

Master Thesis

What really matters: Data-centric learning for core-set selection

Collaborate within an international team with strong ties to the German Aerospace Center and contribute to the development of next-generation AI models for the advanced analysis of remote sensing imagery, making contributions with real-world applications to the sustainable development goals, climate change, and natural hazards.

Background

Apart from accuracy and robustness, efficiency is another very important aspect of machine learning approaches. One simple way to increase the training efficiency of machine learning methods is to subsample the training data, i.e. not using all the available samples. For this, most often a straightforward random sampling approach is used. However, not all samples have the same importance to a classifier. Other samples might even be decreasing the model, e.g. because they are contaminated with strong noise or erroneous annotations. Excluding such samples leads to minimal training sets without any major decrease or even an increase in accuracy.

Project

The goal of this thesis is to implement and evaluate different methods to select samples from a given dataset. The selection is based on an initial estimate computed by a model trained on a very small subset of samples and can take correctness, confidence, agreement etc. into account.

Expected outcome

- Literature research on existing subsampling and core-set selection
- Implementation of different approaches to select samples of a given dataset
- Evaluation of the approaches on benchmarks, comparison with relevant reference methods

Your Profile

- Scientific curiosity, very good ability to work independently, strong communication skills
- High motivation to deeply explore a given research topic
- Background in computer vision and/or machine learning
- Strong programming skills (python)
- Knowledge in remote sensing is not required

Contact

Ronny Hänsch, ronny.haensch@dlr.de

Please feel free to reach out for more topics e.g. on Machine Learning, Deep Learning, Ensemble Learning, Computer Vision, Remote Sensing, Earth Observation, Synthetic Aperture Radar or regarding the option to conduct your Master thesis at the DLR in Oberpfaffenhofen.