Exercise Session
CIVIL-459
Spring 2025
Lab 3

EPFL
School of Civil Engineering
Prof. Alexandre Alahi

Homework 1 part 2 - Transformers

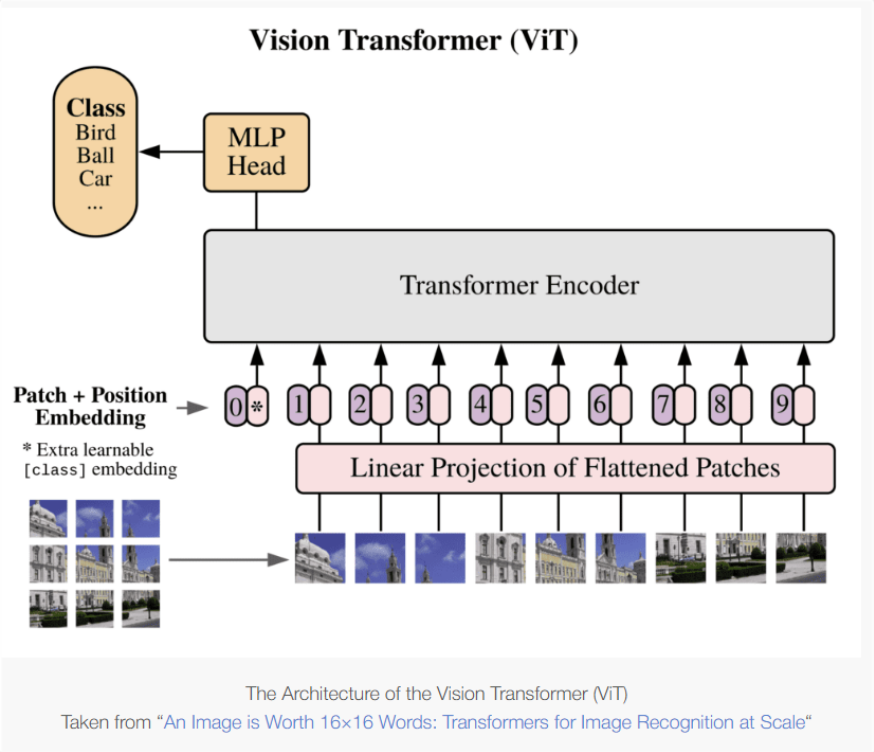

Figure 1: Transformers

Still on the same dataset, we'll check out how to build a Vision transformer network to classify CIFAR10 images (The CIFAR10 dataset consists of 60,000 color images (32x32 pixels) divided into 10 classes of Airplane, Bird, Cat, Dog etc).

What to do

Understand and implement the vision transformers using pytorch (You can get the notebook for this homework from "https://github.com/vita-epfl/DLAV-2025" or on Moodle). There are several theory questions designed to evaluate your understanding. The main goal of this homework is to understand the architecture of vision transformers and compare it against the CNNs. Therefore, at the end of the notebook, you're asked to compare both models. For this task, try to discuss

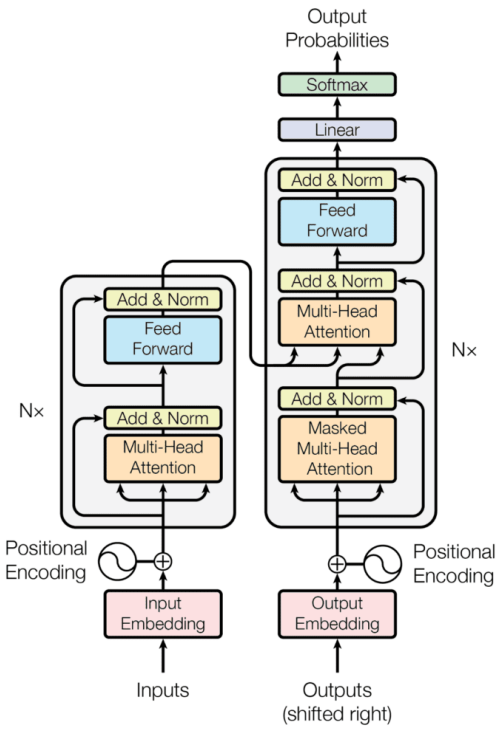

Figure 2: Transformers

the differences (if any) in terms of performance, training schemes as well as advantages and the disadvantages of each model.

1. Follow the notebook, fill in the empty classes to build a vision transformer.
2. Implement the training process and save your best model. You should get full points for around 60%, and bonus points for creative implementations that gets you better accuracy (different than just increasing the number of epochs).
3. Compare the performance on the validation set with the CNN model you trained last week.

Deliverables

You need to submit the jupyter notebooks containing the code and answers to theory questions and the trained models into the moodle.

Grading

* Coding part accounts for 90% of the notebook assessment and remaining 10% will be for the theoretical questions. For the coding section, a student reaching 50

Helpful References

- Pytorch Documentation: https://pytorch.org/docs/stable/
- Pytorch Tutorial: Official link, Collection
- Pytorch Transformer Layers: https://pytorch.org/docs/stable/nn.html#transformer-layers