CSE 586/EE 554Computer Vision II: Homework 4

Professor: Huijuan Xu

Due: March 24, 2022 @ 11:59 EST

1 Assignment Details

In homework 3 we will implementing and examining attention and the Vision Transformer. We allow for the homework to be completed locally (with Jupyter), though we highly recommend students complete this in Google Colab if you do not have a GPU on your machine.

Along with the homework we will also release a short guide to working with the Google Cloud Platform (GCP). This is not relevant for the homework, but may prove helpful for students who are completing a code-oriented project and require a GPU.

2 Vision Transformer (ViT) (100%)

Students will get experience implementing the attention operation, multiheaded attention, transformers, and the vision transformer. We will closely examine ViT's main novelty of splitting an image into patches and obtaining positional embeddings for images.

3 Getting Started + Submission

We suggest students complete the networks in Google Colab. If you'd like to complete the assignments in colab, you can visit the colab website and upload the notebook. To use a GPU, set your runtime to include a hardware accelerator. Students may also complete the homework locally with Jupyter, though training your network will be fairly slow on a CPU.

Submission of this homework will be slightly different than before. You will still turn in your completed ipynb file WITH output cells, and will now generate a submission file after you've completed the assignment. The submission file can be created by running the last cell in the notebook, so you may want to do this last. If you complete the homework on colab, be sure to download it!