

WS25/26-ASL-ConNo

This is a description for one of the WS25/26 -ASL projects. The title is ConNo – Consequences of training with Noise

Project description

The goal of this project is to investigate, in various ways, how the training of deep learning models in the computer audition domain are affected by “noise”, by which we mean in this context either “label noise” or “data noise”.

The general idea behind the training of deep learning models is that by exposing the model to enough data for a given task, the model learns good representations, as well as, in the case of multi-class classification tasks, good classification boundaries based on said features. Intuitively, we assume that the data presented to the model is “accurate” in a sense that each sample contains 1) a relevant pattern based on which the sample can accurately be classified, and 2) the correct label of the sample within the task. In reality, however, this might not always be true, as, for instance, a human labelling process may introduce label errors (which call label noise here) or 2) the samples themselves may contain confounding factors such as noise.

Even though it may seem obvious at first thought that the existence of any noise should have negative consequences, the interplay between model and noise is more nuanced. For instance, some approaches purposefully introduce either label noise (see [label smoothing](#)) or data noise (see [data augmentation](#)) with the aim of improving the model’s generalisation capabilities. In this project, we will try to dig deeper into the interplay between deep learning model training, noise and generalisation. The exact scope of the project is thereby not entirely set in stone, but as is common within research, the research questions will be further refined and adjusted. The underlying idea, however, should be that we train deep learning models in various settings of noise scenarios and we put a particular focus on the computer audition domain. We may, for instance include different aspects such as training

- on samples with small to high levels of noise (pure noise),
- on various levels of label noise (label smoothing up to mislabeling)
- on various computer audition datasets (speech emotion recognition,)
- on different models (CNNs vs transformers, pre-trained vs from scratch, end-to-end vs feature extractor and classifier)

For evaluation, we may consider

- ID performance on the target dataset
- OOD performance on other datasets
- generalisability of learned features
- Impacts on modelled sample difficulty

- explainability or interpretability measures of model states and data representations

References

Frénay, Benoît, and Michel Verleysen. "Classification in the presence of label noise: a survey." *IEEE transactions on neural networks and learning systems* 25, no. 5 (2013): 845-869.

Fonseca, Eduardo, Manoj Plakal, Frederic Font, Daniel PW Ellis, and Xavier Serra. "Audio tagging with noisy labels and minimal supervision." *arXiv preprint arXiv:1906.02975* (2019).

Rampp, Simon, Manuel Milling, Andreas Triantafyllopoulos, and Björn W. Schuller. "Does the definition of difficulty matter? scoring functions and their role for curriculum learning." *arXiv preprint arXiv:2411.00973* (2024).

Workflow

Despite some open questions regarding the definition of the research questions, the project is ready to immediately jump in and start running your model trainings and write your own code.

For model training, we will rely on our tool autrainer (<https://github.com/autrainer/autrainer>), which allows for automated and modular DL training for audio tasks. You will need to get familiar with the coding style and implement a few extra functionalities, but most of the legwork is done.

You can access to a dedicated student partition of our cluster. It can get busy at times on there and it is not most cutting-edge hardware from today's perspective but it should well enough to train some DL models.

It is also mandatory to work with github or gitlab! You will set up a new repo (or autrainer fork) in the first week and you will contribute to the repo via pull requests, which will be reviewed by your fellow students and supervisor to ensure collaboration on the same project with multiple people. We don't want quick and dirty but nice and reproducible code on which other people can base their work.

We expect a high level of commitment that reflects the amount of ECTS (technically 10 ECTS means 300 hours) from every participant and (at least a faked) deep interest in the topic.

In return, I offer a high level of availability for meetings, project definition and offline supervision of what you do (code reviews, results review) or, just in general, a high involvement in the project.

Expectations

It is expected that enrolled students have a solid understanding of deep learning with PyTorch. This means that they have already implemented their own `torch.utils.data.Dataset` class and `torch.nn.Module.forward` method for some non-trivial problem. Given the lack of courses at TUM that corresponds to speech & audio, it is not expected that they have experience with those modalities. However, to gain the most from this course, students should have a solid grasp of deep learning fundamentals and be relatively confident in training their own networks. In addition, they should have experience with git for code management, matplotlib and seaborn for visualizations, and pandas for data exploration.

If students do not have experience with some of those tools, they are expected to acquire it. Material will be given for learning PyTorch. However, there are ample tutorials available online:

https://docs.pytorch.org/tutorials/beginner/deep_learning_60min_blitz.html

<https://git-scm.com/docs/gittutorial>

<https://www.atlassian.com/git/tutorials/using-branches>

<https://www.atlassian.com/git/tutorials/merging-vs-rebasing>

https://pandas.pydata.org/docs/user_guide/10min.html

<https://seaborn.pydata.org/tutorial.html>

Learning Goals

Note that this is an advanced practical course. This is why it is assumed that the students already have experience with deep learning as outlined above. The main goal of this course for the students is to work on an open-ended research problem under the supervision of an experienced researcher. This means, they are expected to contribute and investigate novel research ideas -- i.e. through this course, they are supposed to learn how to do good research.

Supervisor

Name: Manuel Milling

I have been working 7 years now under the supervision of Professor Schuller on deep learning models for various modalities and applications, but with a particular experience in audio data. I do really enjoy understanding things, which is why I studied Physics in the first place, and why this type of problem view captured me more than the commonly employed view: "I apply X to Y and hope it performs better".