Deep Learning Test

May 29, 2019

I implemented the method proposed in the paper "Training Confidence-Calibrated Classifiers For Detecting Out-Of-Distribution Samples" by K. Lee et al. The method was developed in order to be able to train classifiers which are better at deciding whether a test sample is from in-distribution or out-of-distribution. In the paper they present numerous results which demonstrates that their approach performs better than previous attempts. In this case, a successful result is one where a model trained using their Joint Confidence Loss is better at determining whether a sample is from in-distribution or out-of-distribution compared to a model trained using only Cross entropy. I decided that the results presented in figure 4 of the paper where a good way to test my implementation since it contained the most complex training set-up with still easily obtainable datasets. Due to lack of time I only calculated the Detection Accuracy but not the TNR-TPR or AUROC.

In the paper, the proposed Joint Confidence loss performs better than Cross entropy in all experiments shown in figure 4 and I thus expected to obtain similar results. What I observed when training with CIFAR-10 as in-distribution and SVHN as out-of-distribution was somewhat similar to the results in the paper but not at all as clear. The classifier performed about as good when the Joint Confidence loss was used as when only the Cross entropy was used. After training for 3500 gradient steps ($\sim$4 epochs), I got the following results using Joint Confidence Loss:

$$DetectionAccuracy : 72.13\% \tag{1}$$

and the following using only cross entropy:

$$DetectionAccuracy : 70.75\% \tag{2}$$

While these results speaks in favor for the Joint confidence loss, the models were only trained for a fraction of the time done in the paper(yet again due to

lack of time and a weak GPU) and when re-running the training a few times the results where sometimes flipped meaning that they do not actually say much. I ran the same experiment using SVHN as in-distribution and CIFAR-10 as out-of-distribution since the paper's difference in Detection accuracy between the cross entropy and joint confidence loss in this setting is more extreme. Using the Joint Confidence Loss I then got:

$$DetectionAccuracy : 88.82\% \tag{3}$$

and when using only cross entropy:

$$DetectionAccuracy : 86.79\% \tag{4}$$

We can see a similar trend here, the accuracy is better when the Joint confidence loss is used but not by as large a margin as shown in the paper.

To be fair, the model using the joint confidence loss is given a harder task and it is reasonable to assume that it learns slower than the model using the cross entropy loss. If the models are trained for as long as they were in the paper (100 epochs), I believe this effect would be diminished and the network using the Joint Confidence Loss would perform significantly better than the one using cross entropy. In the paper they also performed a hyperparameter search to obtain the best values in each setting, I did not have time to do this and simply chose some values that the authors used in their search. It is quite likely that these values are not the optimal ones and better results could probably have been obtained had a search been done.