Name: Sushruth Danivasa Sridhar

Assignment – 4

Task 1:

1. Describe or draw the two CNN architectures you used for Task 1.

For Task 1 of the assignment, two distinct CNN architectures were constructed and assessed for classifying images within the Monkey Species dataset:

1. **Model 1 Architecture:**
    - Input Layer: (100, 100, 3)
    - Rescaling: 1. /255
    - Conv2D: 32 filters, 3x3 kernel, ReLU activation
    - MaxPooling2D: 2x2 pool size
    - Conv2D: 64 filters, 3x3 kernel, ReLU activation
    - MaxPooling2D: 2x2 pool size
    - Flatten
    - Dense: 128 units, ReLU activation
    - Dropout: 0.5 rate
    - Dense: 10 units, Softmax activation
2. **Model 2 Architecture:**
    - Input Layer: (100, 100, 3)
    - Rescaling: 1. /255
    - Conv2D: 16 filters, 5x5 kernel, ReLU activation
    - MaxPooling2D: 2x2 pool size
    - Conv2D: 64 filters, 5x5 kernel, ReLU activation
    - MaxPooling2D: 3x3 pool size
    - Flatten
    - Dense: 64 units, ReLU activation
    - Dropout: 0.3 rate
    - Dense: 10 units, Softmax activation

2. For each model, show in a table how training accuracy changed over epochs.

Epochs	**Model 1 Accuracy**	**Model 2 Accuracy**
1	0.307484	0.346258
2	0.471897	0.52129
3	0.560565	0.601443
4	0.614968	0.664963
5	0.667669	0.727182
6	0.708246	0.772167
7	0.743913	0.810239
8	0.779982	0.840297
9	0.809739	0.85813
10	0.82096	0.883779
11	0.839295	0.893498
12	0.849915	0.903216
13	0.87366	0.91073
14	0.876465	0.916241
15	0.888488	0.927963
16	0.893498	0.928664
17	0.896904	0.931269
18	0.901613	0.938082
19	0.906522	0.93668
20	0.913736	0.938984
21	0.911231	0.943593
22	0.919748	0.9475
23	0.922352	0.948001
24	0.917844	0.950506
25	0.925358	0.954113
26	0.93177	0.949604
27	0.929065	0.948903
28	0.931971	0.953211
29	0.935377	0.955015
30	0.93698	0.957619

3. Show test accuracy for each model in a table.

	Model 1	**Model 2**
Test accuracy	0.72125	0.70787

4. Show the confusion matrix of the more accurate model on the test set.

Confusion Matrix for Best Model

```
[[ 52   0   0   0   0   0   0   0   0   0]
 [  0 118   2  11   2   2   2   2   5   2]
 [  1   1  15   3   1   2   6   4   2   0]
 [  0   6   0  51   1   1   0   1   8   5]
 [  1   0   1   3  22   3   2   0   0   2]
 [  0   4   0   3   1  77   0   3   1   2]
 [  1   0   2   3   5   3  90   5   3   0]
 [  7   2   4   1   2   8   5 122   4   3]
 [  1   9   3  37  17   7   3   2  79   3]
 [  2  22   1  16  13  21   6  22  15 290]]
```

5. Write some comments on the results comparing the two models, and some comments on the confusion matrix.

In evaluating the performance of the two CNN architectures for image classification, Model 1 with a simpler convolutional architecture, demonstrated superior performance compared to Model 2, both in terms of training and validation accuracy. Throughout the training process, Model 1 maintained a higher accuracy, achieving a final test accuracy of 72.1%, compared to Model 2's 70.7%. Model 2, despite having a deeper and potentially more complex structure with larger filters initially, does not perform as well as Model 1. This could be due to the larger filter sizes in the initial layers, which might be capturing too general features, thus being less effective for fine distinctions needed in this dataset. This suggests that Model 1's configuration is more effective at learning generalizable features from the training dataset without overfitting, despite its relatively higher capacity.

The confusion matrix for Model 1, which achieved higher test accuracy, reveals some interesting insights into its classification behaviour. The matrix shows very high accuracy for some classes, with perfect or near-perfect classification for the first class (100% accuracy). However, there are noticeable misclassifications in other classes, such as between class 2 and class 3, where 11 instances intended for class 3 were incorrectly labelled as class 2. This suggests that while Model 1 is generally robust, it struggles with distinguishing between classes that may have similar features.

Task 2:

1. Mention which pre-trained model you used and what is its size, and what layer(s) you added on top of it.

Used the EfficientNetV2S model as the backbone for the fine-tuned model. The pre-trained network is modified by adding a global average pooling layer, a dense layer with 1024 neurons with ReLU activation, and a final softmax output layer for 10 classes.

2. Show in a table how training accuracy changed over epochs.

Epochs	**Fine-Tune Model Accuracy**
1	0.825168
2	0.893798
3	0.920449
4	0.933373
5	0.946599
6	0.958221
7	0.964733
8	0.964232
9	0.968941
10	0.971646
11	0.97375
12	0.976455
13	0.975654
14	0.978559
15	0.980563
16	0.980162
17	0.979962
18	0.980363
19	0.980663
20	0.980663
21	0.982367
22	0.986474
23	0.985573
24	0.98397
25	0.98407
26	0.982767
27	0.98437
28	0.985673
29	0.985573
30	0.988178

3. Show the test accuracy of the fine-tuned model and the better model of Task 1 in a table.

	Fine-Tune Model	**Better Model**
Test accuracy	0.89448	0.72125

4. Show the confusion matrix of the fine-tuned model.

Confusion Matrix for Fine-tuned Model

```
[[ 52   0   0   0   0   0   0   0   0   0]
 [  0 138   3   3   0   1   0   0   0   1]
 [  0   1  25   1   1   1   3   2   0   1]
 [  0   1   2  60   1   0   0   0   8   1]
 [  0   0   1   0  31   2   0   0   0   0]
 [  0   1   0   0   1  88   0   0   1   0]
 [  0   0   0   0   1   0 107   3   1   0]
 [  4   1   3   4   1   2   0 140   1   2]
 [  0   5   4  15   6   4   4   1 120   2]
 [  0   9   2   4   2   3   2   6   5 375]]
```

5. Write some comments on the results comparing the fine-tuned model and the better model of Task 1, and some comments on the confusion matrix of the fine-tuned model.

Comparing the performance of the fine-tuned model with the best-performing model from Task 1 i.e. Model 1 provides a clear perspective on the advantages of leveraging pre-trained architectures enhanced by fine-tuning. The fine-tuned model demonstrates a substantial increase in accuracy, evident from both the training and validation accuracy curves. The superior performance of the fine-tuned model, with a validation accuracy that remains consistently higher throughout the training epochs, underscores the effectiveness of using pre-trained networks as a foundation, especially when enhanced with additional training on a specific task.

The confusion matrix of the fine-tuned model illustrates its superior classification capability across all classes. For instance, the fine-tuned model shows near-perfect classification in several classes where Model 1 struggled, indicating a significant reduction in false positives and false negatives. The matrix reflects fewer misclassifications among similar categories, suggesting that the fine-tuning process has effectively adapted the pre-trained features to the nuances of our specific dataset.

Task 3:

1. Include the 10 images in the report along with their correct classes, predicted classes by the better model of Task 1, and predicted classes by the fine-tuned model.

Here, the best model is model 1 and true is the correct classes of the dataset with their indices. The indices are as follows,

0 – Bald Uakari

1 – Emperor Tamarin

2 – Golden Monkey

3 – Gray Langur

4 – Hamadryas Baboon

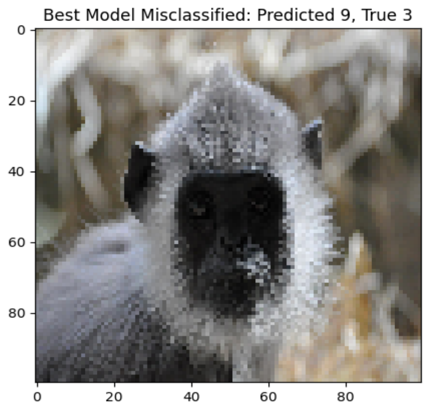
5 – Mandril

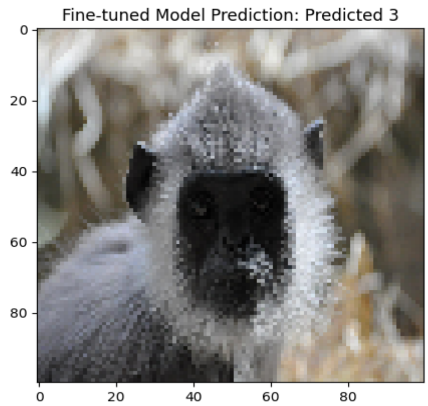
6 – Proboscis Monkey

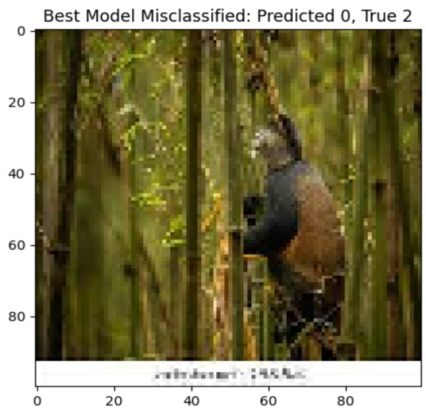
7 – Red Howler

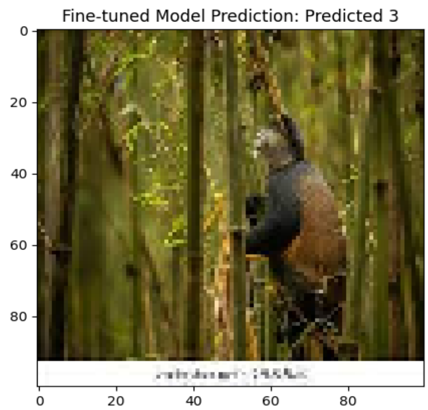
8 – Vervet Monkey

9 – White Faced Saki

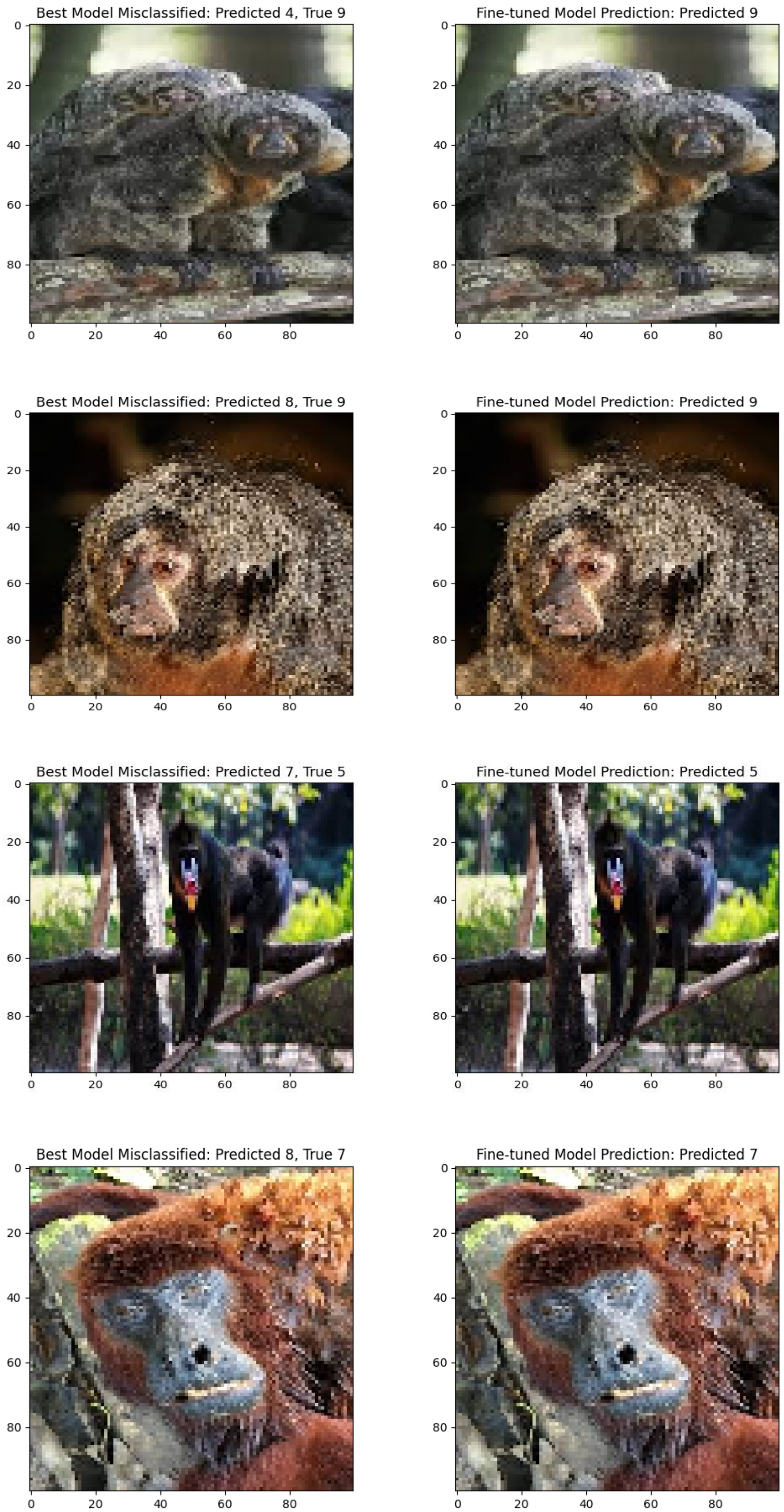









Best Model Misclassified: Predicted 1, True 5
Fine-tuned Model Prediction: Predicted 5
Best Model Misclassified: Predicted 9, True 4
Fine-tuned Model Prediction: Predicted 4
Best Model Misclassified: Predicted 5, True 8
Fine-tuned Model Prediction: Predicted 9
Best Model Misclassified: Predicted 3, True 1
Fine-tuned Model Prediction: Predicted 1


Best Model Misclassified: Predicted 4, True 9
Fine-tuned Model Prediction: Predicted 9
Best Model Misclassified: Predicted 8, True 9
Fine-tuned Model Prediction: Predicted 9
Best Model Misclassified: Predicted 7, True 5
Fine-tuned Model Prediction: Predicted 5
Best Model Misclassified: Predicted 8, True 7
Fine-tuned Model Prediction: Predicted 7

2. Give possible qualitative reasons why the better model of Task 1 may be making those mistakes, and some qualitative reasons why the fine-tuned model may or may not be improving.

Reasons for Misclassification by the Best Model of Task 1:

1. Similar Features Among Classes: Similar physical characteristics that are shared by various classes appear to be the cause of many misclassifications. For example, similar-looking color patterns and body forms could deceive a model with shallow feature extraction layers.

2. Background Noise: It seems that complicated backgrounds have an impact on misclassifications as well, which could be misleading the model. The model may predict the incorrect class when overlapping textures or colors that are similar to the subject are present.

Improvements Seen with the Fine-tuned Model:

1. Enhanced Feature Gathering: Using a pre-trained foundation such as EfficientNetV2S, the fine tune model gains access to more complex architectural elements like scaled depth and breadth, which improve its capacity to identify finer details and class distinctions.

2. Generalization Abilities: Because pre-trained models were exposed to a more diverse dataset during initial training, they are often better at generalizing from the training data to unseen data. The way it performs demonstrates this, since it accurately discovers classes that Task 1's best model misclassified.