GeUmetric Deep Learning 2025
Book of abstracts

Tuesday August 19th

13:00-15:00 Contributed talks, Session 1

Oscar Carlsson

Differential Geometry in Equivariant CNNs via Biprincipal Bundles

The standard convolutional layer in convolutional neural networks (CNNs) arises from a general linear map constrained by translation equivariance. This constraint leads to the classical weight-sharing property of the integration kernel. Interestingly, one can also derive this property from a purely differential geometry viewpoint. In this talk, I will present this alternative derivation for the weight-sharing constraint of the standard CNN and then extend this framework to a more general setting using biprincipal bundles. Finally, I will discuss how this differential geometry framework connects to established approaches for equivariant CNNs on homogeneous spaces.

Elias Nyholm

Unifying transformers and CNNs as equivariant maps

I will present our recent work on a general framework of non-linear equivariant neural networks on homogeneous spaces. The framework generalises both attention-based (transformer) architectures and convolution-based (CNN) architectures. This unifying approach highlights similarities and differences between the two families of architectures, for example by establishing a correspondence between the convolutional kernel in CNN models and relative positional embeddings on the transformer side. It also opens the door for the design of novel non-linear equivariant machine learning models. The presentation is based on joint work with Oscar Carlsson, Maurice Weiler and Daniel Persson (arxiv:2504.20974).

Emma Andersdotter Svensson

A Bundle Formalism for Equivariant Neural ODEs

Previous work by, e.g., M. Weiler et al has shown that using a bundle formalism to describe equivariant CNNs is a useful way to capture the underlying mathematical structures. There, feature maps are described as sections of an associated bundle of a homogeneous vector bundle. In my talk, I will introduce a way to use the bundle formalism for equivariant manifold neural ODEs – a neural network model where the network is defined by a vector field describing how the data evolves continuously with time. By considering neural ODEs on a homogeneous space, we can define a lift of the solution curve to an associated bundle, making it possible to extend previous formulations in terms of a parallel transport. I will include a description of how this formulation might be a generalization of our previous formulation of vector fields being transformed by neural ODEs through the pushforward.

15:30-17:00 Contributed talks, Session 2

Max Guillen

Finite-width Neural Tangent Kernels from Feynman Diagrams

Neural tangent kernels (NTKs) are a powerful tool for analyzing deep, non-linear neural networks. In the infinite-width limit, NTKs can easily be computed for most common architectures, yielding full analytic control over the training dynamics. However, at infinite width, important properties of training such as NTK evolution or feature learning are absent. Nevertheless, finite width effects can be included by computing corrections to the Gaussian statistics at infinite width. We introduce Feynman diagrams for computing finite-width corrections to NTK statistics. These dramatically simplify the necessary algebraic manipulations and enable the computation of layer-wise recursive relations for arbitrary statistics involving preactivations, NTKs and certain higher-derivative tensors (dNTK and ddNTK) required to predict the training dynamics at leading order. We demonstrate the feasibility of our framework by extending stability results for deep networks from preactivations to NTKs and proving the absence of finite-width corrections for scale-invariant nonlinearities such as ReLU on the diagonal of the Gram matrix of the NTK. We validate our results with numerical experiments.

Phillip Misof

Equivariant Neural Tangent Kernels - Connecting data augmentation and equivariant architectures

In recent years, the neural tangent kernel (NTK) has proven to be a valuable tool to study training dynamics of neural networks (NN) analytically. In this talk, I will present how this NTK framework can be extended to equivariant NNs based on group convolutional NNs (GCNNs). Not only does this enable the analytic study of influences of hyperparameters, training biases etc. in equivariant NNs, but it also allows us to draw an interesting connection between data augmentation and manifestly equivariant architectures. In particular, we show that the mean predictions of an ensemble of data augmented non-equivariant networks coincide with the mean predictions of an ensemble of specific GCNNs at all training times in the infinite-width limit. We further provide explicit implementations of the equivariant NTK for roto-translations in the plane and 3d rotations. To evaluate the performance of the equivariant infinite width solution, we benchmark the models on quantum mechanical property prediction and medical image classification.

Oskar Nordenfors

Data augmentation yields equivariance for ensembles of neural networks

There is an ongoing debate in Deep Learning, whether one should restrict neural network architectures to equivariant ones when the task calls for an equivariant model, or whether one should simply train a more expressive model and hope to learn the equivariance through, for example, data augmentation. It turns out that iterations of SGD under data augmentation are equivariant, which has the effect that invariantly initialized weights keep their invariant distribution during training. Isotropic Gaussians are invariant to orthogonal transformations. Thus, Gaussian initialization is enough to guarantee that the distribution of our network's weights will be invariant to representations of any compact group throughout training with data augmentation. In particular, the mean network will be equivariant whenever we stop training, given that the network's architecture is in some sense compatible with the geometry given by the group. This result generalizes earlier work which guarantees equivariance for ensembles of MLPs.

Wednesday August 20th

9.00-10.00 Invited talk

Elisenda Grigsby

Local complexity measures in modern parameterized function classes for supervised learning

The parameter space for any fixed architecture of neural networks serves as a proxy during training for the associated class of functions - but how faithful is this representation? For any fixed feedforward ReLU network architecture, it is well-known that many different parameter settings can determine the same function. It is less well-known that the degree of this redundancy is inhomogeneous across parameter space. I'll discuss two locally-applicable complexity measures for ReLU network classes and what we know about the relationship between them: (1) the local functional dimension, and (2) a local version of VC dimension called persistent pseudodimension. The former is easy to compute on finite batches of points, the latter should give local bounds on the generalization gap. I'll speculate about how this circle of ideas might help guide our understanding of the double descent phenomenon. All of the work described in this talk is joint with Kathryn Lindsey. Some portions are also joint with Rob Meyerhoff, David Rolnick, and Chenxi Wu.

10.30-12:00 Contributed talks, Session 3

Vahid Shahverdi

Neuroalgebraic Cartography: Geometry Under the Surface

In this talk, I will introduce neuroalgebraic geometry, the study of neural networks through the lens of algebraic geometry. The main object of interest is the neuromanifold, which is the space of all functions realizable by a given architecture. I will explain how two fundamental invariants of the neuromanifold, its dimension and algebraic degree, are closely linked to the network’s expressivity and sample complexity. I will then discuss how singularities, often associated with subnetworks, arise naturally in these sets. These lower-dimensional regions tend to attract training dynamics, contributing to the network’s implicit bias. Finally, I will show how studying the fibers of the parameterization map gives insight into identifiability and the symmetries encoded by the architecture.

Stefano Mereta

The geometry of the neuromanifold of ReLU networks

Neural networks define a space of functions, often called "neuromanifold", as their parameters vary. Studying these spaces can lead to a better understanding of the statistical and computational aspects of the network itself. Algebraic geometry (among other tools) has proven to be a valuable tool to study neuromanifolds, as recent works on polynomial neural networks prove. In this talk we will approach the same study for ReLU neural networks. I will focus on explaining how, in order to identify the neuromanifold of such networks, it is to necessary to characterise all the ways in which a continuous piecewise linear function can be written as the difference of two convex piecewise linear functions, and some preliminary results in this direction. This is joint work with A. Flinth and M. Pernice.

Alex Massarenti

On the dimension of neuro varieties

We study polynomial neural networks with a single-node output layer.
We relate the thickness of the associated neuro varieties to the secant defectiveness of Veronese varieties,
propose a conjecture for the thickness, and outline a strategy toward its proof.

13:30-15:00 Contributed Talks, Session 4

Felix Stollenwerk

Research at AI Sweden

This talk will first give a broad overview of AI Sweden's activities and projects, with an emphasis on the work of the Natural Language Understanding Group. Subsequently, I will present our latest research on the geometry of LLM word embeddings and the theoretical foundation of dynamic activation functions.

Hampus Linander

PEAR: Equal Area Weather Forecasting on the Sphere

Machine learning methods for global medium-range weather forecasting have recently received immense attention. Following the publication of the Pangu Weather model, the first deep learning model to outperform traditional numerical simulations of the atmosphere, numerous models have been published in this domain, building on Pangu's success. However, all of these models operate on input data and produce predictions on the Driscoll--Healy discretization of the sphere which suffers from a much finer grid at the poles than around the equator. In contrast, in the Hierarchical Equal Area iso-Latitude Pixelization (HEALPix) of the sphere, each pixel covers the same surface area, removing unphysical biases. Motivated by a growing support for this grid in meteorology and climate sciences, we propose to perform weather forecasting with deep learning models which natively operate on the HEALPix grid. To this end, we introduce Pangu Equal ARea (PEAR), a transformer-based weather forecasting model which operates directly on HEALPix-features and outperforms the corresponding model on Driscoll--Healy without any computational overhead.

Tino Paulsen

Hyperbolic Music Representations

As music is inherently hierarchical, Euclidean geometry fails to capture this explicitly. Utilizing hyperbolic geometry not only encodes different features clearly, especially keys and musical richness. This enables novel interpolation approaches, which allow for different control over music generation.

Thursday August 21st

09.00-10.00 Invited talk

Luca Cosmo

Graph Generative Models for Interpretable Graph Neural Networks

Graph neural networks are effective tools for learning from structured data, but in many cases, the structure most relevant to the task is unknown, either because it is not observed during data collection or because it must be discovered during learning. In this talk, I will present how graph generative models can be used to learn these task-relevant (sub)structures and to design graph convolutional operators that improve both performance and interpretability. We will explore the principles behind this approach, along with examples showing how uncovering the right structure can make GNNs more transparent and insightful.

10.30-12.00 Contributed talks, Session 5

Longde Huang

Learning Chern Numbers of Topological Insulators With Gauge Equivariant Neural Networks

Equivariant network architectures are a well-established tool for predicting invariant or equivariant quantities. However, almost all learning problems considered in this context feature a global symmetry, i.e. each point of the underlying space is transformed with the same group element, as opposed to a local "gauge" symmetry, where each point is transformed with a different group element, exponentially enlarging the size of the symmetry group. Gauge equivariant networks have so far mainly been applied to problems in quantum chromodynamics. Here, we introduce a novel application domain for gauge-equivariant networks in the theory of topological condensed matter physics. We use gauge equivariant networks to predict topological invariants (Chern numbers) of multiband topological insulators. The gauge symmetry of the network guarantees that the predicted quantity is a topological invariant. We introduce a novel gauge equivariant normalization layer to stabilize the training and prove a universal approximation theorem for our setup. We train on samples with trivial Chern number only but show that our models generalize to samples with non-trivial Chern number. We provide various ablations of our setup. Our code is available at https://github.com/sitronsea/GENet/tree/main.

Alexander Friedrich

Autoencoding via Neural ODEs on M-polyfolds

Neural ordinary differential equations (NODEs) describe the dynamics of information propagating through the model using a system of ordinary differential equations (ODEs) defined on a manifold. In the Euclidean case the NODE corresponds to a (recurrent) neural networks in the limit of infinite depth, but non-Euclidean geometry can be also be accommodated. This results in manifold NODEs that describe flows generated by vector fields. These models offer several attractive properties, and have been used successfully to model probability densities in different geometries by constructing continuous normalizing flows (CNFs).

However, current NODE models are fundamentally constrained by the fact that the dimension of the state vector in the dynamical system is fixed due to the intrinsic nature of the dimension of the manifold. On the other hand the encoder-decoder architecture famously extracts a lowerdimensional latent representation from which the original data can be efficiently reconstructed. There is currently no way to incorporate such variable dimension dynamics into the existing NODE framework.

In this talk, we show how to extend NODEs from manifolds to Manifold like polyfolds, M-polyfolds for short, in order to incorporate variable dimension dynamics in geometric deep learning. We present a brief overview of M-polyfolds, together with generalized notions of continuity and differentiability, which endow stratified topological spaces with a smooth structure. Further, we construct explicit M-polyfolds featuring the dimensional bottleneck characteristic of autoencoders and define NODEs capable of encoding and decoding geometric objects like curves and surfaces in theses spaces. In particular, the encoding is accomplished through the construction of compressing vector fields, which give rise to flows that traverse the bottleneck. Finally, we show how to use the adjoint method for training the resulting models for reconstruction and classifications tasks.