Sun 2024.04.14

Sparse Laneformer

Ji Liu, Zifeng Zhang, Mingjie Lu, Hongyang Wei, Dong Li, Yile Xie, Jinzhang Peng, Lu Tian, Ashish Sirasao, Emad Barsoum

Lane detection is a fundamental task in autonomous driving, and has achieved great progress as deep learning emerges. Previous anchor-based methods often design dense anchors, which highly depend on the training dataset and remain fixed during inference. We analyze that dense anchors are not necessary for lane detection, and propose a transformer-based lane detection framework based on a sparse anchor mechanism. To this end, we generate sparse anchors with position-aware lane queries and angle queries instead of traditional explicit anchors. We adopt Horizontal Perceptual Attention (HPA) to aggregate the lane features along the horizontal direction, and adopt Lane-Angle Cross Attention (LACA) to perform interactions between lane queries and angle queries. We also propose Lane Perceptual Attention (LPA) based on deformable cross attention to further refine the lane predictions. Our method, named Sparse Laneformer, is easy-to-implement and end-to-end trainable. Extensive experiments demonstrate that Sparse Laneformer performs favorably against the state-of-the-art methods, e.g., surpassing Laneformer by 3.0% F1 score and O2SFormer by 0.7% F1 score with fewer MACs on CULane with the same ResNet-34 backbone.

link: http://arxiv.org/abs/2404.07821v1

Heron-Bench: A Benchmark for Evaluating Vision Language Models in Japanese

Yuichi Inoue, Kento Sasaki, Yuma Ochi, Kazuki Fujii, Kotaro Tanahashi, Yu Yamaguchi

Vision Language Models (VLMs) have undergone a rapid evolution, giving rise to significant advancements in the realm of multimodal understanding tasks. However, the majority of these models are trained and evaluated on English-centric datasets, leaving a gap in the development and evaluation of VLMs for other languages, such as Japanese. This gap can be attributed to the lack of methodologies for constructing VLMs and the absence of benchmarks to accurately measure their performance. To address this issue, we introduce a novel benchmark, Japanese Heron-Bench, for evaluating Japanese capabilities of VLMs. The Japanese Heron-Bench consists of a variety of imagequestion answer pairs tailored to the Japanese context. Additionally, we present a baseline Japanese VLM that has been trained with Japanese visual instruction tuning datasets. Our Heron-Bench reveals the strengths and limitations of the proposed VLM across various ability dimensions. Furthermore, we clarify the capability gap between strong closed models like GPT-4V and the baseline model, providing valuable insights for future research in this domain. We release the benchmark dataset and training code to facilitate further developments in Japanese VLM research.

link: http://arxiv.org/abs/2404.07824v1

On the Sample Efficiency of Abstractions and Potential-Based Reward Shaping in Reinforcement Learning

Giuseppe Canonaco, Leo Ardon, Alberto Pozanco, Daniel Borrajo

The use of Potential Based Reward Shaping (PBRS) has shown great promise in the ongoing research effort to tackle sample inefficiency in Reinforcement Learning (RL). However, the choice of the potential function is critical for this technique to be effective. Additionally, RL techniques are usually constrained to use a finite horizon for computational limitations. This introduces a bias when using PBRS, thus adding an additional layer of complexity. In this paper, we leverage abstractions to automatically produce a "good" potential function. We analyse the bias induced by finite horizons in the context of PBRS producing novel insights. Finally, to asses sample efficiency and performance impact, we evaluate our approach on four environments including a goal-oriented navigation task and three Arcade Learning Environments (ALE) games demonstrating that we can reach the same level of performance as CNN-based solutions with a simple fully-connected network.

link: http://arxiv.org/abs/2404.07826v1

Streamlined Photoacoustic Image Processing with Foundation Models: A Training-Free Solution

Handi Deng, Yucheng Zhou, Jiaxuan Xiang, Liujie Gu, Yan Luo, Hai Feng, Mingyuan Liu, Cheng Ma

Foundation models have rapidly evolved and have achieved significant accomplishments in computer vision tasks. Specifically, the prompt mechanism conveniently allows users to integrate image prior information into the model, making it possible to apply models without any training. Therefore, we propose a method based on foundation models and zero training to solve the tasks of photoacoustic (PA) image segmentation. We employed the segment anything model (SAM) by setting simple prompts and integrating the model's outputs with prior knowledge of the imaged objects to accomplish various tasks, including: (1) removing the skin signal in three-dimensional PA image rendering; (2) dual speed-of-sound reconstruction, and (3) segmentation of finger blood vessels. Through these demonstrations, we have concluded that deep learning can be directly applied in PA imaging without the requirement for network design and training. This potentially allows for a hands-on, convenient approach to achieving efficient and accurate segmentation of PA images. This letter serves as a comprehensive tutorial, facilitating the mastery of the technique through the provision of code and sample datasets.

link: http://arxiv.org/abs/2404.07833v1

Question Generation in Knowledge-Driven Dialog: Explainability and Evaluation

Juliette Faille, Quentin Brabant, Gwenole Lecorve, Lina M. Rojas-Barahona, Claire Gardent

We explore question generation in the context of knowledge-grounded dialogs focusing on explainability and evaluation. Inspired by previous work on planning-based summarisation, we present a model which instead of directly generating a question, sequentially predicts first a fact then a question. We evaluate our approach on 37k test dialogs adapted from the KGConv dataset and we show that, although more demanding in terms of inference, our approach performs on par with a standard model which solely generates a question while allowing for a detailed referenceless evaluation of the model behaviour in terms of relevance, factuality and pronominalisation.

link: http://arxiv.org/abs/2404.07836v1

The Role of Confidence for Trust-based Resilient Consensus (Extended Version)

Luca Ballotta, Michal Yemini

We consider a multi-agent system where agents aim to achieve a consensus despite interactions with malicious agents that communicate misleading information. Physical channels supporting communication in cyberphysical systems offer attractive opportunities to detect malicious agents, nevertheless, trustworthiness indications coming from the channel are subject to uncertainty and need to be treated with this in mind. We propose a resilient consensus protocol that incorporates trust observations from the channel and weighs them with a parameter that accounts for how confident an agent is regarding its understanding of the legitimacy of other agents in the network, with no need for the initial observation window T_0 that has been utilized in previous works. Analytical and numerical results show that (i) our protocol achieves a resilient consensus in the presence of malicious agents and (ii) the steady-state deviation from nominal consensus can be minimized by a suitable choice of the confidence parameter that depends on the statistics of trust observations.

link: http://arxiv.org/abs/2404.07838v1

RecurrentGemma: Moving Past Transformers for Efficient Open Language Models

Aleksandar Botev, Soham De, Samuel L Smith, Anushan Fernando, George-Cristian Muraru, Ruba Haroun, Leonard Berrada, Razvan Pascanu, Pier Giuseppe Sessa, Robert Dadashi, Léonard

Hussenot, Johan Ferret, Sertan Girgin, Olivier Bachem, Alek Andreev, Kathleen Kenealy, Thomas Mesnard, Cassidy Hardin, Surya Bhupatiraju, Shreya Pathak, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, Pouya Tafti, Armand Joulin, Noah Fiedel, Evan Senter, Yutian Chen, Srivatsan Srinivasan, Guillaume Desjardins, David Budden, Arnaud Doucet, Sharad Vikram, Adam Paszke, Trevor Gale, Sebastian Borgeaud, Charlie Chen, Andy Brock, Antonia Paterson, Jenny Brennan, Meg Risdal, Raj Gundluru, Nesh Devanathan, Paul Mooney, Nilay Chauhan, Phil Culliton, Luiz GUStavo Martins, Elisa Bandy, David Huntsperger, Glenn Cameron, Arthur Zucker, Tris Warkentin, Ludovic Peran, Minh Giang, Zoubin Ghahramani, Clément Farabet, Koray Kavukcuoglu, Demis Hassabis, Raia Hadsell, Yee Whye Teh, Nando de Frietas

We introduce RecurrentGemma, an open language model which uses Google's novel Griffin architecture. Griffin combines linear recurrences with local attention to achieve excellent performance on language. It has a fixed-sized state, which reduces memory use and enables efficient inference on long sequences. We provide a pre-trained model with 2B non-embedding parameters, and an instruction tuned variant. Both models achieve comparable performance to Gemma-2B despite being trained on fewer tokens.

link: http://arxiv.org/abs/2404.07839v1

On Training Data Influence of GPT Models

Qingyi Liu, Yekun Chai, Shuohuan Wang, Yu Sun, Keze Wang, Hua Wu

Amidst the rapid advancements in generative language models, the investigation of how training data shapes the performance of GPT models is still emerging. This paper presents GPTfluence, a novel approach that leverages a featurized simulation to assess the impact of training examples on the training dynamics of GPT models. Our approach not only traces the influence of individual training instances on performance trajectories, such as loss and other key metrics, on targeted test points but also enables a comprehensive comparison with existing methods across various training scenarios in GPT models, ranging from 14 million to 2.8 billion parameters, across a range of downstream tasks. Contrary to earlier methods that struggle with generalization to new data, GPTfluence introduces a parameterized simulation of training dynamics, demonstrating robust generalization capabilities to unseen training data. This adaptability is evident across both fine-tuning and instruction-tuning scenarios, spanning tasks in natural language understanding and generation. We will make our code and data publicly available.

link: http://arxiv.org/abs/2404.07840v1

TBSN: Transformer-Based Blind-Spot Network for Self-Supervised Image Denoising

Junyi Li, Zhilu Zhang, Wangmeng Zuo

Blind-spot networks (BSN) have been prevalent network architectures in self-supervised image denoising (SSID). Existing BSNs are mostly conducted with convolution layers. Although transformers offer potential solutions to the limitations of convolutions and have demonstrated success in various image restoration tasks, their attention mechanisms may violate the blind-spot requirement, thus restricting their applicability in SSID. In this paper, we present a transformer-based blind-spot network (TBSN) by analyzing and redesigning the transformer operators that meet the blind-spot requirement. Specifically, TBSN follows the architectural principles of dilated BSNs, and incorporates spatial as well as channel self-attention layers to enhance the network capability. For spatial self-attention, an elaborate mask is applied to the attention matrix to restrict its receptive field, thus mimicking the dilated convolution. For channel self-attention, we observe that it may leak the blind-spot information when the channel number is greater than spatial size in the deep layers of multi-scale architectures. To eliminate this effect, we divide the channel into several groups and perform channel attention separately. Furthermore, we introduce a knowledge distillation strategy that distills TBSN into smaller denoisers to improve computational efficiency while maintaining performance. Extensive experiments on real-world image denoising datasets show that TBSN largely extends the receptive field and exhibits favorable performance against state-of-the-art SSID methods. The code and pre-trained models will be publicly available at https://github.com/nagejacob/TBSN.

link: http://arxiv.org/abs/2404.07846v1

Fuss-Free Network: A Simplified and Efficient Neural Network for Crowd Counting

Lei Chen, Xingen Gao

In the field of crowd-counting research, many recent deep learning based methods have demonstrated robust capabilities for accurately estimating crowd sizes. However, the enhancement in their performance often arises from an increase in the complexity of the model structure. This paper introduces the Fuss-Free Network (FFNet), a crowd counting deep learning model that is characterized by its simplicity and efficiency in terms of its structure. The model comprises only a backbone of a neural network and a multi-scale feature fusion structure.The multi-scale feature fusion structure is a simple architecture consisting of three branches, each only equipped with a focus transition module, and combines the features from these branches through the concatenation operation.Our proposed crowd counting model is trained and evaluated on four widely used public datasets, and it achieves accuracy that is comparable to that of existing complex models.The experimental results further indicate that excellent performance in crowd counting tasks can also be achieved by utilizing a simple, low-parameter, and computationally efficient neural network structure.

link: http://arxiv.org/abs/2404.07847v1

Overparameterized Multiple Linear Regression as Hyper-Curve Fitting

E. Atza, N. Budko

The paper shows that the application of the fixed-effect multiple linear regression model to an overparameterized dataset is equivalent to fitting the data with a hyper-curve parameterized by a single scalar parameter. This equivalence allows for a predictor-focused approach, where each predictor is described by a function of the chosen parameter. It is proven that a linear model will produce exact predictions even in the presence of nonlinear dependencies that violate the model assumptions. Parameterization in terms of the dependent variable and the monomial basis in the predictor function space are applied here to both synthetic and experimental data. The hyper-curve approach is especially suited for the regularization of problems with noise in predictor variables and can be used to remove noisy and "improper" predictors from the model.

link: http://arxiv.org/abs/2404.07849v1

MindBridge: A Cross-Subject Brain Decoding Framework

Shizun Wang, Songhua Liu, Zhenxiong Tan, Xinchao Wang

Brain decoding, a pivotal field in neuroscience, aims to reconstruct stimuli from acquired brain signals, primarily utilizing functional magnetic resonance imaging (fMRI). Currently, brain decoding is confined to a per-subject-per-model paradigm, limiting its applicability to the same individual for whom the decoding model is trained. This constraint stems from three key challenges: 1) the inherent variability in input dimensions across subjects due to differences in brain size; 2) the unique intrinsic neural patterns, influencing how different individuals perceive and process sensory information; 3) limited data availability for new subjects in real-world scenarios hampers the performance of decoding models. In this paper, we present a novel approach, MindBridge, that achieves cross-subject brain decoding by employing only one model. Our proposed framework establishes a generic paradigm capable of addressing these challenges by introducing biological-inspired aggregation function and novel cyclic fMRI reconstruction mechanism for subject-invariant representation learning. Notably, by cycle reconstruction of fMRI, MindBridge can enable novel fMRI synthesis, which also can serve as pseudo data augmentation. Within the framework, we also devise a novel reset-tuning method for adapting a pretrained model to a new subject. Experimental results demonstrate MindBridge's ability to reconstruct images for multiple subjects, which is competitive with dedicated subject-specific models. Furthermore, with limited data for a new subject, we achieve a high level of decoding accuracy, surpassing that of subject-specific models. This advancement in cross-subject brain decoding suggests promising directions for wider applications in neuroscience and indicates potential for more efficient utilization of limited fMRI data in real-world scenarios. Project page: https://littlepure2333.github.io/MindBridge

link: http://arxiv.org/abs/2404.07850v1

Guiding Large Language Models to Post-Edit Machine Translation with Error Annotations

Dayeon Ki, Marine Carpuat

Machine Translation (MT) remains one of the last NLP tasks where large language models (LLMs) have not yet replaced dedicated supervised systems. This work exploits the complementary strengths of LLMs and supervised MT by guiding LLMs to automatically post-edit MT with external feedback on its quality, derived from Multidimensional Quality Metric (MQM) annotations. Working with LLaMA-2 models, we consider prompting strategies varying the nature of feedback provided and then fine-tune the LLM to improve its ability to exploit the provided guidance. Through experiments on Chinese-English, English-German, and English-Russian MQM data, we demonstrate that prompting LLMs to post-edit MT improves TER, BLEU and COMET scores, although the benefits of fine-grained feedback are not clear. Fine-tuning helps integrate fine-grained feedback more effectively and further improves translation quality based on both automatic and human evaluation.

link: http://arxiv.org/abs/2404.07851v1

Resolve Domain Conflicts for Generalizable Remote Physiological Measurement

Weiyu Sun, Xinyu Zhang, Hao Lu, Ying Chen, Yun Ge, Xiaolin Huang, Jie Yuan, Yingcong Chen

Remote photoplethysmography (rPPG) technology has become increasingly popular due to its non-invasive monitoring of various physiological indicators, making it widely applicable in multimedia interaction, healthcare, and emotion analysis. Existing rPPG methods utilize multiple datasets for training to enhance the generalizability of models. However, they often overlook the underlying conflict issues across different datasets, such as (1) label conflict resulting from different phase delays between physiological signal labels and face videos at the instance level, and (2) attribute conflict stemming from distribution shifts caused by head movements, illumination changes, skin types, etc. To address this, we introduce the DOmain-HArmonious framework (DOHA). Specifically, we first propose a harmonious phase strategy to eliminate uncertain phase delays and preserve the temporal variation of physiological signals. Next, we design a harmonious hyperplane optimization that reduces irrelevant attribute shifts and encourages the model's optimization towards a global solution that fits more valid scenarios. Our experiments demonstrate that DOHA significantly improves the performance of existing methods under multiple protocols. Our code is available at https://github.com/SWY666/rPPG-DOHA.

link: http://arxiv.org/abs/2404.07855v1

Konnektor: Connection Protocol for Ensuring Peer Uniqueness in Decentralized P2P Networks

Onur Ozkan

Konnektor is a connection protocol designed to solve the challenge of managing unique peers within distributed peer-to-peer networks. By prioritizing network integrity and efficiency, Konnektor offers a comprehensive solution that safeguards against the spread of duplicate peers while optimizing resource utilization. This paper provides a detailed explanation of the protocol's key components, including peer addressing, connection initialization, detecting peer duplications and mitigation strategies against potential security threats.

link: http://arxiv.org/abs/2404.07861v1