Assignment 6

- Aishwarya Jayagopal, A0228508B

Introduction:

The Learning Rate Range test helps identify a learning rate to train networks faster. Using this test, a set of learning rates were experimented with, alongwith two schedulers. All experiments following the learning range test were run with 30 epochs.

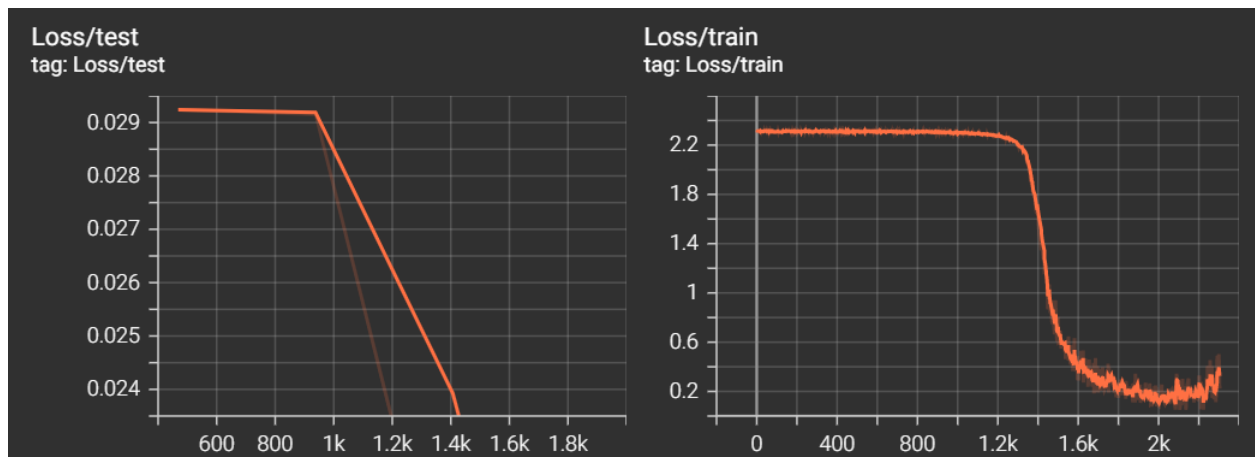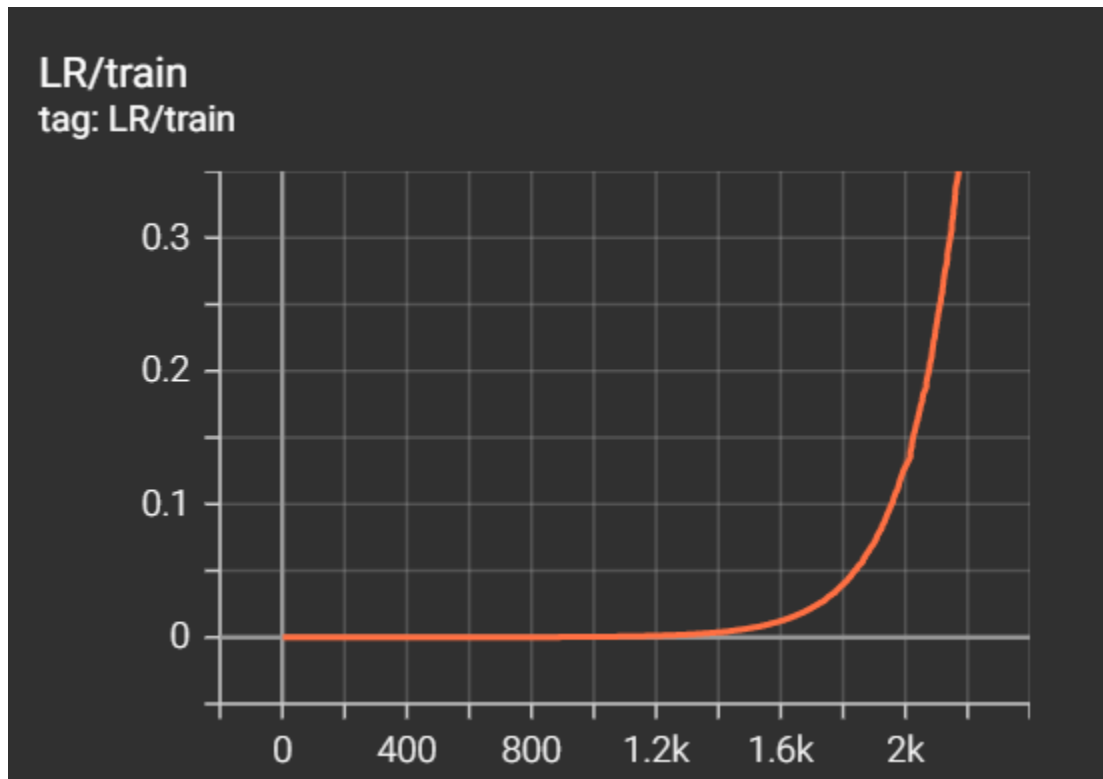| Optimizer	Scheduler	Learning Rate
SGD	OneCycle	0.1158
	MultiStep	0.1158
	OneCycle	2.05e-3
	OneCycle	0.04128
	OneCycle	0.2389

Based on these experiments, the following learning rate range and scheduler are proposed, for the SGD optimizer:

Learning Rate can vary from 0.04128 to 0.1158

Scheduler = OneCycle
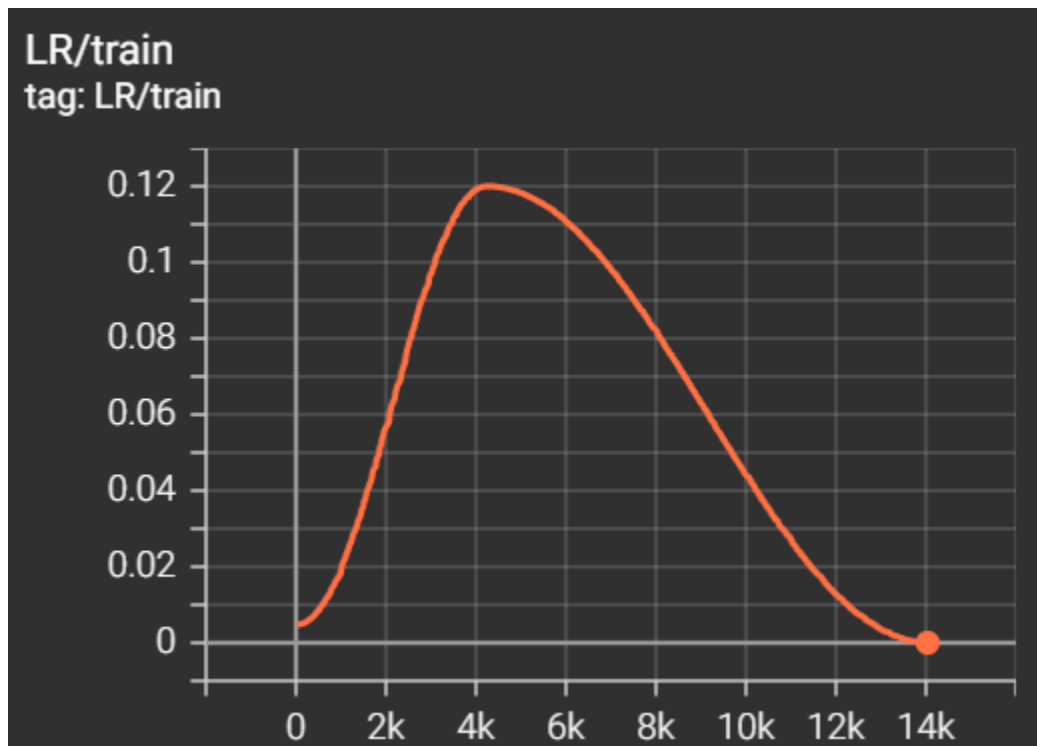
Experiments:

Initial LR Range test (5 epochs):

In the LR range test, the loss for the train dataset is minimum at about 1.977k steps, which corresponds to a learning rate of approximately 0.1158. The dip in loss/train starts at around 1.293k steps, corresponding to a LR of 2.05e-3. The minima of the loss curve lies in the step range of 1.8k to 2.1k, which corresponds to a LR range of 0.04128 to 0.2389.

The test loss dips down from 0.029 after approximately 938 steps.

OneCycle (LR=0.1158, 30 epochs) - LR chosen from lowest point
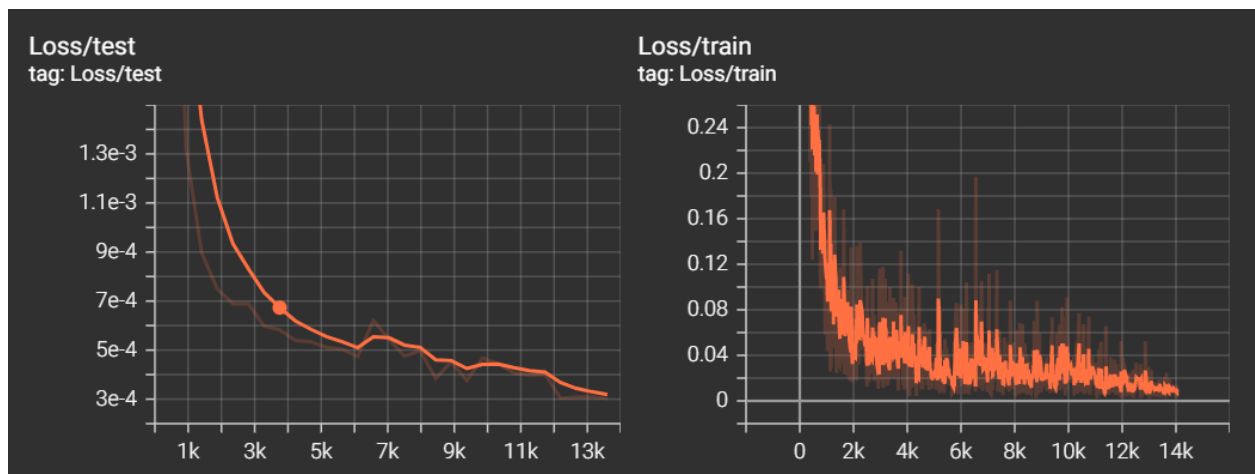
As can be seen in the above graphs, the loss/test is much lower in a fewer number of steps, indicating that the loss reduces much more rapidly. At 1.4k steps, the loss/test is approximately 8.96e-4 as compared to 0.024 in the earlier experiment.

MultiStep(LR=0.1158, 30 epochs)
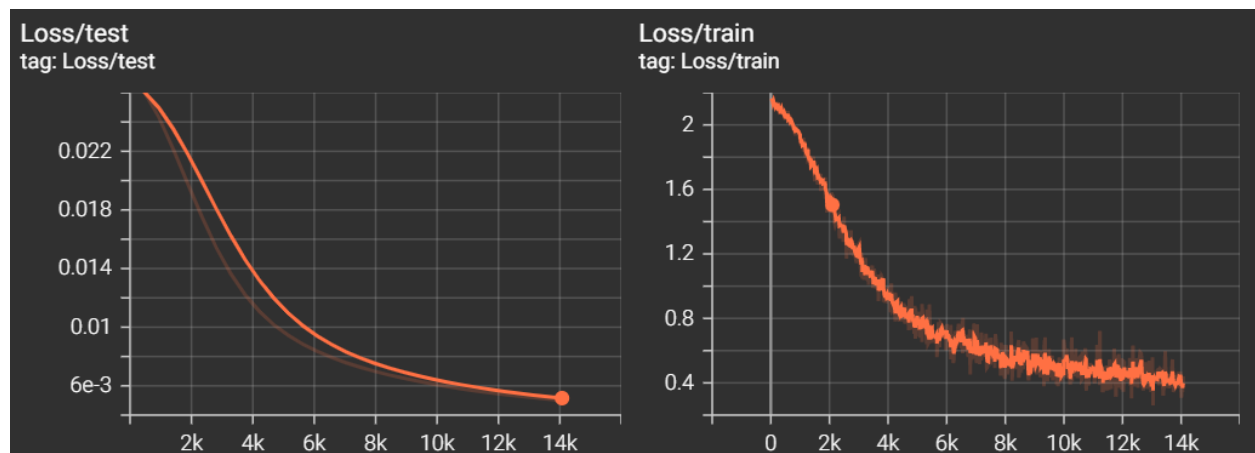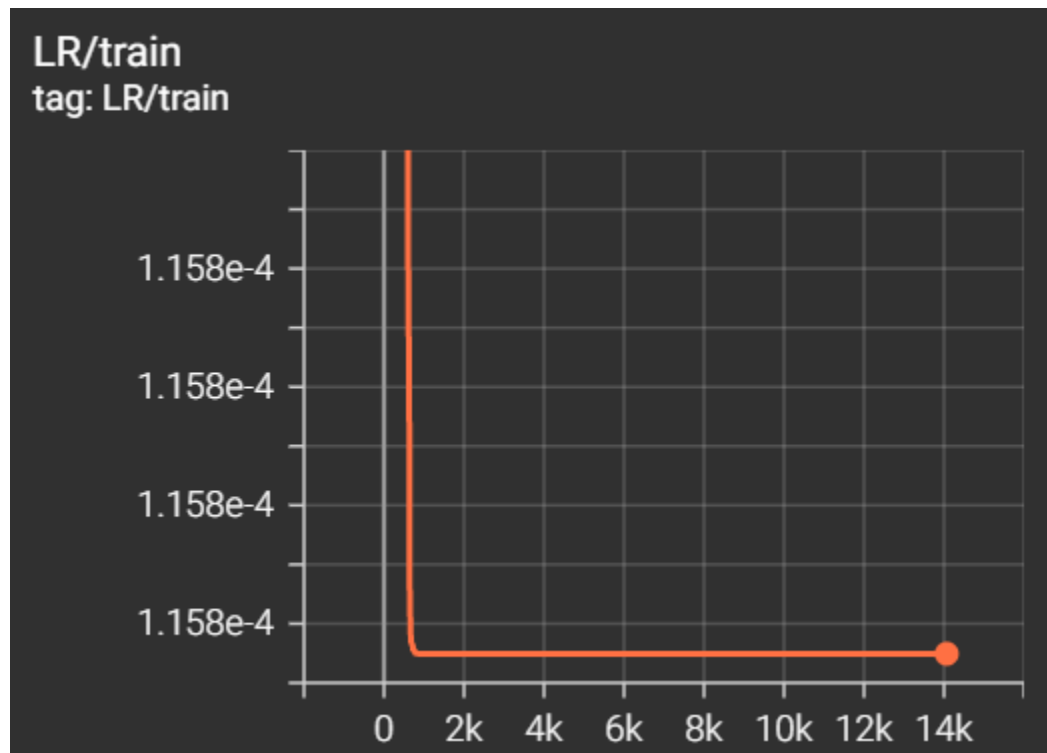
Although the loss curve is much smoother than the one from OneCycle, loss/test is approximately 0.02215 as compared to 0.024 in the initial LR range test. Since they are almost comparable, this does not help prove the LR range test. OneCycle appears to do better in training, and is used for the remaining experiments.
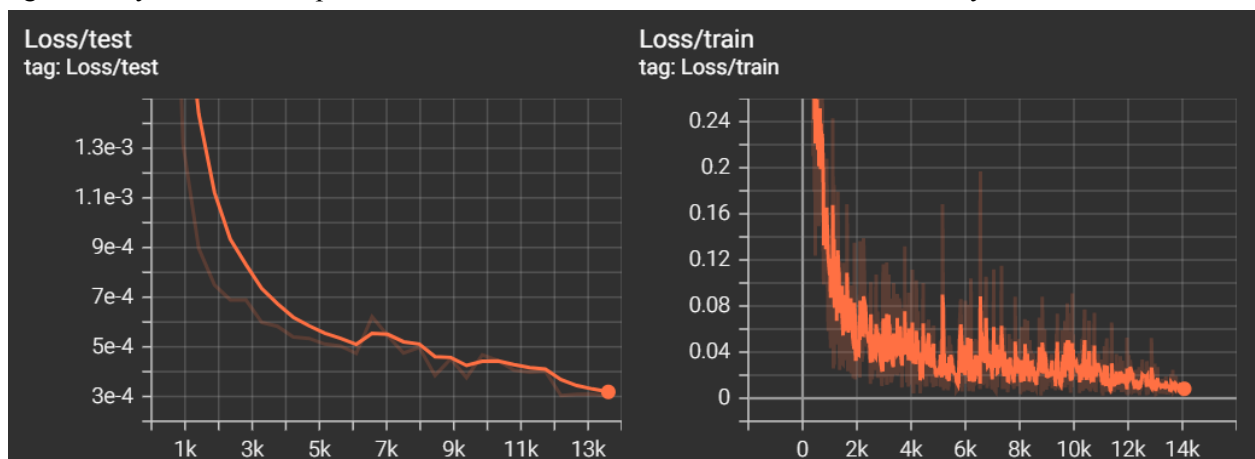
OneCycle (LR=2.05x1e-3, 30 epochs) - LR chosen from start of dip in loss curve

Trying with LR=2.05e-3, i.e. the start of the dip in loss curve in the LR range test does not appear significantly different compared to the test conducted with LR = 0.1158 and OneCycle scheduler.
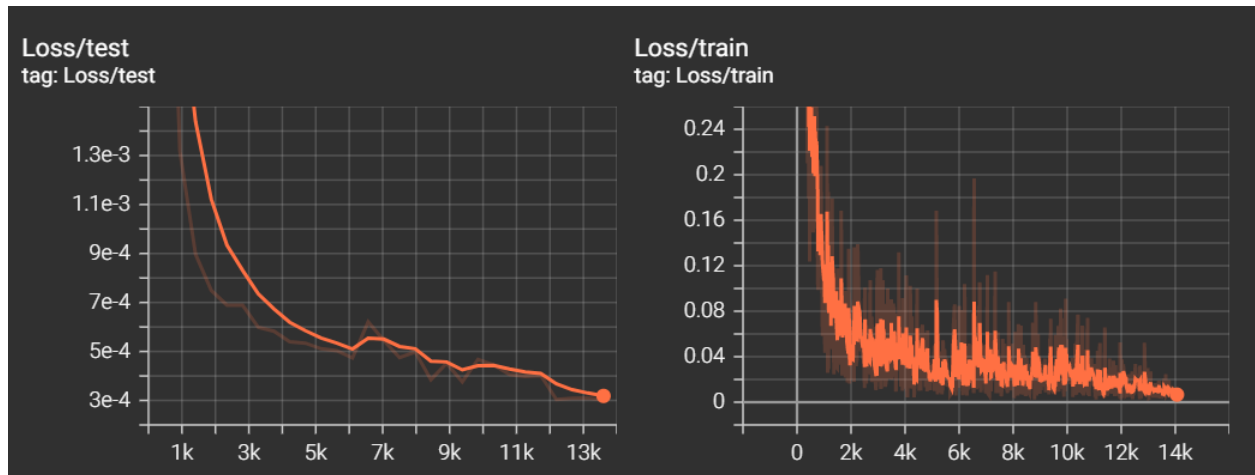
OneCycle (LR=0.04128, 30 epochs) - LR chosen from start of low loss range

Trying with LR=0.04128, i.e. the start of the dip in loss curve in the LR range test does not appear significantly different compared to the test conducted with LR = 0.1158 and OneCycle scheduler.

OneCycle (LR=0.2389, 30 epochs) - LR chosen from end of low loss range

Trying with LR=0.2389, has more noise in loss/train compared to the experiment conducted with LR = 0.1158 and OneCycle scheduler. This may not be a good value for LR.
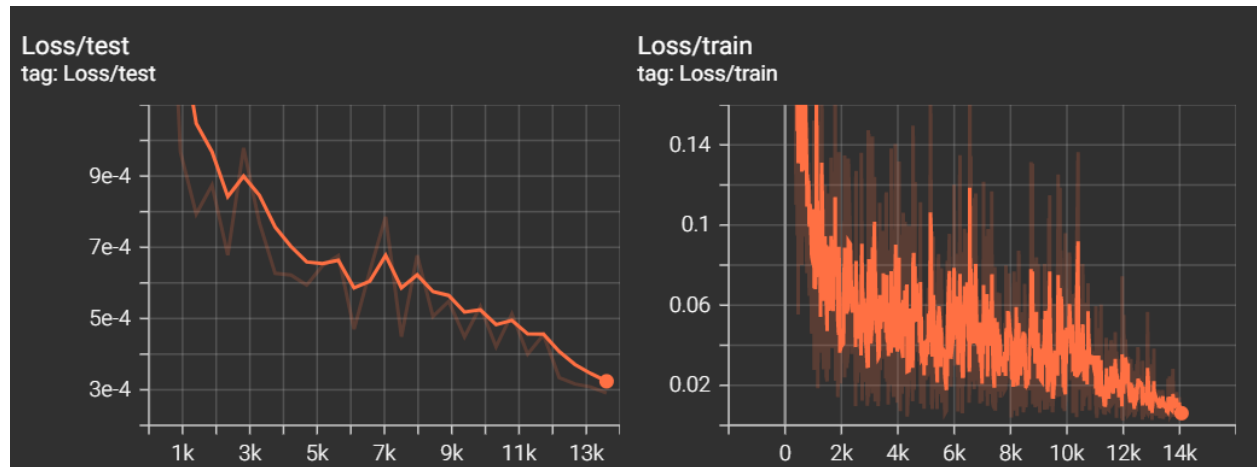
Github Repo:

Please refer to the README file at https://github.com/ajayago/cs5260 for details on the notebooks implementing the above experiments.