Species Identification with Deep Convolutional Networks

Fabian Otto

I. Scores

The currently best performing ResNet-50 Architecture achieves a test accuracy of 94.92% and a corresponding training accuracy of 97.63% after fine-tuning based on ImageNet for the provided Species data set. The accuracy's improvement throughout the training progress can be seen in figure 1.

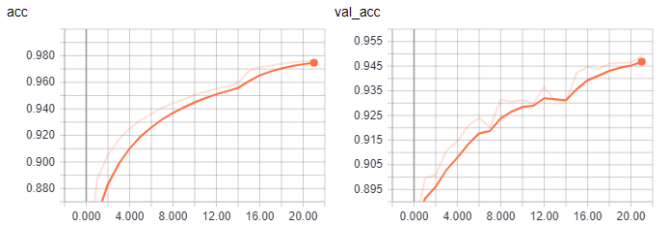

Fig. 1. Accuracy improvement during training.

Further, table I provides an overview about the performance on a reduced coarse grained level.

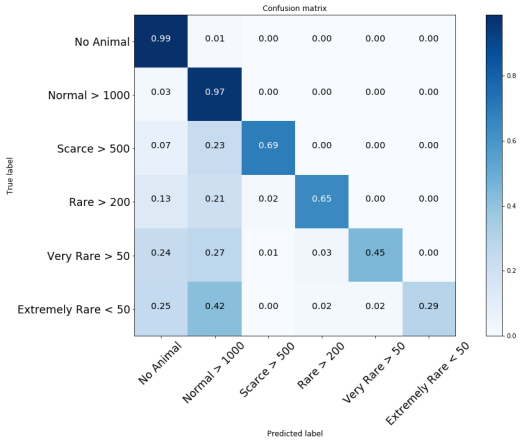

Fig. 2. Confusion matrix for reduced classes.

II. Correct Attention, Correct Prediction

As seen in figure 2 the network mainly focuses on the animal itself, this can be seen for most correct predictions.

III. Correct Attention, Wrong Prediction

The network often finds the correct attention in the images (as in 3), i.e. the animals. However, especially rare species are still predicted incorrectly based on the small sample size during training. This can be seen in the last image pair 3, it shows a Crested Serpent Eagle, which is classified as Elephant.

Advisor: Dr. Anirban Mukhopadhyay, GRIS, TU Darmstadt, SS 2018.

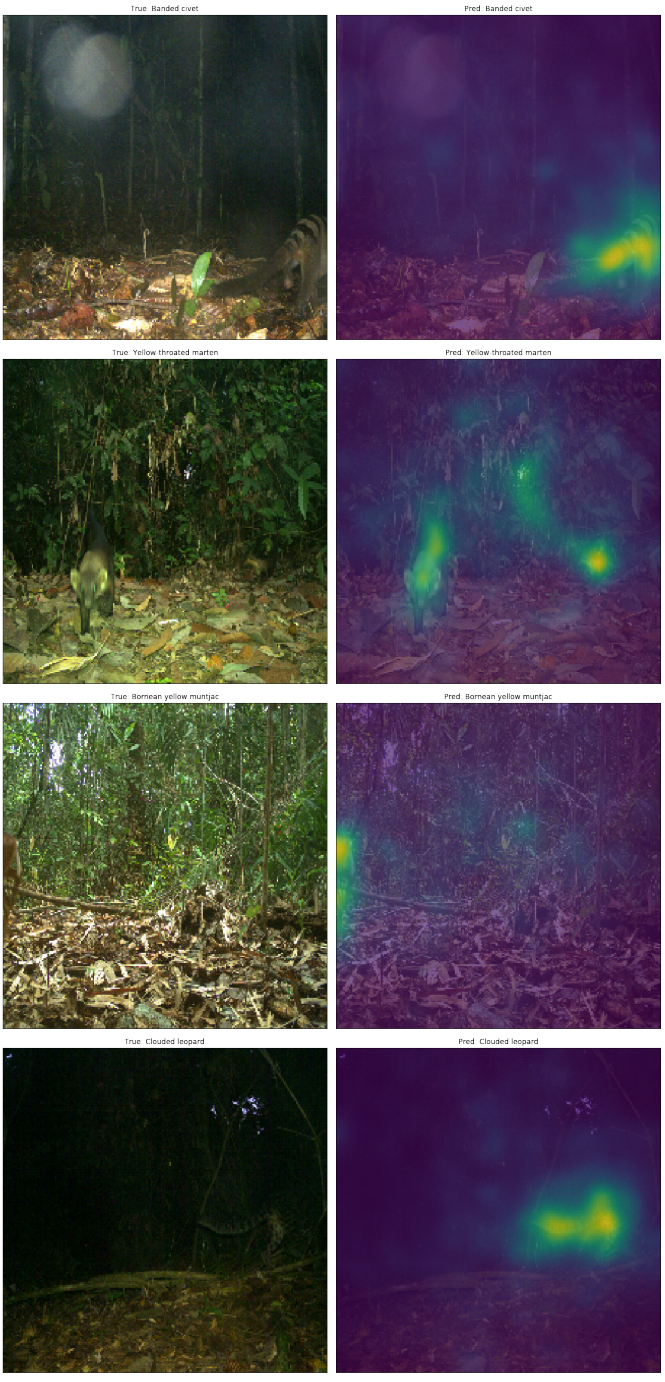

Fig. 3. Correct prediction and correct attention of the network.

IV. Wrong Attention, Correct Prediction

In contrast to the first set of image from 2, in some cases the attention is slightly more spread around the animal, this

TABLE I
PERFORMANCE EVALUATION FOR SIMPLIFIED CLASSES.

	Precision	Recall	F1-Score	Images	Classes
No Animal	0.98	0.99	0.98	60231	9
Normal > 1000	0.97	0.97	0.97	39932	16
Scarce > 500	0.90	0.69	0.78	1077	6
Rare > 200	0.92	0.65	0.76	623	8
Very Rare > 50	0.98	0.45	0.62	317	15
Extremely Rare < 50	0.45	0.28	0.34	108	3
avg / total	0.97	0.97	0.97	102288	

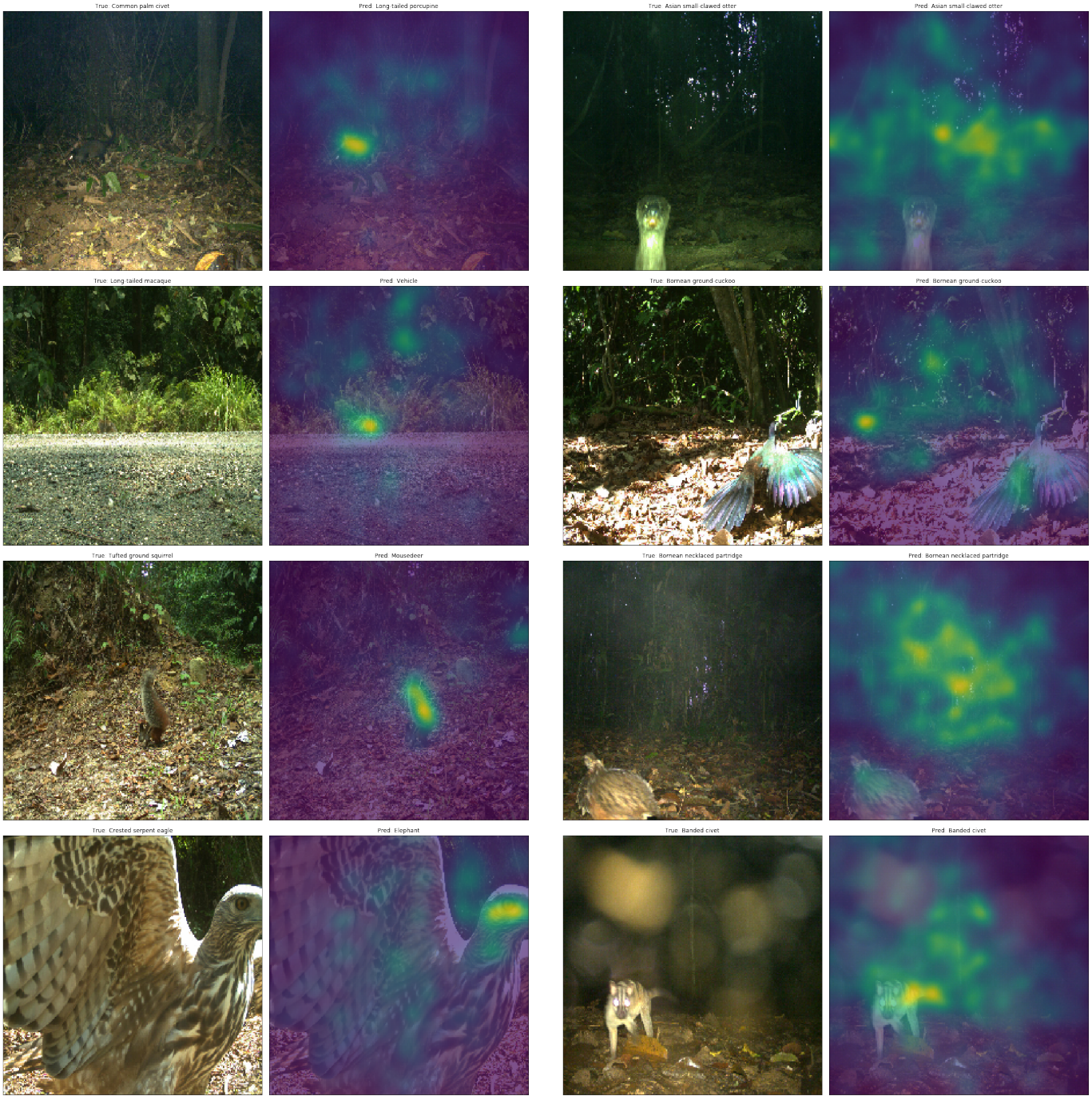

Fig. 4. Wrong prediction and correct attention of the network.

Fig. 5. Correct prediction and wrong attention of the network.

happens with bad lighting or similar background colors. This should not be seen as actual learning of the network and probably occurs due to the 3-image sequences, which are shot and provide similar context without the animal.

V. Wrong Attention, Wrong Prediction

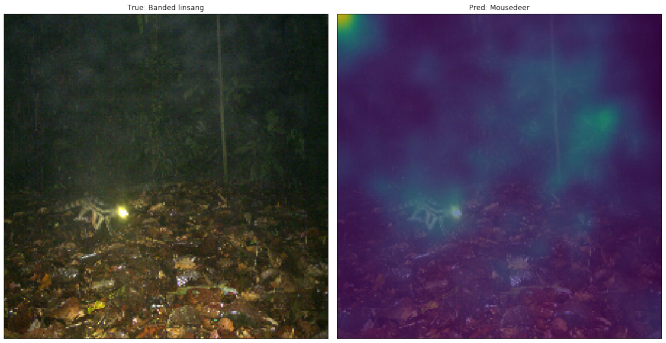

Fig. 6. Wrong prediction and wrong attention of the network.

The amount of cases in which attention and prediction are both wrong are quite limited.

VI. Notable Cases

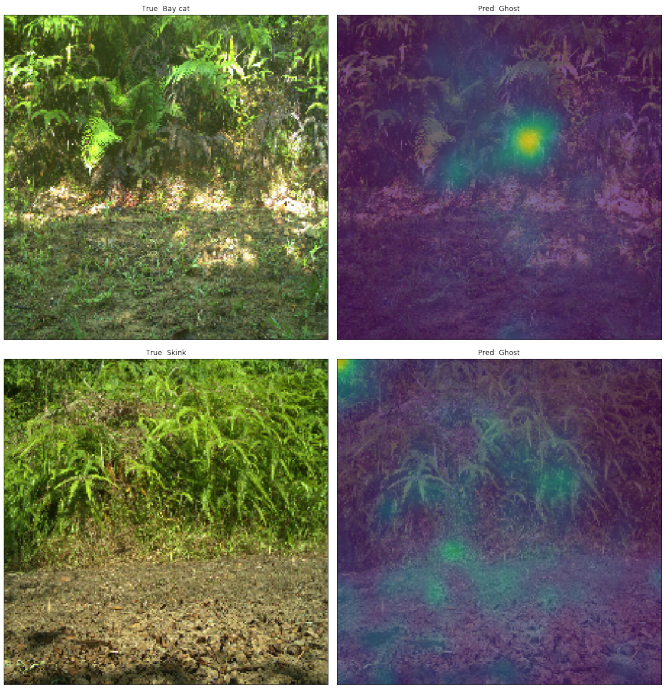

Fig. 7. Wrong prediction due to noisy label (Bay Cat and Skink are actually Ghost and predicted correctly).

Some addtional mistakes to highlight are shown in Figure 7 and 6. In figure 7 is one example that mismatches happend due to similarities between the classes, gere specifically Banded Civet and Malay Civet. Additionally, the provided labels include some noise, i.e. some labels might be incorrect as seen in figure 6. This probably can be attributed to process

Fig. 8. Wrong prediction due to class similarity.

of taking 3-image shots and classifying all image the same. Besides those issues, camera occlusions due to rain are often influencing the prediction as already mentioned earlier (figure 4). Apart from those more specific issues, as always, frequent classes are more often assigned to an incorret label, which can also be seen in the confusion matrix in figure **??**.

VII. Confusion Matrix

The reduced confusion matrix also shows that the network is able to distinguish between animals and non animals quite well as well as frequent animals, which make up the majority of the data set. This differentiation between animals and non animals is also present for all other categories, the network is more likely to assign a wrong label from the common animal classe than from a non animal class.