AiAp 2022 Graded Miniproject 1: Deep Learning, multiclass image classification

Your AI-Applications grade will be composed of 3 grades:

- 15% weight: Miniproj 1
- 35% weight: Miniproj 2
- 50% weight: written Exam

The specification of the first miniproject is relatively narrow. Think of it like a graded exercise. In the second miniproject, you will have more freedom to choose technology, language and problem domain. This document specifies the first miniproject.

1. Important Dates:

Start: Monday, March 14th 2022
Deadline: Sunday, April 10th 2022 (upload your report before 23:59)

2. Teamwork!

- We recommend to work in teams of two people. Larger teams are not allowed. Working alone is possible but not recommended!
- Collaborate effectively! You learn more and you are faster when you discuss your ideas or problems.
- Both team members will get the same grade. We reserve the right to interview (randomly selected) team members in order to make sure no one gets a free ride! Note that fraud is not accepted at OST and may have severe consequences.

3. Specification

In this miniproject you **train a deep convolutional network to classify images (at least 3 classes)**. You have to use the following technologies:

- Python
- Keras/TensorFlow
- Jupyter notebook

The goal of this miniproject is to become familiar with Keras and experiment with different network architectures and hyperparameters. As a starting point, you can use one of the Keras/TensorFlow tutorials seen in the exercises.

3.1. Deliverables

Implement your project in a **Jupyter notebook**. Use Markdown-blocks to explain/discuss the code and results in the same document. Make sure the first block in your ipynb file contains the names of the team members.
Upload the following documents on Moodle:

- Your .ipynb file
- A PDF generated from that .ipynb file.
  - Do NOT upload any other files. Do NOT upload data.

3.2. Tasks

- Find a suitable **dataset** on internet.
  You are NOT allowed to use the MNIST or fashion-MNIST data. Important: be aware that searching, analyzing and preprocessing a dataset is a potentially time-consuming task! Choose something relatively simple and make sure you have annotated (labelled) data! Make sure the dataset fits the task (multi-class classification!, not segmentation). Also, just because a dataset consists of millions of images in hundreds of classes does not mean that you have to use all of that! You are free to use only a (small) subset of the data. The goal is not to create a perfect classifier!
  - use a search engine and look for public datasets. You are likely to find things like:
  - https://imerit.net/blog/top-13-machine-learning-image-classification-datasets-all-pbm/
  - https://www.kaggle.com/datasets?search=image
  - https://www.v7labs.com/blog/computer-vision-datasets
  - https://www.kaggle.com/moltean/fruits
  - ...

a) In the notebook, cite and describe your dataset properly (data source, preprocessing)
b) Use Keras to define and train a deep convolutional network for multi-class image classification (at least 3 classes, at most 20 classes).
c) Monitoring: Plot how loss and accuracy evolve over time.
d) Plot a confusion matrix
e) Define and train a second model by changing at least one of the following: number of layers, number of filters (=kernels), regularization method, etc.
f) compare the two models (for example: number of parameters, nr. of epochs needed to train the model, test-performance of the models etc.)
g) Add a short discussion section where you comment your observations and draw conclusions.

- It is NOT important to reach best performance! But: it is important to quantify the quality of your model. It is even ok if training overfits – but make sure you comment on this.
- It is not necessary to do cross-validation.

4. How to submit your miniproject

There will be an assignment called " Project 1, Submission" on Moodle. If you work in a team, both team members will have to upload the (same) files.

5. Grading

The following criteria will be used to evaluate the notebook:

- All tasks implemented (complete, correct, appropriate methods)
- Quality of the comments and discussion (concise, clear, correct terminology)

6. Manage your resources!

This project is worth 15% of your final grade. Manage your resources accordingly. There's no gain in using a complicated dataset or in optimizing for best performance. There are no extra points for going beyond the specification.

7. Respect the deadline

If you do not hand-in a Jupyter notebook (original and PDF) by the deadline, we have to give you a grade 1.0. If you are unable to meet the deadline for important reason (illness, military service, etc.), the deadline can be discussed. Contact us as early as possible!