MLNS A1 Part 2 Report

Keshav Gupta

2021101018

Without using Pre-trained models

I took 4 query embeddings, and a transformer block for attending onto the image embeddings. The 4 query embeddings predict a logit vector for the 10 classes (0-9) for the ith digit present. Then I take all the permutations of the normalized probability scores corresponding to the class label.

The loss function is

$$-log(\sum_{i+j+k+l=label} p_i * p_j * p_k * p_l)$$

I split the training data into 2 parts and used 20000 samples for training, and 10000 samples for validating. The model converges much faster than the baseline CNN, and goes to an accuracy of 52.6%

The hyperparameters used are

learning_rate: 0.001
batch_size: 256
num_transformer_layers : 4
embedding_dim: 256
devices: [0, 1, 2, 3]

The optimizer used is Adam. The effective batch size used is 1024.

Using Pre-trained models

I used a resnet50 model pretrained on ImageNet, for predicting classification scores for 37 (0-36) labels. Due to the large size of the model, and the added benefit of pretraining on ImageNet, the model converges much faster and generalizes pretty well.

The hyperparameters used are

learning_rate: 0.001
batch_size: 256
devices: [0, 1, 2, 3]

The optimizer used is Adam. The effective batch size used is 1024.

The model achieves an accuracy of 96.53%