COMPONENT 2

Multiple Object Recognizer

Neural Network	Pooling	Accuracy (validation)	Loss (validation)	Accuracy (train)	Loss (train)	Time
Shallow	Max pooling Average pooling	0.56 0.54	1.82 1.72	0.96 0.94	0.11 0.16	35m 11s 33m 16s
Deep	Average pooling	0.60	0.97	0.74	0.62	43m 21s
Augmented	Average pooling	0.81	0.50	0.76	0.63	48m 10s

1. **How long does the network need to train until reaching an accuracy of 95% (or does it not reach this level at all)**
   From the above table, we can see that the highest validation accuracy recorded was 81% after passing the model via 10 epochs. It took the network 58 minutes and 10 seconds to train up to this level. However, the model did train up to 96% accuracy under 45 minutes and 11 seconds using max pooling.

2. **What is the trade-off between using many layers (i.e., having a "deeper" network) and accuracy? And layers and time?**
   The more the layers in a neural network, the more non-linearity it brings to the network. Technically this means, that the network gets powerful enough to learn complex pattern in the data set. As the data is a complex one with a large number of input variables, it is better to increase the number of layers. Conversely, it is advised not to use too many layers in the model as it can consequently lead to bias due to more neurons involved. Bias can occur when there are too many neurons and only a few are associated with heavy weights and biases such that other neurons are never used. This will lead to a low accuracy because the model has not learned properly. The solution to this is to include a drop out layer to allow every neuron have equal weights and biases.
   From the table above, it is observed that having a deeper network increases the time the model will take to train due to the increased number of parameters.

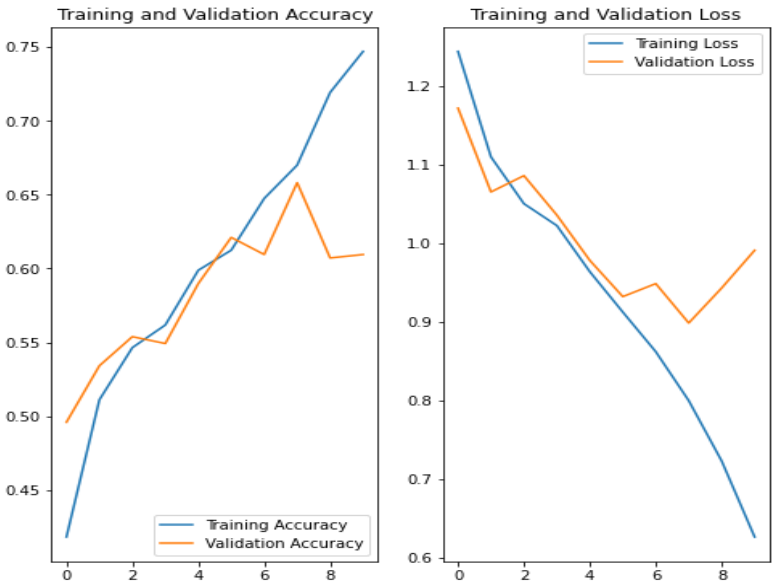

Fig 1: Deeper network

3. **What is the effect of changing the pooling mechanism, e.g., average vs max?**

Pooling is a regularisation technique that improves the performance of CNN while reducing the computation time. A max pooling layer retains the most important features of an image. In order words, it is more focused on the maximum values of the pixels. Average pooling on the other hand retains the average values of features in the image. By changing the pooling method, the image is decreased to a set of features.

Since we are more interested in detecting the overall features of the image rather than whether a specific feature appears or not average pooling did a little better than max pooling. This might be because average pooling prevents the network from learning the image structures such as edges and textures.

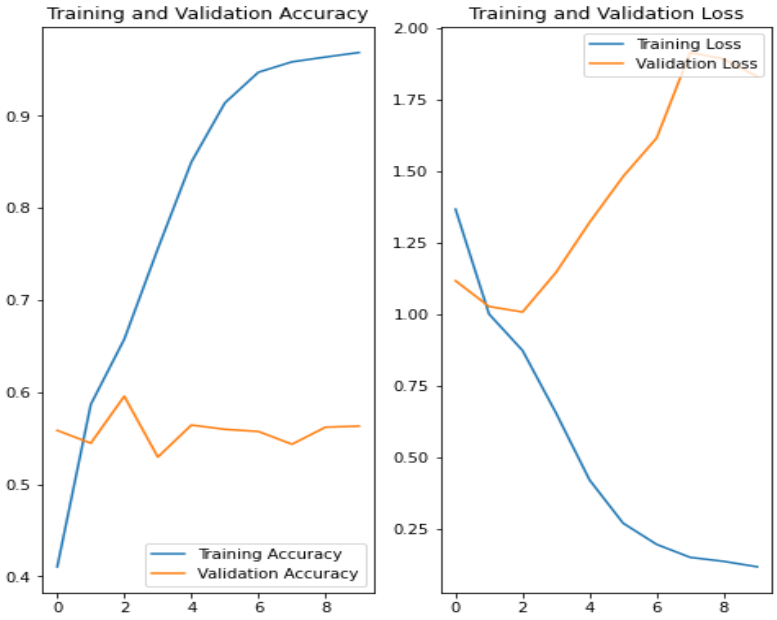

Fig 2: Average pooling

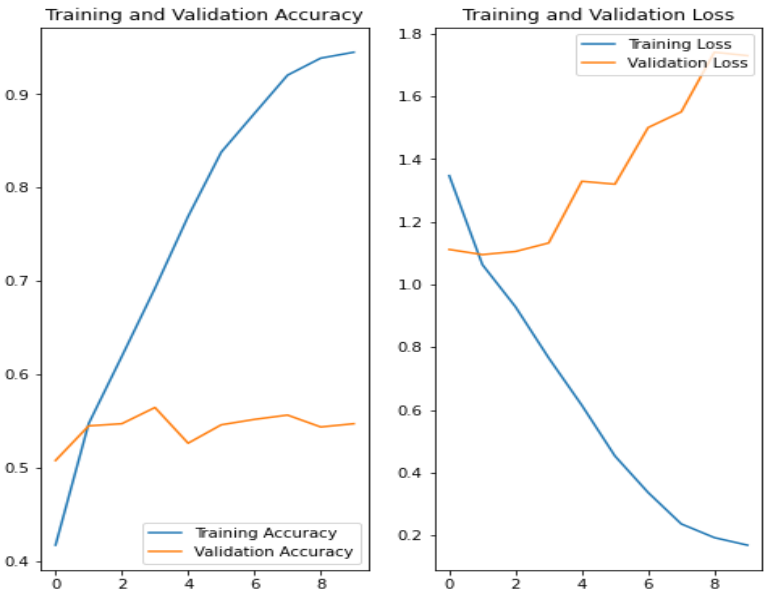

Fig 3: Max pooling

4. **How well does your network do at classifying these images?**

The model did well in classifying the images accurately. Below here is a plot showing how the model did at classifying the images.

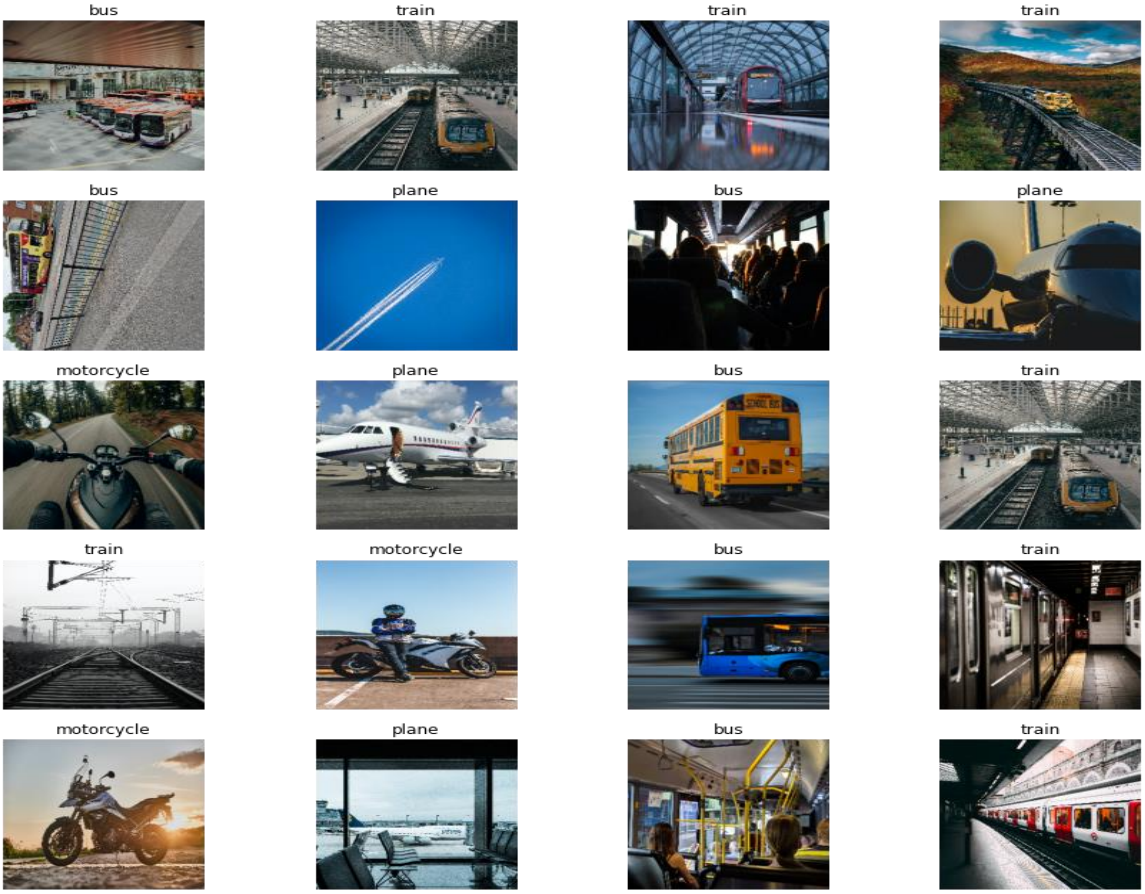

Fig 4: Test images

5. **Does fine-tuning make a difference?**

The model did a nice job at classifying the test images so fine-tuning was not used. Here is a correlation matrix to show how the model did at classifying the images.

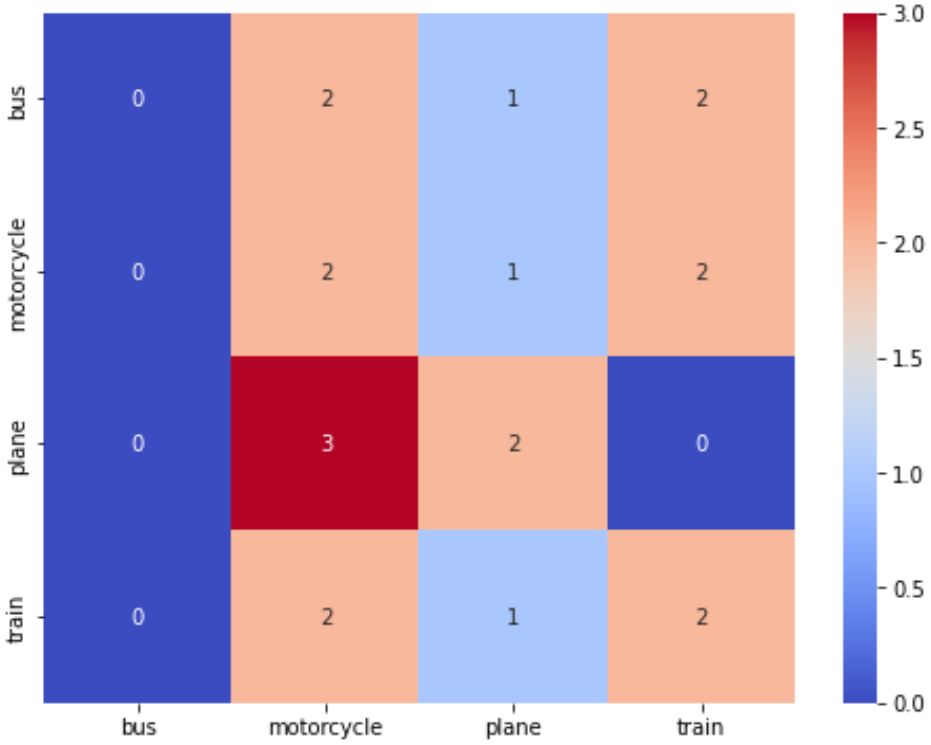

Fig 5: Correlation matrix

Extra Challenge - Model Explainability:

Explainability in machine learning is the process of analysing ML models parameters so as to understand the result produced. This is an important concept with *black box* ML models when dealing with unsupervised learning (Seldon, 2022). The 20 new images collected were passed through tf-explain library to evaluate the interpretability of the model.

The bright yellow spots represent the part of the image that influenced the model's decision to make predictions.

Fig 6: Model Explainability