README

By Rahul Sheth

This is an implementation of the Cost Effective Active Learning algorithm.

Background:

Cost Effective Active Learning is a version of Active Learning which utilizes

the high confidence samples as well as the low confidence samples.

Our CNN is implemented using 3 convolutional layers with 1 fully connected (dense) layer.

To calculate the high confidence samples, we calculate the entropy of the probabilities, and for

the low confidence samples, we filter based on the least confidence. After that, for each sample,

we add in all of the low confidence samples and then add the most confident samples based on the percentage

of the full training set we want for that iteration (either add all of most confidence or enough to hit

the percentage). Below is a graph of percent labelled training data versus test set accuracy for a standard CNN

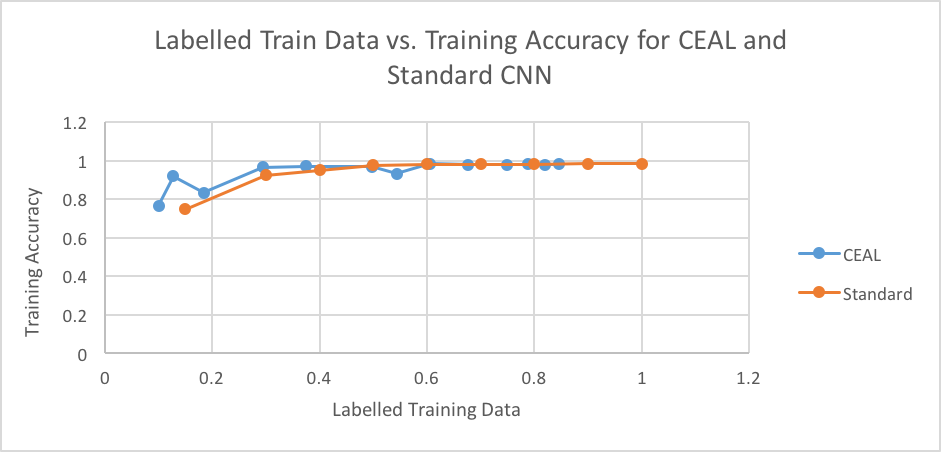
and for our Active Learning implementation. If you would like to see these results for yourself, please feel free

to run the following commands (preferably in different terminals so you can see the print statements in a logical manner

but it will work as two processes in one terminal window).

python3 active\_learning\_cnn.py

python3 regular\_classifier.py



*Figure 1: Sample data from one run of the CEAL implementation vs a Standard implementation.*