

# REEVALUATING AUTOMATED WILDLIFE SPECIES DETECTION: A REPRODUCIBILITY STUDY ON A CUSTOM IMAGE DATASET

PREPRINT, COMPILED SEPTEMBER 11, 2025

Tobias Abraham Haider <sup>1, 2\*</sup>

<sup>1</sup>Vienna University of Technology

<sup>2</sup>University of Veterinary Medicine Vienna

## ABSTRACT

This experiment attempts to reproduce the results of the paper [1], which tests the pretrained Google Inception-ResNet-v2 model for predicting animal species. We describe the required software, image loading processes, model outputs. Furthermore we calculate prediction global and per-class prediction accuracies and compare them to the metrics from the original paper.

**Keywords** machine learning, reproducibility, animal species classification, computer vision, neural networks, cnn, resnet, tensorflow, wildlife monitoring

## 1 INTRODUCTION

## 2 EXPERIMENT SETUP

Carl et al. do not supply great insight into the runtime environment. This is likely due to the standardized way the model can be used through the TensorFlow library. 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. State-of-the-art Python packages are chosen, installed, and imported. The exact versions are shown in Table 3. The Jupyter notebook is run locally on a Thinkpad T14 with an AMD Ryzen 5 PRO 5650U processor, 16 GB of memory, and Linux Mint 22.1 installed. No GPU was used, but it can be expected that the results would not change if one were used.

```
!python3.12 -m pip install -r requirements.txt
```

## 3 MODEL

Once the Python runtime is set up and the packages are imported, the neural network can be instantiated and directly fit to the ImageNet dataset. This eliminates all model design and training work.

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

## 4 DATA

For the experiment, 90 common animals are used. They are sourced from Google Images and provided in a labeled format in [2]. The Kaggle dataset is rather large. 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

The images are loaded with three color channels (RGB), resized to 299 by 299 pixels and converted into an 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_normalized_image(path, target_size):  
    image = Image.open(path).convert("RGB")  
    image = image.resize(target_size)  
    return np.array(image) / 255.0
```

The testing data is constructed by stacking the normalized image vectors and using the folder names as the label.

```
animal_images = [load_normalized_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 model yields a probability for each of the 1000 classes. The classes represent 1000 different classes taken from the ImageNet database. For this experiment, we use the output from the top neuron of the final softmax layer and compare its label to the true label.

Table 1: Subset of Inception-ResNet-v2 raw predictions

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

```

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 results, 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. This shows that the InceptionResNetV2 is generalizable to some extent.

### 5.1 Label Mapping

The main issue with this experiment is the set of classes known to the model, which do not match the dataset used for testing. This is not specific to this dataset, but it is very likely to happen in any kind of realistic setup. We manually define a mapping table to relate the model output label to the labels from the dataset.

This mapping is done manually, as a best-effort approach following the Linnean system of taxonomy, and we acknowledge some shortcomings of it:

- Some semantic information is lost as species are sometimes mapped to their families. (e.g., all bear species are mapped to bear)
- The dataset contains images of species that are not directly related to any class from the model. (e.g., all bats, deers)

We document all defined mappings in Table 4.

## 6 EVALUATION

Carl et al. provide two kinds of performance metrics: overall model accuracy and the accuracy for each species. By grouping the samples by the true species, it is straightforward to calculate metrics that can be directly compared to the results from Carl et al.

```
accuracy = accuracy_score(y_true, y_pred_mapped)
```

Refer to Table 6 for the group accuracies for each species.

Table 2: Prediction accuracy for 5 different species and the total accuracy

species	accuracy
bison	1.0
goose	1.0
gorilla	0.9
dolphin	0.0
woodpecker	0.0
...	...
TOTAL	0.62

Table 3: Runtime dependencies

package	version
pathlib	1.0.1
tabulate	0.9.0
Pillow	11.3.0
numpy	2.1.3
pandas	2.3.1
tensorflow	2.19.0
scikit-learn	1.7.1
matplotlib	3.10.5
seaborn	0.13.2

## 7 SUMMARY

## 8 FUTURE WORK

## 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] 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.

Table 4: Imagenet label mapping

mapped label	imagenet label
antelope	gazelle, impala
bear	American_black_bear, brown_bear
beetle	ground_beetle, leaf_beetle, rhinoceros_beetle, dung_beetle
boar	wild_boar
butterfly	ringlet, monarch, sulphur_butterfly, lycaenid
cat	Egyptian_cat, tabby, Siamese_cat, Persian_cat, lynx
cow	water_buffalo
crab	Dungeness_crab, fiddler_crab, rock_crab, king_crab
crow	magpie, jay
deer	red_deer, elk
dog	Labrador_retriever, basset, Border_collie, Chihuahua, Bouvier_des_Flandres, Brittany_spaniel, English_setter, Greater_Swiss_Mountain_dog, Ibizan_hound, Mexican_hairless, tox_terrier, Pekinese, Pomeranian, golden_retriever, pug
donkey	ass
duck	mallard, drake
eagle	bald_eagle, golden_eagle
elephant	African_elephant, Indian_elephant, tusker
fox	Arctic_fox, red_fox, grey_fox, kit_fox
goat	ibex, mountain_goat
horse	Arabian_horse, Appaloosa
kangaroo	wallaby
lizard	agama, alligator_lizard, banded_gecko, Komodo_dragon, whiptail
lobster	American_lobster, spiny_lobster
mouse	house_mouse
owl	great_grey_owl
panda	giant_panda
parrot	African_grey, macaw, sulphur-crested_cockatoo
pelecaniformes	pelican, black_stork
penguin	king_penguin
pig	hog
sandpiper	red-backed_sandpiper, red-shank, dowitcher
shark	great_white_shark, hammerhead, tiger_shark
sheep	ram, bighorn
snake	horned_viper, vine_snake, king_snake, night_snake, Indian_cobra, ring-neck_snake, rock_python, thunder_snake
squirrel	fox_squirrel
swan	black_swan
turtle	box_turtle, logger-

Table 5: Inception-ResNet-v2 predictions

truth	mapped tion	predic- tion	model prediction
antelope	antelope		gazelle
antelope	antelope		impala
antelope	antelope		impala
antelope	antelope		gazelle
antelope	antelope		gazelle
antelope	antelope		impala
antelope	goat		ibex
antelope	antelope		gazelle
antelope	antelope		impala
antelope	antelope		impala
badger	badger		badger
badger	badger		badger
badger	badger		badger
badger	badger		badger
badger	bear		American_black_bear
badger	badger		badger
badger	badger		badger
badger	badger		badger
badger	badger		badger
badger	badger		badger
bat	hummingbird		hummingbird
bat	wood_rabbit		wood_rabbit
bat	hook		hook
bat	hummingbird		hummingbird
bat	cowboy_boot		cowboy_boot
bat	barracouta		barracouta
bat	house_finch		house_finch
bat	chime		chime
bat	cat		tabby
bat	mink		mink
bear	bear		brown_bear
bear	bear		brown_bear
bear	bear		American_black_bear
bear	bear		brown_bear
bear	bear		American_black_bear
bear	bear		brown_bear
bear	bear		brown_bear
bear	bear		American_black_bear
bear	bear		brown_bear
bear	bear		brown_bear
bee	bee		bee
bee	bee		bee
bee	honeycomb		honeycomb
bee	bee		bee
bee	bee		bee
bee	bee		bee
bee	bee		bee
bee	bee		bee
bee	bee		bee
bee	bee		bee
beetle	honeycomb		honeycomb
beetle	beetle		ground_beetle
beetle	fly		fly
beetle	beetle		ground_beetle
beetle	cockroach		cockroach
beetle	beetle		leaf_beetle
beetle	beetle		ground_beetle
beetle	beetle		rhinoceros_beetle
beetle	beetle		rhinoceros_beetle
beetle	beetle		dung_beetle
bison	bison		bison
bison	bison		bison
bison	bison		bison

Table 6: The prediction accuracy for each animal species

species	accuracy
bison	1
bear	1
boar	1
crab	1
elephant	1
eagle	1
dog	1
chimpanzee	1
cockroach	1
snake	1
panda	1
pelecaniformes	1
pig	1
koala	1
orangutan	1
ladybug	1
leopard	1
lobster	1
hornbill	1
jellyfish	1
hyena	1
hummingbird	1
goose	1
goldfish	1
fox	1
fly	1
sandpiper	1
zebra	1
wombat	1
turtle	1
parrot	0.9
shark	0.9
starfish	0.9
squirrel	0.9
otter	0.9
kangaroo	0.9
penguin	0.9
coyote	0.9
butterfly	0.9
flamingo	0.9
badger	0.9
bee	0.9
antelope	0.9
hare	0.9
gorilla	0.9
porcupine	0.9
tiger	0.9
hamster	0.9
sheep	0.8
lizard	0.8
lion	0.8
cat	0.8
dragonfly	0.8
wolf	0.8
beetle	0.7
hippopotamus	0.7
ox	0.7
grasshopper	0.7
whale	0.6
duck	0.4
owl	0.3
goat	0.3
crow	0.3
swan	0.1