Funded by
the European Union

www.elias-ai.eu
eds2024.github.io

POSTER CATALOGUE

ELLIS Doctoral Symposium #EDS2024

AI & Sustainability

26-30 August, 2024
Paris, France

TABLE OF CONTENTS

The ELLIS Doctoral Symposium is an annual conference for ELLIS PhD students and other PhD students to meet in person and share knowledge about machine learning. The ELLIS Doctoral Symposium 2024 (EDS2024) is the fourth edition, which is co-organised by the University of Amsterdam, ELSA and the Institut Polytechnique de Paris, and will be held in Paris, France. It is expected to host 150 attendees during one week from Monday, August 26 - to Friday, August 30.
The focus of this year’s symposium is AI & Sustainability.

Abhishek Saroha

GAUSSIAN SPLATTING IN STYLE

3D scene stylization extends the work of neural style transfer to 3D. A vital challenge in this problem is to maintain the uniformity of the stylized appearance across multiple views. A vast majority of the previous works achieve this by training a 3D model for every stylized image and a set of multi-view images. In contrast, we propose a novel architecture trained on a collection of style images that, at test time, produces real time high-quality stylized novel views. We choose the underlying 3D scene representation for our model as 3D Gaussian splatting. We take the 3D Gaussians and process them using a multi-resolution hash grid and a tiny MLP to obtain stylized views. The MLP is conditioned on different style codes for generalization to different styles during test time. The explicit nature of 3D Gaussians gives us inherent advantages over NeRF-based methods, including geometric consistency and a fast training and rendering regime. This enables our method to be useful for various practical use cases, such as augmented or virtual reality. We demonstrate that our method achieves state-of-the-art performance with superior visual quality on various indoor and outdoor real-world data.

Abourayya Amr

LITTLE IS ENOUGH: IMPROVING PRIVACY BY SHARING LABELS IN FEDERATED SEMI-SUPERVISED LEARNING

In many critical applications, sensitive data is inherently distributed and cannot be centralized due to privacy concerns. A wide range of federated learning approaches have been proposed in the literature to train models locally at each client without sharing their sensitive local data. Most of these approaches either share local model parameters, soft predictions on a public dataset, or a combination of both. This, however, still discloses private information and restricts local models to those that lend themselves to training via gradient-based methods. To reduce the amount of shared information, we propose to share only hard labels on a public unlabeled dataset, and use a consensus over the shared labels as a pseudo-labeling to be used by clients. The resulting federated co-training approach empirically improves privacy substantially, without compromising on model quality. At the same time, it allows us to use local models that do not lend themselves to the parameter aggregation used in federated learning, such as (gradient boosted) decision trees, rule ensembles, and random forests.

Aditya Gulati

WHAT IS BEAUTIFUL IS STILL GOOD: THE ATTRACTIVENESS HALO EFFECT IN THE ERA OF BEAUTY FILTERS

The impact of cognitive biases on decision-making in the digital world remains under-explored despite its well-documented effects in physical contexts. This study addresses this gap by investigating the attractiveness halo effect using AI-based beauty filters. We conduct a large-scale online user study involving 2,748 participants who rated facial images from a diverse set of 462 distinct individuals in two conditions: original and attractive after applying a beauty filter.
Our study reveals that the same individuals receive statistically significantly higher ratings of attractiveness and other traits, such as intelligence and trustworthiness, in the attractive condition. We also study the impact of age, gender, and ethnicity and identify a weakening of the halo effect in the beautified condition, resolving conflicting findings from the literature and suggesting that filters could mitigate this cognitive bias. Finally, our findings raise ethical concerns regarding the use of beauty filters.

Ahmad Dawar Hakimi

CITANCE-CONTEXTUALIZED SUMMARIZATION OF SCIENTIFIC PAPERS

Current approaches to automatic summarization of scientific papers generate informative summaries in the form of abstracts. However, abstracts are not intended to show the relationship between a paper and the references cited in it. We propose a new contextualized summarization approach that can generate an informative summary conditioned on a given sentence containing the citation of a reference (a so-called "citance"). This summary outlines the content of the cited paper relevant to the citation location. Thus, our approach extracts and models the citances of a paper, retrieves relevant passages from cited papers, and generates abstractive summaries tailored to each citance. We evaluate our approach using Webis-Context-SciSumm-2023, a new dataset containing 540K~computer science papers and 4.6M~citances therein.

Alisa Sheinkman

VARIATIONAL BAYESIAN NEURAL NETWORKS WITH SHRINKAGE

Despite the dominant role of deep models in machine learning, limitations persist, including overconfident predictions, susceptibility to adversarial attacks, and underestimation of variability in predictions.

The Bayesian paradigm provides a natural framework to overcome such issues and has become the gold standard for uncertainty estimation with deep models, also providing improved accuracy and tuning of critical hyperparameters. However, exact Bayesian inference is challenging, typically involving variational algorithms that impose strong independence and distributional assumptions. Moreover, existing methods are sensitive to the architectural choice of the network. We address these issues and construct a relaxed version of the standard feed-forward rectified neural network, employing Polya-Gamma data augmentation tricks to render a conditionally linear and Gaussian model. Additionally, we use sparsity-promoting priors on the weights of the neural network for data-driven architectural design. To approximate the posterior, we derive a variational inference algorithm that avoids distributional assumptions and independence across layers and is a faster alternative to the usual Markov Chain Monte Carlo schemes.

Andre Cruz

EVALUATE CALIBRATION OF LANGUAGE MODELS WITH FOLKTEXTS

While large language models have increased dramatically in accuracy on numerous tasks, they are still lacking in their ability to express uncertainty about outcomes. Calibration is a fundamental form of uncertainty quantification. A calibrated risk score, on average, reflects the true frequency of outcomes in a population.

We introduce folktexts, a software package that provides datasets and tools to evaluate and benchmark the calibration properties of large language models. Our goal is to strengthen the evaluation ecosystem in a previously underserved direction, specifically, the systematic evaluation of uncertainty quantification in large language models.

We demonstrate the necessity and utility of our package through a large-scale evaluation of popular large language models. Our empirical results show that, despite having surprisingly strong predictive capabilities, model outputs are wildly miscalibrated.

Andreas Müller

THE IMPACT OF UNIFORM INPUTS ON ACTIVATION SPARSITY AND ENERGY-LATENCY ATTACKS IN COMPUTER VISION

Resource efficiency plays an important role for machine learning nowadays. The energy and decision latency are two critical aspects to ensure a sustainable and practical application.

Unfortunately, the energy consumption and decision latency are not robust against adversaries. Researchers have recently demonstrated that attackers can compute and submit so-called sponge examples at inference time to increase the energy consumption and decision latency of neural networks. In computer vision, the proposed strategy crafts inputs with less activation sparsity which could otherwise be used to accelerate the computation.

In this paper, we analyze the mechanism how these energy-latency attacks reduce activation sparsity. In particular, we find that input uniformity is a key enabler. A uniform image, that is, an image with mostly flat, uniformly colored surfaces, triggers more activations due to a specific interplay of convolution, batch normalization, and ReLU activation. Based on these insights, we propose two new simple, yet effective strategies for crafting sponge examples: sampling images from a probability distribution and identifying dense, yet inconspicuous inputs in natural datasets. We empirically examine our findings in a comprehensive evaluation with multiple image classification models and show that our attack achieves the same sparsity effect as prior sponge-example methods, but at a fraction of computation effort. We also show that our sponge examples transfer between different neural networks. Finally, we discuss applications of our findings for the good by improving efficiency by increasing sparsity.

Andrei Lixandru

FRODO: FRACTIONAL ORDER DISTRIBUTED OPTIMIZATION

Distributed optimization (DO) is critical in machine learning, especially for multi-agent coordination and federated learning for healthcare applications. DO involves networked agents, each with a local objective function, collaboratively optimizing a shared variable to minimize their collective functions. However, when these functions have ill-conditioned Hessians, DO struggles with optimal convergence rates. To address this, we introduce fractional calculus in DO to incorporate long-term memory, enhancing stability and convergence in scenarios with challenging objective functions. We assess the efficacy of this approach in several settings and compare it with other state-of-the-art methods.

Angel David Reyero Lobo

EXPLAINABLE AI THROUGH VARIABLE IMPORTANCE TEST

Study of a model-agnostic, statistically consistent variable importance measure with accommodation for unknown groups to handle highly-correlated covariates.

Animesh Awasthi

DEEP LEARNING-DRIVEN CELL STATE-SPECIFIC DNA DESIGN

Engineering immune cells is an emerging way to potentially treat diseases like cancer. One of the major challenges of such cell-based therapeutics is the design of specific functions and avoiding immune response. A promising approach to achieving specific functions is by inducing cell-state-specific gene expression of the inserted cassette with the desired gene. Cis-regulatory elements like enhancers and promoters regulate cell-state-specific gene expression. Cell-state specificity prevents indiscriminate expression of genes can lead to inefficacies and off-target effects. Current methods, predominantly built on bioinformatics analyses of over-expressed genes, struggle with the challenges of differentially expressed genes and enhancer-gene association, compounded by limitations in data availability for rare cell states.

Recently, deep learning methods have shown the potential to design cell-type/tissue-type specific enhancers. However, these methods require large-scale datasets and/or functional assays to design novel enhancers and hence have limited biological application. This project aims to introduce a novel, scalable deep-learning approach to design enhancers that are specific to cell states with limited data by leveraging multimodal integration of ATACseq and RNAseq data to fine-tune pre-trained models, allowing for the accurate decoding of unique regulatory patterns.

The approach is divided into three main objectives: Learning the regulatory code using genomic sequence deep-learning models, AI-driven DNA design using generative models, and validation of designed elements through experimental assays.

The proposed method aims to extend beyond typical cell type-specific design to accommodate dynamic states of cell differentiation or disease, thereby paving the way for more targeted and effective therapeutic interventions. By integrating advanced machine learning techniques with biological insights, this project seeks to advance the field of gene regulation and therapeutic DNA design.

Antonio Alliegro

3D VISUAL LEARNING FOR REAL-WORLD SCENARIOS

Recent advancements in deep learning, particularly in natural language and vision-language models, have brought us closer to creating artificial intelligence agents capable of conversing and interpreting their surroundings. Despite these advancements, the ability of these agents to navigate and interact seamlessly in real-world environments remains uncertain. This challenge is primarily due to the complex nature of 3D environments, which continue to pose significant challenges for autonomous systems.

ROne of the critical barriers to success in this field is the issue of distributional shifts, which occur when there is a significant disparity between an autonomous system's training environment and its deployment environment. These shifts can cause decreased performance and limit the system's applicability in scenarios where safety is paramount. Such shifts may involve changes in weather, lighting, style, or the introduction of new semantic concepts.

The collection of works presented herein tackles the pivotal issue of distributional shifts in 3D computer vision, with a focus on the domain and semantic variations that autonomous systems encounter in real-world 3D settings, where the variability and complexity far exceed those of controlled, synthetic training environments.
It begins with the formulation of a hybrid learning approach that integrates supervised and self-supervised learning to enhance robustness in cross-domain 3D shape recognition. This is particularly pertinent as 3D sensors frequently generate noisy and incomplete data.

Further, this work addresses point cloud completion challenges by developing a category-agnostic deep learning model that effectively generalizes to new semantic categories during testing. Moving beyond mere perception, the research proposes an end-to-end methodology for generating accurate grasp poses from real-world point clouds, demonstrating superior performance and generalization capabilities across unseen object categories.

Additionally, this work pioneers the areas of Semantic Novelty Detection and Open Set Recognition for 3D data. It establishes the first comprehensive benchmark complete with a dedicated testbed and critically examines the limitations of existing approaches, which are mainly designed for 2D images. A novel and efficient strategy for 3D Semantic Novelty Detection is also introduced, utilizing the representational power of large-scale pre-trained models to enable swift deployment in a variety of real-world applications without the need for a customized learning phase.

Altogether, this body of work lays a solid foundation for future research into managing 3D distributional shifts and significantly bridges the gap between theoretical advancements in 3D vision and their practical applications. The methodologies and insights offered are crucial steps toward developing a new generation of reliable, autonomous agents that can seamlessly operate in real-world environments.

Cajetan Emeka Oriekezie

ENHANCING THE ACCURACY AND EFFICIENCY OF LIFE CYCLE ASSESSMENT THROUGH MACHINE LEARNING: A COMPREHENSIVE REVIEW

Understanding and assessing the environmental impact of a product is difficult due to the complex nature of analysing the entire life cycle, involving various stages and data sources. In recent years, integrating machine learning (ML) into life cycle assessment (LCA) practices has grown significantly across different sectors.
However, existing reviews are outdated or narrowly focused on specific applications or phases, leading to a fragmented understanding of ML's potential in LCA practices.

This literature review aims to provide a comprehensive overview of ML's integration across all LCA phases and sectors, offering insight into the current state and future directions of ML-enhanced LCA methodologies. Current applications of ML in LCA include data integration, predictive modelling, scope optimisation, data collection and preparation, classification, enhanced predictive accuracy, scenario analysis, data augmentation, optimisation and decision support, aggregation and analysis, visualisation, uncertainty, and hotspot analysis.

The review found an increasing trend in ML applications for LCA decision-making, with the life cycle impact assessment category having the most publications. Despite promising developments, challenges exist in the widespread adoption of ML in LCA, including data-related issues such as availability, quality, heterogeneity, and uncertainty, as well as methodological challenges related to model selection, interpretability, and validation. Domain-specific challenges like integrating ML with existing LCA methods and addressing ethical considerations also remain significant obstacles.

Cecilia Curreli

NONISOTROPIC DIFFUSION FOR 3D HUMAN MOTION PREDICTION

The task of probabilistic human motion prediction aims to predict multiple future motions from past observations. Current diffusion approaches do not exploit the structured nature of the human skeleton, resulting in unrealistic motion sequences. We propose a novel nonisotropic diffusion model that defines the forward transition as an interpolation between isotropic noise and the correlation matrix of body joints. We preserve the notion of body joint in every layer of our network and design our architecture to explicitly exploit the skeleton graph structure, amplifying the impact of our nonisotropic formulation on the hierarchical latent variables. Our model achieves state-of-the-art performance on real-world benchmarks by generating realistic and diverse motions.

Célia Benquet

A UNIFIED ENCODER FOR MEASURING NEURAL DYNAMICS WITH SELF-SUPERVISED CONTRASTIVE LEARNING

Authors: Célia Benquet, Steffen Schneider, Mackenzie W. Mathis

Building encoding models that capture faithful representations of time series data, such as those recorded in the brain, is a challenging task. Ideally, one is able to develop unified encoder models that can be used in downstream tasks, such as neural decoding for brain machine interfaces. Datasets can be disjoint and collected from different animals, and the recordings are only a subspace of the full neural space, thus, building encoder models from multiple sessions is an attractive approach to represent the latent neural dynamics better.

Here, we build on CEBRA [1], a self-supervised framework for generalized contrastive learning with auxiliary variables. We developed a new sampling scheme, tied to the optimization function, and showed that this produces high-performance encoders that can be used in downstream decoding tasks. Our approach aligns information from multiple potentially noisy sessions with limited number of neurons, and integrates it into a unified embedding. We demonstrate that this unified embedding improved behavioral decoding performance over single sessions and a prior multi-session sampling scheme on two benchmark datasets [2, 3, 1], suggesting that our approach could be used for downstream decoding tasks in neuroscience.

References: [1] Schneider, Lee, and Mathis, Nature, 2023. [2] Grosmark & Buzś aki, Science, 2016. [3] de Vries et al., Nat. Neurosci., 2020.

Christian Koke

HOLONETS: SPECTRAL CONVOLUTIONS DO EXTEND TO DIRECTED GRAPHS

Within the graph learning community, conventional wisdom dictates that spectral convolutional networks may only be deployed on undirected graphs: Only there could the existence of a well-defined graph Fourier transform be guaranteed, so that information may be translated between spatial- and spectral domains. Here we show this traditional reliance on the graph Fourier transform to be superfluous and -- making use of certain advanced tools from complex analysis and spectral theory -- extend spectral convolutions to directed graphs.

We provide a frequency-response interpretation of newly developed filters, investigate the influence of the basis used to express filters and discuss the interplay with characteristic operators on which networks are based. In order to thoroughly test the developed theory, we conduct experiments in real world settings, showcasing that directed spectral convolutional networks provide new state of the art results for heterophilic node classification on many datasets and -- as opposed to baselines -- may be rendered stable to resolution-scale varying topological perturbations.

Daniel Musekamp

ACTIVE LEARNING FOR NEURAL PDE SOLVERS

Numerically solving Partial Differential Equations (PDEs) is a fundamental task in physical sciences. While neural PDE solvers can be very efficient, they often require costly upfront training data generation through classical numerical solvers. By querying these solvers with more informative initial conditions and PDE parameters, Active Learning (AL) promises to reach the same accuracy with smaller training sets and, consequently, less upfront costs. While AL is common in other domains of Scientific ML, it has yet to be studied extensively for autoregressive neural PDE solvers. To address this gap, we introduce a modular and extensible active learning benchmark called AL4PDE. The benchmark adopts a range of parametric PDEs for the solver-in-the-loop setting as well as modern neural operators to enable easy evaluation of AL methods. We apply multiple batch AL algorithms, ranging from classical Query-by-Committee to feature-based selection strategies such as largest cluster maximum distance (LCMD). Our experiments show that AL can reduce the average error by up to 65% compared to random sampling, especially improving the worst-case error. Under the investigated approaches, stochastic batch AL and LCMD prove to be the most effective. AL is reliable regarding the distribution of the selected inputs and produces reusable datasets, which are also advantageous
for models that are not used to select the inputs.

Dimitris Michailidis

FAIRNESS IN URBAN TRANSPORT DESIGN USING REINFORCEMENT LEARNING

Public transport networks are the foundation of modern urban living. Designing transport networks, however, is a complex task that involves physical, social, political, and legal constraints. This complexity is further compounded when considering the trade-off between efficiency and fairness. Efficient routes can enhance ridership and decrease reliance on cars, thereby promoting environmental sustainability. However, they might also favour densely populated central areas, potentially neglecting other underserved communities and intensifying existing inequalities. It is therefore crucial to develop tools that address these challenges and prioritize fairness. Recent advancements in Artificial Intelligence (AI) provide promising solutions. My research focuses on employing Reinforcement Learning to train planning agents capable of generating public transport networks while optimizing for different notions of fairness and social good.

Edoardo Caldarelli

HETEROSCEDASTIC GAUSSIAN PROCESSES AND RANDOM FEATURES: SCALABLE MOTION PRIMITIVES WITH GUARANTEES

Heteroscedastic Gaussian processes (HGPs) are kernel-based, non-parametric models that can be used to infer nonlinear functions with time-varying noise. In robotics, they can be employed for learning from demonstration as motion primitives, ie as a model of the trajectories to be executed by the robot. HGPs provide variance estimates around the reference signal modeling the trajectory, capturing both the predictive uncertainty and the motion variability. However, similarly to standard Gaussian processes they suffer from a cubic complexity in the number of training points, due to the inversion of the kernel matrix. The uncertainty can be leveraged for more complex learning tasks, such as inferring the variable impedance profile required from a robotic manipulator. However, suitable approximations are needed to make HGPs scalable, at the price of potentially worsening the posterior mean and variance profiles. Motivated by these observations, we study the combination of HGPs and random features, which are a popular, data-independent approximation strategy of kernel functions. In a theoretical analysis, we provide novel guarantees on the approximation error of the HGP posterior due to random features. Moreover, we validate this scalable motion primitive on real robot data, related to the problem of variable impedance learning. In this way, we show that random features offer a viable and theoretically sound alternative for speeding up the trajectory processing, without sacrificing accuracy.

Elies Gil-Fuster

ON THE RELATION BETWEEN TRAINABILITY AND DEQUANTIZATION OF VARIATIONAL QUANTUM LEARNING MODELS

One hurdle in the design of parametrized quantum circuits is to find trainable Ansätze, as it is known that it is hard to train one-size-fits-all variational algorithms. The standard way to avoid this issue is to be very deliberate in the design of quantum circuits. However, in the context of variational optimization algorithms for parametrized quantum circuits, e.g. the variational quantum eigensolver, recent works have shown that very often the same conditions that ensure trainability can be used to prove that the same parametrized quantum circuits are efficiently classically simulable. These results raise the question whether, or to what extent, trainability implies dequantization in variational quantum algorithms in general. We formalise this question in the context of machine learning tasks.

We highlight the pivotal role of the precise notions of trainability and dequantization, we resolve the relationship, and prove that trainability does not imply dequantization in variational quantum machine learning, even when training via the most standard methods, i.e. gradient descent. Results of this type had been proven for quantum kernel methods, but these are not illustrative when compared to truly variational Ansätze, as kernel-based machine learning is arguably non-variational. Thus, we discuss the degrees of variationalness of different quantum learning models, and prove that trainability does not imply dequantization also for genuinely variational models.

Erfan Mirzaei

EMPIRICAL BERNSTEIN INEQUALITY FOR LEARNING DYNAMICAL SYSTEMS

Many data-driven algorithms have been used to study dynamical systems by learning the corresponding transfer, or Koopman operator, but their relationship with statistical learning is largely unexplored. We formalize a framework to learn the Koopman operator from finite data trajectories of the dynamical system, modeled as a Markov Chain. We consider the restriction of this operator to a reproducing kernel Hilbert space and introduce a notion of risk, from which different estimators naturally arise. We link the risk with the estimation of the spectral decomposition of the Koopman operator. Since sampled data in the general case are non-independent and non-identically distributed, a non-iid statistical learning analysis is needed. Thus, we introduce a novel Bernstein inequality tailored specifically for random variables in Hilbert spaces. This inequality exploits mixing coefficients and employs the method of blocks to leverage the rapid decrease in correlation between temporally separated variables. Additionally, we present two data-dependent inequalities applicable to non-stationary and stationary processes, respectively.

We demonstrate the utility of these bounds by showcasing their application in estimating covariance operators in the Hilbert-Schmidt norm, a topic of potential independent interest. Moreover, these bounds are relevant for statistical learning of transfer operators, particularly when access to samples from invariant distributions is unfeasible. Finally, we conduct numerical experiments to illustrate the practical implications of these bounds for both applications.

Eryng Luca

RENO: ENHANCING ONE-STEP TEXT-TO-IMAGE MODELS THROUGH REWARD-BASED NOISE OPTIMIZATION

Text-to-Image (T2I) models have made significant advancements in recent years, but they still struggle to accurately capture intricate details specified in complex compositional prompts. While fine-tuning T2I models with reward objectives has shown promise, it suffers from "reward hacking" and may not generalize well to unseen prompt distributions. In this work, we propose \textbf{Re}ward-based \textbf{N}oise \textbf{O}ptimization (\textbf{ReNO}), a novel approach that enhances T2I models at inference by optimizing the initial noise based on the signal from one or multiple human preference reward models. Remarkably, solving this optimization problem with gradient ascent for 50 iterations yields impressive results on four different one-step models across two competitive benchmarks, T2I-CompBench and GenEval. Within a computational budget of 20-50 seconds, ReNO-enhanced one-step models consistently surpass the performance of all current open-source Text-to-Image models. Extensive user studies demonstrate that our model is preferred nearly twice as often compared to the popular SDXL model and is on par with the proprietary Stable Diffusion 3 with 8B parameters. Moreover, given the same computational resources, a ReNO-optimized one-step model outperforms widely-used open-source models such as SDXL and PixArt-α, highlighting the efficiency and effectiveness of ReNO in enhancing T2I model performance at inference time.

Federico Tiblias

QUANTUM HAMILTONIAN ENCODING FOR CLASSIFICATION

Quantum computing holds great promise for expanding the realm of efficiently solvable problems due to the theoretical advantages of some quantum algorithms over their classical counterparts in terms of runtime and resource usage. Quantum machine learning (QML) seeks to extend these advantages to data-driven methods.

Despite evidence suggesting potential improvements, the development of practically advantageous QML algorithms is hindered by hardware limitations and the high costs of data manipulation on current Near Intermediate Scale Quantum (NISQ) devices.
These limitations restrict researchers to small-scale tasks and minimal datasets, impeding broader insights into QML performance. In response, we propose an efficient scheme for encoding and measuring dense vector representations and show its effectiveness on simple but realistic classification tasks.
Inspired by quantum chemistry's ground state energy optimization, our Hamiltonian encoding uses inputs to determine a finite set of Pauli strings for measurements, achieving linear sample complexity. We additionally introduce two variants with different scaling in terms of parameters and sample complexity.

Felix Sarnthein

ON REPRESENTATION LEARNING IN SELF-DISTILLATION WITH NO LABELS

In this project, I explore the learning dynamics induced by the self-distillation component in the self-supervised learning method DINO. I investigate this by distilling a randomly initialized teacher network into a student network. Surprisingly, the student’s internal representations improve over the teacher’s according to the linear classification accuracy on a labeled downstream task. I further analyze this phenomenon under the light of supervised fine-tuning. The results suggest that label information might be redundant for certain aspects of the learning dynamics in deep neural networks.

Filip Szatkowski

D2DMOE: EXPLOITING ACTIVATION SPARSITY WITH DENSE TO DYNAMIC-K MIXTURE-OF-EXPERTS CONVERSION

Transformers are powerful but computationally expensive. At the same time, previous works demonstrated that these models exhibit significant activation sparsity, which suggests that most of their computations are redundant. As first proposed in MoEfication, activation sparsity can be leveraged to reduce inference costs by converting network parts into equivalent MoE layers. This is done by grouping similar FFN weights into experts and training router networks with a subset of training data. In our work, we identify a series of weaknesses of the MoEfication process limiting the resulting cost-performance trade-off and propose the corresponding mitigation steps.

Florian Hübler

TUNING-FREE OPTIMIZATION UNDER RELAXED SMOOTHNESS

Tuning hyperparameters, such as the stepsize, presents a major challenge of training machine learning models. To address this challenge, numerous adaptive optimization algorithms have been developed that achieve near-optimal complexities, even when stepsizes are independent of problem-specific parameters, provided that the loss function is L-smooth. However, as the assumption is relaxed to the more realistic (L_0, L_1)-smoothness, all existing convergence results still necessitate tuning of the stepsize. In this study, we demonstrate that Normalized Stochastic Gradient Descent with Momentum (NSGD-M) can achieve a (nearly) rate-optimal complexity without prior knowledge of any problem parameter, though this comes at the cost of introducing an exponential term dependent on L_1 in the complexity. We further establish that this exponential term is inevitable to such schemes by introducing a theoretical framework of lower bounds tailored explicitly for parameter-agnostic algorithms. Interestingly, in deterministic settings, the exponential factor can be neutralized by employing Gradient Descent with a Backtracking Line Search. To the best of our knowledge, these findings represent the first parameter-agnostic convergence results under the generalized smoothness condition. Our empirical experiments further confirm our theoretical insights.

Francesca Drummer

LONG-RANGE INTERACTION MODELLING OF HETEROGENOUS NICHES WITH GRAPH TRANSFORMERS

Spatial transcriptomics has revolutionized our understanding of cellular interactions by providing spatial context to gene expression data. This spatially resolved approach allows for the incorporation of cell proximity into cell-cell communication (CCC) models, accelerating research about CCC within complex heterogenous biological systems. We propose a novel method, the GNNTransformer, to address the limitations of existing CCC models that are restricted to model intercellular signalling within local neighbourhoods. Our method integrates graph neural networks (GNNs) with transformer-based self-attention mechanisms to capture both local and global interactions among cells. The GNN component focuses on capturing local interactions within the spatial graph, while the transformer component models long-range interactions, allowing for the consideration of cell interactions across varying spatial distances. We show that our GNNTransformer model improves its ability to capture diverse cell-cell interactions and identify heterogeneous niches within tissues on multiple spatial transcriptomics techniques and tissues. For instance, our analysis revealed that different cell types exhibit signalling patterns at various spatial scales, ranging from juxtracrine to paracrine signalling spanning distances from 0 µm to 100 µm.

Gbetondji Jean-Sebastien Alexandre Dovonon

SETTING THE RECORD STRAIGHT ON TRANSFORMER OVERSMOOTHING

Transformer-based models have recently become wildly successful across a diverse set of domains. At the same time, recent work has shown empirically and theoretically that Transformers are inherently limited. Specifically, they argue that as model depth increases all features become more and more similar. A natural question is: How can Transformers achieve these successes given this shortcoming? In this work we test these observations empirically and theoretically and uncover a number of surprising findings. We find that there are cases where feature similarity increases but, contrary to prior results, this is not inevitable, even for existing pre-trained models. Theoretically, we show that smoothing behavior depends on the eigenspectrum of the value and projection weights and potentially the sign of the layer normalization weights (if using Pre-LN). Our analysis reveals a simple way to parameterize the weights of the Transformer update equations to influence smoothing behavior. We hope that our findings give ML researchers and practitioners additional insight into how to develop future Transformer models.

Geng Mingmeng

IS CHATGPT TRANSFORMING ACADEMICS' WRITING STYLE?

Based on one million arXiv papers submitted from May 2018 to January 2024, we assess the textual density of ChatGPT's writing style in their abstracts by means of a statistical analysis of word frequency changes. Our model is calibrated and validated on a mixture of real abstracts and ChatGPT-modified abstracts (simulated data) after a careful noise analysis. We find that ChatGPT is having an increasing impact on arXiv abstracts, especially in the field of computer science, where the fraction of ChatGPT style abstracts is estimated to be approximately 35%, if we take the output of one of the simplest prompts, "revise the following sentences", as a baseline. We conclude with an analysis of both positive and negative aspects of the penetration of ChatGPT into academics' writing style.

Gergely Dániel Németh

OBSERVATIONS OF USING MODEL COMPLEXITY REDUCTION AS A DEFENSE AGAINST MEMBERSHIP RETRIEVAL

Federated Learning (FL) has been proposed as a privacy-preserving solution for machine learning. However, recent works have reported that FL can leak private client data through membership inference attacks. In this paper, we show that the effectiveness of these attacks on the clients negatively correlates with the size of the client's datasets and model complexity. Based on this finding, we study the capabilities of model-agnostic Federated Learning to preserve privacy, as it enables the use of models of varying complexity in the clients.

To systematically study this topic, we first propose a taxonomy of model-agnostic FL methods according to the strategies adopted by the clients to select the sub-models from the server's model. This taxonomy provides a framework for existing model-agnostic FL approaches and leads to the proposal of new FL methods to fill the gaps in the taxonomy. Next, we analyze the privacy-performance trade-off of all the model-agnostic FL architectures as per the proposed taxonomy when subjected to 3 different membership inference attacks on the CIFAR-10 and CIFAR-100 vision datasets. In our experiments, we find that randomness in the strategy used to select the server's sub-model to train the clients' models can control the clients' privacy while keeping competitive performance on the server's side.

Griselda Vilar Sastre

APPROACH TO THE REGULATORY FRAMEWORK OF AI IN THE EU: IMPLICATIONS AND OPPORTUNITIES FOR INNOVATION AND RESEARCH IN THE IMMEDIATE HORIZON

This article examines the regulatory framework of the European Union (EU) on artificial intelligence (AI), machine learning, and Big Data, focusing on personal data protection. The objective is to clearly determine the regulatory margins of the EU for companies and research groups to use AI tools in their innovative processes and knowledge contribution. The General Data Protection Regulation (GDPR) of 2018 establishes strict rules on the processing of personal data, emphasizing explicit consent, transparency, and the right not to be subjected to automated decisions (GDPR, 2018).

The European Commission complements the GDPR with the "White Paper on Artificial Intelligence" (2020) and the proposed "AI Act" in 2021, promoting transparency and accountability in AI (European Commission, 2020; 2021). The "Recommendation on Artificial Intelligence and Data Protection" by the Council of Europe (2021) underscores the need for impact assessments and supervision mechanisms (Council of Europe, 2021).

In 2023, the EU reached a provisional agreement on the "AI Act," establishing a comprehensive legal framework for AI. This agreement categorizes AI systems based on their risk and includes strict requirements for high-impact AI models, such as risk assessments and the obligation to report serious incidents (Mayer Brown, 2023; World Economic Forum, 2023). In 2024, the European Council formally adopted the "AI Act," regulating aspects such as transparency, developer and provider responsibility, and the conditions for using AI in public spaces (Artificial Intelligence Act, 2024).

This article also analyzes the use of emerging tools in social, communication, arts, educational, and health fields. In education, AI is revolutionizing learning processes by providing personalized learning experiences (Smith, 2022). In the healthcare sector, AI and Big Data are improving patient outcomes and operational efficiencies (Johnson et al., 2023). In the communication sector, AI is transforming the way content is managed and distributed, improving personalization and efficiency in digital communication (Lee & Martin, 2023; Brown et al., 2023). Additionally, the positive view of AI in research is highlighted, showing how these technologies can drive innovation and efficiency in various fields (AI Index Report, 2024; Nature, 2023).

Guanhua Zhang

INHERENT TRADE-OFFS BETWEEN DIVERSITY AND STABILITY IN MULTI-TASK BENCHMARKS

We examine multi-task benchmarks in machine learning through the lens of social choice theory. We draw an analogy between benchmarks and electoral systems, where models are candidates and tasks are voters. This suggests a distinction between cardinal and ordinal benchmark systems. The former aggregate numerical scores into one model ranking; the latter aggregate rankings for each task. We apply Arrow's impossibility theorem to ordinal benchmarks to highlight the inherent limitations of ordinal systems, particularly their sensitivity to the inclusion of irrelevant models. Inspired by Arrow's theorem, we empirically demonstrate a strong trade-off between diversity and sensitivity to irrelevant changes in existing multi-task benchmarks. Our result is based on new quantitative measures of diversity and sensitivity that we introduce. Sensitivity quantifies the impact that irrelevant changes to tasks have on a benchmark. Diversity captures the degree of disagreement in model rankings across tasks. We develop efficient approximation algorithms for both measures, as exact computation is computationally challenging. Through extensive experiments on seven cardinal benchmarks and eleven ordinal benchmarks, we demonstrate a clear trade-off between diversity and stability:

The more diverse a multi-task benchmark, the more sensitive to trivial changes it is. Additionally, we show that the aggregated rankings of existing benchmarks are highly unstable under irrelevant changes. The codes and data are available at https://socialfoundations.github.io/benchbench/.

Hazel Kim

MODEL ANSWERABILITY

Do LLMs behave differently when the model does not know or is not confident about their answers than when they are certain about their answers? Do LLM behaviors have different patterns? Do LLM behaviors provide us clues if the model produces answers without confidence? We explore model behaviors when the given context is not enough for models to answer questions. Because the model answers do not offer us information how much the answers are trustworthy, this analysis would help us understand models better for their confidence in themselves. Both answerable and unanswerable questions have the similar feature similarities with the context, it seems non-trivial to predict distinguishable behaviors of models that heavily rely on self-attention that operates by correlations or similarities of tokens.

Ingo Ziegler

CRAFT YOUR DATASET: TASK-SPECIFIC SYNTHETIC DATASET GENERATION THROUGH CORPUS RETRIEVAL AND AUGMENTATION

Building high-quality datasets for specialized tasks remains time-consuming and resource-intensive. To address this challenge, we propose CRAFT (Corpus Retrieval and Augmentation for Fine-Tuning), a method that generates task-specific synthetic datasets. Our approach requires a minimal number of user-written few-shots that demonstrate the task to be performed. Subsequently, we leverage existing large-scale pre-training corpora and similarity-based document retrieval to find other relevant human-written documents. Lastly, instruction-tuned large language models (LLMs) are employed to augment free-text documents into custom-formatted task samples, which can readily be used for fine-tuning. We demonstrate that CRAFT can efficiently generate large-scale task-specific training datasets for 4 diverse tasks: biology, medical, and commonsense question answering, as well as summarization. Our experiments show that CRAFT-based models outperform or achieve comparable performance to general LLMs for classification tasks, while enabling CRAFT-based summarization models to outperform models trained on human-annotated data by 46 preference points.

Jan Schneider

IDENTIFYING POLICY GRADIENT SUBSPACES

Policy gradient methods hold great potential for solving complex continuous control tasks. Still, their training efficiency can be improved by exploiting structure within the optimization problem. Recent work indicates that supervised learning can be accelerated by leveraging the fact that gradients lie in a low-dimensional and slowly-changing subspace. In this paper, we conduct a thorough evaluation of this phenomenon for two popular deep policy gradient methods on various simulated benchmark tasks. Our results demonstrate the existence of such gradient subspaces despite the continuously changing data distribution inherent to reinforcement learning. These findings reveal promising directions for future work on more efficient reinforcement learning, e.g., through improving parameter-space exploration or enabling second-order optimization.

Jana Zeller

FULL CIRCLE: LEARNING VISUAL PROMPTS FOR VLMS

Vision-language models (VLMs) like CLIP are highly effective at zero-shot image understanding. However, they struggle with referring expression comprehension, a key task in computer vision, largely due to their training on multimodal datasets that primarily include captions without explicit spatial references. Manual visual prompts, such as highlighting objects with red circles, have shown success in guiding these models' attention but require significant domain expertise and manual effort. This paper introduces an automatic visual prompting method, which optimizes a patch-independent prompt using the model's gradients. In a self-supervised setting, our method achieves up to an 8.2% improvement over previous optimised prompting techniques. In a supervised setting, our learned prompt surpasses current ensemble methods.

José Maria Pombal

TOWER: AN OPEN MULTILINGUAL LARGE LANGUAGE MODEL FOR TRANSLATION-RELATED TASKS

While general-purpose large language models (LLMs) demonstrate proficiency on multiple tasks within the domain of translation, approaches based on open LLMs are competitive only when specializing on a single task. In this paper, we propose a recipe for tailoring LLMs to multiple tasks present in translation workflows.

We perform continued pretraining on a multilingual mixture of monolingual and parallel data, creating TowerBase, followed by finetuning on instructions relevant for translation processes, creating TowerInstruct. Our final model surpasses open alternatives on several tasks relevant to translation workflows and is competitive with general-purpose closed LLMs. To facilitate future research, we release the Tower models, our specialization dataset, an evaluation framework for LLMs focusing on the translation ecosystem, and a collection of model generations, including ours, on our benchmark.

Julien COLIN

LOCAL VS DISTRIBUTED REPRESENTATIONS: WHAT IS THE RIGHT BASIS FOR INTERPRETABILITY?

Much of the interpretability of deep neural networks has focused on methods to interpret the visual features that maximally activate individual neurons. However, recent work has cast doubts on the usefulness of these local representations to study the behavior of deep neural networks. Dictionary learning methods have been proposed as a promising direction to recover sparsely distributed vector representations instead of local ones. Yet, it remains an open question whether feature visualizations derived from these methods constitute a better basis for interpretability than the ones derived from single neuron activation. To answer this question, we conducted three psychophysics experiments where we collected 15,720 responses from a pool of 560 participants. We compared the suitability of maximally activating stimuli derived from local representations (encoded in individual neurons) vs. distributed representations (encoded via populations of neurons) to support the interpretability of deep neural networks. We find strong evidence in support of distributed representations learned via sparse dictionary learning methods as a superior basis for interpretability, especially when interpreting the deepest layer of a neural network. Our results suggest a need for the field to move past the interpretation of local neural codes in favor of sparsely distributed ones.

Kairan Zhao

WHAT MAKES UNLEARNING HARD AND WHAT TO DO ABOUT IT

Machine unlearning is the problem of removing the effect of a subset of training data (the "forget set") from a trained model without damaging the model's utility e.g. to comply with users' requests to delete their data, or remove mislabeled, poisoned or otherwise problematic data.

With unlearning research still being at its infancy, many fundamental open questions exist: Are there interpretable characteristics of forget sets that substantially affect the difficulty of the problem? How do these characteristics affect different state-of-the-art algorithms? With this paper, we present the first investigation aiming to answer these questions. We identify two key factors affecting unlearning difficulty and the performance of unlearning algorithms. Evaluation on forget sets that isolate these identified factors reveals previously-unknown behaviours of state-of-the-art algorithms that don't materialize on random forget sets. Based on our insights, we develop a framework coined Refined-Unlearning Meta-algorithm (RUM) that encompasses: (i) refining the forget set into homogenized subsets, according to different characteristics; and (ii) a meta-algorithm that employs existing algorithms to unlearn each subset and finally delivers a model that has unlearned the overall forget set. We find that RUM substantially improves top-performing unlearning algorithms. Overall, we view our work as an important step in (i) deepening our scientific understanding of unlearning and (ii) revealing new pathways to improving the state-of-the-art.

Kajetan Schweighofer

ON INFORMATION-THEORETIC MEASURES OF PREDICTIVE UNCERTAINTY

Accurately estimating predictive uncertainty is crucial for machine learning applications, particularly in safety-critical scenarios where knowledge about risks is essential. Despite its importance, there is no consensus on how to correctly measure predictive uncertainty. Different information-theoretic measures have been proposed, each claiming to accurately capture the predictive uncertainty. We address this discrepancy by positing that all of these measures are valid under different assumptions and propose a unified framework for their categorization. Our framework categorizes predictive uncertainty measures according to two simple questions: (1) Which model predicts? (2) How is the true model approximated? By exploring the possible combinations of answers to these questions, we derive various measures of predictive uncertainty applicable to different scenarios, encompassing existing measures and introducing new ones. We conduct a comprehensive quantitative analysis of these measures in common uncertainty estimation settings to validate our proposed framework. Our findings indicate, that some measures are more suitable than others depending on (a) the specific task and (b) the method to sample models from the posterior. Thus, our work offers guidance on how to assess uncertainty under constraints of a given problem at hand.

Karnik Ram

ACCELERATING MOLECULAR SIMULATION USING NEURAL FREE-ENERGY FUNCTIONALS

Molecular simulation is a fundamental and powerful tool in chemical physics, but in many cases it can be too computationally expensive. We show how we can leverage the framework of dynamical density functional theory from statical mechanics, together with deep learning, to accelerate these calculations. We show preliminary results of our approach in comparison with molecular dynamic simulations, towards enabling high-throughout computational screening for novel material discovery in the context of carbon capture.

Kirtan Padh

YOUR ASSUMED DAG IS WRONG AND HERE'S HOW TO DEAL WITH IT

The starting point for cause-effect estimation is typically to assume that the underlying causal graph is known. Existing literature commonly points to domain experts or causal discovery algorithms to provide the causal graph. In practice, neither may have perfect confidence: causal discovery can be brittle or provide only a Markov equivalence class of directed acyclic graphs (DAGs), and there is usually dispute among experts about the absence of certain edges. Exhaustive enumeration of all plausible DAGs quickly becomes infeasible due to super-exponential growth. We develop a framework to bound causal queries from data when there is uncertainty in the underlying DAG, i.e., when the presence of some edges is unknown. In synthetic and real-world experiments, we demonstrate that our bounds cover the true value of common causal queries, such as the average treatment effect. Our approach is an easy-to-use and widely applicable rebuttal to the valid critique of `What if you assumed the wrong DAG?'

Korbinian Pöppel

XLSTM: EXTENDED LONG SHORT-TERM MEMORY

In the 1990s, the constant error carousel and gating were introduced as the central ideas of the Long Short-Term Memory (LSTM). Since then, LSTMs have stood the test of time and contributed to numerous deep learning success stories, in particular they constituted the first Large Language Models (LLMs).

However, the advent of the Transformer technology with parallelizable self-attention at its core marked the dawn of a new era, outpacing LSTMs at scale. We now raise a simple question: How far do we get in language model- ing when scaling LSTMs to billions of parameters, leveraging the latest techniques from modern LLMs, but mitigating known limitations of LSTMs? Firstly, we introduce exponential gating with appropriate normalization and stabilization techniques. Secondly, we modify the LSTM memory structure, obtaining: (i) sLSTM with a scalar memory, a scalar update, and new memory mixing, (ii) mLSTM that is fully parallelizable with amatrix memory and a covariance update rule. Integrating these LSTM extensions into residual block backbones yields xLSTM blocks that are then residually stacked into xLSTM architectures. Exponential gating and modified memory structures boost xLSTM capabilities to perform favorably when compared to state-of-the-art Transformers and State Space Models, both in performance and scaling.

Linara Adilova

INFORMATION-THEORETIC ANALYSIS OF GENERALIZATION IN DEEP LEARNING

We employ an information-theoretic analysis in the form of information planes to deep neural networks in order to predict its generalization, and potentially to optimize its performance and size. We select an approach for mutual information estimation in the context of deep neural networks, which is challenging due to the high dimensionality and computational complexity. Using this approach, the next step is to analyze the capabilities of the information plane analysis to understand a network's generalization through a large but carefully selected set of experiments. We finally propose and evaluate an approach for optimizing network architectures for a given task.

Lucas Resck

EXPLORING THE TRADE-OFF BETWEEN MODEL PERFORMANCE AND EXPLANATION PLAUSIBILITY OF TEXT CLASSIFIERS USING HUMAN RATIONALES (NAACL FINDINGS 2024)

Saliency post-hoc explainability methods are important tools for understanding increasingly complex NLP models. While these methods can reflect the model’s reasoning, they may not align with human intuition, making the explanations not plausible. In this work, we present a methodology for incorporating rationales, which are text annotations explaining human decisions, into text classification models.

This incorporation enhances the plausibility of post-hoc explanations while preserving their faithfulness. Our approach is agnostic to model architectures and explainability methods. We introduce the rationales during model training by augmenting the standard cross-entropy loss with a novel loss function inspired by contrastive learning. By leveraging a multi-objective optimization algorithm, we explore the trade-off between the two loss functions and generate a Pareto-optimal frontier of models that balance performance and plausibility. Through extensive experiments involving diverse models, datasets, and explainability methods, we demonstrate that our approach significantly enhances the quality of model explanations without causing substantial (sometimes negligible) degradation in the original model's performance.

Lucas Schorling

META-LEARNING OF QUANTUM PROCESSES

To enable several quantum technologies, quantum systems need to be controlled for which no good models exist. This is further complicated by large device variabilities, in particularly for spin-based

quantum computers. In this work, we present the power of common, but adapted, meta-learning algorithms to predict the dynamics of unknown quantum processes. We train the machine-learning

algorithm on several instances of quantum processes from the same class with simulated data. After accessing very little data for new unseen system instances of the same class, we predict the

dynamics for those new systems. We benchmark this method for three different quantum systems with the unchanged meta-learning algorithm, a vanilla transformer, and a Multi-layer-perceptron.

Lucas Ventura

COVR: COMPOSED VIDEO RETRIEVAL LEARNING COMPOSED VIDEO RETRIEVAL FROM WEB VIDEO CAPTIONS

Composed Image Retrieval (CoIR) has recently gained popularity as a task that considers both text and image queries together, to search for relevant images in a database. Most CoIR approaches require manually annotated datasets, comprising image-text-image triplets, where the text describes a modification from the query image to the target image. However, manual curation of CoIR triplets is expensive and prevents scalability.

In this work, we instead propose a scalable automatic dataset creation methodology that generates triplets given video-caption pairs, while also expanding the scope of the task to include composed video retrieval (CoVR). To this end, we mine paired videos with a similar caption from a large database, and leverage a large language model to generate the corresponding modification text. Applying this methodology to the extensive WebVid2M collection, we automatically construct our WebVid-CoVR dataset, resulting in 1.6 million triplets. Moreover, we introduce a new benchmark for CoVR with a manually annotated evaluation set, along with baseline results. Our experiments further demonstrate that training a CoVR model on our dataset effectively transfers to CoIR, leading to improved state-of-the-art performance in the zero-shot setup on both the CIRR and FashionIQ benchmarks. Our code, datasets, and models are publicly available.

Lucile Alys Favero Montero

ENHANCING CRITICAL THINKING IN EDUCATION BY MEANS OF A SOCRATIC CHATBOT

While large language models (LLMs) are increasingly playing a pivotal role in education by providing instantaneous, adaptive responses, their potential to promote critical thinking remains understudied. In this paper, we fill such a gap and present an innovative educational chatbot designed to foster critical thinking through Socratic questioning. Unlike traditional intelligent tutoring systems, including educational chatbots, that tend to offer direct answers, the proposed Socratic tutor encourages students to explore various perspectives and engage in self-reflection by posing structured, thought-provoking questions. Our Socratic questioning is implemented by fine and prompt-tuning the open-source pre-trained LLM with a specialized dataset that stimulates critical thinking and offers multiple viewpoints. In an effort to democratize access and to protect the students' privacy, the proposed tutor is based on small LLMs (Llama2 7B and 13B-parameter models) that are able to run locally on off-the-shelf hardware. We validate our approach in a battery of experiments consisting of interactions between a simulated student and the chatbot to evaluate its effectiveness in enhancing critical thinking skills. Results indicate that the Socratic tutor supports the development of reflection and critical thinking significantly better than standard chatbots. Our approach opens the door for improving educational outcomes by cultivating active learning and encouraging intellectual autonomy.

Lukas Aichberger

SEMANTICALLY DIVERSE LANGUAGE GENERATION FOR UNCERTAINTY ESTIMATION IN LANGUAGE MODELS

Large language models (LLMs) can suffer from hallucinations when generating text. These hallucinations impede various applications in society and industry by making LLMs untrustworthy. Current LLMs generate text in an autoregressive fashion by predicting and appending text tokens. When an LLM is uncertain about the semantic meaning of the next tokens to generate, it is likely to start hallucinating. Thus, it has been suggested that hallucinations stem from predictive uncertainty. We introduce Semantic Diverse Language Generation (SDLG) to quantify predictive uncertainty in LLMs. SDLG steers the LLM to generate semantically diverse yet likely alternatives for an initially generated text. This approach provides a precise measure of aleatoric semantic uncertainty, detecting whether the initial text is likely to be hallucinated. Experiments on question-answering tasks demonstrate that SDLG consistently outperforms existing methods while being the most computationally efficient, setting a new standard for uncertainty estimation in LLMs.

Maciej Pióro

STATE SOUP: IN-CONTEXT SKILL LEARNING, RETRIEVAL AND MIXING

A new breed of gated-linear recurrent neural networks has reached state-of-the-art performance on a range of sequence modeling problems. Such models naturally handle long sequences efficiently, as the cost of processing a new input is independent of sequence length. Here, we explore another advantage of these stateful sequence models, inspired by the success of model merging through parameter interpolation. Building on parallels between fine-tuning and in-context learning, we investigate whether we can treat internal states as task vectors that can be stored, retrieved, and then linearly combined, exploiting the linearity of recurrence. We study this form of fast model merging on Mamba-2.8b, a pretrained recurrent model, and present preliminary evidence that simple linear state interpolation methods suffice to improve next-token perplexity as well as downstream in-context learning task performance.

Maris Galesloot

LEARNING MEMORY-BASED POLICIES FOR ROBUST POMDPS

Robust partially observable Markov decision processes (robust POMDPs) extend classical POMDPs by accounting for additional uncertainty on the transition and observation probabilities via so-called uncertainty sets. Policies for robust POMDPs must not only be memory-based to account for partial observability, but they must also be robust against model uncertainty by assuming an adversarial selection of all potential probabilities from the uncertainty sets. We propose rFSCNet: an iterative framework to find such robust memory-based policies for robust POMDPs. rFSCNet has three key ingredients: (1) select an adversarial (non-robust) POMDP via worst-case probability instances from the uncertainty sets; (2) compute a finite-state controller (FSC) as a memory-based policy for this adversarial POMDP; (3) evaluate the performance of this FSC on the original robust POMDP and use this evaluation to derive a new adversarial (non-robust) POMDP. In each iteration, we find FSCs through a recurrent neural network that we train using supervision policies derived for these adversarial POMDPs. We select new adversarial POMDPs until the FSCs reach a desired performance. We use four benchmark environments to showcase improved robustness against a baseline method in an ablation study and competitive performance compared to a state-of-the-art solver that computes robust finite-memory policies for robust POMDPs.

Marlon Tobaben

ON THE IMPACT OF DATASET PROPERTIES ON MEMBERSHIP PRIVACY OF DEEP LEARNING

We apply a state-of-the-art membership inference attack (MIA) to systematically test the practical privacy vulnerability of fine-tuning large image classification models. We focus on understanding the properties of data sets and samples that make them vulnerable to membership inference. In terms of data set properties, we find a strong power law dependence between the number of examples per class in the data and the MIA vulnerability, as measured by true positive rate of the attack at a low false positive rate. We train a linear model to predict true positive rate based on data set properties and observe good fit for MIA vulnerability on unseen data. To analyse the phenomenon theoretically, we reproduce the result on a simplified model of membership inference that behaves similarly to our experimental data. We prove that in this model, the logarithm of the difference of true and false positive rates depends linearly on the logarithm of the number of examples per class.For an individual sample, the gradient norm is predictive of its vulnerability.

Matias Patricio Pizarro Bustamante

DISTRIBLOCK: IDENTIFYING ADVERSARIAL AUDIO SAMPLES BY LEVERAGING CHARACTERISTICS OF THE OUTPUT DISTRIBUTION

Adversarial attacks can mislead automatic speech recognition (ASR) systems into predicting an arbitrary target text, thus posing a clear security threat. To prevent such attacks, we propose DistriBlock, an efficient detection strategy applicable to any ASR system that predicts a probability distribution over output tokens in each time step. We measure a set of characteristics of this distribution: the median, maximum, and minimum over the output probabilities, the entropy of the distribution, as well as the Kullback-Leibler and the Jensen-Shannon divergence with respect to the distributions of the subsequent time step. Then, by leveraging the characteristics observed for both benign and adversarial data, we apply binary classifiers, including simple threshold-based classification, ensembles of such classifiers, and neural networks. Through extensive analysis across different state-of-the-art ASR systems and language data sets, we demonstrate the supreme performance of this approach, with a mean area under the receiver operating characteristic for distinguishing target adversarial examples against clean and noisy data of 99% and 97%, respectively. To assess the robustness of our method, we show that adaptive adversarial examples that can circumvent DistriBlock are much noisier, which makes them easier to detect through filtering and creates another avenue for preserving the system's robustness.

Max Cairney-Leeming

DEMYSTIFYING AMORTIZED CAUSAL DISCOVERY WITH TRANSFORMERS

Supervised learning approaches for causal discovery from observational data often achieve competitive performance despite seemingly avoiding explicit assumptions that traditional methods make for identifiability. In this work, we investigate CSIvA \citep{ke2023learning}, a transformer-based model promising to train on synthetic data and transfer to real data. First, we bridge the gap with existing identifiability theory and show that constraints on the training data distribution implicitly define a prior on the test observations. Consistent with classical approaches, good performance is achieved when we have a good prior on the test data, and the underlying model is identifiable. At the same time, we find new trade-offs. Training on datasets generated from different classes of causal models, unambiguously identifiable in isolation, improves the test generalization. Performance is still guaranteed, as the ambiguous cases resulting from the mixture of identifiable causal models are unlikely to occur (which we formally prove). Overall, our study finds that amortized causal discovery still needs to obey identifiability theory, but it also differs from classical methods in how the assumptions are formulated, trading more reliance on assumptions on the noise type for fewer hypotheses on the mechanisms.

Maximilian Beck

XLSTM: EXTENDED LONG SHORT-TERM MEMORY

In the 1990s, the constant error carousel and gating were introduced as the central ideas of the Long Short-Term Memory (LSTM). Since then, LSTMs have stood the test of time and contributed to numerous deep learning success stories, in particular they constituted the first Large Language Models (LLMs). However, the advent of the Transformer technology with parallelizable self-attention at its core marked the dawn of a new era, outpacing LSTMs at scale. We now raise a simple question: How far do we get in language modeling when scaling LSTMs to billions of parameters, leveraging the latest techniques from modern LLMs, but mitigating known limitations of LSTMs? Firstly, we introduce exponential gating with appropriate normalization and stabilization techniques. Secondly, we modify the LSTM memory structure, obtaining: (i) sLSTM with a scalar memory, a scalar update, and new memory mixing, (ii) mLSTM that is fully parallelizable with a matrix memory and a covariance update rule. Integrating these LSTM extensions into residual block backbones yields xLSTM blocks that are then residually stacked into xLSTM architectures. Exponential gating and modified memory structures boost xLSTM capabilities to perform favorably when compared to state-of-the-art Transformers and State Space Models, both in performance and scaling.

Merel Kuijs

PERTURBATION ANALYSIS USING VARIATIONAL INFERENCE

A given compound or pathogen interacts with certain cells directly, while also inducing indirect effects on neighboring cells. Indirect effects are mediated by cell-cell communication: direct responders warn the broader cell population about the perturbation by releasing ligands that, in turn, modulate signaling cascades in adjacent cells, resulting in changes in gene expression and, ultimately, cell state trajectories. By applying variational inference, we infer cell state trajectories between two conditions, while taking into account inter-cell dependencies that arise from inter-cell interactions.

Michael Dorkenwald

PIN: POSITIONAL INSERT UNLOCKS OBJECT LOCALISATION ABILITIES IN VLMS

Vision-Language Models (VLMs), such as Flamingo and GPT-4V, have shown immense potential by integrating large language models with vision systems. Nevertheless, these models face challenges in the fundamental computer vision task of object localisation, due to their training on multimodal data containing mostly captions without explicit spatial grounding. While it is possible to construct custom, supervised training pipelines with bounding box annotations that integrate with VLMs, these result in specialized and hard-to-scale models. In this paper, we aim to explore the limits of caption-based VLMs and instead propose to tackle the challenge in a simpler manner by i) keeping the weights of a caption-based VLM frozen and ii) not utilizing any supervised detection data. To this end, we introduce an input-agnostic Positional Insert (PIN), a learnable spatial prompt, containing a minimal set of parameters that are slid inside the frozen VLM, unlocking object localisation capabilities. Our PIN module is trained with a simple next-token prediction task on synthetic data without requiring the introduction of new output heads. Our experiments demonstrate strong zero-shot localisation performances on a variety of images, including Pascal VOC, COCO, LVIS, and diverse images like paintings or cartoons.

Michał Krutul

MIXTURE OF TOKENS: CONTINUOUS MOE THROUGH CROSS-EXAMPLE AGGREGATION

Mixture of Experts (MoE) models based on Transformer architecture are pushing the boundaries of language and vision tasks. The allure of these models lies in their ability to substantially increase the parameter count without a corresponding increase in FLOPs. Most widely adopted MoE models are discontinuous with respect to their parameters - often referred to as sparse. At the same time, existing continuous MoE designs either lag behind their sparse counterparts or are incompatible with autoregressive decoding. Motivated by the observation that the adaptation of fully continuous methods has been an overarching trend in deep learning, we develop Mixture of Tokens (MoT), a simple, continuous architecture that is capable of scaling the number of parameters similarly to sparse MoE models. Unlike conventional methods, MoT assigns mixtures of tokens from different examples to each expert. This architecture is fully compatible with autoregressive training and generation. Our best models not only achieve a 3× increase in training speed over dense Transformer models in language pretraining but also match the performance of state-of-the-art MoE architectures. Additionally, a close connection between MoT and MoE is demonstrated through a novel technique we call transition tuning.

Miriam Rateike

DESIGNING LONG-TERM GROUP FAIR POLICIES IN DYNAMICAL SYSTEMS

Neglecting the effect that decisions have on individuals (and thus, on the underlying data distribution) when designing algorithmicdecision-making policies may increase inequalities and unfairness in the long term—even if fairness considerations were taken intoaccount in the policy design process. In this paper, we propose a novel framework for studying long-term group fairness in dynamicalsystems, in which current decisions may affect an individual's features in the next step, and thus, future decisions. Specifically, ourframework allows us to identify a time-independent policy that converges, if deployed, to the targeted fair stationary state of thesystem in the long-term, independently of the initial data distribution. We model the system dynamics with a time-homogeneousMarkov chain and optimize the policy leveraging the Markov Chain Convergence Theorem to ensure unique convergence. Ourframework enables the utilization of historical temporal data to tackle challenges associated with delayed feedback when learninglong-term fair policies in practice. Importantly, our framework shows that interventions on the data distribution (e.g., subsidies) canbe used to achieve policy learning that is both short- and long-term fair. We provide examples of different targeted fair states of thesystem, encompassing a range of long-term goals for society and policymakers. In semi-synthetic simulations based on real-worlddatasets, we show how our approach facilitates identifying effective interventions for long-term fairness.

Mohammadreza Salehi

SIGMA: SINKHORN-GUIDED MASKED VIDEO MODELING

Video-based pretraining offers immense potential for learning strong visual representations on an unprecedented scale. Recently, masked video modeling methods have shown promising scalability, yet fall short in capturing higher-level semantics due to reconstructing predefined low-level targets such as pixels. To tackle this, we present Sinkhorn-guided Masked Video Modelling (SIGMA), a novel video pretraining method that jointly learns the video model in addition to a target feature space using a projection network. However, this simple modification means that the regular L2 reconstruction loss will lead to trivial solutions as both networks are jointly optimized. As a solution, we distribute features of space-time tubes evenly across a limited number of learnable clusters. By posing this as an optimal transport problem, we enforce high entropy in the generated features across the batch, infusing semantic and temporal meaning into the feature space.

The resulting cluster assignments are used as targets for a symmetric prediction task where the video model predicts cluster assignment of the projection network and vice versa. Experimental results on ten datasets across three benchmarks validate the effectiveness of SIGMA in learning more performant, temporally-aware, and robust video representations improving upon state-of-the-art methods.

Niclas Popp

ZERO-SHOT DISTILLATION FOR IMAGE ENCODERS: HOW TO MAKE EFFECTIVE USE OF SYNTHETIC DATA

Multi-modal foundation models such as CLIP have showcased impressive zero-shot capabilities. However, their applicability in resource-constrained environments is limited due to their large number of parameters and high inference time. While existing approaches have scaled down the entire CLIP architecture, we focus on training smaller variants of the image encoder, which suffices for efficient zero-shot classification. The use of synthetic data has shown promise in distilling representations from larger teachers, resulting in strong few-shot and linear probe performance. However, we find that this approach surprisingly fails in true zero-shot settings when using contrastive losses. We identify the exploitation of spurious features as being responsible for poor generalization between synthetic and real data. However, by using the image feature-based L2 distillation loss, we mitigate these problems and train students that achieve zero-shot performance which on four domain-specific datasets is on-par with a ViT-B/32 teacher model trained on DataCompXL, while featuring up to 92% fewer parameters.

Nicolò Penzo

EVALUATING LLMS ON MULTI-PARTY CONVERSATIONS: A DIAGNOSTIC PIPELINE

Evaluating the efficacy of Large Language Models (LLMs) in Multi-Party Conversations (MPCs) presents several challenges. Tasks on MPCs range from purely linguistic objectives (e.g. generating responses, classification of sentence/conversations) to more “structural” objectives (e.g., addressee recognition, turn taking). Conventional evaluation methodologies, primarily reliant on overarching metrics, often miss the details of model behavior, tied to specific features of the test dataset. To this matter, we propose a methodological pipeline aimed at providing a deeper understanding of model performance, as compared to the mere aggregate evaluations provided by standard approaches.

By evaluating model performance across specific structural attributes of the conversations, our approach aims to elucidate nuanced insights, thereby promoting a deeper understanding of model capabilities in managing user interactions in a MPC.

Niklas Kormann

APPLICATION OF GRAPH REPRESENTATION LEARNING IN HISTOPATHOLOGY

In the field of medical diagnosis, histopathology is the "gold standard" for understanding biological phenomena for many diseases on the microscopic level by tissue examination under a light microscope. In this project, we use graph-based deep learning approaches to capture spatial information from relevant structures/cells and to support diagnosis by classifying unhealthy or dead tissue. In particular, we first obtain an abstract representation of the histopathological images in the form of an attributed graph, we then process these attributed graphs with a variety of Graph Neural Networks for the tissue classification. Our graph-based approach helps to generalize over sources and image modalities, which can limit the performance of image-based approaches.

The presented work is part of the HistoGraph project which is a collaboration of Ecole Polytechnique and the Universite de Strasbourg with the goal to study glomeruli structures in slices from human nephrectomies.

Niladri Shekhar Dutt

DIFFUSION 3D FEATURES (DIFF3F): DECORATING UNTEXTURED SHAPES WITH DISTILLED SEMANTIC FEATURES

We present Diff3F as a simple, robust, and class-agnostic feature descriptor that can be computed for untextured input shapes (meshes or point clouds). Our method distills diffusion features from image foundational models onto input shapes. Specifically, we use the input shapes to produce depth and normal maps as guidance for conditional image synthesis. In the process, we produce (diffusion) features in 2D that we subsequently lift and aggregate on the original surface. Our key observation is that even if the conditional image generations obtained from multi-view rendering of the input shapes are inconsistent, the associated image features are robust and, hence, can be directly aggregated across views. This produces semantic features on the input shapes, without requiring additional data or training.

We perform extensive experiments on multiple benchmarks (SHREC'19, SHREC'20, FAUST, and TOSCA) and demonstrate that our features, being semantic instead of geometric, produce reliable correspondence across both isometric and non-isometrically related shape families. Code is available via the project page at https://diff3f.github.io/

Olaf Dünkel

GENERATIVE MODELS FOR ROBUST VISION

Machine learning models are validated and tested on fixed datasets under the assumption of independent and identically distributed samples, which may not fully reflect the models' true capabilities and potential vulnerabilities. These vulnerabilities can become evident when the model is tested in real-world scenarios. On the other hand, generative models, having witnessed substantial advancements in recent years, can produce realistic out-of-distribution samples. We leverage such models to perform adversarial testing of vision models.

Onno Niemann

REGULARIZATION-AWARE KNOWLEDGE DISTILLATION

Research on explaining the success of Knowledge Distillation (KD) has identified three main driving factors, but no existing framework leverages the full potential of this knowledge. By splitting the classical KD loss into target and non-target loss SOTA techniques successfully control the "focus effect". However, both "similarity effect" and "regularization effect" are not addressed explicitly. This work presents two novel distillation frameworks allowing for a more effective control of the three effects making KD. Regularization-Aware Knowledge Distillation (RAKD) is a naive re-formulation of classical KD that allows for regularization control and Re-scaled Regularization-Aware Distillation (RRAD) additionally includes elements found to be beneficial by the best SOTA method. RAKD is shown to slightly outperform the SOTA and RRAD offers a further lift.

Prasanna Mayilvahanan

IN SEARCH OF FORGOTTEN DOMAIN GENERALIZATION

Out-of-Domain (OOD) generalization is the ability of a model trained on one or more domains to generalize to unseen domains. In the ImageNet era of computer vision, evaluation sets for measuring a model's OOD performance were designed to be strictly OOD with respect to their style. However, the emergence of foundation models and expansive web-scale datasets has obfuscated this evaluation process, as datasets cover a broad range of domains and risk test data contamination. In search of the forgotten domain generalization, we create large-scale datasets subsampled from LAION—LAION-Natural and LAION-Rendition—that are strictly OOD to corresponding ImageNet and DomainNet test sets in terms of style. By training CLIP models on these datasets, we find that OOD generalization challenges from the ImageNet era still prevail. Furthermore, through a systematic exploration of combining natural and stylistic datasets at varying proportions, we identify optimal ratios for model generalization across several style domains. Our datasets and results re-enable meaningful assessment of OOD robustness at scale—a crucial prerequisite for improving model robustness. Overall, we make the sobering point that large-scale data merely obscures the issue of OOD generalization, which remains an unsolved problem.

Rémi Marsal

MOTION ANALYSIS IN VIDEOS \NEWLINE WITH DEEP SELF-SUPERVISED LEARNING

This work explores self-supervised learning methods based on motion in videos to reduce the reliance on costly annotated datasets for the tasks of optical flow and monocular depth estimation. In the absence of ground truth, both tasks are mainly learned with an image reconstruction loss, which relies on the brightness constancy hypothesis. In practice, this assumption may not be verified due to brightness changes often caused by moving shadows or non-Lambertian surfaces, which prevents some reconstructions.

On the one hand, solutions can be implemented to limit the impact of these brightness changes. Thus, our first contribution improves the performance of self-supervised optical flow estimation methods thanks to a neural network designed to compensate for any brightness change at the training only, so that the running time at inference is not affected.

On the other hand, since the reconstruction loss limits make some cases poorly supervised and therefore difficult to estimate for a depth estimation neural network, they are a source of aleatoric uncertainty that can be estimated. In our second contribution, we show that using our new probabilistic formulation of the problem of self-supervised learning of monocular depth provides both better depth and uncertainty predictions.

Ricardo Dominguez-Olmedo

TRAINING ON THE TEST TASK CONFOUNDS EVALUATION AND EMERGENCE

We study a fundamental problem in the evaluation of large language models that we call training on the test task. Unlike wrongful practices like training on the test data, leakage, or data contamination, training on the test task is not a malpractice. Rather, the term describes a growing set of techniques to include task-relevant data in the pretraining stage of a language model. We demonstrate that training on the test task confounds both relative model evaluations and claims about emergent capabilities. We argue that the seeming superiority of one model family over another may be explained by a different degree of training on the test task. To this end, we propose an effective method to adjust for training on the test task by fine-tuning each model under comparison on the same task-relevant data before evaluation. We then show that instances of emergent behavior largely vanish once we adjust for training on the test task. This also applies to reported instances of emergent behavior that cannot be explained by the choice of evaluation metric. Our work promotes a new perspective on the evaluation of large language models with broad implications for benchmarking and the study of emergent capabilities.

Samiran Gode

ZERO-SHOT SONAR BASED OBJECT DETECTION FOR UNDERWATER PERCEPTION

There have been huge improvements in camera based perception, largely lead by deep learning based methods. These methods haven't translated to underwater perception because of the unique challenges underwater- not enough light, turbidity etc. Sonars help solve this problem and work underwater, however they are also noisy. This work uses sonar images for underwater perception.

Sander Boelders

PREDICTING COGNITIVE FUNCTION AFTER SURGERY IN PATIENTS WITH A GLIOMA USING PREOPERATIVE CLINICAL VARIABLES

Introduction: Patients with a glioma often suffer from cognitive impairments both before and after anti-tumor treatment. Ideally, clinicians would be able to rely on predictions of cognitive functioning after surgery for individual patients based on information that is commonly available before surgery. Such predictions would facilitate selecting the optimal treatment considering patients’ onco-functional balance (balancing survival against the preservation of function) and could improve patient counseling.

Method: Post-operative cognitive functioning was predicted for 317 patients with a glioma across eight cognitive tests using a machine-learning approach. This was done using nine multivariate Bayesian regression models that used pre-operative cognitive functioning and a comprehensive set of clinical predictors commonly available before surgery. Bayesian estimates of out-of-sample prediction accuracy (ELPD-LOO) were compared against one another and against models using only pre-operative cognitive functioning as predictors , The accuracy of point-wise predictions of the best-performing model was evaluated using cross validation (R2 and MAE). Model parameters were interpreted and examples of how Bayesian prediction models can be applied in clinical practice were provided.

Results: The best performing model had an R^2 of 34.20% and an ELPD-LOO of -1624.3±46.6. Individual predictions, however, were uncertain. Pre-operative cognitive functioning had the greatest impact on post-operative cognitive outcomes. Models including only pre-operative functioning performed worse than those with clinical predictors, with a difference in ELPD-LOO of -14.4±10.0. Interestingly, models without pre-operative functioning had better point estimates with an R^2 of 34.89% . The top model included interactions between clinical predictors and tumor histology, though models without these interactions were within the 95% credibility interval (ELPD-LOO).

Conclusion: The role of clinical predictors is limited when predicting post-operative cognitive functioning. Consequently, clinicians should not rely on these parameters to infer patients' postoperative cognitive functioning. Our findings highlight the importance of uncertainty estimates to facilitate trust in prediction models and prevent unreliable predictions from being used to make important decisions. Moreover, they stress the need to collect larger cross-center multimodal datasets including the same predictors and outcome measures to improve predictions for individual patients.

Sebastian Sanokowski

A DIFFUSION MODEL FRAMEWORK FOR UNSUPERVISED NEURAL COMBINATORIAL OPTIMIZATION

Learning to sample from intractable distributions over discrete sets without relying on corresponding training data is a central problem in a wide range of fields, including Combinatorial Optimization. Currently, popular deep learning-based approaches rely primarily on generative models that yield exact sample likelihoods. This work introduces a method that lifts this restriction and opens the possibility to employ highly expressive latent variable models like diffusion models. Our approach is conceptually based on a loss that upper bounds the reverse Kullback-Leibler divergence and evades the requirement of exact sample likelihoods. We experimentally validate our approach in data-free Combinatorial Optimization and demonstrate that our method achieves a new state-of-the-art on a wide range of benchmark problems.

Simone Antonelli

DATA VALUATION FOR GRAPHS

Data valuation is emerging as a framework to estimate the influence of training data on model predictions. For i.i.d. data, we measure the difference in model performance when removing some training instances. However, for graphs we also need to consider unlabeled nodes that additionally influence the predictions via the network structure. We propose to estimate the influence of different subsets of nodes by approximating the entire training pipeline with a proxy model. Our proxies accurately predict the output after training for any potential subset of nodes, without the need to re-train from scratch. We investigate variuos proxies including linear models and different semi-values. Our proxies enable various applications such as identifying influential nodes and quantifying the brittleness of the predictions.

Stefano Esposito

BRIDGING THE GAP BETWEEN REAL-TIME SURFACE AND VOLUME RENDERING ON MOBILE DEVICES

Neural radiance fields (NeRF) achieve unprecedented quality for novel view synthesis, but their volumetric formulation remains expensive to evaluate, requiring many samples per ray to render high-resolution images. Volumetric-based methods are able to realistically synthetize soft and fuzzy geometries such as foliage and hair or even complex view-dependent lighting effects as the result of the integration of many sample's appearance predictions. Surface-based methods, on the other hand, tend to condense all information on a single (solid) implicit surface, usually represented as an SDF. While being much faster to render, such representations struggle in non-rigid scenarios and can be suboptimal for novel-view synthesis.

Our hybrid neural representation approximates volume rendering by means of alpha-composition of an ordered set of implicit semi-transparent surfaces. We greatly reduce - down to a fixed number - the number of samples required over volume-based method while improving visual fidelity over single-surface methods. Finally, our representation can easily be baked into lightweight textured meshes and rendered at high frame rates on mobile hardware.

Takeru Miyato

GTA: A GEOMETRY-AWARE ATTENTION MECHANISM FOR MULTI-VIEW TRANSFORMERS

As transformers are equivariant to the permutation of input tokens, encoding the positional information of tokens is necessary for many tasks. However, since existing positional encoding schemes have been initially designed for NLP tasks, their suitability for vision tasks, which typically exhibit different structural properties in their data, is questionable. We argue that existing positional encoding schemes are suboptimal for 3D vision tasks, as they do not respect their underlying 3D geometric structure. Based on this hypothesis, we propose a geometry-aware attention mechanism that encodes the geometric structure of tokens as relative transformation determined by the geometric relationship between queries and key-value pairs. By evaluating on multiple novel view synthesis (NVS) datasets in the sparse wide-baseline multi-view setting, we show that our attention, called Geometric Transform Attention (GTA), improves learning efficiency and performance of state-of-the-art transformer-based NVS models without any additional learned parameters and only minor computational overhead.

Tim Weiland

SCALING UP PROBABILISTIC PDE SIMULATORS WITH STRUCTURED VOLUMETRIC INFORMATION

Modeling real-world problems with partial differential equations (PDEs) is a prominent topic in scientific machine learning. Classic solvers for this task continue to play a central role, e.g. to generate training data for deep learning analogues. Any such numerical solution is subject to multiple sources of uncertainty, both from limited computational resources and limited data (including unknown parameters). Gaussian process analogues to classic PDE simulation methods have recently emerged as a framework to construct fully probabilistic estimates of all these types of uncertainty. So far, much of this work focused on theoretical foundations, and as such is not particularly data efficient or scalable. Here we propose a framework combining a discretization scheme based on the popular Finite Volume Method with complementary numerical linear algebra techniques. Practical experiments, including a spatiotemporal tsunami simulation, demonstrate substantially improved scaling behavior of this approach over previous collocation-based techniques.

Tobias Lorenz

FULLCERT: DETERMINISTIC END-TO-END CERTIFICATION FOR TRAINING AND INFERENCE OF NEURAL NETWORKS

Modern machine learning models are sensitive to the manipulation of both the training data (poisoning attacks) and inference data (adversarial examples). Recognizing this issue, the community has developed many empirical defenses against both attacks and, more recently, certification methods with provable guarantees against inference-time attacks. However, such guarantees are still largely lacking for training-time attacks. In this work, we present FullCert, the first end-to-end certifier with sound, deterministic bounds, which proves robustness against both training-time and inference-time attacks. We first bound all possible perturbations an adversary can make to the training data under the considered threat model. Using these constraints, we bound the perturbations' influence on the model's parameters. Finally, we bound the impact of these parameter changes on the model's prediction, resulting in joint robustness guarantees against poisoning and adversarial examples. To facilitate this novel certification paradigm, we combine our theoretical work with a new open-source library BoundFlow, which enables model training on bounded datasets. We experimentally demonstrate FullCert's feasibility on two different datasets.

Victor Manuel Yeom Song

LEARNING HOW HUMANS LEARN TO PLAY BOARD GAMES WITH GPT-4IAR

We present GPT-4IAR, a transformer neural network architecture for modeling and predicting human behavior in the board game four-in-a-row (4IAR). Experiments show that conditioning action predictions on longer histories of previous moves leads to improved accuracy over prior state-of-the-art models, hinting at longer-term strategic biases in human gameplay. Reaction time prediction is also explored, showing promise in capturing meaningful gameplay statistics beyond raw actions. This work ultimately aims to produce a faithful emulator of human cognition to afford detailed investigation into how humans plan and make decisions.

Viet Anh Khoa Tran

FAST AND SLOW CONTINUAL LEARNING BY CONSOLIDATING DENDRITIC MODULATIONS

The brain is remarkably adept at learning from a continuous stream of data without significantly forgetting previously learnt skills. Conventional machine learning models struggle at continual learning, as weight updates that optimize the current task interfere with previously learnt tasks, i.e. catastrophic forgetting. However, recent work suggests that this issue arises due to an overemphasis on constraining memory requirements, despite memory being cheap and generative replay being a plausible mechanism for regenerating memory samples. We therefore focus on compute-constrained continual learning, aiming to achieve positive backward and forward transfer, while avoiding compute-inefficient full retraining.

Our approach employs a dual learning system: a "wake" phase for rapid online learning on new data (learning fast) and a "sleep" phase for consolidating learned representations into a shared backbone model, learning useful inductive biases across tasks (learning slow). During the wake phase, task-specific dendritic modulations – shaping feedforward activations - are learned via supervised learning, allowing for efficient adaptation to new information. This is followed by a sleep phase where a task-agnostic consolidation process, implemented through Task-Modulated Contrastive Learning (TMCL), integrates the learned modulations into the backbone. This consolidated model acts as a powerful representation learner, benefiting from all previously encountered tasks.

We evaluated our method on class-incremental learning via 1-vs-rest CIFAR-100 tasks, showing that our paradigm leads to a continual learning with positive forward and backward transfer. For future work, we aim to demonstrate the effectiveness of our paradigm towards continually improving foundational models for certain domains in a label-sparse semi-supervised setting.

Weirong Chen

LEAP-VO: LONG-TERM EFFECTIVE ANY POINT TRACKING FOR VISUAL ODOMETRY

Visual odometry estimates the motion of a moving camera based on visual input. Existing methods, mostly focusing on two-view point tracking, often ignore the rich temporal context in the image sequence, thereby overlooking the global motion patterns and providing no assessment of the full trajectory reliability. These shortcomings hinder performance in scenarios with occlusion, dynamic objects, and low-texture areas. To address these challenges, we present the Long-term Effective Any Point Tracking (LEAP) module. LEAP innovatively combines visual, inter-track, and temporal cues with mindfully selected anchors for dynamic track estimation. Moreover, LEAP's temporal probabilistic formulation integrates distribution updates into a learnable iterative refinement module to reason about point-wise uncertainty. Based on these traits, we develop LEAP-VO, a robust visual odometry system adept at handling occlusions and dynamic scenes. Our mindful integration showcases a novel practice by employing long-term point tracking as the front-end. Extensive experiments demonstrate that the proposed pipeline significantly outperforms existing baselines across various visual odometry benchmarks.

Wolfgang Boettcher

SCRIBBLES FOR ALL: BENCHMARKING SCRIBBLE SUPERVISED SEGMENTATION ACROSS DATASETS

In this work, we introduce Scribbles for All, a label and training data generation algorithm for semantic segmentation trained on scribble labels. Training or fine-tuning semantic segmentation models with weak supervision has become an important topic recently and was subject to significant advances in model quality.

In this setting, scribbles are a promising label type to achieve high quality segmentation results while requiring a much lower annotation effort than usual pixel-wise dense semantic segmentation annotations. The main limitation of scribbles as source for weak supervision is the lack of challenging datasets for scribble segmentation, which hinders the development of novel methods and conclusive evaluations. To overcome this limitation, Scribbles for All provides scribble labels for several popular segmentation datasets and provides an algorithm to automatically generate scribble labels for any dataset with dense annotations, paving the way for new insights and model advancements in the field of weakly supervised segmentation. In addition to providing datasets and algorithm, we evaluate state-of-the-art segmentation models on our datasets and show that models trained with our synthetic labels perform competitively with respect to models trained on manual labels. Thus, our datasets enable state-of-the-art research into methods for scribble-labeled semantic segmentation. Moreover, we document diverging robustness against scribble length with the methods.

Yasir Zubayr Barlas

IMPROVING REINFORCEMENT LEARNING FOR BAYESIAN EXPERIMENTAL DESIGN

Bayesian experimental design (BED) plays a critical role in various fields, including engineering and healthcare, by enabling informed decision-making through the systematic planning and execution of experiments over time. Conducting experiments can be expensive and time-consuming, rendering many approaches to BED impractical. Reinforcement learning (RL) has emerged as a powerful tool for addressing the challenges of BED, where the goal is to maximise the expected information gain from experiments. However, there is no definitive basis for favouring one algorithm over another in the context of BED. In our study, we investigate several RL algorithms to solve BED problems of varying complexities, aiming to establish a basis for selecting the most appropriate algorithm for different types of BED problems. We focus on the soft actor-critic algorithm, which is renowned for its sample efficiency and stability, and explore potential variants for enhanced performance. Generalisability is a key criterion when measuring the performance of an algorithm, and so we place emphasis on generalisable agents to a range of similar experimental setups.

Yavuz Durmazkeser

DO WE ACTUALLY NEED THAT MANY PARAMETERS FOR LLMS?

Smaller language models offer several advantages over larger ones, including increased efficiency due to lower computational power and memory requirements, which enables deployment on devices with limited resources like smartphones or edge devices. They are more cost-effective, reducing both hardware expenses and energy consumption, making AI solutions more accessible and sustainable. Smaller models also provide faster inference times, crucial for real-time applications, and are easier to deploy across various platforms. They contribute to reduced energy usage, supporting green computing initiatives, and enhance privacy and security by allowing local data processing. Additionally, their maintainability is higher, simplifying updates and fine-tuning. To determine the necessity of relying on large models, we will compare the performances of small and large models, evaluating the trade-offs and identifying when smaller models might be sufficient or even preferable.

Zhaochong An

RETHINKING FEW-SHOT 3D POINT CLOUD SEMANTIC SEGMENTATION

This paper revisits few-shot 3D point cloud semantic segmentation (FS-PCS) with a focus on two significant issues in the state-of-the-art: foreground leakage and sparse point distribution. The former arises from non-uniform point sampling allowing models to distinguish the density disparities between foreground and background for easier segmentation. The latter results from sampling only 2048 points limiting semantic information and deviating from the real-world practice. To address these issues we introduce a standardized FS-PCS setting upon which a new benchmark is built. Moreover we propose a novel FS-PCS model. While previous methods are based on feature optimization by mainly refining support features to enhance prototypes our method is based on correlation optimization referred to as Correlation Optimization Segmentation (COSeg). Specifically we compute Class-specific Multi-prototypical Correlation (CMC) for each query point representing its correlations to category prototypes. Then we propose the Hyper Correlation Augmentation (HCA) module to enhance CMC. Furthermore tackling the inherent property of few-shot training to incur base susceptibility for models we propose to learn non-parametric prototypes for the base classes during training. The learned base prototypes are used to calibrate correlations for the background class through a Base Prototypes Calibration (BPC) module. Experiments on popular datasets demonstrate the superiority of COSeg over existing methods. The code is available at github. com/ZhaochongAn/COSeg.

GET IN TOUCH

✉ eds24-help@telecom-paris.fr

🌐 https://eds2024.github.io/