Deep Learning: Lab 6

Johnny Joyce (jvjj1u19@soton.ac.uk)

1 Transfer Learning

1.1 Fine-tuning

The pre-trained ResNet50 network was fine-tuned by "freezing" the weights and biases of the final layer. All other layers were then optimised using the training data of the given problem.

To make our results comparable to those of BetterCNN, the same parameters were used: learning rate λ=0.001 and num_epochs=10. Another trial was also performed with a reduced learning rate $\lambda = 0.0001$ to see the performance of fine-tuning when only smaller changes are made. The results were as follows:

	Cross-entropy Loss	Accuracy
BetterCNN ($\lambda = 0.001$)	0.737	0.758
Fine-tuning (λ=0.001)	0.427	0.861
Fine-tuning (λ=0.0001)	1.042	0.713

We can see that using a low value of λ for fine-tuning produces worse results than training BetterCNN from scratch, but if we use the same value of λ as we did with BetterCNN, then fine-tuning produces better results.

1.2 Reflect on the two different approaches

We can compare each of the approaches we have used by examining their performance on the testing data

Since the second approach outputs a single integer as a prediction, we can use one-hot encoding to calculate the cross-entropy loss on the given data. The results were as follows:

	Cross-entropy Loss	Accuracy
BetterCNN	0.737	0.758
Approach 1 (Fine-tuning, λ=0.001)	0.427	0.861
Approach 2 (SVM)	2.872	0.872

We can see that approach 2 achieved the best accuracy, but also had the worst loss value. This is likely due to the one-hot encoding as the SVM approach is only able to make a single prediction, making the loss value somewhat of an estimate. Approach 2 was also significantly faster to train than approach 1, making it much more time efficient.

However, the accuracy difference between approaches 1 and 2 was relatively small and may not be statistically significant. Each of the accuracy values could be improved with further training, but if we were to give the same amount of training time to each approach, it is likely that approach 2 would achieve significantly better accuracy (without considering overfitting).