Graduate School Event

Thesis Defense: Robust Image Classification with 1-Lipschitz Networks

Bernd Prach (Lampert Group)

Lampert Group

Host: Michael Sammler

Despite generating remarkable results in various computer vision tasks, deep learning comes with some surprising shortcomings. For example, tiny perturbations, often imperceptible to the human eye, can completely change the predictions of image classifiers. Despite a decade of research, the field has made limited progress in developing image classifiers that are both accurate and robust. This thesis aims to address this gap.As our first contribution, we simplify the process of training certifiably robust image classifiers. We do this by designing a convolutional layer that does not require executing an iterative procedure in every forward pass, but relies on an explicit bound instead. We also propose a loss function that allows optimizing for a particular margin more precisely.Next, we provide an overview and comparison of various methods that create robust image classifiers by constraining the Lipschitz constant. This is important since generally longer training times and more parameters improve the performance of robust classifiers, making it challenging to determine the most practical and effective methods from existing literature.In 1-Lipschitz classification, the performance of current methods is still much worse than what we expect on the simple tasks we consider. Therefore, we next investigate potential causes of this shortcoming. We first consider the role of the activation function. We prove a theoretical shortcoming of the commonly used activation function, and provide an alternative without it. However this theoretical improvement does barely translate to the empirical performance of robust classifiers, suggesting a different bottleneck.Therefore, in the final part, we study how the performance depends on the amount of training data. We prove that in the worst case, we might require far more data to train a robust classifier compared to a normal one. We furthermore find that the amount of training data is a key determinant of the performance current methods achieve on popular datasets. Additionally, we show that linear subspaces exist with tiny data variance, and yet we can still train very accurate classifiers after projecting into those subspaces. This shows that on the datasets considered, enforcing robustness in classification makes the task strictly more challenging.

Monday, March 31, 2025 02:00pm - 03:00pm

Office Bldg West / Ground floor / Heinzel Seminar Room (I21.EG.101) and Zoom

This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station. Please find a schedule of the ISTA Shuttle on our webpage: https://ista.ac.at/en/campus/how-to-get-here/ The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.