

Happywhale - Whale and Dolphin Identification

Matthew Harding

April 2024

1 Introduction

This project explores the problem of classifying images of dolphin and whales to identify individuals. This work is based on the *Happywhale - Whale and Dolphin Identification* Kaggle competition.

[1]

Data for this competition contains images of over 15,000 unique individual marine mammals from 30 different species collected from 28 different research organizations. Individuals have been manually identified and given an individual_id by marine researches.

Unlike a typical classification problem in which there is a fixed set of classes which examples of all classes within the training data, this problem required the ability to classify individuals not contained within the training dataset.

2 Training Data

Within the training data there are 51,033 labelled images. Looking at the histogram of species Figure 1, there is a clear imbalance in the number of iamges per species with the vast majority of labelled images being for bottlenose dolphins, beluga whales, humpback whales and blue whales. For some species the number of examples it so low it will be difficult to gain a high level of accuracy within classification.

Look at Figure 2, the mean count of images per individual by species we see another imbalance with some species on average containing a large number of images per indivudal whilst others only contain one or two images per individual.

Need to talk about performance if dataset was limited to just bottlenose dolphins

3 Image Preprocessing

The training images for the bottlenose dolphins came in a variety of sizes. Figure 3 shows an example unprocessed image. Large amounts of each image are just

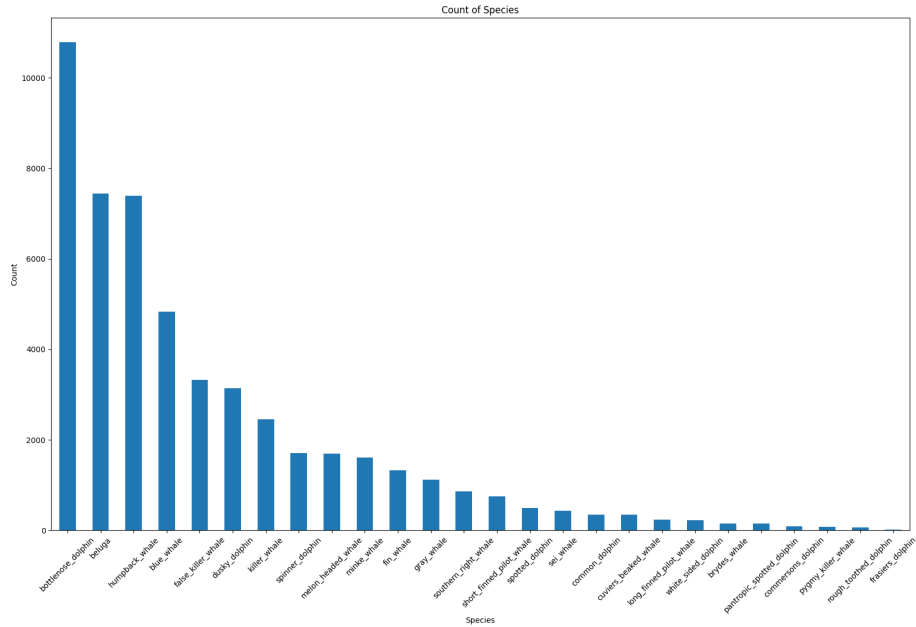


Figure 1: Number of training images per species

of water which is irrelevant to the classification task. We are only interested in the parts of the images that contain the animals.

A solution to this issue was proposed within the discussions of the Kaggle competition. YoloV5 was trained on a dataset of 1200 pictures of whale flukes and the corresponding location of points on the edge of the fluke for those pictures. This was then assumed to generalise to finding the bounding box for any part of a whale or dolphin within the water. However, it is worth noting that finding the bounding box of any part of the animal is considered an Out of Distribution problem.

From this technique a cropped dataset was created with all images having a dimension of 256 x 256.

<https://www.kaggle.com/datasets/martinpiotte/humpback-whale-identification-fluke-location>
<https://www.kaggle.com/code/awsaf49/happywhale-cropped-data-prepare-yolov5>

MAT - Need to talk about the transformation logic. Why do we convert to tensor and then normalize?

A train/test split of 80/20 was used (need to reference this decision).

What is random state, why 42?

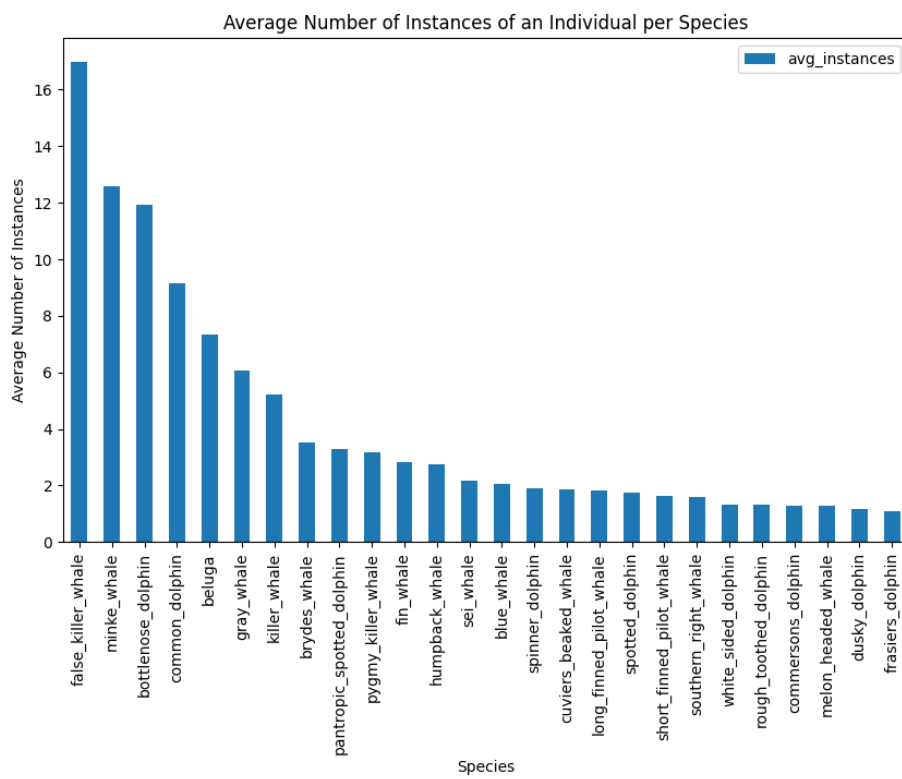


Figure 2: Mean count of images per individual by species



Figure 3: Example image of a bottlenose dolphin

4 Measuring Model Performance

Performance of the model was measured via the Mean Average Precision @ 5 (MAP@5). Instead of using just accuracy as a measure, MAP@5 was used to account of the complexity of the classification task. Accuracy only considers whether the top predicted class is correct or not. It does not take into account the ranking of predictions. MAP@5, on the other hand, considers the ranking of predictions within the top 5 predicted classes. It rewards models that rank the correct classes higher in the prediction list.

5 Network Architecture

As this was an image classification task, a network architecture based around Convolutional layers seemed logical.

The first layer was a 2D Convolutional layer. This layer takes the input image with 3 channels (RGB) and applies 32 different 3x3 convolutional filters to extract low-level features like edges, corners and simple patterns. The stride of 1 ensures that the filters are applied to every pixel, and the padding of 1 preserves the spatial dimension.

Next, a ReLU activation function was run to add non-linearity to the model (????)

Finally a pooling layer was added.

MATT - Why do we then do two additional convolutional layers? Look at forward method, we reuse relu and maxpool

MATT - Is fully connected 2 considered a SOFTMAX?

6 Model Training

Despite the massive reduction in file size that came from using the cropped dataset instead of the original dataset, training times on the local machine was still prohibitively long. Therefore, training was done using AWS Sagemaker.

7 Conclusions

References

- [1] A. N. Author. Hello world. *Journal Name*, 5(2):123–456, 2022.