Training results – ViT Light



Dataset: MNIST

Hyperparams:

5 epochs, batch size 32

Adam optimizer lr 0.002

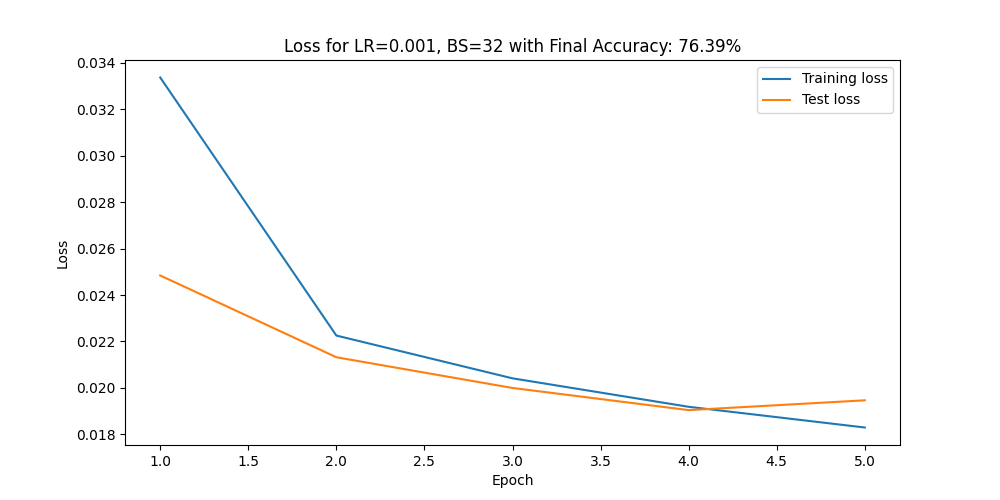
Dataset: Fashion MNIST

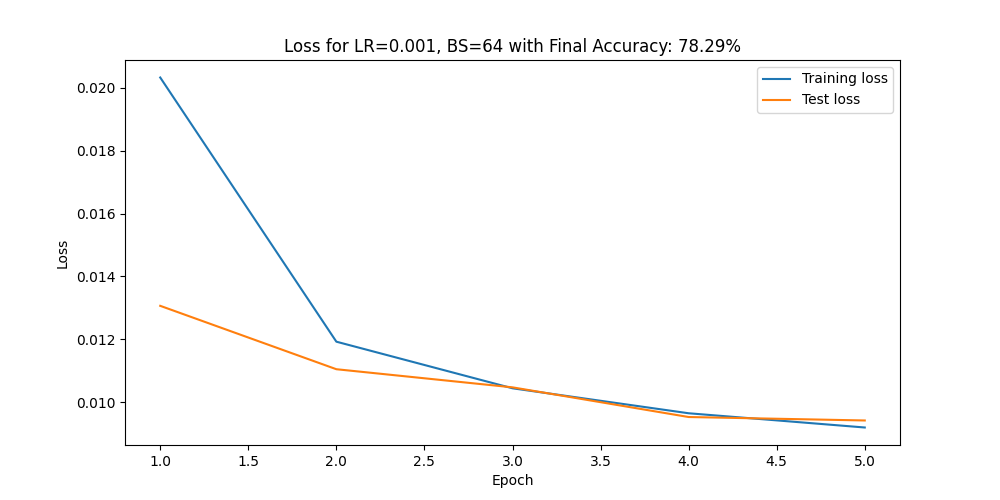
I trained different models with different hyperparameters:

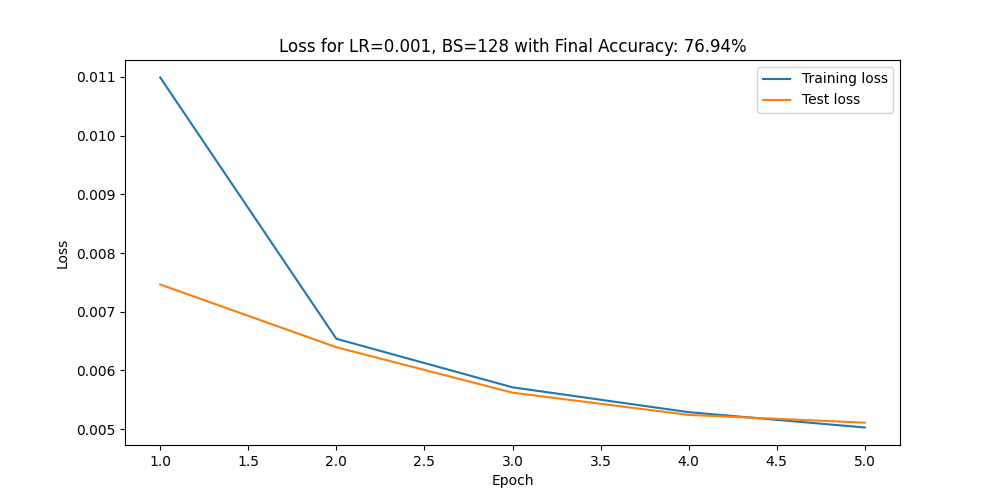
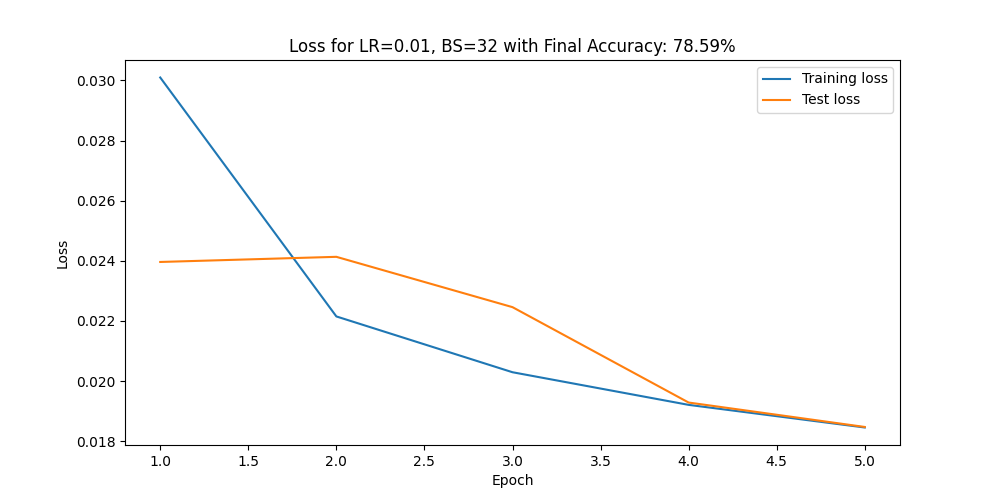
1. Learning rate [0.001, 0.01, 0.1]
2. Batch size [16, 32, 64]

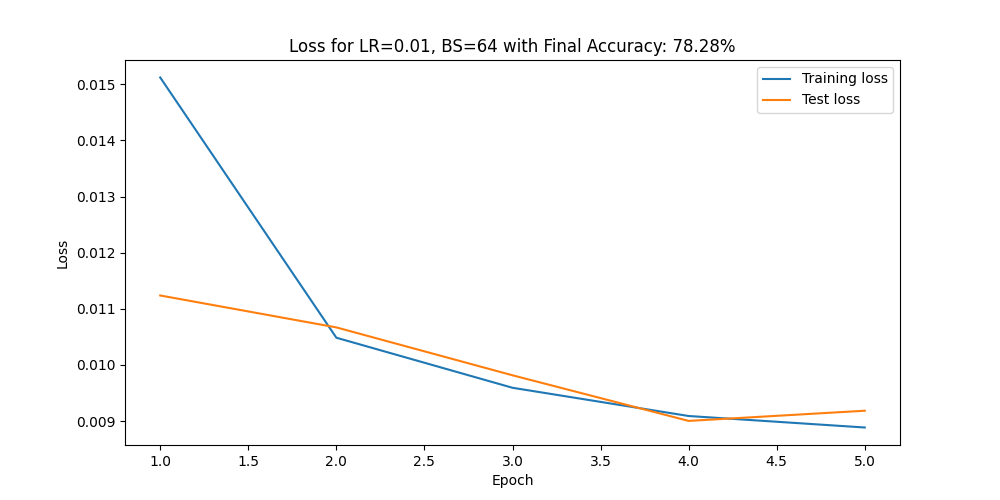
I plotted each of them and saved the best model. I only used 5 epochs because of computational resources I had.

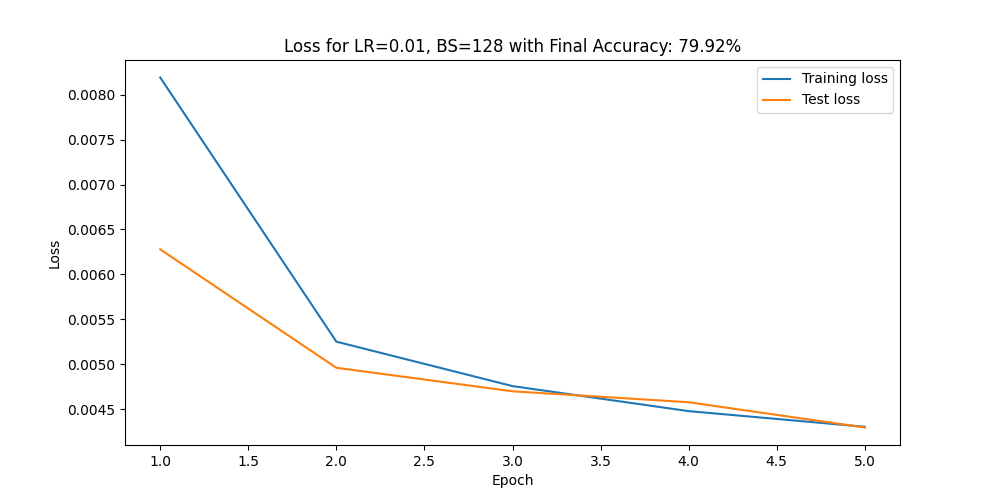
Results:



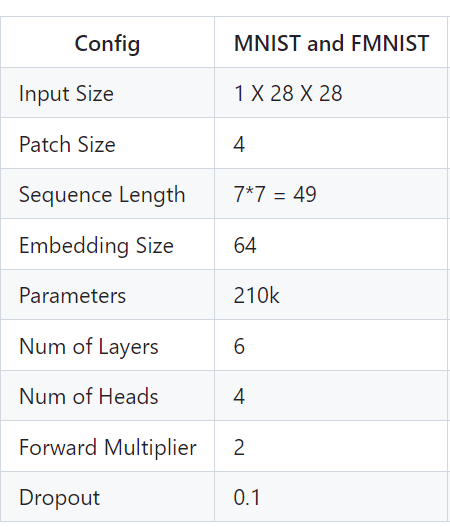








The results I achieve are low in comparison to ViT, but I believe that is because the structure we use is too simple. They achieve 99% accuracy on MNIST and 92% on Fashion MNIST, but they use more heads and more attention layer. They also use dropout for regularization:



Original ViT-Base has 86 million, the [ViT I use for comparison](https://github.com/s-chh/PyTorch-Vision-Transformer-ViT-MNIST-CIFAR10) has  200k-800k