Université Abdelmalek Essaadi
Faculté ses Sciences et techniques de Tanger
Département Génie Informatique
Master : MBD
Deep Learning
Pr . ELAACHAk LOTFI

Lab 2

Objective : The main purpose behind this lab is to get familiar with Pytorch library, to build neural architectures, CNN, RCNN, FCNN, Vit, etc for computer vision.

Work to do :

Part 1: CNN Classifier

DataSet MNIST Dataset : https://www.kaggle.com/datasets/hojjatk/mnist-dataset

1. Establish a CNN Architecture (Based on Pytorch Library) to classify MINST Dataset, by defining layers (Convolution, pooling, fully connect layer), the hyper-parameters (Kernels, Padding , stride, optimizers, regularization, etc) and running the model in GPU mode.

2. Do the same thing with Faster R-CNN.

3. Compare the two models (By using several metrics (Accuracy, F1 score, Loss, Training time))

4. By using retrained models (VGG16 and AlexNet) fine tune your model to the new dataSet, then compare the obtained results to CNN and Faster R-CNN, what is your conclusion.

Part 2: Vision Transformer (VIT)

Vision Transformers (ViT), since their introduction by Dosovitskiy et. al. [reference] in 2020, have dominated the field of Computer Vision, obtaining state-of-the-art performance in image classification first, and later on in other tasks as well.

1. By following this tutorial : https://medium.com/mlearning-ai/vision-transformers-from-scratch-pytorch-a-step-by-step-guide-96c3313c2e0c, establish a Vit model architecture from scratch, then do classification task on MINST Dataset.

2. interpret the obtained results then compare them with the results obtained in the first part.

Notes :

- **At the end each student must give a brief synthesis about what he has learn during the proposed lab.**

- **Push the work in the Github repository and write a brief report in Github readme file.**

Tools:

Google colab or Kaggle, gitlab/github.