CMPSC 497 - Deep Learning for Computer Vision
Professor: Huijuan Xu
Homework 4
Due: March 20, 2023 @ 11:59 PM EST

1 Assignment Details

In homework 4 we will implement and examine the attention and the Vision Transformer. We allow for the homework to be completed locally (with Jupyter), though we highly recommend students to complete this homework in Google Colab if you do not have a GPU on your machine.

2 Vision Transformer (ViT) (100%)

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

3 Rubric for Assignment Grading

Attention (MLP + multi-head self-attention module): 30%

Transformer (residual connection, forward pass): 20%

ViT (splitting images into patches, positional embedding, scheduler): 50%

4 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.

For submitting the assignment, simply upload the completed ipynb file. Be sure the cells have output from running your code. You do not need to include any other files (checkpoints, images, or h5py).