Reevaluating Automated Wildlife Species Detection: A Reproducibility Study on a Custom Image Dataset

Preprint, compiled October 1, 2025

Tobias Abraham Haider [1, 2*]

[1]Vienna University of Technology
[2]University of Veterinary Medicine Vienna

Abstract

This study revisits the findings of Carl et al. [1], who evaluated the pre-trained Google Inception-ResNet-v2 model for automated detection of European wild mammal species in camera trap images. To assess the reproducibility and generalizability of their approach, we reimplemented the experiment from scratch using openly available resources and a different dataset consisting of 900 images spanning 90 species. After minimal preprocessing, we obtained an overall classification accuracy of 62%, closely aligning with the 71% reported in the original work despite differences in datasets. As in the original study, per-class performance varied substantially, as indicated by a macro F1 score of 0.28, highlighting limitations in generalization when labels do not align directly with ImageNet classes. Our results confirm that pretrained convolutional neural networks can provide a practical baseline for wildlife species identification but also reinforce the need for species-specific adaptation or transfer learning to achieve consistent, high-quality predictions.

Keywords machine learning, reproducibility, camera trap, pre-trained model, animal species classification, computer vision, neural networks, cnn, resnet, tensorflow, wildlife monitoring

1 Introduction

While biodiversity is decreasing at a rapid pace, the rise of specific species, be they invasive or predatory, concerns societies around the world. As a consequence, researchers and conservationists are interested in continuously monitoring wildlife populations in terms of their geographical distribution, size, and behavior. Researchers successfully deploy camera traps that can take photographs of passing animals without disturbing them [2]. The photos are typically manually collected from the traps and annotated with the name of the species present in the image [3].

Deep convolutional neural networks (CNNs) have emerged as a promising solution to automate this process, offering robust image classification capabilities. Building on prior work, this study evaluates the reproducibility of results reported by Carl et al. [1], who applied a pre-trained Inception-ResNet-v2 model for European mammal species detection in camera trap images. By reconstructing their experiment using a different dataset and a reproducible, open-source workflow, we examine both the reliability of the original findings and the generalizability of pre-trained CNNs to broader wildlife monitoring scenarios.

2 Experiment Setup

We reimplemented the Python code for the experiment from scratch because all the necessary components (data [4], model [5], and metrics [6]) can be taken from stable public sources. To maximize the readability and reproducibility of the experiment, a minimal setup was chosen, defining all necessary code, data, and requirements in one GitHub project [7]. State-of-the-art Python packages are chosen, installed, and imported. The exact versions are shown in Table 1. The Jupyter notebook is run locally on a Thinkpad T14 with an AMD Ryzen 5 PRO 5650U processor and 16 GB of memory but no GPU. The operating system is Linux Mint 22.1 and the Python kernel is version 3.12. We expect there to be no deviation in the results, even if different hardware or runtime is chosen.

Table 1: Runtime dependencies

package	version
pathlib	1.0.1
Pillow	11.3.0
numpy	2.1.3
pandas	2.3.1
tensorflow	2.19.0
scikit-learn	1.7.1

3 Model

After setting up the Python runtime and importing the required packages, we load the Inception-ResNet-v2 model from the TensorFlow model repository [8]. We use the publicly available pretrained weights, which were obtained by training the model on the ImageNet dataset [9]. This approach eliminates the model design and training phase completely but limits the model prediction space to the 1000 classes from the ImageNet dataset.

```
model = InceptionResNetV2(weights="imagenet")
```

4 DATA

Carl et al. provide the source of the wildlife images used in their dataset [10]. This source is no longer available, requiring us to run the experiment on a different dataset. To test the generalizability of the model, we take a larger public dataset containing images of 90 different species [4]. To mimic the original experiment setup, only 10 samples are used for each species, resulting in a total test sample size of 900 images.

4.1 Data Preprocessing

We load the images, respecting all three color channels (RGB), resize them to 299 by 299 pixels, and convert them into a 1-dimensional vector. The color intensities are scaled to be floating-point numbers from 0 to 1. This is the minimal preprocessing required to fit the required input size of the neural network.

```
def load_image(path, target_size):
    image = Image.open(path).convert("RGB")
    image = image.resize(target_size)
    return np.array(image) / 255.0
```

Then we construct the testing dataset by stacking all normalized image vectors and using the folder names as the labels.

```
animal_images = [load_image(p, input_shape)
                 for p in wildlife_image_paths]
animal_species = [p.parent.name
                  for p in wildlife_image_paths]

X_test = np.stack(animal_images, axis=0)
y_true = animal_species
```

5 TEST

The Inception-ResNet-v2 model outputs a probability distribution over 1,000 classes, corresponding to the categories defined in the ImageNet dataset. For this study, we use only the top-1 prediction (the class with the highest softmax probability) as the model's output and compare it to the ground-truth label from our test dataset.

```
y_pred = model.predict(X_test)
y_pred = [pred[0][1] # take output label
          for pred
          in decode_predictions(y_pred, top=1)]
```

When looking at the predictions, it becomes apparent that the model yields usable results. Almost all inference outputs are animal species somehow related to the one present in the image. The data already shows that the InceptionResNetV2 is generalizable to some extent.

Table 2: Subset of Inception-ResNet-v2 raw predictions

y_pred	y_true
gazelle	antelope
badger	badger
hummingbird	bat
brown_bear	bear
bee	bee
honeycomb	beetle
bison	bison
wild_boar	boar
ringlet	butterfly
Egyptian_cat	cat

5.1 Label Mapping

A key challenge in this setup is that ImageNet classes do not align directly with the wildlife species in our dataset. To enable evaluation, we constructed a manual mapping table linking model output labels to the target species. This mapping followed a best-effort approach based on the Linnean system of taxonomy.

Several limitations arise from this process:

- When we collapse multiple species into higher-level taxa (e.g., mapping all bear species to "bear"), we lose some species-level detail.
- Certain species in the dataset are not represented in ImageNet classes at all (e.g., bats, deer), preventing meaningful predictions for these cases.

While this mapping introduces ambiguity, it reflects a realistic challenge when applying pretrained ImageNet models to ecological data and illustrates the need for task-specific model adaptation.

6 EVALUATION

Carl et al. reported two performance metrics for their study: overall classification accuracy across the dataset and per-species accuracy. To enable a direct comparison, we adopt the same evaluation strategy. After applying the label mapping described above, predictions are grouped by true species, and accuracy is computed at both the global and class levels. Additionally, we compute the macro F1 score to quantify the per-class imbalance observed in the results.

7 SUMMARY

Our reproduced results confirm the findings of Carl et al. We achieve an overall top-1 prediction accuracy of 62% by using a larger dataset that includes many species not present in the original study. This result is comparable to the 71% reported by Carl et al. Consistent with their study, we observe substantial variation in per-species accuracies, signaled by a macro F1

Table 3: Mapping rules between ImageNet classes and test data classes

ImageNet label	dataset label
gazelle, impala	antelope
American_black_bear, brown_bear	bear
ground_beetle, leaf_beetle, rhinoceros_beetle, dung_beetle	beetle
wild_boar	boar
ringlet, monarch, sulphur_butterfly, lycaenid	butterfly
Egyptian_cat, tabby, Siamese_cat, Persian_cat, lynx	cat
water_buffalo	cow
Dungeness_crab, fiddler_crab, rock_crab, king_crab	crab
magpie, jay	crow
red_deer, elk	deer

Table 4: Prediction accuracy per species and the total accuracy

species	accuracy
bison, bear, boar, crab, elephant, eagle, dog, chimpanzee, cockroach, snake, panda, pelecaniformes, pig, koala, orangutan, ladybug, leopard, lobster, hornbill, jellyfish, hyena, hummingbird, goose, goldfish, fox, fly, sandpiper, zebra, wombat, turtle	1
parrot, shark, starfish, squirrel, otter, kangaroo, penguin, coyote, butterfly, flamingo, badger, bee, antelope, hare, gorilla, porcupine, tiger, hamster	0.9
sheep, lizard, lion, cat, dragonfly, wolf	0.8
beetle, hippopotamus, ox, grasshopper	0.7
whale	0.6
duck	0.4
owl, goat, crow	0.3
swan	0.1
caterpillar, bat, dolphin, donkey, cow, deer, mosquito, horse, hedgehog, okapi, moth, mouse, octopus, seal, raccoon, rat, possum, pigeon, oyster, seahorse, rhinoceros, reindeer, squid, sparrow, turkey, woodpecker	0
TOTAL	0.62

score of 0.28. 48 species out of 90 are predicted with an accuracy greater than or equal to 90%, and 26 species are predicted without any success (0% accuracy). A detailed summary of per-species prediction accuracies is provided in Table 4.

The experiment shows that pretrained convolutional neural networks, like the Inception-ResNet-v2, are a viable option for the annotation of camera trap images. The network design allows for detailed pattern recognition and robust identification of a large number of animal species. The main issue encountered is the model weights, which are fit for a fixed set of animal species but unsuitable for many classes in our dataset.

8 FUTURE WORK

Previous research has explored transfer learning of convolutional neural networks for recognizing animals and their facial features, showing that retraining pretrained networks for a specific use case can substantially improve prediction accuracy [11]. We expect this to be a key way to overcome low prediction accuracies for certain species.

We emphasize that, although highly powerful, the models investigated in most studies remain very deep and too large for large-scale deployment in nature. The Inception-ResNet-v2 uses about 55 million parameters, requiring a significant amount of memory and energy. Other models such as MobileNet [11] and EfficientNet [12] use far fewer layers and are optimized for edge deployment.

For future work, we suggest combining these two adaptations (transfer learning and smaller neural networks) and reevaluating this experiment. This effort may yield a highly efficient and accurate model suitable for scalable deployment in real-world wildlife monitoring.

REFERENCES

[1] Christin Carl, Fiona Schönfeld, Ingolf Profft, Alisa Klamm, and Dirk Landgraf. Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *European Journal of Wildlife Research*, 66(4), 7 2020. ISSN 1439-0574. doi: 10.1007/s10344-020-01404-y. URL http://dx.doi.org/10.1007/s10344-020-01404-y.

[2] Franck Trolliet, Marie-Claude Huynen, Cédric Vermeulen, and Alain Hambuckers. Use of camera traps for wildlife studies. A review. *Biology Agriculture Science Environnement*, 18:446–454, 1 2014.

[3] Dhruv Tulasi, Alys Granados, Prabath Gunawardane, Abhay Kashyap, Zara McDonald, and Sunil Thulasidasan. Smart Camera Traps: Enabling Energy-Efficient Edge-AI for Remote Monitoring of Wildlife. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on AI-Driven Spatio-Temporal Data Analysis for Wildlife Conservation*, GeoWildLife '23, pages 9–16, Hamburg, Germany, 2023. Association for Computing Machinery. ISBN 9798400703553. doi: 10.1145/3615893.3628760. URL https://doi.org/10.1145/3615893.3628760.

[4] Sourav Banerjee. Animal Image Dataset (90 Different Animals). https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals, 2024. URL https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals. Accessed: 2025-09-03.

[5] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, 2016. URL https://arxiv.org/abs/1602.07261.

[6] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[7] Tobias Haider. tobsel7/research-vetmedwien-animal-species-identification: More detailed evaluation and description of experiment results, 2025. URL https://zenodo.org/doi/10.5281/zenodo.17116549.

[8] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 6 2009. doi: 10.1109/cvpr.2009.5206848. URL http://dx.doi.org/10.1109/CVPR.2009.5206848.

[10] Nationalparkverwaltung Hainich, FFK Gotha. Schwarzwildforschung im Hainich. https://www.schwarzwild-hainich.de, 2019. URL https://www.schwarzwild-hainich.de.

[11] Andrew Howard, Mark Sandler, Bo Chen, Weijun Wang, Liang-Chieh Chen, Mingxing Tan, Grace Chu, Vijay Vasudevan, Yukun Zhu, Ruoming Pang, Hartwig Adam, and Quoc Le. Searching for MobileNetV3. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324. IEEE, 10 2019. doi: 10.1109/iccv.2019.00140. URL http://dx.doi.org/10.1109/ICCV.2019.00140.

[12] Mingxing Tan and Quoc V. Le. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks. 2019. doi: 10.48550/ARXIV.1905.11946. URL https://arxiv.org/abs/1905.11946.