Federico Ferretti, Matteo Pasotti, Salerno Fabio

Birds signals

Task 1

Audio classification

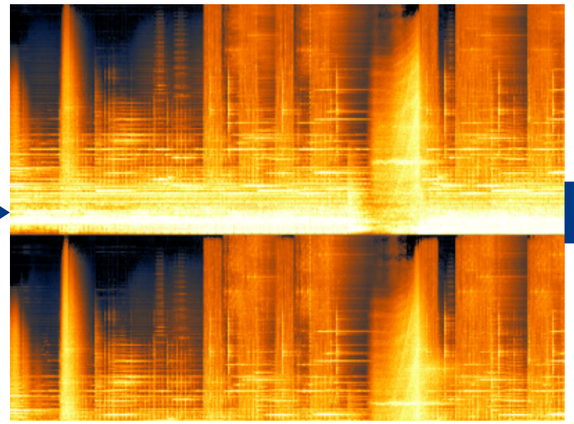
Dataset: BirdCLEF 25 SPECIES

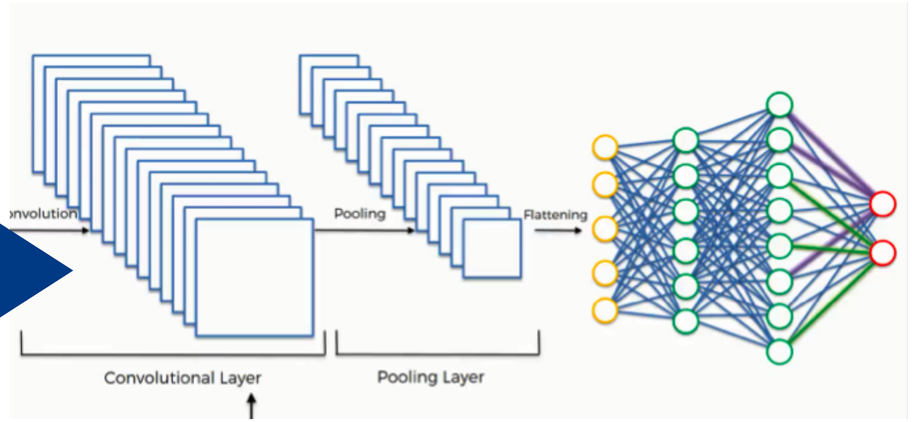
Dataset of:

25 bird species

- 1.700 training audio (~68 audio per species),
- 50 test audio (2 audio per species)

The architecture





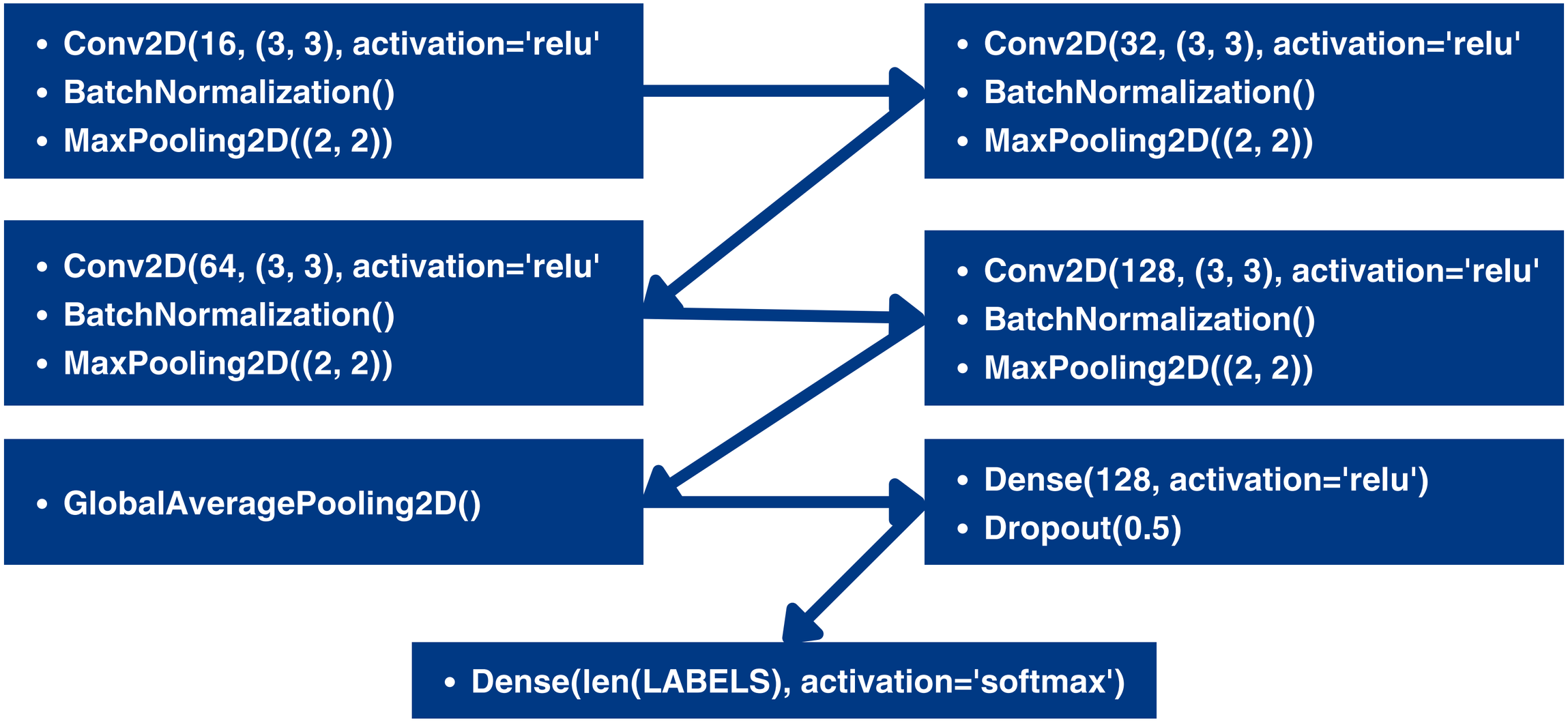
Audio bird

Spectrogram

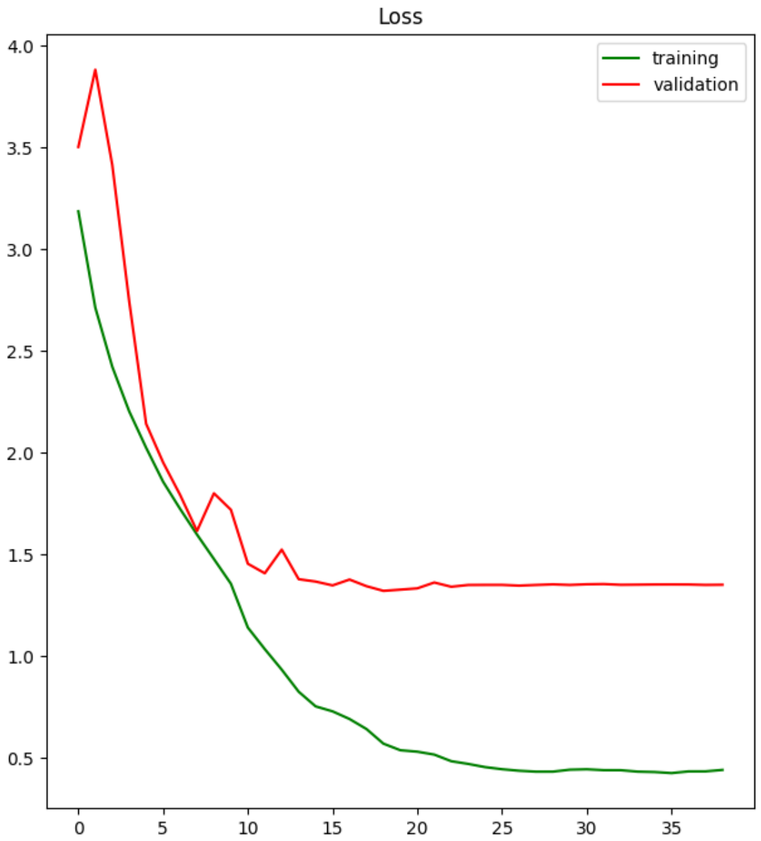
48 x 128

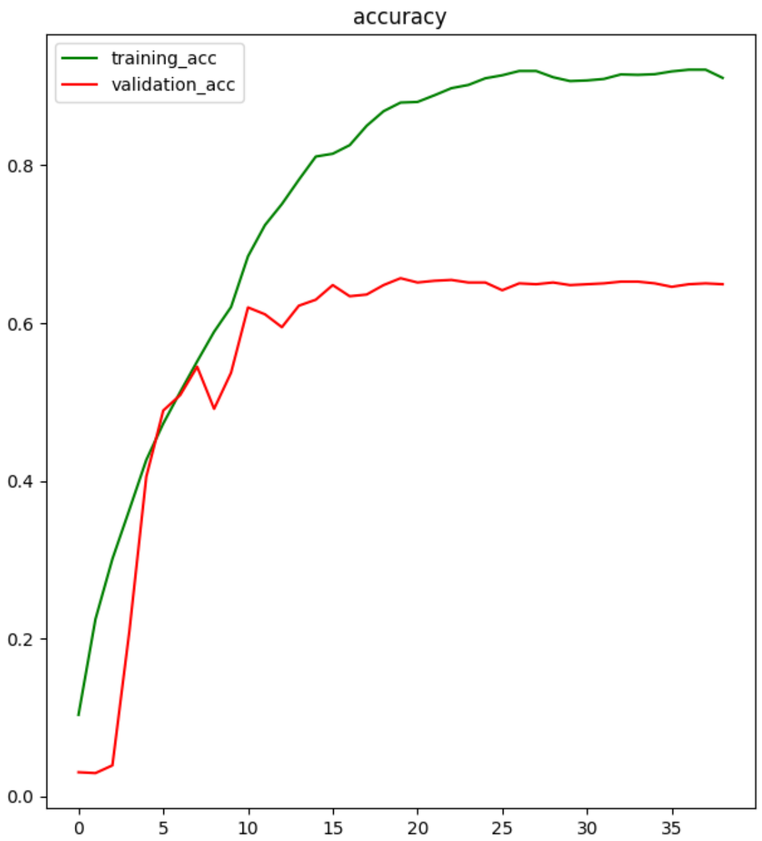
Convolutional Neural Networks

Convolutional Neural Network



Training





Epoch 39/50 with Early stopping

loss: 0.4425 - accuracy: 0.9109 - f1_score: 0.9111

val_loss: 1.3519 - val_accuracy: 0.6492 - val_f1_score: 0.6425

Test set performance

Overall

Accuracy: 0.60

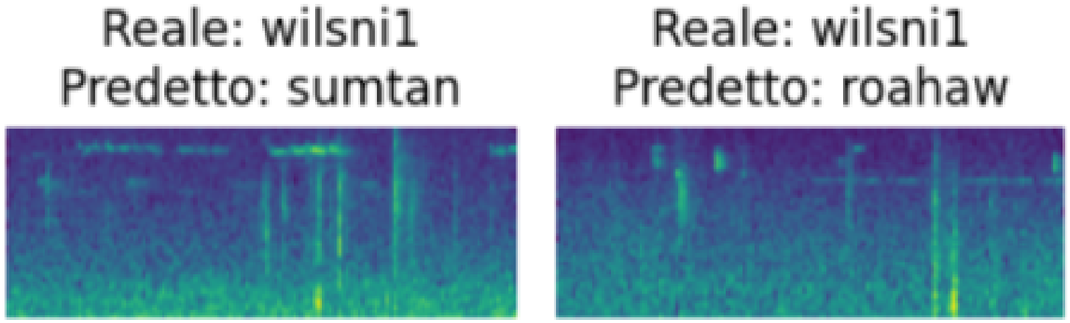
Precision: 0.60

Recall: 0.60

F1score: 0.56

Classe Reale	Classe Predetta
yehbla	yehbla
strfly1	bnhcow
butsal1	butsal1
yelwar	yelwar
rumfly1	rumfly1
blkpho	eastow
roahaw	roahaw
pilwoo	pilwoo
pilwoo	sumtan
ducfly	roahaw
yelwar	yelwar

Error Analysis

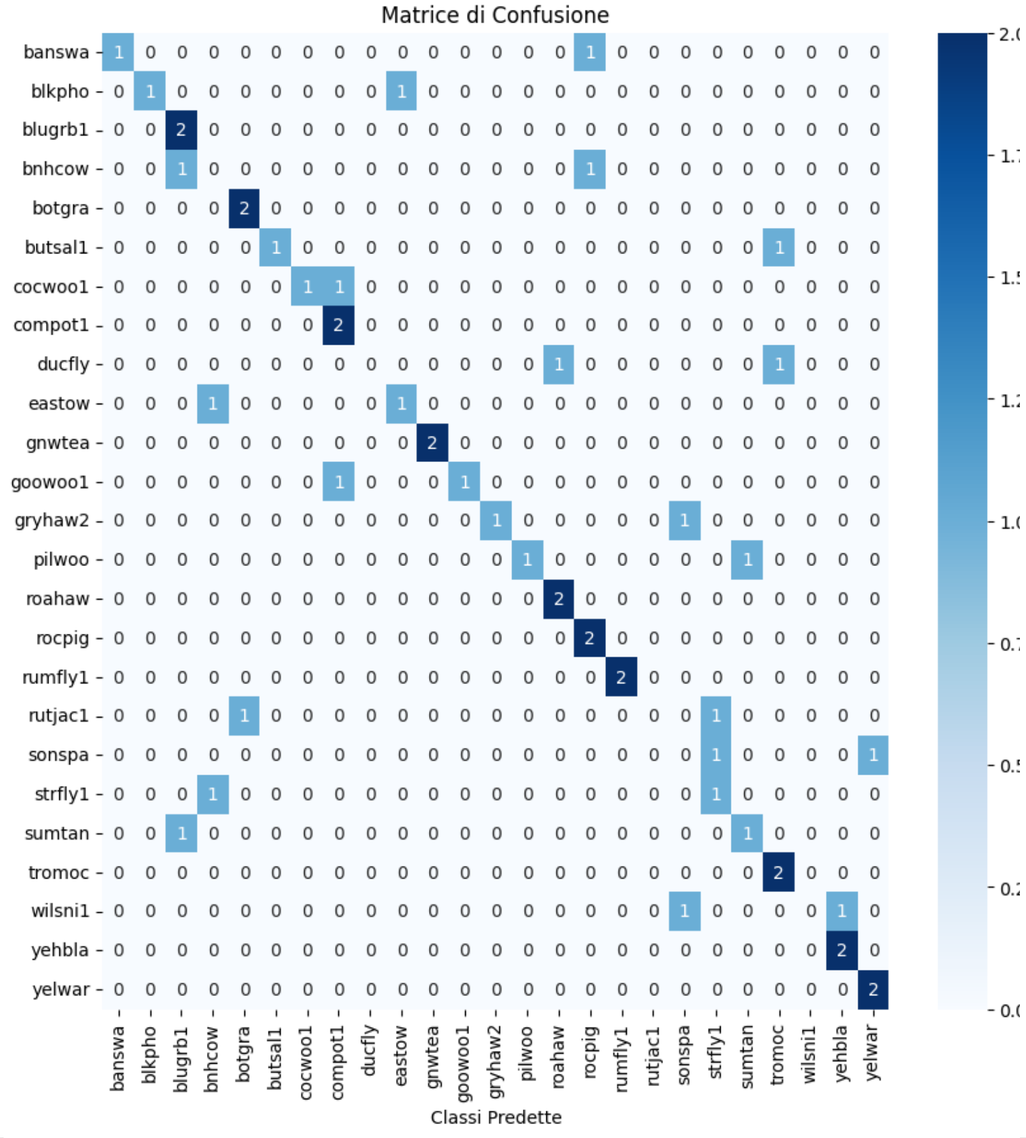


Strategies to advance and refine the model's capabilities

Dataset Expansion

Ensemble Learning

Post-Prediction Processing



Task 2

Image classification

Dataset: BIRDS 525 SPECIES

Dataset of:

525 bird species

- 84.635 training images (~160 img per species),
- 2.625 test images(5 img per species) and
- 2.625 validation images(5 img per species).

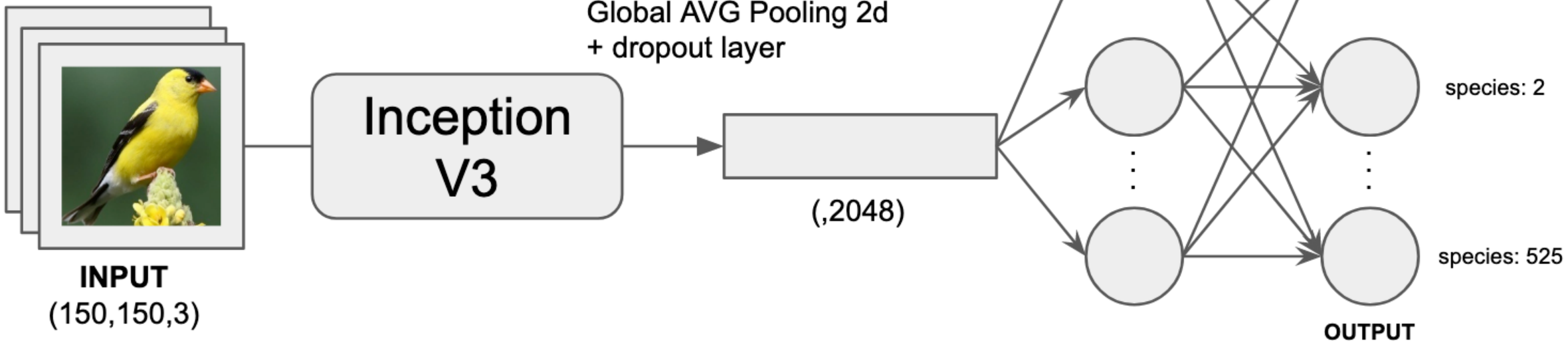
150x150x3

The architecture

InceptionV3
pre-processing

Data augmentation:

- Horizontal Flip
- Rotation Range=15
- Brightness Range=[0.8, 1.2]



Training

10 epochs training

4 epochs fine-tuning

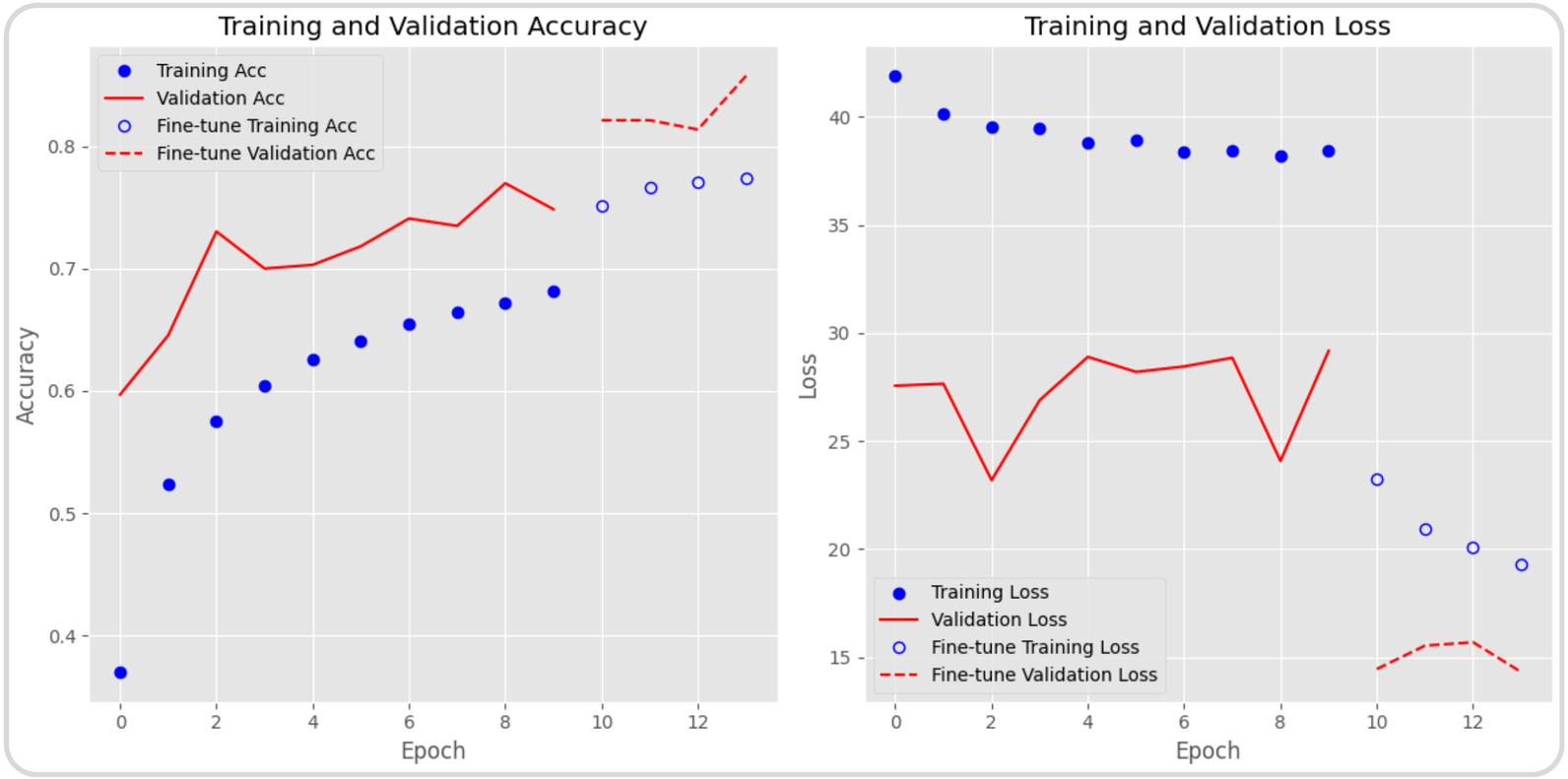
(last 12 layers unfreezed)

Accuracy

test (10 epochs): 76.6%

test (fine-tuning): 85.5%

8.9% improvement



Test set performance

Overall

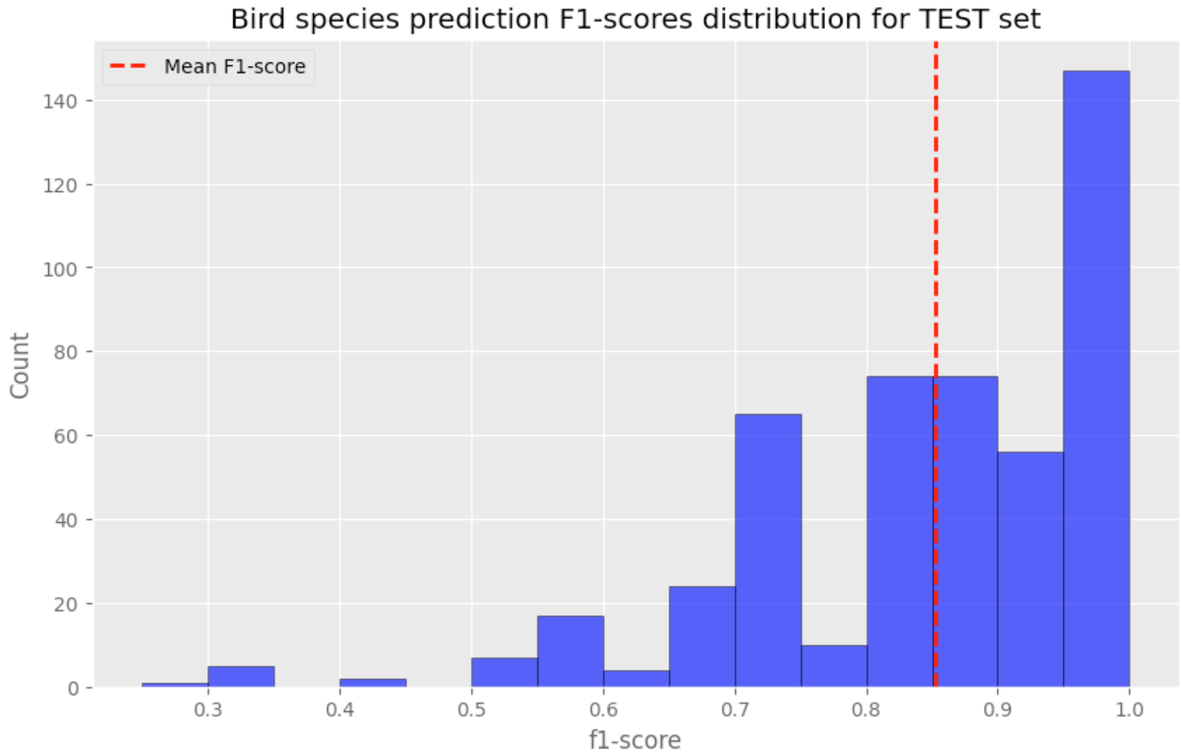
Accuracy: 85.5%

Precision: 88%

Recall: 85.5%

F1score: 85.2%

By species



By plotting each bird species' F1-Score on a distribution plot, we can see that there are **165** (30.4% of total) **bird species with F1-Scores smaller than 0.8**.

How the model makes prediction?

12 random birds classified

True: EMU
Predicted: EMU

True: BAR-TAILED GODWIT
Predicted: BAR-TAILED GODWIT

True: BAY-BREASTED WARBLER
Predicted: EASTERN BLUEBIRD

True: IWI
Predicted: IWI

True: SAMATRAN THRUSH
Predicted: SAMATRAN THRUSH

True: NORTHERN BEARDLESS TYRANNULET
Predicted: BROWN HEADED COWBIRD

True: YELLOW CACIQUE
Predicted: YELLOW CACIQUE

True: ROADRUNNER
Predicted: ROADRUNNER

True: CRIMSON SUNBIRD
Predicted: CRIMSON SUNBIRD

True: WHITE EARED HUMMINGBIRD
Predicted: ELEGANT TROGON

True: UMBRELLA BIRD
Predicted: UMBRELLA BIRD

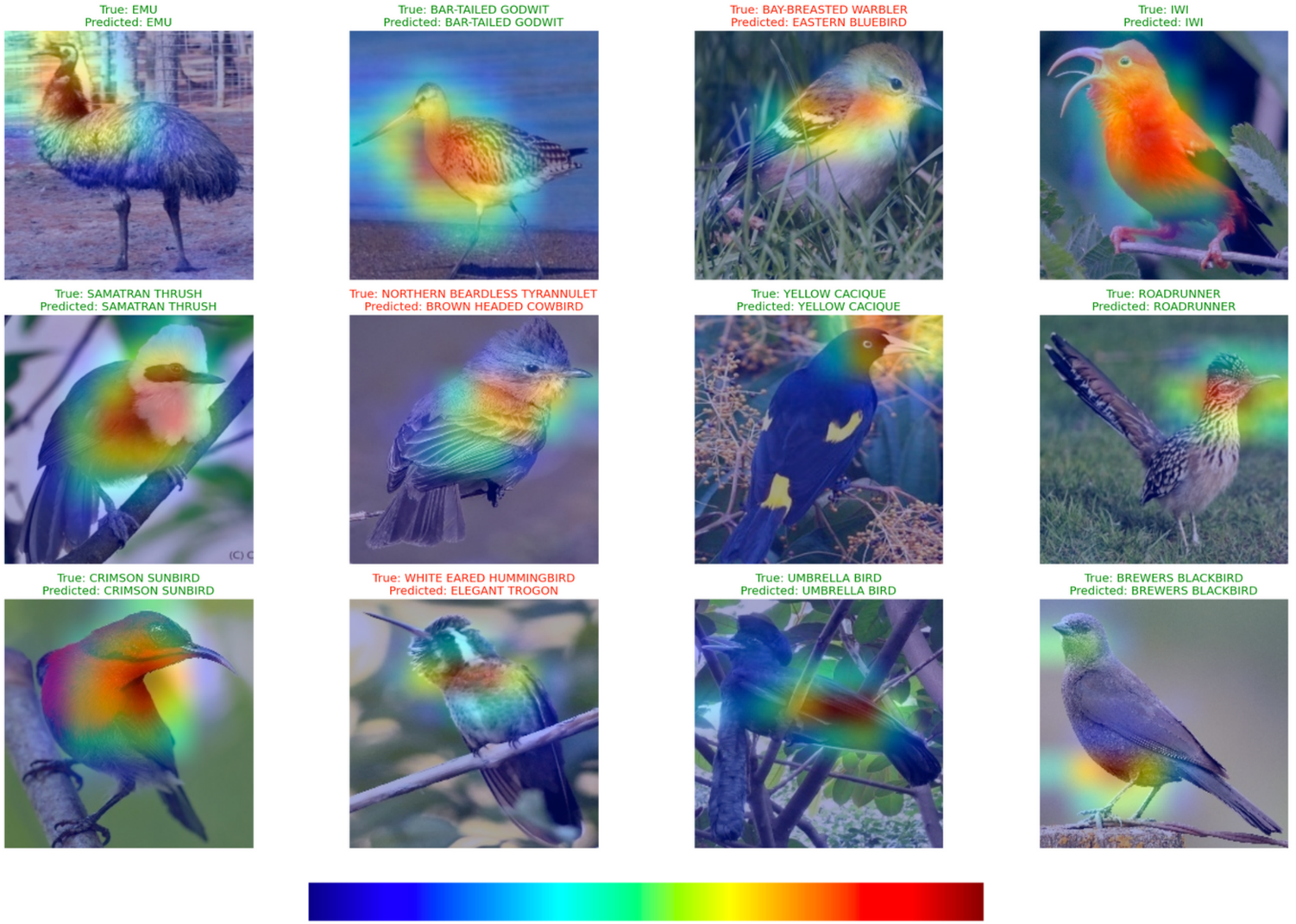
True: BREWERS BLACKBIRD
Predicted: BREWERS BLACKBIRD

Grad-CAM

Grad-CAM viz:
In every bird image, we find the areas of the image that are important for the CNN to make its class predictions.

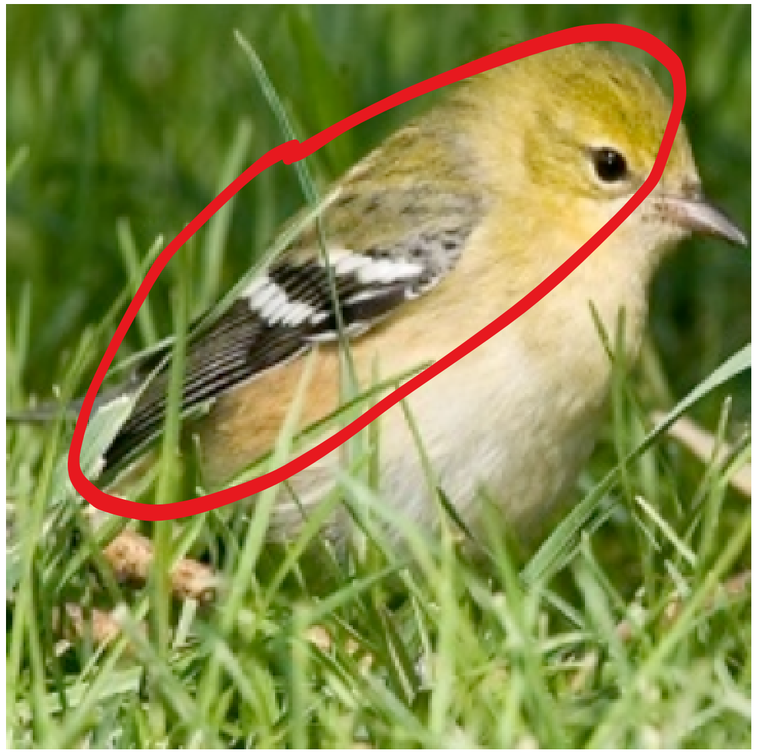
IWI species
(first row, forth column)
The CNN has made the correct prediction because of its red neck

BAY-BREASTED WARBLER species
(first row, third column)
The CNN focused on the wing. Why the mistake?

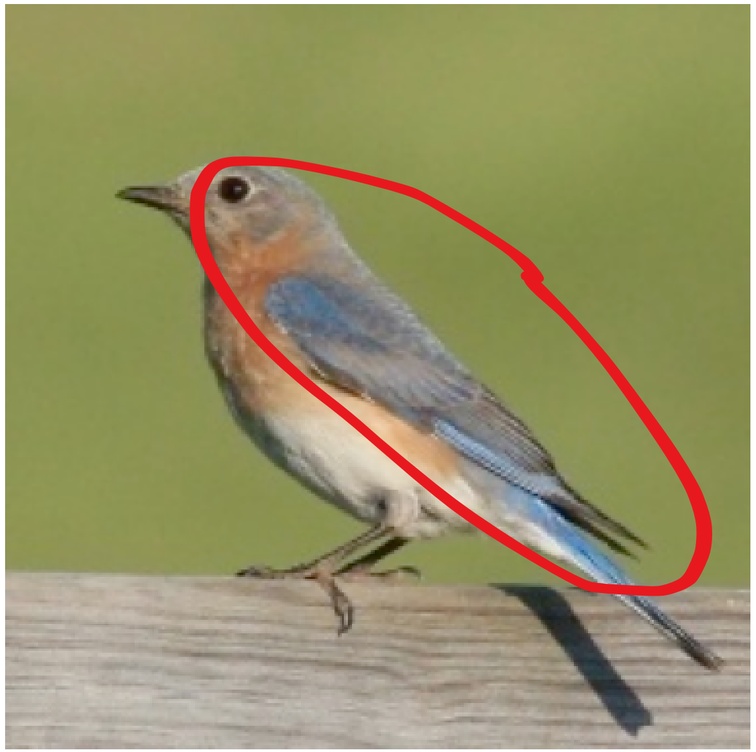


Why the mistake?

True: BAY-BREASTED WARBLER



Predicted: EASTERN BLUEBIRD

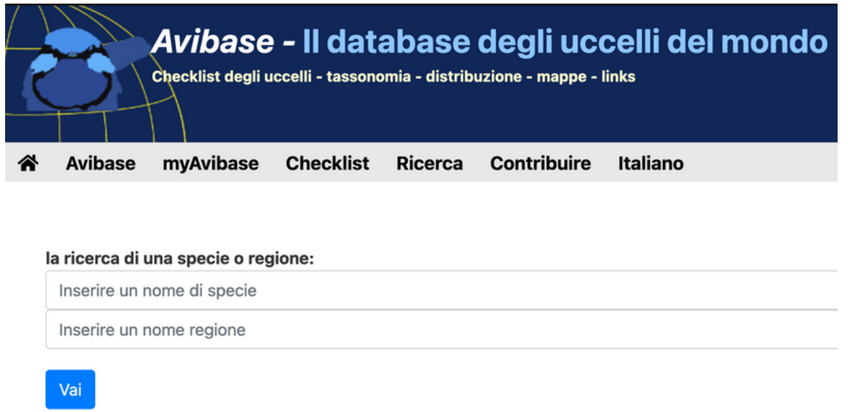


Outside the test set

50 birds images
from 17 species

Image quality:

- not centered
- less than 50% of the image
- hostile background



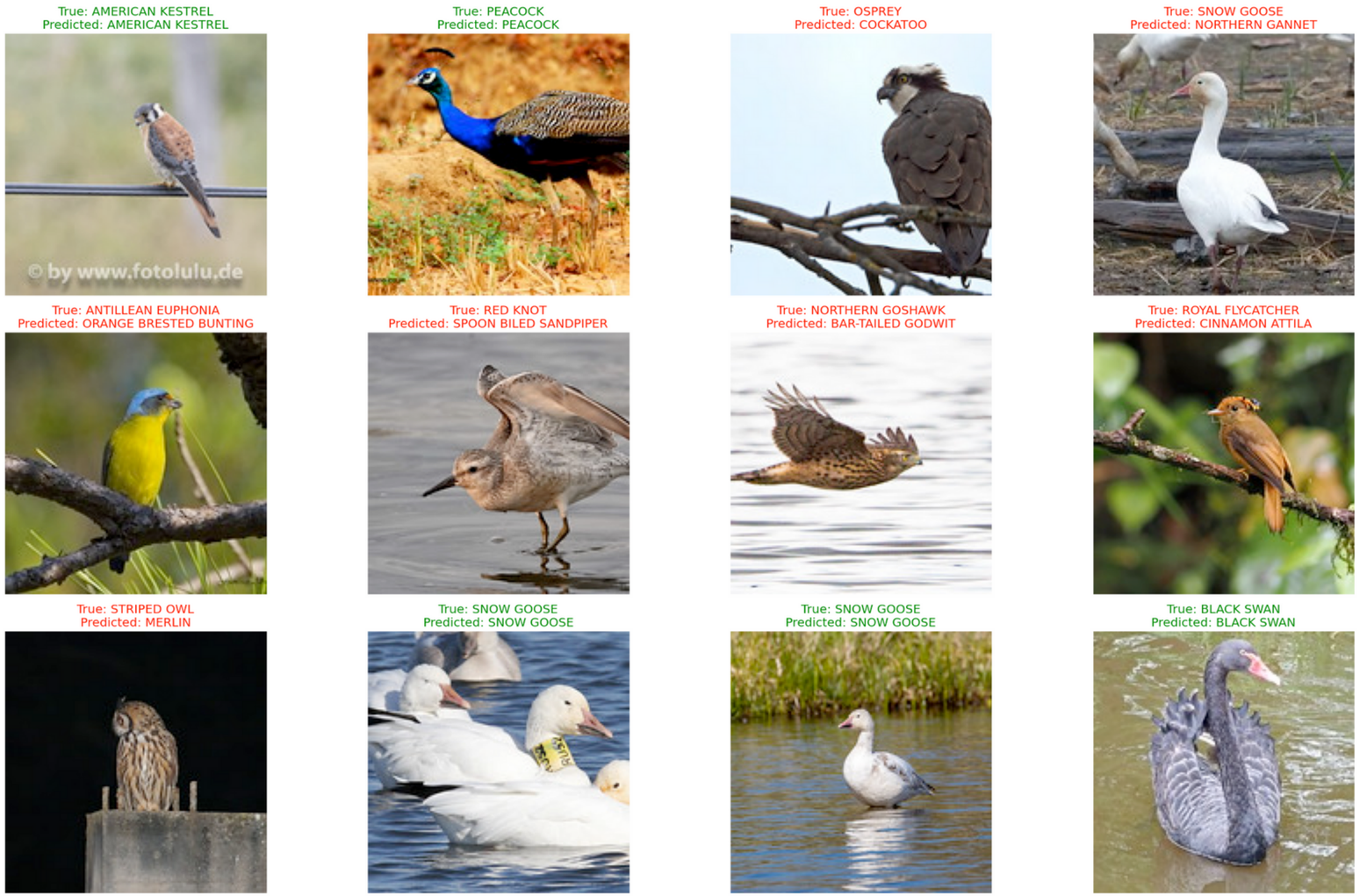
Outside the test set

Accuracy: 34%
Precision: 58%
Recall: 34%
F1score: 40%

Centered images (Black swan) or with more than one animal (Snow Goose). Good predictions

Not centered images (Northern Goshawk) or distant images (Stripped Owl) tends to be mistaken

12 random birds classified



Task 3

Image retrieval

Dataset: BIRDS 525 SPECIES

Dataset of:

525 bird species

- 84.635 training images (~160 img per species),
- 2.625 test images(5 img per species) and
- 2.625 validation images(5 img per species).

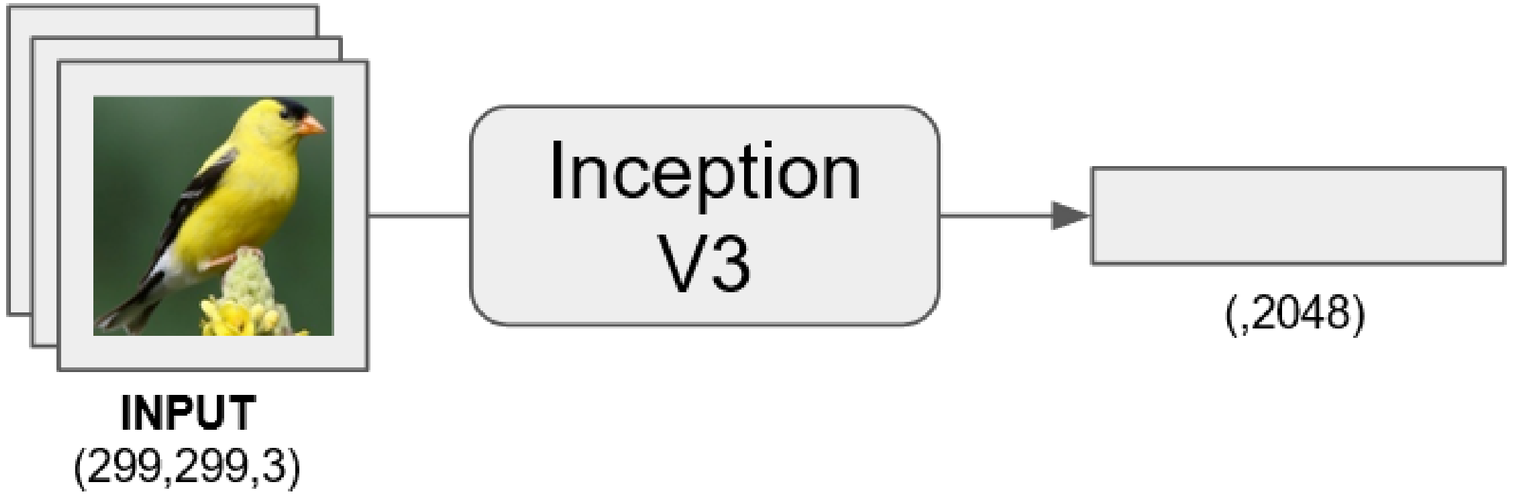
299x299x3

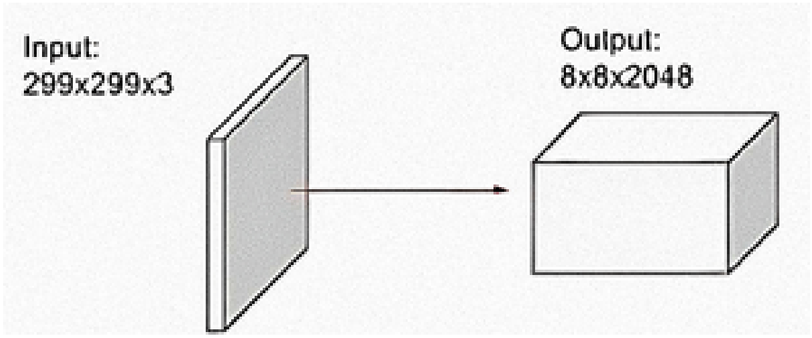
The Architecture

InceptionV3 pre-processing

No output Layer

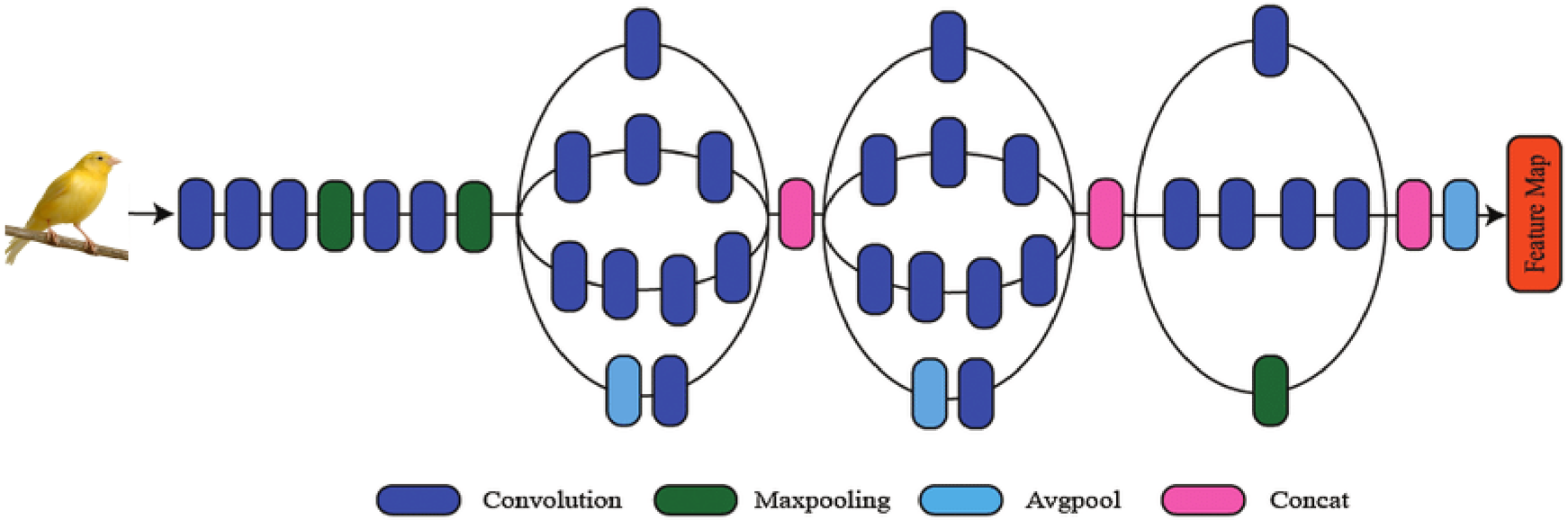
Only Convulation part





Inside the InceptionV3

Convolution
Applying filters to extract low-level features such as edges and textures.



Maxpooling
Reducing feature map sizes by retaining the maximum values within pooling windows, aiding in data dimensionality reduction.

Avgpool
Compresses feature maps by calculating the average values in pooling windows.

Concat
Merging features from different paths or layers to preserve multi-scale information.

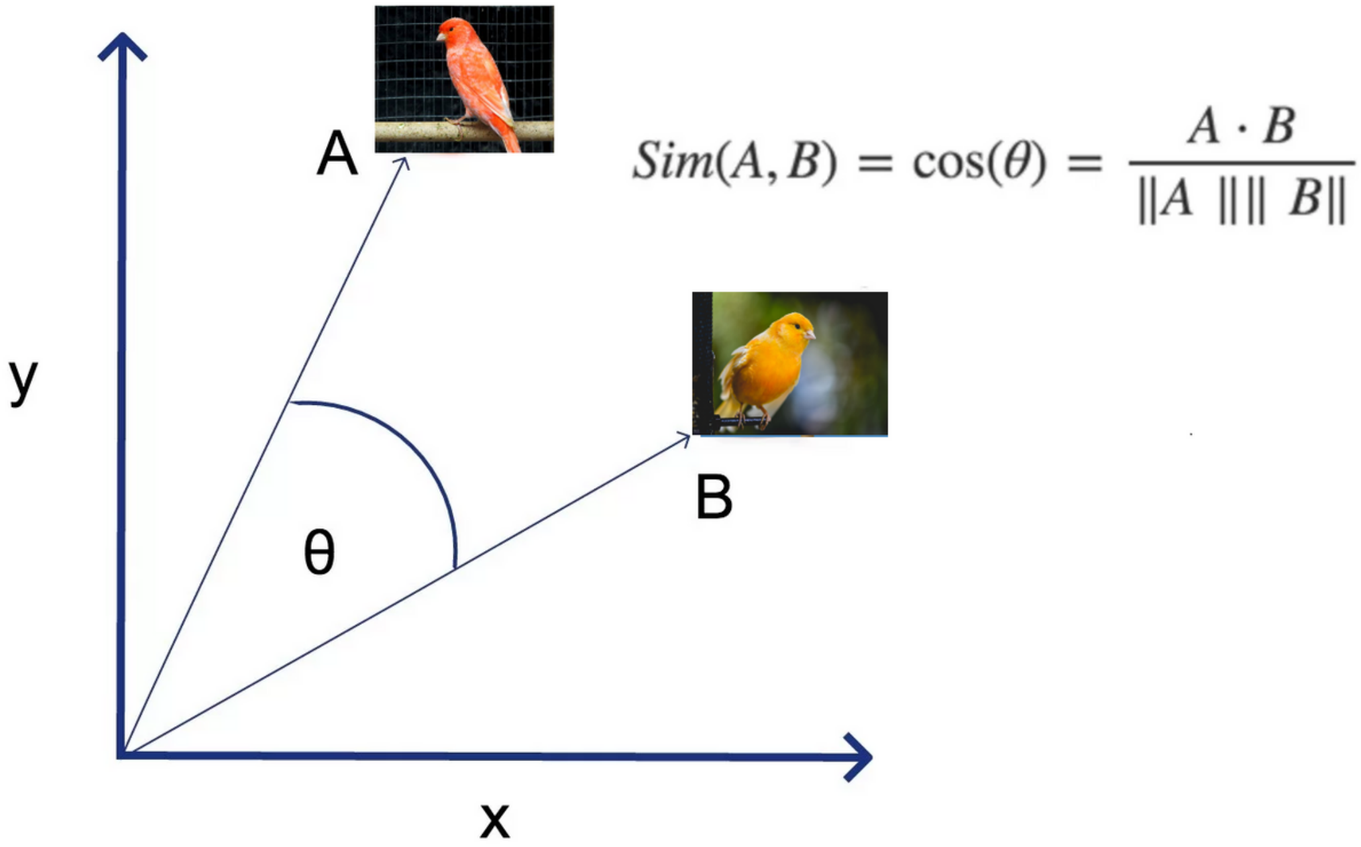
Feature Map
Comprising high-dimensional feature maps that encapsulate the image's extracted information.

Cosine Similarity

Cosine similarity measures how similar two image feature vectors are

These vectors represent the visual contents of the images, like colors and shapes

Score ranges from -1 to 1, with 1 meaning the images are very similar.

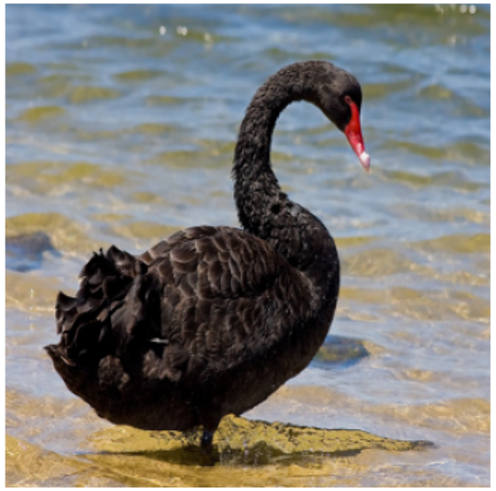


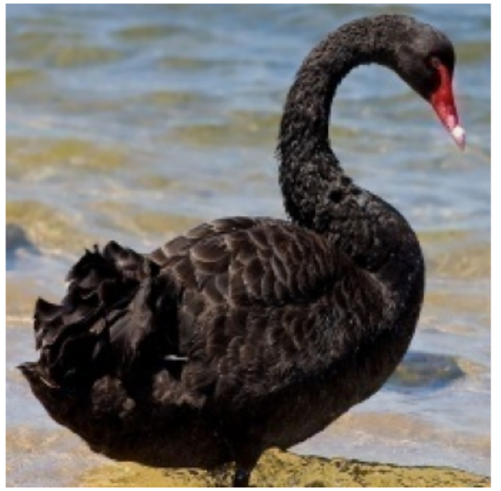
Identifying Similar Images

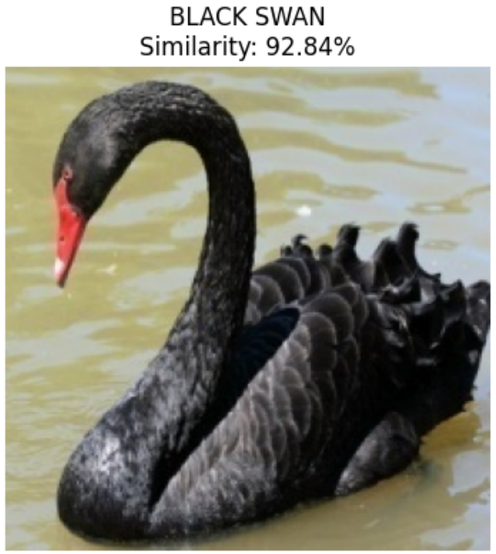
A query image is converted into a feature vector and compared to a dataset's vectors via cosine similarity

Images are then ranked by similarity, with top matches returned

The output is a list of the most similar images. Similarity is the in %





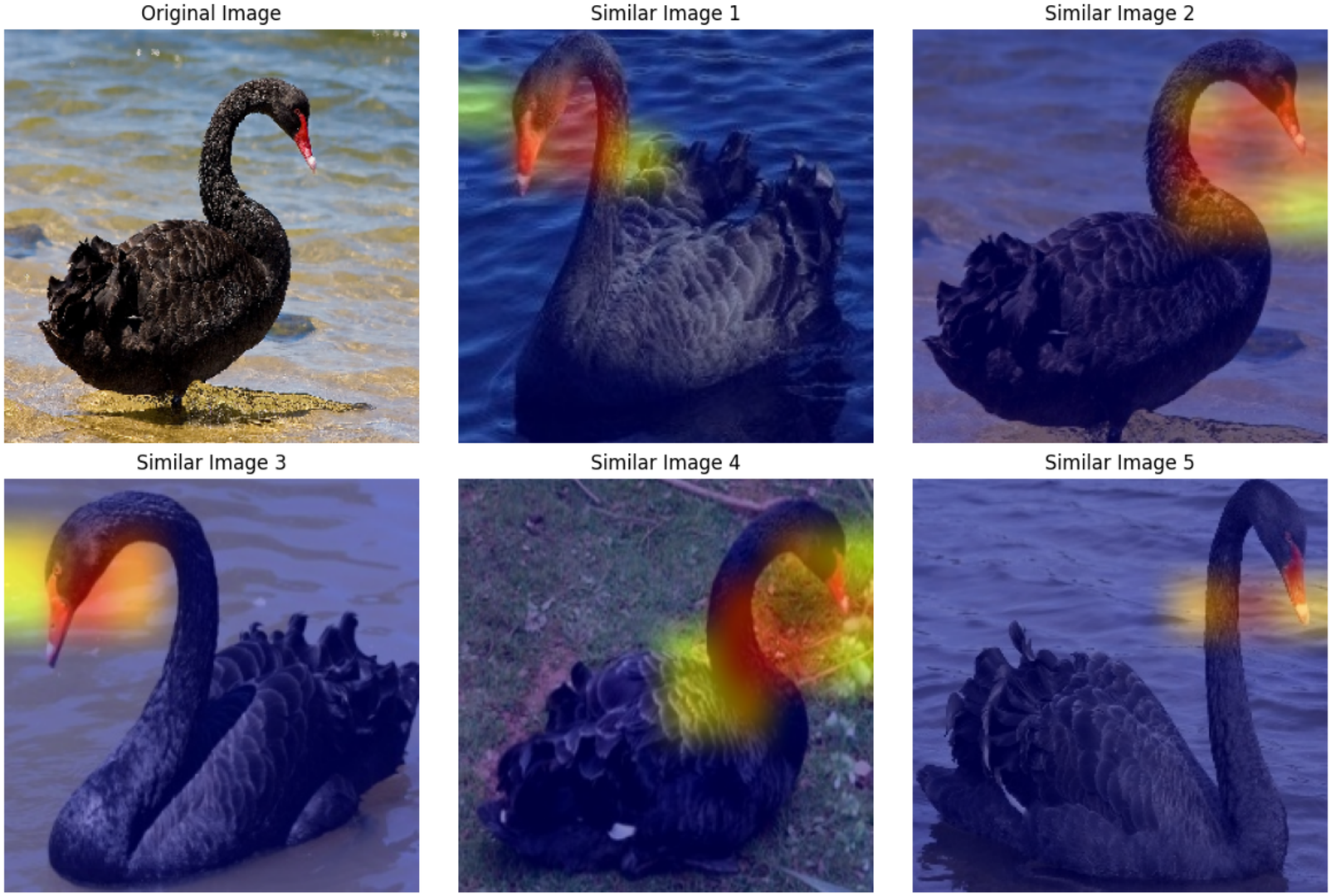


Visualizing Image Features with Grad-CAM

Show heatmaps overlaid on similar images to highlight the features that contributed most to the neural network's decision.

Each heatmap indicates areas of high importance in red to yellow gradients, with cooler colors indicating lesser importance.

The most intense areas (warmest colors) usually correspond to the distinctive features of the swan that the model focused on when assessing similarity.



Thank you