**CVIA (11376) Assignment 1 report – u3206488**

**Methodology/ Approaches**

**Handcrafted Feature and Classic Machine Learning**

Histogram of Oriented Gradients (HOG) and Support Vector Machines (SVMs) were used for experiments 1 and 3. All images were resized to 64x64 while being ingested into the datastore. 64x64 was chosen as any size larger than that would take an extensive amount of time to train. [4 4] cell size was chosen to extract HOG features, as [2 2] encodes a lot of shape information but increases the dimensionality of the HOG feature vector significantly, whereas [8 8] does not encode much shape information. Before extracting HOG features, the images would be turned into grayscale, and then binarized into black and white images. After that, SVMs were applied to further enhance the training.

**Deep Learning**

Convolutional Neural Network was used for experiments 2 and 4. The CNN had 15 layers in total including 1 for each input and output layer, 3 convolutional layers, 3 batch normalisation layers, 3 reluLayers, 2 max pooling layers, 1 fully connected layer, and 1 softmax layer. Experiment of 3 convolutional layers and 4 convolutional layers were tested but only results of the former were reported as it gave better accuracies (the report will still include the accuracy for 4 convolutional layers). All images were resized to 244x244 before training, this ensured validation and consistency when inputting images into the network. The frequently experimented learning rates were 0.001 and 0.005, as they gave similar accuracies and would not take an excessive amount of time to train. No specific number of epochs was set, as the training would be terminated once the training accuracy hit its peak and the validation accuracy was stable for a few iterations.

**Experiment 1**

The overall accuracy of experiment 1 is 2.9%. With the low accuracy, we can see extracting HOG features might not be the best handcraft feature for this classification task. It is comprehensible that it would be impossible for the classifier to recognise the bird species correctly depending solely on their shapes and the noises (backgrounds).

Classes 44 (Frigatebird) were confused with class 143 (Caspian Tern) where 33.3% of class 44 images were misclassified as class 143. If we look at the images of both classes, we can see these two kinds of birds have similar poses (e.g. wings spread with blue sky as the background), which makes the classifying task even trickier for the classifier with only HOG features and SVMs.

**Experiment 2**

The overall accuracy of experiment 2 is 6.4% (5.3% for 4 convolutional layers).

The class that has the highest recognition rate is class 17 (Cardinal) with 45.5%, which is believed because of its relatively distinct feature that not many other bird species, that is, most of the Cardinals being almost entirely in red. However, on some occasions, the network wrongfully classified Cardinal as other bird species, such as class 42 (Vermilion Flycatcher) and class 140 (Summer Tanager), due to having similar features.

On the other hand, from the confusion matrix and accuracy, it is not hard to see that bird species were mistakenly classified. One of the most worth mentioning classes is class 2 (Laysan Albatross), which has a 0% recognition rate. Due to not having enough unique features to distinguish it from class 45 (Northern Fulmar).

**Experiment 3**

The overall accuracy of experiment 3 is 5.9%. With the help of bounding boxes, the accuracy had a slight improvement of 3% compared to experiment 1.

The most correctly classified class was class 51 (Horned Grebe) with a 66.7% recognition rate but also one of the most easily confused classes at the same time. From the images of classes 50 (Eared Grebe), 51, and 52 (Pied Billed Grebe), we can see almost all birds in these classes have a highly identical pose (i.e. floating on the water). As a result, HOG features make these three classes easy to classify but at the same time, easy to get wrong.

**Experiment 4**

The overall accuracy of experiment 4 is 18.7%. The accuracy has improved by approximately 10%. Most classes now have at least one image recognised correctly. This improvement was reasonable as removing most of the backgrounds makes the tasks simpler.

However, the most surprising result of the experiment was class 17 (Cardinal) has a very low recognition rate of 9%, considering its performance in experiment 2. Images of class 140 (Summer Tanager) were vastly classified as class 17. An assumption for this scenario is, in experiment 2, the network was trained with the birds and noises (backgrounds), it used not only the bird itself but also the noises to correctly classify Cardinal as class 17. As a result, we might need to rethink what the network is learning and change its way of learning to achieve the classification goal.

**Lessons Learned**

As a result, extracting HOG features plus SVMs were not as good as Convolutional Neural Network, and so the deep learning approach is better at this classification task. However, it might not always be the case, considering we only experimented with HOG plus SVMs and there are not sufficient data (images) to properly train the deep learning model.

Before the experiments, the more convolutional layers the higher accuracies were expected but it turned out to be false. An assumption is that the model did not need to extract that many features resulting as over-trained as the complexity of the data was not too high.

If we had more time to experiment on this dataset or have a chance to experiment with it in the future, we could or can apply techniques (e.g. SIFT, KNN, or RCNN) and parameters (e.g. different numbers of convolution layers) to observe their differences and behaviours.