

AAI3001 - the main project

Alex

October 28, 2023

Submission due Week 13 , Sunday 11:59pm

1 The main project 2

Up to Teams of 4

Learning goals:

- running a project from the start, including
 - Data splitting
 - writing a custom data set
 - training using Finetuning
 - evaluation of the results

1.1 What to consider before starting to train

- to split the data into training, validation and testing
- to verify that your splits are disjoint after creating them
- to set up the code so that it can run when the split is re-generated

2 an overview over tasks

- either you run your own project
- or you do semantic segmentation (easier) or instance segmentation on the MonuSEG dataset which is provided in here
https://sitsingaporetechedu-my.sharepoint.com/:f:/g/personal/a102463_singaporetech_edu_sg/Er7ZZhyALDhAhX6qTz8h1PsBnRqXf8T3XarV0WaPDAnSfw?e=4ZkpLw – you need to access that from your SIT email .

- If you use MonuSEG, then reserve for the test set some tissue types which are not included in train and val. Note: the test set should also contain tissue types included in train and val - for comparison.

This allows to compare generalization to unseen tissue types!!

- evaluate in an accuracy-type measure (e.g. mIoU for segmentation) and in a second category.

2.1 Deliverables

- training phase: code for training on the dataset with finetuning
- a chosen trained model
- validation phase: code which uses the trained model to predict on the test set images and saves those predictions, and which computes the accuracy-type evaluation.
- **a reproduction routine: the scores from the pretrained model computed on the fly when we run your code should be compared against the scores which you saved when you ran your code - for the accuracy type evaluation**
- code and results for the second category evaluation
- curves showing for every epoch loss and performance on validation and training set - for the run which generates the model you saved (you could have tried multiple runs with different hyperparameters)
- a brief pdf-report about the above (first and second category evaluation: description and results, the curves) and
 - your name and your matriculation number
 - describes the experimental parameters of the training (learning rate batch size, seed values and anything else necessary to reproduce the training)
 - novels longer than 10 pages will not be entertained.
- put everything: codes, saved model, saved predictions, the pdf and everything else you want to add into one single zip file.

2.2 Coding Guidelines

- path portability: all paths (dataset, pretrained model, saved predictions) must be relative to the root path of the main .py-file or relative to some path variable in the code
- no absolute paths
- reproducibility: set all involved seeds to fixed values (python, numpy, torch)

- one or more python files for the code
- it should run using the following steps:
 - unpacking the zip files
 - set **one single path** for the root of the dataset. This must be documented. Nothing else should be needed to set it up
- code should run without typing tons/dozens/piles of parameters on the command line!! `python blafle.py`. Configuration parameters should be inside the code in one place
- python scripts. No jupyter notebooks.