Sun 2024.05.12

MAD-ICP: It Is All About Matching Data -- Robust and Informed LiDAR Odometry

Simone Ferrari, Luca Di Giammarino, Leonardo Brizi, Giorgio Grisetti

LiDAR odometry is the task of estimating the ego-motion of the sensor from sequential laser scans. This problem has been addressed by the community for more than two decades, and many effective solutions are available nowadays. Most of these systems implicitly rely on assumptions about the operating environment, the sensor used, and motion pattern. When these assumptions are violated, several well-known systems tend to perform poorly. This paper presents a LiDAR odometry system that can overcome these limitations and operate well under different operating conditions while achieving performance comparable with domain-specific methods. Our algorithm follows the well-known ICP paradigm that leverages a PCA-based kd-tree implementation that is used to extract structural information about the clouds being registered and to compute the minimization metric for the alignment. The drift is bound by managing the local map based on the estimated uncertainty of the tracked pose. To benefit the community, we release an open-source C++ anytime real-time implementation.

link: http://arxiv.org/abs/2405.05828v1

Mask-TS Net: Mask Temperature Scaling Uncertainty Calibration for Polyp Segmentation

Yudian Zhang, Chenhao Xu, Kaiye Xu, Haijiang Zhu

Lots of popular calibration methods in medical images focus on classification, but there are few comparable studies on semantic segmentation. In polyp segmentation of medical images, we find most diseased area occupies only a small portion of the entire image, resulting in previous models being not well-calibrated for lesion regions but well-calibrated for background, despite their seemingly better Expected Calibration Error (ECE) scores overall. Therefore, we proposed four-branches calibration network with Mask-Loss and Mask-TS strategies to more focus on the scaling of logits within potential lesion regions, which serves to mitigate the influence of background interference. In the experiments, we compare the existing calibration methods with the proposed Mask Temperature Scaling (Mask-TS). The results indicate that the proposed calibration network outperforms other methods both qualitatively and quantitatively.

link: http://arxiv.org/abs/2405.05830v1

Informed Decision-Making through Advancements in Open Set Recognition and Unknown Sample Detection

Atefeh Mahdavi, Marco Carvalho

Machine learning-based techniques open up many opportunities and improvements to derive deeper and more practical insights from data that can help businesses make informed decisions. However, the majority of these techniques focus on the conventional closed-set scenario, in which the label spaces for the training and test sets are identical. Open set recognition (OSR) aims to bring classification tasks in a situation that is more like reality, which focuses on classifying the known classes as well as handling unknown classes effectively. In such an open-set problem the gathered samples in the training set cannot encompass all the classes and the system needs to identify unknown samples at test time. On the other hand, building an accurate and comprehensive model in a real dynamic environment presents a number of obstacles, because it is prohibitively expensive to train for every possible example of unknown items, and the model may fail when tested in testbeds. This study provides an algorithm exploring a new representation of feature space to improve classification in OSR tasks. The efficacy and efficiency of business processes and decision-making can be improved by integrating OSR, which offers more precise and insightful predictions of outcomes. We demonstrate the performance of the proposed method on three established datasets. The results indicate that the proposed model outperforms the baseline methods in accuracy and F1-score.

link: http://arxiv.org/abs/2405.05836v1

Self-Supervised Pre-training with Symmetric Superimposition Modeling for Scene Text Recognition

Zuan Gao, Yuxin Wang, Yadong Qu, Boqiang Zhang, Zixiao Wang, Jianjun Xu, Hongtao Xie

In text recognition, self-supervised pre-training emerges as a good solution to reduce dependence on expansive annotated real data. Previous studies primarily focus on local visual representation by leveraging mask image modeling or sequence contrastive learning. However, they omit modeling the linguistic information in text images, which is crucial for recognizing text. To simultaneously capture local character features and linguistic information in visual space, we propose Symmetric Superimposition Modeling (SSM). The objective of SSM is to reconstruct the direction-specific pixel and feature signals from the symmetrically superimposed input. Specifically, we add the original image with its inverted views to create the symmetrically superimposed inputs. At the pixel level, we reconstruct the original and inverted images to capture character shapes and texture-level linguistic context. At the feature level, we reconstruct the feature of the same original image and inverted image with different augmentations to model the semantic-level linguistic context and the local character discrimination. In our design, we disrupt the character shape and linguistic rules. Consequently, the dual-level reconstruction facilitates understanding character shapes and linguistic information from the perspective of visual texture and feature semantics. Experiments on various text recognition benchmarks demonstrate the effectiveness and generality of SSM, with 4.1% average performance gains and 86.6% new state-of-the-art average word accuracy on Union14M benchmarks.

link: http://arxiv.org/abs/2405.05841v1

Could It Be Generated? Towards Practical Analysis of Memorization in Text-To-Image Diffusion Models

Zhe Ma, Xuhong Zhang, Qingming Li, Tianyu Du, Wenzhi Chen, Zonghui Wang, Shouling Ji

The past few years have witnessed substantial advancement in text-guided image generation powered by diffusion models. However, it was shown that text-to-image diffusion models are vulnerable to training image memorization, raising concerns on copyright infringement and privacy invasion. In this work, we perform practical analysis of memorization in text-to-image diffusion models. Targeting a set of images to protect, we conduct quantitive analysis on them without need to collect any prompts. Specifically, we first formally define the memorization of image and identify three necessary conditions of memorization, respectively similarity, existence and probability. We then reveal the correlation between the model's prediction error and image replication. Based on the correlation, we propose to utilize inversion techniques to verify the safety of target images against memorization and measure the extent to which they are memorized. Model developers can utilize our analysis method to discover memorized images or reliably claim safety against memorization. Extensive experiments on the Stable Diffusion, a popular open-source text-to-image diffusion model, demonstrate the effectiveness of our analysis method.

link: http://arxiv.org/abs/2405.05846v1

Learned feature representations are biased by complexity, learning order, position, and more

Andrew Kyle Lampinen, Stephanie C. Y. Chan, Katherine Hermann

Representation learning, and interpreting learned representations, are key areas of focus in machine learning and neuroscience. Both fields generally use representations as a means to understand or improve a system's computations. In this work, however, we explore surprising dissociations between representation and computation that may pose challenges for such efforts. We create datasets in which we attempt to match the computational role that different features play, while manipulating other properties of the features or the data. We train various deep learning architectures to compute these multiple abstract features about their inputs. We find that their learned feature representations are systematically biased towards representing some features

more strongly than others, depending upon extraneous properties such as feature complexity, the order in which features are learned, and the distribution of features over the inputs. For example, features that are simpler to compute or learned first tend to be represented more strongly and densely than features that are more complex or learned later, even if all features are learned equally well. We also explore how these biases are affected by architectures, optimizers, and training regimes (e.g., in transformers, features decoded earlier in the output sequence also tend to be represented more strongly). Our results help to characterize the inductive biases of gradient-based representation learning. These results also highlight a key challenge for interpretability $-$ or for comparing the representations of models and brains $-$ disentangling extraneous biases from the computationally important aspects of a system's internal representations.

link: http://arxiv.org/abs/2405.05847v1

Age of Information and Energy Consumption in IoT: an Experimental Evaluation

Federico Cristofani, Valerio Luconi, Alessio Vecchio

The Age of Information (AoI) is an end-to-end metric frequently used to understand how "fresh" the information about a remote system is. In this paper, we present an experimental study of the relationship between AoI and the energy spent by the device that produces information, e.g. an IoT device or a monitoring sensor. Such a relationship has been almost neglected so far, but it is particularly important whenever the sensing side is battery-operated. The study is carried out in a scenario where access is achieved via the cellular network and information is transferred using MQTT, a popular messaging protocol in the IoT domain. Numerous parameters of operation are considered, and the most efficient solutions in all configurations are provided.

link: http://arxiv.org/abs/2405.05849v1

Pre-trained Text-to-Image Diffusion Models Are Versatile Representation Learners for Control

Gunshi Gupta, Karmesh Yadav, Yarin Gal, Dhruv Batra, Zsolt Kira, Cong Lu, Tim G. J. Rudner

Embodied AI agents require a fine-grained understanding of the physical world mediated through visual and language inputs. Such capabilities are difficult to learn solely from task-specific data. This has led to the emergence of pre-trained vision-language models as a tool for transferring representations learned from internet-scale data to downstream tasks and new domains. However, commonly used contrastively trained representations such as in CLIP have been shown to fail at enabling embodied agents to gain a sufficiently fine-grained scene understanding -- a capability vital for control. To address this shortcoming, we consider representations from pre-trained text-to-image diffusion models, which are explicitly optimized to generate images from text prompts and as such, contain text-conditioned representations that reflect highly fine-grained visuo-spatial information. Using pre-trained text-to-image diffusion models, we construct Stable Control Representations which allow learning downstream control policies that generalize to complex, open-ended environments. We show that policies learned using Stable Control Representations are competitive with state-of-the-art representation learning approaches across a broad range of simulated control settings, encompassing challenging manipulation and navigation tasks. Most notably, we show that Stable Control Representations enable learning policies that exhibit state-of-the-art performance on OVMM, a difficult open-vocabulary navigation benchmark.

link: http://arxiv.org/abs/2405.05852v1

Robust and Explainable Fine-Grained Visual Classification with Transfer Learning: A Dual-Carriageway Framework

Zheming Zuo, Joseph Smith, Jonathan Stonehouse, Boguslaw Obara

In the realm of practical fine-grained visual classification applications rooted in deep learning, a common scenario involves training a model using a pre-existing dataset. Subsequently, a new dataset becomes available, prompting the desire to make a pivotal decision for achieving enhanced and leveraged inference performance on both sides: Should one opt to train datasets from scratch

or fine-tune the model trained on the initial dataset using the newly released dataset? The existing literature reveals a lack of methods to systematically determine the optimal training strategy, necessitating explainability. To this end, we present an automatic best-suit training solution searching framework, the Dual-Carriageway Framework (DCF), to fill this gap. DCF benefits from the design of a dual-direction search (starting from the pre-existing or the newly released dataset) where five different training settings are enforced. In addition, DCF is not only capable of figuring out the optimal training strategy with the capability of avoiding overfitting but also yields built-in quantitative and visual explanations derived from the actual input and weights of the trained model. We validated DCF's effectiveness through experiments with three convolutional neural networks (ResNet18, ResNet34 and Inception-v3) on two temporally continued commercial product datasets. Results showed fine-tuning pathways outperformed training-from-scratch ones by up to 2.13% and 1.23% on the pre-existing and new datasets, respectively, in terms of mean accuracy. Furthermore, DCF identified reflection padding as the superior padding method, enhancing testing accuracy by 3.72% on average. This framework stands out for its potential to guide the development of robust and explainable AI solutions in fine-grained visual classification tasks.

link: http://arxiv.org/abs/2405.05853v1

Compressed Bayesian Federated Learning for Reliable Passive Radio Sensing in Industrial IoT

Luca Barbieri, Stefano Savazzi, Monica Nicoli

Bayesian Federated Learning (FL) has been recently introduced to provide well-calibrated Machine Learning (ML) models quantifying the uncertainty of their predictions. Despite their advantages compared to frequentist FL setups, Bayesian FL tools implemented over decentralized networks are subject to high communication costs due to the iterated exchange of local posterior distributions among cooperating devices. Therefore, this paper proposes a communication-efficient decentralized Bayesian FL policy to reduce the communication overhead without sacrificing final learning accuracy and calibration. The proposed method integrates compression policies and allows devices to perform multiple optimization steps before sending the local posterior distributions. We integrate the developed tool in an Industrial Internet of Things (IIoT) use case where collaborating nodes equipped with autonomous radar sensors are tasked to reliably localize human operators in a workplace shared with robots. Numerical results show that the developed approach obtains highly accurate yet well-calibrated ML models compatible with the ones provided by conventional (uncompressed) Bayesian FL tools while substantially decreasing the communication overhead (i.e., up to 99%). Furthermore, the proposed approach is advantageous when compared with state-of-the-art compressed frequentist FL setups in terms of calibration, especially when the statistical distribution of the testing dataset changes.

link: http://arxiv.org/abs/2405.05855v1

Free-Moving Object Reconstruction and Pose Estimation with Virtual Camera

Haixin Shi, Yinlin Hu, Daniel Koguciuk, Juan-Ting Lin, Mathieu Salzmann, David Ferstl

We propose an approach for reconstructing free-moving object from a monocular RGB video. Most existing methods either assume scene prior, hand pose prior, object category pose prior, or rely on local optimization with multiple sequence segments. We propose a method that allows free interaction with the object in front of a moving camera without relying on any prior, and optimizes the sequence globally without any segments. We progressively optimize the object shape and pose simultaneously based on an implicit neural representation. A key aspect of our method is a virtual camera system that reduces the search space of the optimization significantly. We evaluate our method on the standard HO3D dataset and a collection of egocentric RGB sequences captured with a head-mounted device. We demonstrate that our approach outperforms most methods significantly, and is on par with recent techniques that assume prior information.

link: http://arxiv.org/abs/2405.05858v1

The Perspectivist Paradigm Shift: Assumptions and Challenges of Capturing Human Labels

Eve Fleisig, Su Lin Blodgett, Dan Klein, Zeerak Talat

Longstanding data labeling practices in machine learning involve collecting and aggregating labels from multiple annotators. But what should we do when annotators disagree? Though annotator disagreement has long been seen as a problem to minimize, new perspectivist approaches challenge this assumption by treating disagreement as a valuable source of information. In this position paper, we examine practices and assumptions surrounding the causes of disagreement--some challenged by perspectivist approaches, and some that remain to be addressed--as well as practical and normative challenges for work operating under these assumptions. We conclude with recommendations for the data labeling pipeline and avenues for future research engaging with subjectivity and disagreement.

link: http://arxiv.org/abs/2405.05860v1

ExACT: An End-to-End Autonomous Excavator System Using Action Chunking With Transformers

Liangliang Chen, Shiyu Jin, Haoyu Wang, Liangjun Zhang

Excavators are crucial for diverse tasks such as construction and mining, while autonomous excavator systems enhance safety and efficiency, address labor shortages, and improve human working conditions. Different from the existing modularized approaches, this paper introduces ExACT, an end-to-end autonomous excavator system that processes raw LiDAR, camera data, and joint positions to control excavator valves directly. Utilizing the Action Chunking with Transformers (ACT) architecture, ExACT employs imitation learning to take observations from multi-modal sensors as inputs and generate actionable sequences. In our experiment, we build a simulator based on the captured real-world data to model the relations between excavator valve states and joint velocities. With a few human-operated demonstration data trajectories, ExACT demonstrates the capability of completing different excavation tasks, including reaching, digging and dumping through imitation learning in validations with the simulator. To the best of our knowledge, ExACT represents the first instance towards building an end-to-end autonomous excavator system via imitation learning methods with a minimal set of human demonstrations. The video about this work can be accessed at https://youtu.be/NmzR_Rf-aEk.

link: http://arxiv.org/abs/2405.05861v1

Faster Linear Systems and Matrix Norm Approximation via Multi-level Sketched Preconditioning

Micha■ Derezi■ski, Christopher Musco, Jiaming Yang

We present a new class of preconditioned iterative methods for solving linear systems of the form $Ax = b$. Our methods are based on constructing a low-rank Nystr\"om approximation to A using sparse random sketching. This approximation is used to construct a preconditioner, which itself is inverted quickly using additional levels of random sketching and preconditioning. We prove that the convergence of our methods depends on a natural average condition number of A, which improves as the rank of the Nystr\"om approximation increases. Concretely, this allows us to obtain faster runtimes for a number of fundamental linear algebraic problems: 1. We show how to solve any $n\times n$ linear system that is well-conditioned except for k outlying large singular values in $\tilde{O}(n^{2.065} + k^\omega)$ time, improving on a recent result of [Derezi\'nski, Yang, STOC 2024] for all $k \gtrsim n^{0.78}$. 2. We give the first $\tilde{O}(n^2 + {d_\lambda}^{\omega})$ time algorithm for solving a regularized linear system $(A + \lambda I)x = b$, where A is positive semidefinite with effective dimension d_λ. This problem arises in applications like Gaussian process regression. 3. We give faster algorithms for approximating Schatten p-norms and other matrix norms. For example, for the Schatten 1 (nuclear) norm, we give an algorithm that runs in $\tilde{O}(n^{2.11})$ time, improving on an $\tilde{O}(n^{2.18})$ method of [Musco et al., ITCS 2018]. Interestingly, previous state-of-the-art algorithms for most of the problems above relied on stochastic iterative methods, like stochastic coordinate and gradient descent. Our work takes a completely different approach, instead leveraging tools from matrix

sketching.

link: http://arxiv.org/abs/2405.05865v1

Selecting the Most Conflicting Pair of Candidates

Théo Delemazure, ■ukasz Janeczko, Andrzej Kaczmarczyk, Stanis■aw Szufa

We study committee elections from a perspective of finding the most conflicting candidates, that is, candidates that imply the largest amount of conflict, as per voter preferences. By proposing basic axioms to capture this objective, we show that none of the prominent multiwinner voting rules meet them. Consequently, we design committee voting rules compliant with our desiderata, introducing conflictual voting rules. A subsequent deepened analysis sheds more light on how they operate. Our investigation identifies various aspects of conflict, for which we come up with relevant axioms and quantitative measures, which may be of independent interest. We support our theoretical study with experiments on both real-life and synthetic data.

link: http://arxiv.org/abs/2405.05870v1