Chair of Computer Architecture and Parallel Systems
TUM Department of Informatics
Technische Universität München

Master Thesis: Compare Dataset Distillation's Impacts with Long-Rage Dataset Distillation Methods

Many projects and research endeavors encounter challenges when dealing with the training of extensive datasets. Dataset distillation emerges as a solution to substantially decrease the volume of data while maintaining comparable test accuracy. A common practice in dataset distillation involves the widespread use of data augmentation techniques.
However, there remains a gap in understanding the impact of integrating two distinct data augmentation pipelines in various sections. Therefore, in this master's thesis, we aim to investigate the effects of employing a two-stage data augmentation approach with different combinations.
It's important to note that prior knowledge of dataset distillation is not a prerequisite; a foundational understanding of deep learning and a Python framework for deep learning is sufficient for this research.

Tasks:

- Get familier with dataset distillation method MTT and reproduce its result on images
- Design experiments with data augmentation on images
- Apply the method on time-series dataset HPC-ODA
- Design experiments with data augmentation on the time-series dataset
- Investigate into the difference

Recommended knowledge and experience:

- Experience in programming with `Python`
- Experience in deep learning(classes and projects)

Benefits:

- Involve in the academic environment of chair of Computer Architecture and Parallel Systems
- Research with new research topic dataset distillation

Application:

If you are interested in this topic, get in contact with Dai Liu (find the contact details below) through email.

Technische Universität München
Chair of Computer Architecture and Parallel Systems (Prof. Schulz)
dai.liu(at)tum.de
www.caps.in.tum.de