Homework 16

Image Classification using Vision Transformer

Description

In this assignment, you will train and evaluate the Vision Transformer (Keras starter code in [1]) on the Horses vs. Camels dataset [2].

Part A

The specific steps for this task are:

1. Prepare your dataset (choose any split).
2. Build the ViT model.

Part B

3. Train the model. How long does it take? (Simply record the start and finish time of the training. Feel free to use any Python tool to record the training time).
4. Evaluate the performance by reporting the confusion matrix.
5. Compared to the CNN that you evaluated on the same dataset in HW07, which model performed better? Explain Why.
6. How can the Vision Transformer outperform state-of-the-art CNNs?

Submission Guidelines

1. Submit your working code (.py or .ipynb files)
2. Upload any .zip file or folder if your code refers to the paths of those files.
3. A pdf of your report (name: HW16-Part(A or B)-Report-Firstname-Lastname.pdf) with your output and comments.

References

[1] “Image Classification using Vision Transformer”, https://keras.io/examples/vision/image_classification_with_vision_transformer/

[2] “Horses vs. Camels dataset”, https://www.kaggle.com/datasets/akrsnv/horses-and-camels