Wed 2024.03.13

Genetic Learning for Designing Sim-to-Real Data Augmentations

Bram Vanherle, Nick Michiels, Frank Van Reeth

Data augmentations are useful in closing the sim-to-real domain gap when training on synthetic data. This is because they widen the training data distribution, thus encouraging the model to generalize better to other domains. Many image augmentation techniques exist, parametrized by different settings, such as strength and probability. This leads to a large space of different possible augmentation policies. Some policies work better than others for overcoming the sim-to-real gap for specific datasets, and it is unclear why. This paper presents two different interpretable metrics that can be combined to predict how well a certain augmentation policy will work for a specific sim-to-real setting, focusing on object detection. We validate our metrics by training many models with different augmentation policies and showing a strong correlation with performance on real data. Additionally, we introduce GeneticAugment, a genetic programming method that can leverage these metrics to automatically design an augmentation policy for a specific dataset without needing to train a model.

link: http://arxiv.org/abs/2403.06786v1

SPLADE-v3: New baselines for SPLADE

Carlos Lassance, Hervé Déjean, Thibault Formal, Stéphane Clinchant

A companion to the release of the latest version of the SPLADE library. We describe changes to the training structure and present our latest series of models -- SPLADE-v3. We compare this new version to BM25, SPLADE++, as well as re-rankers, and showcase its effectiveness via a meta-analysis over more than 40 query sets. SPLADE-v3 further pushes the limit of SPLADE models: it is statistically significantly more effective than both BM25 and SPLADE++, while comparing well to cross-encoder re-rankers. Specifically, it gets more than 40 MRR@10 on the MS MARCO dev set, and improves by 2% the out-of-domain results on the BEIR benchmark.

link: http://arxiv.org/abs/2403.06789v1

Next4: Snapshots in Ext4 File System

Aditya Dani, Shardul Mangade, Piyush Nimbalkar, Harshad Shirwadkar

The growing value of data as a strategic asset has given rise to the necessity of implementing reliable backup and recovery solutions in the most efficient and cost-effective manner. The data backup methods available today on linux are not effective enough, because while running, most of them block I/Os to guarantee data integrity. We propose and implement Next4 - file system based snapshot feature in Ext4 which creates an instant image of the file system, to provide incremental versions of data, enabling reliable backup and data recovery. In our design, the snapshot feature is implemented by efficiently infusing the copy-on-write strategy in the write-in-place, extent based Ext4 file system, without affecting its basic structure. Each snapshot is an incremental backup of the data within the system. What distinguishes Next4 is the way that the data is backed up, improving both space utilization as well as performance.

link: http://arxiv.org/abs/2403.06790v1

Boosting Image Restoration via Priors from Pre-trained Models

Xiaogang Xu, Shu Kong, Tao Hu, Zhe Liu, Hujun Bao

Pre-trained models with large-scale training data, such as CLIP and Stable Diffusion, have demonstrated remarkable performance in various high-level computer vision tasks such as image understanding and generation from language descriptions. Yet, their potential for low-level tasks such as image restoration remains relatively unexplored. In this paper, we explore such models to enhance image restoration. As off-the-shelf features (OSF) from pre-trained models do not directly

serve image restoration, we propose to learn an additional lightweight module called Pre-Train-Guided Refinement Module (PTG-RM) to refine restoration results of a target restoration network with OSF. PTG-RM consists of two components, Pre-Train-Guided Spatial-Varying Enhancement (PTG-SVE), and Pre-Train-Guided Channel-Spatial Attention (PTG-CSA). PTG-SVE enables optimal short- and long-range neural operations, while PTG-CSA enhances spatial-channel attention for restoration-related learning. Extensive experiments demonstrate that PTG-RM, with its compact size ($<$1M parameters), effectively enhances restoration performance of various models across different tasks, including low-light enhancement, deraining, deblurring, and denoising.

link: http://arxiv.org/abs/2403.06793v1

Leveraging Internal Representations of Model for Magnetic Image Classification

Adarsh N L, Arun P V, Alok Porwal, Malcolm Aranha

Data generated by edge devices has the potential to train intelligent autonomous systems across various domains. Despite the emergence of diverse machine learning approaches addressing privacy concerns and utilizing distributed data, security issues persist due to the sensitive storage of data shards in disparate locations. This paper introduces a potentially groundbreaking paradigm for machine learning model training, specifically designed for scenarios with only a single magnetic image and its corresponding label image available. We harness the capabilities of Deep Learning to generate concise yet informative samples, aiming to overcome data scarcity. Through the utilization of deep learning's internal representations, our objective is to efficiently address data scarcity issues and produce meaningful results. This methodology presents a promising avenue for training machine learning models with minimal data.

link: http://arxiv.org/abs/2403.06797v1

Dynamic Perturbation-Adaptive Adversarial Training on Medical Image Classification

Shuai Li, Xiaoguang Ma, Shancheng Jiang, Lu Meng

Remarkable successes were made in Medical Image Classification (MIC) recently, mainly due to wide applications of convolutional neural networks (CNNs). However, adversarial examples (AEs) exhibited imperceptible similarity with raw data, raising serious concerns on network robustness. Although adversarial training (AT), in responding to malevolent AEs, was recognized as an effective approach to improve robustness, it was challenging to overcome generalization decline of networks caused by the AT. In this paper, in order to reserve high generalization while improving robustness, we proposed a dynamic perturbation-adaptive adversarial training (DPAAT) method, which placed AT in a dynamic learning environment to generate adaptive data-level perturbations and provided a dynamically updated criterion by loss information collections to handle the disadvantage of fixed perturbation sizes in conventional AT methods and the dependence on external transference. Comprehensive testing on dermatology HAM10000 dataset showed that the DPAAT not only achieved better robustness improvement and generalization preservation but also significantly enhanced mean average precision and interpretability on various CNNs, indicating its great potential as a generic adversarial training method on the MIC.

link: http://arxiv.org/abs/2403.06798v1

MambaMIL: Enhancing Long Sequence Modeling with Sequence Reordering in Computational Pathology

Shu Yang, Yihui Wang, Hao Chen

Multiple Instance Learning (MIL) has emerged as a dominant paradigm to extract discriminative feature representations within Whole Slide Images (WSIs) in computational pathology. Despite driving notable progress, existing MIL approaches suffer from limitations in facilitating comprehensive and efficient interactions among instances, as well as challenges related to time-consuming computations and overfitting. In this paper, we incorporate the Selective Scan Space State Sequential Model (Mamba) in Multiple Instance Learning (MIL) for long sequence modeling with linear complexity, termed as MambaMIL. By inheriting the capability of vanilla

Mamba, MambaMIL demonstrates the ability to comprehensively understand and perceive long sequences of instances. Furthermore, we propose the Sequence Reordering Mamba (SR-Mamba) aware of the order and distribution of instances, which exploits the inherent valuable information embedded within the long sequences. With the SR-Mamba as the core component, MambaMIL can effectively capture more discriminative features and mitigate the challenges associated with overfitting and high computational overhead. Extensive experiments on two public challenging tasks across nine diverse datasets demonstrate that our proposed framework performs favorably against state-of-the-art MIL methods. The code is released at https://github.com/isyangshu/MambaMIL.

link: http://arxiv.org/abs/2403.06800v1

CT2Rep: Automated Radiology Report Generation for 3D Medical Imaging

Ibrahim Ethem Hamamci, Sezgin Er, Bjoern Menze

Medical imaging plays a crucial role in diagnosis, with radiology reports serving as vital documentation. Automating report generation has emerged as a critical need to alleviate the workload of radiologists. While machine learning has facilitated report generation for 2D medical imaging, extending this to 3D has been unexplored due to computational complexity and data scarcity. We introduce the first method to generate radiology reports for 3D medical imaging, specifically targeting chest CT volumes. Given the absence of comparable methods, we establish a baseline using an advanced 3D vision encoder in medical imaging to demonstrate our method's effectiveness, which leverages a novel auto-regressive causal transformer. Furthermore, recognizing the benefits of leveraging information from previous visits, we augment CT2Rep with a cross-attention-based multi-modal fusion module and hierarchical memory, enabling the incorporation of longitudinal multimodal data. Access our code at: https://github.com/ibrahimethemhamamci/CT2Rep

link: http://arxiv.org/abs/2403.06801v1

Joint Source-and-Channel Coding for Small Satellite Applications

Olga Kondrateva, Stefan Dietzel, Björn Scheuermann

Small satellites are widely used today as cost effective means to perform Earth observation and other tasks that generate large amounts of high-dimensional data, such as multi-spectral imagery. These satellites typically operate in low earth orbit, which poses significant challenges for data transmission due to short contact times with ground stations, low bandwidth, and high packet loss probabilities. In this paper, we introduce JSCC-Sat, which applies joint source-and-channel coding using neural networks to provide efficient and robust transmission of compressed image data for satellite applications. We evaluate our mechanism against traditional transmission schemes with separate source and channel coding and demonstrate that it outperforms the existing approaches when applied to Earth observation data of the Sentinel-2 mission.

link: http://dx.doi.org/10.1109/LCN58197.2023.10223379

Data-Independent Operator: A Training-Free Artifact Representation Extractor for Generalizable Deepfake Detection

Chuangchuang Tan, Ping Liu, RenShuai Tao, Huan Liu, Yao Zhao, Baoyuan Wu, Yunchao Wei

Recently, the proliferation of increasingly realistic synthetic images generated by various generative adversarial networks has increased the risk of misuse. Consequently, there is a pressing need to develop a generalizable detector for accurately recognizing fake images. The conventional methods rely on generating diverse training sources or large pretrained models. In this work, we show that, on the contrary, the small and training-free filter is sufficient to capture more general artifact representations. Due to its unbias towards both the training and test sources, we define it as Data-Independent Operator (DIO) to achieve appealing improvements on unseen sources. In our framework, handcrafted filters and the randomly-initialized convolutional layer can be used as the training-free artifact representations extractor with excellent results. With the data-independent operator of a popular classifier, such as Resnet50, one could already reach a new state-of-the-art without bells and whistles. We evaluate the effectiveness of the DIO on 33 generation models, even

DALLE and Midjourney. Our detector achieves a remarkable improvement of 13.3%, establishing a new state-of-the-art performance. The DIO and its extension can serve as strong baselines for future methods. The code is available at \url{https://github.com/chuangchuangtan/Data-Independent-Operator}.

link: http://arxiv.org/abs/2403.06803v1

Shape Non-rigid Kinematics (SNK): A Zero-Shot Method for Non-Rigid Shape Matching via Unsupervised Functional Map Regularized Reconstruction

Souhaib Attaiki, Maks Ovsjanikov

We present Shape Non-rigid Kinematics (SNK), a novel zero-shot method for non-rigid shape matching that eliminates the need for extensive training or ground truth data. SNK operates on a single pair of shapes, and employs a reconstruction-based strategy using an encoder-decoder architecture, which deforms the source shape to closely match the target shape. During the process, an unsupervised functional map is predicted and converted into a point-to-point map, serving as a supervisory mechanism for the reconstruction. To aid in training, we have designed a new decoder architecture that generates smooth, realistic deformations. SNK demonstrates competitive results on traditional benchmarks, simplifying the shape-matching process without compromising accuracy. Our code can be found online: https://github.com/pvnieo/SNK

link: http://arxiv.org/abs/2403.06804v1

On the Robustness of Lexicase Selection to Contradictory Objectives

Shakiba Shahbandegan, Emily Dolson

Lexicase and epsilon-lexicase selection are state of the art parent selection techniques for problems featuring multiple selection criteria. Originally, lexicase selection was developed for cases where these selection criteria are unlikely to be in conflict with each other, but preliminary work suggests it is also a highly effective many-objective optimization algorithm. However, to predict whether these results generalize, we must understand lexicase selection's performance on contradictory objectives. Prior work has shown mixed results on this question. Here, we develop theory identifying circumstances under which lexicase selection will succeed or fail to find a Pareto-optimal solution. To make this analysis tractable, we restrict our investigation to a theoretical problem with maximally contradictory objectives. Ultimately, we find that lexicase and epsilon-lexicase selection each have a region of parameter space where they are incapable of optimizing contradictory objectives. Outside of this region, however, they perform well despite the presence of contradictory objectives. Based on these findings, we propose theoretically-backed guidelines for parameter choice. Additionally, we identify other properties that may affect whether a many-objective optimization problem is a good fit for lexicase or epsilon-lexicase selection.

link: http://arxiv.org/abs/2403.06805v1

On the Global Convergence of Policy Gradient in Average Reward Markov Decision Processes

Navdeep Kumar, Yashaswini Murthy, Itai Shufaro, Kfir Y. Levy, R. Srikant, Shie Mannor

We present the first finite time global convergence analysis of policy gradient in the context of infinite horizon average reward Markov decision processes (MDPs). Specifically, we focus on ergodic tabular MDPs with finite state and action spaces. Our analysis shows that the policy gradient iterates converge to the optimal policy at a sublinear rate of $O\left({\frac{1}{T}}\right),$ which translates to $O\left({\log(T)}\right)$ regret, where T represents the number of iterations. Prior work on performance bounds for discounted reward MDPs cannot be extended to average reward MDPs because the bounds grow proportional to the fifth power of the effective horizon. Thus, our primary contribution is in proving that the policy gradient algorithm converges for average-reward MDPs and in obtaining finite-time performance guarantees. In contrast to the existing discounted reward performance bounds, our performance bounds have an explicit dependence on constants that capture the complexity of the underlying MDP. Motivated by this observation, we reexamine and improve the existing performance bounds for discounted reward

MDPs. We also present simulations to empirically evaluate the performance of average reward policy gradient algorithm.

link: http://arxiv.org/abs/2403.06806v1

Multistep Consistency Models

Jonathan Heek, Emiel Hoogeboom, Tim Salimans

Diffusion models are relatively easy to train but require many steps to generate samples. Consistency models are far more difficult to train, but generate samples in a single step. In this paper we propose Multistep Consistency Models: A unification between Consistency Models (Song et al., 2023) and TRACT (Berthelot et al., 2023) that can interpolate between a consistency model and a diffusion model: a trade-off between sampling speed and sampling quality. Specifically, a 1-step consistency model is a conventional consistency model whereas we show that a ∞-step consistency model is a diffusion model. Multistep Consistency Models work really well in practice. By increasing the sample budget from a single step to 2-8 steps, we can train models more easily that generate higher quality samples, while retaining much of the sampling speed benefits. Notable results are 1.4 FID on Imagenet 64 in 8 step and 2.1 FID on Imagenet128 in 8 steps with consistency distillation. We also show that our method scales to a text-to-image diffusion model, generating samples that are very close to the quality of the original model.

link: http://arxiv.org/abs/2403.06807v1

Deep Learning Approaches for Human Action Recognition in Video Data

Yufei Xie

Human action recognition in videos is a critical task with significant implications for numerous applications, including surveillance, sports analytics, and healthcare. The challenge lies in creating models that are both precise in their recognition capabilities and efficient enough for practical use. This study conducts an in-depth analysis of various deep learning models to address this challenge. Utilizing a subset of the UCF101 Videos dataset, we focus on Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Two-Stream ConvNets. The research reveals that while CNNs effectively capture spatial features and RNNs encode temporal sequences, Two-Stream ConvNets exhibit superior performance by integrating spatial and temporal dimensions. These insights are distilled from the evaluation metrics of accuracy, precision, recall, and F1-score. The results of this study underscore the potential of composite models in achieving robust human action recognition and suggest avenues for future research in optimizing these models for real-world deployment.

link: http://arxiv.org/abs/2403.06810v1