COMP6248 Lab 6 Exercise

Adrian Azzareli, aab1g18@soton.ac.uk

1 Fine-tuning a pre-trained ResNet50

Fine-tuning is a method which supposedly improves the performance of a CNN, such as ResNet50 network, by continuing back propagation over weights (specific layers/weights can be selected to change while other can be fixed). It is common place to fix earlier layers as otherwise there concerns of over-fitting may arise. ResNet50, as seen in the labs, can be bad at classifying similar images, however if fine-tuned with new data (perhaps containing only a subset of the original 1000-classes of the pre-trained ResNet50) we could improve its relative performance in differentiating similar classes.

With respect to our fine-tuned ResNet50, we firstly modified the *global average pooling* layer to work with the non-square images in our new data set by changing the kernel size 1x1. Considering the convolutional layers understand and locate features (we assume this is completed during the initial training phase) we don't want to modify their behaviour[1]. Instead, we inspect the dense layers; specifically we inspect the final fully connect layer. Hence, we fix the weights associated with all layers apart from the final fc layer and run the Adam optimiser (with Cross Entropy Loss as it does not require a target value). We will use a small learning rate as we expect the weights in the dense layers to be somewhat close to their global optimum, thus don't want to reduce the risk of over-shooting. In finding a good approximation for learning we tested three exponentially increasing rates and evaluated their outcomes. A part of the results are shown Table 1, thus we concluded out fine tuning by setting the learning rate to 0.001.

Learning Rate	Accuracy	Average F1-Score
1e-5	0.53	0.11
1e-4	0.73	0.18
1e-3	0.78	0.35
1e-2	0.71	0.34

Table 1. Learning rate effect on accuracy of a pre-trained ResNet50 network

2 Compare Performance to SVM approach

We can expect the SVM method of improving the performance of the ResNet50 to be more beneficial when the new data set is very different from the original set. Thus, any comparative difference between SVM approach and fine-tuning could be explained by the differences which exist between the new and prior data sets.

Table 2 exemplifies the results of this comparison. What is evident is that SVM provides the best overall accuracy and maintains a higher average F1-Score. Do these results suggest that SVM is also better at differentiating between similar classes (which is a problem the fine-tuning method had)? No, however we can look at the break down of accuracies over all the imposed (16) classes to demonstrate this behaviour (Table 3). We see that the SVM approach is better in almost every class (with exceptions for the *Barchino* and *Polizia* images). Despite this, both models lack in absolute recognition of four classes of images, this represents a quarter of the classes, and perform poorly or several other classes. Consequently, we can imagine there is still a large amount of improvement which can be made to both the fine-tuning and SVM approaches.

We also not that the run-time (measured using *Colab*'s run-time-clock) shows that fine-tuning took roughly 3 minutes to complete while the SVM method took 15 seconds. For the considerable performance boost this is very significant.

Improvements	Accuracy	Average F1-Score
Optimal Fine Tuning	0.78	0.35
SVM	0.87	0.56

Table 2. Difference between fine-tuning and SVM improvements on ResNet50

Class	SVM Accuracy	Fine-tuning Accuracy
Optimal Fine Tuning	0.78	0.35
Alilaguna	1.00	0.43
Ambulanza	0.82	0.51
Barchino	0.25	0.30
Gondola	0.67	0.00
Lanciafino10m	0.0	0.00
Motobarca	0.31	0.05
Motoponto...	1.00	0.40
MotoscafoACTV	0.00	0.00
Mototopo	0.99	0.75
Pantanella	0.84	0.43
Polizia	0.13	0.29
Raccolta...	0.74	0.43
Sandoloaremi	0.00	0.00
Topa	0.00	0.00
VaporettoACTV	1.00	0.98
Water	0.97	0.94

Table 3. Difference between fine-tuning and SVM for each class on ResNet50

[1] We note that doing so with a small data set such as ours would bias (over-fit) the network's understanding of features to the newly introduced data