks.

link: http://arxiv.org/abs/2405.03553v1

A Comprehensive Overview and Survey of O-RAN: Exploring Slicing-aware Architecture, Deployment Options, and Use Cases

Khurshid Alam, Mohammad Asif Habibi, Matthias Tammen, Dennis Krummacker, Walid Saad, Marco Di Renzo, Tommaso Melodia, Xavier Costa-Pérez, Mérouane Debbah, Ashutosh Dutta, Hans D. Schotten

Open-radio access network (O-RAN) seeks to establish principles of openness, programmability, automation, intelligence, and hardware-software disaggregation with interoperable interfaces. It advocates for multi-vendorism and multi-stakeholderism within a cloudified and virtualized wireless infrastructure, aimed at enhancing the deployment, operation, and maintenance of RAN architecture. This enhancement promises increased flexibility, performance optimization, service innovation, energy efficiency, and cost efficiency in fifth-generation (5G), sixth-generation (6G), and future networks. One of the key features of the O-RAN architecture is its support for network slicing, which entails interaction with other slicing domains within a mobile network, notably the transport network (TN) domain and the core network (CN) domain, to realize end-to-end (E2E) network slicing. The study of this feature requires exploring the stances and contributions of diverse standards development organizations (SDOs). In this context, we note that despite the ongoing industrial deployments and standardization efforts, the research and standardization communities have yet to comprehensively address network slicing in O-RAN. To address this gap, this survey paper provides a comprehensive exploration of network slicing in O-RAN through an in-depth review of specification documents from O-RAN Alliance and research papers from leading industry and academic institutions. The paper commences with an overview of the ongoing standardization efforts and open-source contributions associated with O-RAN, subsequently delving into the latest O-RAN architecture with an emphasis on its slicing aspects. Further, the paper explores deployment scenarios for network slicing within O-RAN, examining options for the deployment and orchestration of O-RAN and TN network slice subnets...

link: http://arxiv.org/abs/2405.03555v3

Liberating Seen Classes: Boosting Few-Shot and Zero-Shot Text Classification via Anchor Generation and Classification Reframing

Han Liu, Siyang Zhao, Xiaotong Zhang, Feng Zhang, Wei Wang, Fenglong Ma, Hongyang Chen, Hong Yu, Xianchao Zhang

Few-shot and zero-shot text classification aim to recognize samples from novel classes with limited labeled samples or no labeled samples at all. While prevailing methods have shown promising performance via transferring knowledge from seen classes to unseen classes, they are still limited by (1) Inherent dissimilarities among classes make the transformation of features learned from seen classes to unseen classes both difficult and inefficient. (2) Rare labeled novel samples usually cannot provide enough supervision signals to enable the model to adjust from the source distribution to the target distribution, especially for complicated scenarios. To alleviate the above issues, we propose a simple and effective strategy for few-shot and zero-shot text classification. We aim to liberate the model from the confines of seen classes, thereby enabling it to predict unseen categories without the necessity of training on seen classes. Specifically, for mining more related unseen category knowledge, we utilize a large pre-trained language model to generate pseudo novel samples, and select the most representative ones as category anchors. After that, we convert the multi-class classification task into a binary classification task and use the similarities of query-anchor pairs for prediction to fully leverage the limited supervision signals. Extensive experiments on six widely used public datasets show that our proposed method can outperform other strong baselines significantly in few-shot and zero-shot tasks, even without using any seen class samples.

link: http://arxiv.org/abs/2405.03565v1

Deep Space Separable Distillation for Lightweight Acoustic Scene Classification

ShuQi Ye, Yuan Tian

Acoustic scene classification (ASC) is highly important in the real world. Recently, deep learning-based methods have been widely employed for acoustic scene classification. However, these methods are currently not lightweight enough as well as their performance is not satisfactory. To solve these problems, we propose a deep space separable distillation network. Firstly, the network performs high-low frequency decomposition on the log-mel spectrogram, significantly reducing computational complexity while maintaining model performance. Secondly, we specially design three lightweight operators for ASC, including Separable Convolution (SC), Orthonormal Separable Convolution (OSC), and Separable Partial Convolution (SPC). These operators exhibit highly efficient feature extraction capabilities in acoustic scene classification tasks. The experimental results demonstrate that the proposed method achieves a performance gain of 9.8% compared to the currently popular deep learning methods, while also having smaller parameter count and computational complexity.

link: http://arxiv.org/abs/2405.03567v1

Majority consensus thresholds in competitive Lotka--Volterra populations

Matthias Függer, Thomas Nowak, Joel Rybicki

One of the key challenges in synthetic biology is devising robust signaling primitives for engineered microbial consortia. In such systems, a fundamental signal amplification problem is the majority consensus problem: given a system with two input species with initial difference of Δ in population sizes, what is the probability that the system reaches a state in which only the initial majority species is present? In this work, we consider a discrete and stochastic version of competitive Lotka--Volterra dynamics, a standard model of microbial community dynamics. We identify new threshold properties for majority consensus under different types of interference competition: - We show that under so-called self-destructive interference competition between the two input species, majority consensus can be reached with high probability if the initial difference satisfies $\Delta \in \Omega(\log^2 n)$, where n is the initial population size. This gives an exponential improvement compared to the previously known bound of $\Omega(\sqrt{n \log n})$ by Cho et al. [Distributed Computing, 2021] given for a special case of the competitive Lotka--Volterra model. In contrast, we show that an initial gap of $\Delta \in \Omega(\sqrt{\log n})$ is necessary. - On the other hand, we prove that under non-self-destructive interference competition, an initial gap of $\Omega(\sqrt{n})$ is necessary to succeed with high probability and that a $\Omega(\sqrt{n \log n})$ gap is sufficient. This shows a strong qualitative gap between the performance of self-destructive and non-self-destructive interference competition. Moreover, we show that if in

addition the populations exhibit interference competition between the individuals of the same species, then majority consensus cannot always be solved with high probability, no matter what the difference in the initial population counts.

link: http://arxiv.org/abs/2405.03568v1

Understanding Read-Write Wait-Free Coverings in the Fully-Anonymous Shared-Memory Model

Giuliano Losa, Eli Gafni

In the fully-anonymous (shared-memory) model, inspired by a biological setting, processors have no identifiers and memory locations are anonymous. This means that there is no pre-existing agreement among processors on any naming of the memory locations. In this work, we ask fundamental questions about the fully-anonymous model in the hope to obtain a better understanding of the role of naming and anonymity in distributed computing. First, we ask what it means to solve a task under processor anonymity. With tasks such as renaming, the traditional notion obviously does not apply. Instead of restricting ourselves to colorless tasks, we propose using the notion of group solvability, which allows transferring any task to processor-anonymous models. Second, the difficulty with anonymity is that processors can hardly avoid covering and then overwriting each other's writes, erasing information written by their predecessors. To get to the bottom of this phenomenon, we ask what system configurations are stable when processors keep reading and writing ad infinitum. Resolving this question leads us to a wait-free solution to the snapshot task, which then allows us to solve renaming and obstruction-free consensus.

link: http://arxiv.org/abs/2405.03573v2

ILILT: Implicit Learning of Inverse Lithography Technologies

Haoyu Yang, Haoxing Ren

Lithography, transferring chip design masks to the silicon wafer, is the most important phase in modern semiconductor manufacturing flow. Due to the limitations of lithography systems, Extensive design optimizations are required to tackle the design and silicon mismatch. Inverse lithography technology (ILT) is one of the promising solutions to perform pre-fabrication optimization, termed mask optimization. Because of mask optimization problems' constrained non-convexity, numerical ILT solvers rely heavily on good initialization to avoid getting stuck on sub-optimal solutions. Machine learning (ML) techniques are hence proposed to generate mask initialization for ILT solvers with one-shot inference, targeting faster and better convergence during ILT. This paper addresses the question of \textit{whether ML models can directly generate high-quality optimized masks without engaging ILT solvers in the loop}. We propose an implicit learning ILT framework: ILILT, which leverages the implicit layer learning method and lithography-conditioned inputs to ground the model. Trained to understand the ILT optimization procedure, ILILT can outperform the state-of-the-art machine learning solutions, significantly improving efficiency and quality.

link: http://arxiv.org/abs/2405.03574v1

Some Statistical and Data Challenges When Building Early-Stage Digital Experimentation and Measurement Capabilities

C. H. Bryan Liu

Digital experimentation and measurement (DEM) capabilities -- the knowledge and tools necessary to run experiments with digital products, services, or experiences and measure their impact -- are fast becoming part of the standard toolkit of digital/data-driven organisations in guiding business decisions. Many large technology companies report having mature DEM capabilities, and several businesses have been established purely to manage experiments for others. Given the growing evidence that data-driven organisations tend to outperform their non-data-driven counterparts, there has never been a greater need for organisations to build/acquire DEM capabilities to thrive in the current digital era. This thesis presents several novel approaches to statistical and data challenges for organisations building DEM capabilities. We focus on the fundamentals associated with building DEM capabilities, which lead to a richer understanding of the underlying assumptions

and thus enable us to develop more appropriate capabilities. We address why one should engage in DEM by quantifying the benefits and risks of acquiring DEM capabilities. This is done using a ranking under lower uncertainty model, enabling one to construct a business case. We also examine what ingredients are necessary to run digital experiments. In addition to clarifying the existing literature around statistical tests, datasets, and methods in experimental design and causal inference, we construct an additional dataset and detailed case studies on applying state-of-the-art methods. Finally, we investigate when a digital experiment design would outperform another, leading to an evaluation framework that compares competing designs' data efficiency.

link: http://dx.doi.org/10.25560/110307

Functional Latent Dynamics for Irregularly Sampled Time Series Forecasting

Christian Klötergens, Vijaya Krishna Yalavarthi, Maximilian Stubbemann, Lars Schmidt-Thieme

Irregularly sampled time series with missing values are often observed in multiple real-world applications such as healthcare, climate and astronomy. They pose a significant challenge to standard deep learn- ing models that operate only on fully observed and regularly sampled time series. In order to capture the continuous dynamics of the irreg- ular time series, many models rely on solving an Ordinary Differential Equation (ODE) in the hidden state. These ODE-based models tend to perform slow and require large memory due to sequential operations and a complex ODE solver. As an alternative to complex ODE-based mod- els, we propose a family of models called Functional Latent Dynamics (FLD). Instead of solving the ODE, we use simple curves which exist at all time points to specify the continuous latent state in the model. The coefficients of these curves are learned only from the observed values in the time series ignoring the missing values. Through extensive experi- ments, we demonstrate that FLD achieves better performance compared to the best ODE-based model while reducing the runtime and memory overhead. Specifically, FLD requires an order of magnitude less time to infer the forecasts compared to the best performing forecasting model.

link: http://arxiv.org/abs/2405.03582v1