Master thesis: Compare Dataset Distillation's impact on different dataset

A lot of projects or researches are facing problems with training huge amount of data. Dataset distillation can help to significantly reduce the number of data aiming the same test accuracy. The other subject data augmentation has been widely used in dataset distillation. But the effect of folding two different data augmentation pipeline in different sections is not explored. So in this master thesis, we would like to analyse the effect of two stage data augmentation with different kinds of combination.

This master thesis does not require you to know dataset distillation beforehand, but only fundamental knowledge of deep learning and deep learning python framework.

Tasks:

- Get familier with dataset distillation method CAFE 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