CSE 586/EE 554Computer Vision II: Homework 3

Professor: Huijuan Xu

Due: March 15, 2022 @ 11:59 EST

1 Assignment Details

In homework 3 we will be creating an image captioning system using convnets and a RNN (LSTM in this case). 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 Roar system and how studnets can submit jobs to it which use a GPU. This is not relevant for the homework, but for students who are completing a code-oriented project and require a GPU.

2 Image Captioner (100%)

Students will create and test an image captioning system in Pytorch which works with Microsoft's COCO dataset. We will get to train the network, and also load in weights from a already trained network whose weight's have been uploaded.

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.

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).