DIGITAL FUTURES DATA SCIENCE CAPSTONE

WHALE AND DOLPHIN IMAGE CLASSIFICATION

Amisha Bhojwani

Agenda

- Background
- EDA and processing
- Modelling
- Limitations and future efforts

Background

Sections

- How do we study whale and dolphin ecology?
- How can you identify a species from a photo?
- Project objective

Studying whale and dolphin ecology

OBSERVE

Photograph sightings of individuals

GATHER DATA

Note location of sighting, species and identify individual

RESEARCH

Use spatial and individual information to track movement and social patterns

Dorsal Fins

Bottlenose dolphin

Dolphins

Curved dorsal fins

Humpback whale

Whales

Many different shapes,
sometimes not present

Learning objectives

- To understand image classification using convolutional neural networks
- To incorporate a Bayesian perspective into modelling

Project objective

- To automate classifying whale and dolphin images, to reduce the time taken to gather data for research purposes

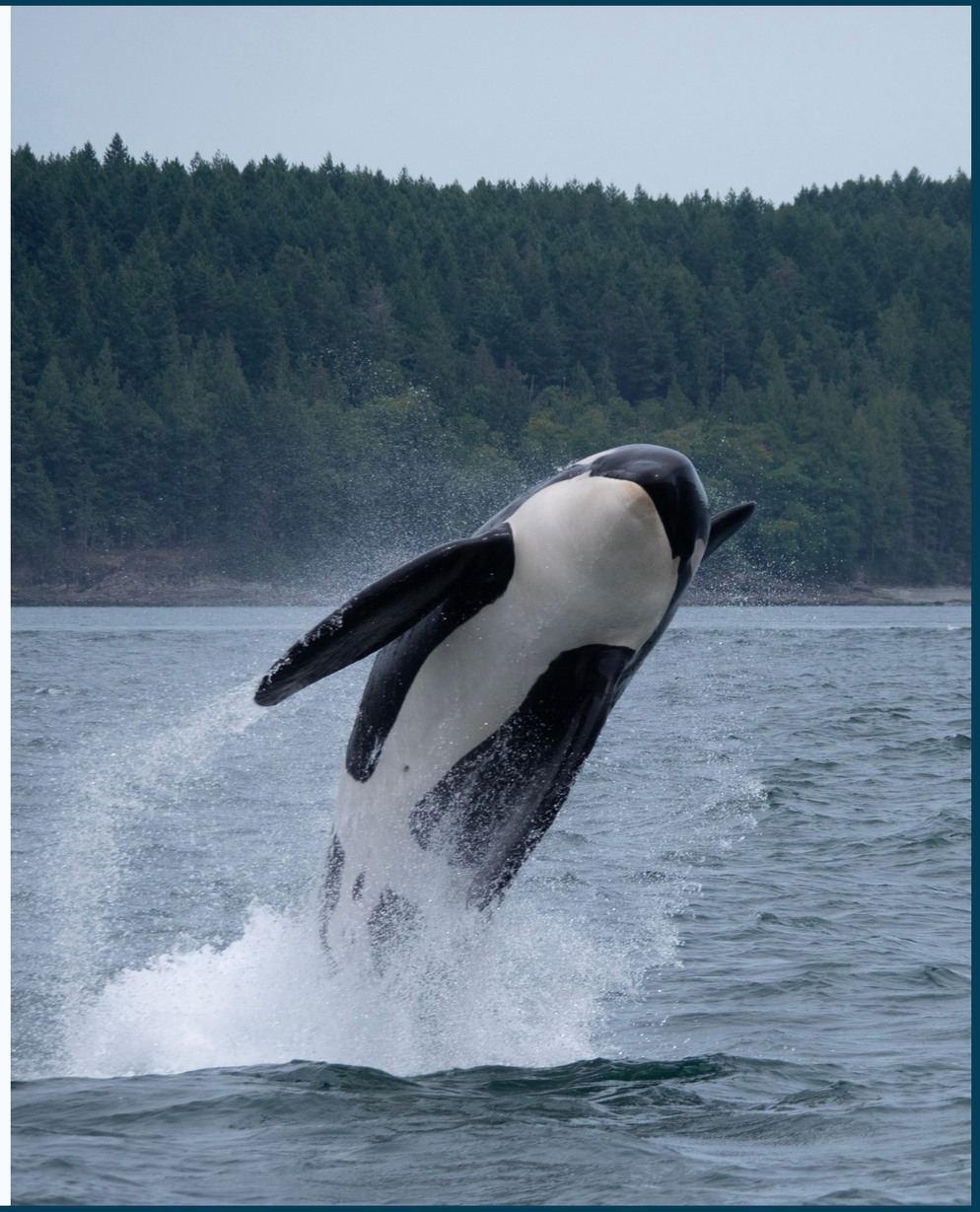
EDA and processing

Sections

- Cleaning
- Original data description
- Sampled data description
- Example images

Data Cleaning

- Spelling mistakes in species names
  - 'bottlenose_dolpin'
  - 'kiler_whale'
- The case of pilot whales
  - 'globis', 'pilot_whale' = genus
  - 'short_finned_pilot_whale', 'long_finned_pilot_whale' = species
- Taxonomic mapping



Taxonomic mapping

- 26 species
- 5 families

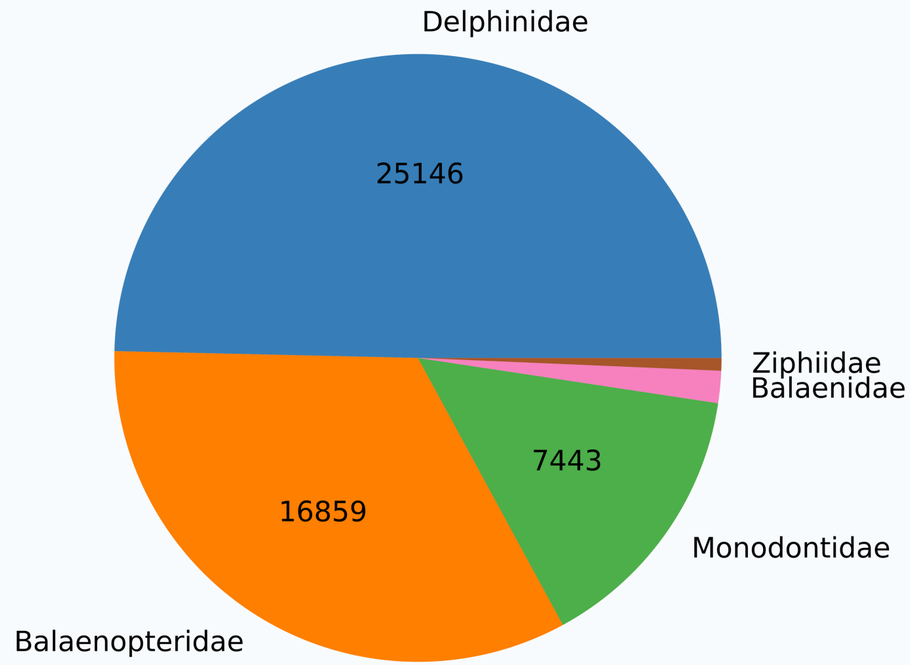
Species	Family
southern_right_whale	Balaenidae
humpback_whale	Balaenopteridae
blue_whale	Balaenopteridae
brydes_whale	Balaenopteridae
minke_whale	Balaenopteridae
fin_whale	Balaenopteridae
gray_whale	Balaenopteridae
sei_whale	Balaenopteridae
beluga	Monodontidae
cuviers_beaked_whale	Ziphiidae

Species	Family
pygmy_killer_whale	Delphinidae
commersons_dolphin	Delphinidae
pantropic_spotted_dolphin	Delphinidae
white_sided_dolphin	Delphinidae
long_finned_pilot_whale	Delphinidae
common_dolphin	Delphinidae
short_finned_pilot_whale	Delphinidae
bottlenose_dolphin	Delphinidae
rough_toothed_dolphin	Delphinidae
killer_whale	Delphinidae
melon_headed_whale	Delphinidae
spinner_dolphin	Delphinidae
dusky_dolphin	Delphinidae
false_killer_whale	Delphinidae
spotted_dolphin	Delphinidae
frasiers_dolphin	Delphinidae

Original dataset

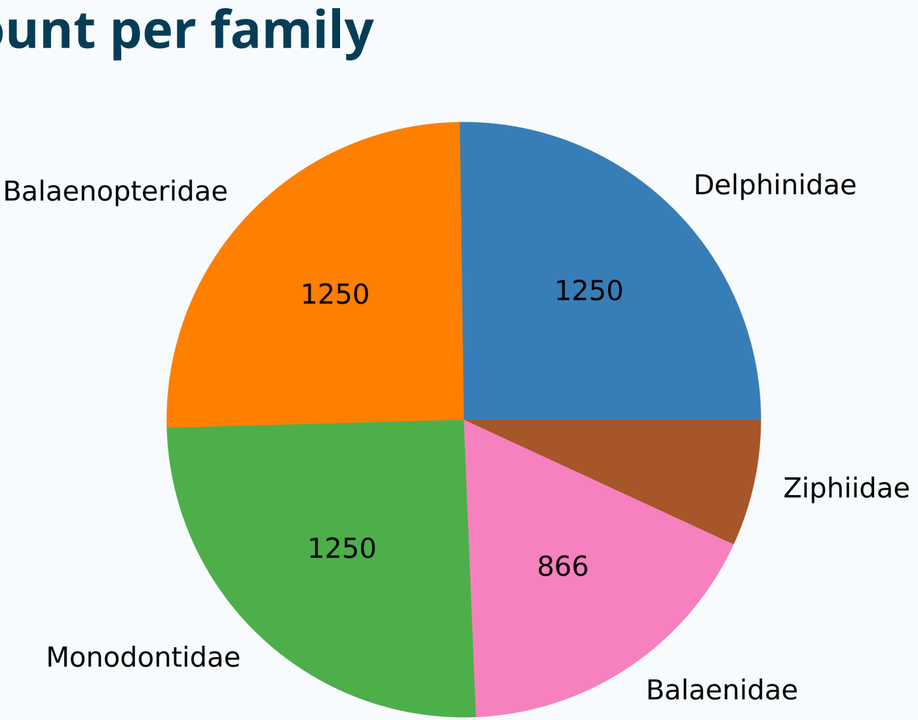
- 50,655 observations
- Image sizes from 200x200 to 5000x3500 pixels
- Size > 50GB

Count per family



Sampled dataset

- Family class balance
- 4957 observations
- More on sizes later!

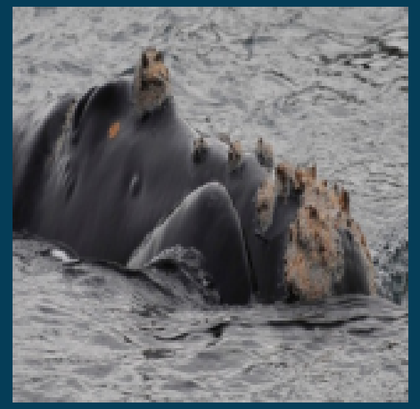


Balenidae

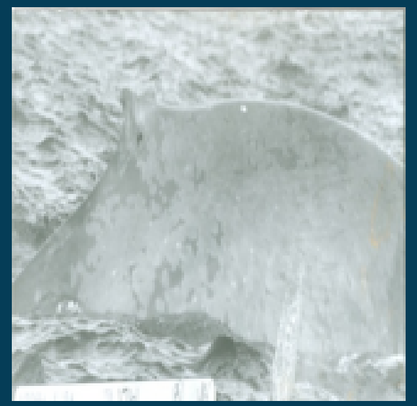








Balaenopteridae



Delphinidae









Monodontidae









Ziphiidae









Modelling

Sections

- Feature engineering
- Convolutional Neural Networks (CNN)

FEATURE ENGINEERING

IMAGE RESIZING	DATA AUGMENTATION	IMAGE ARRAYS	NORMALISING DATA
All images should be the same size, resized to 356x356	Images are randomly flipped and recoloured to introduce variability in the data and reduce overfitting	Images are read as NumPy 4D arrays (1, 356, 356, 3) for modelling	All numbers in the array are between 0 and 255 (RGB values). These are normalised to be between 0 and 1.

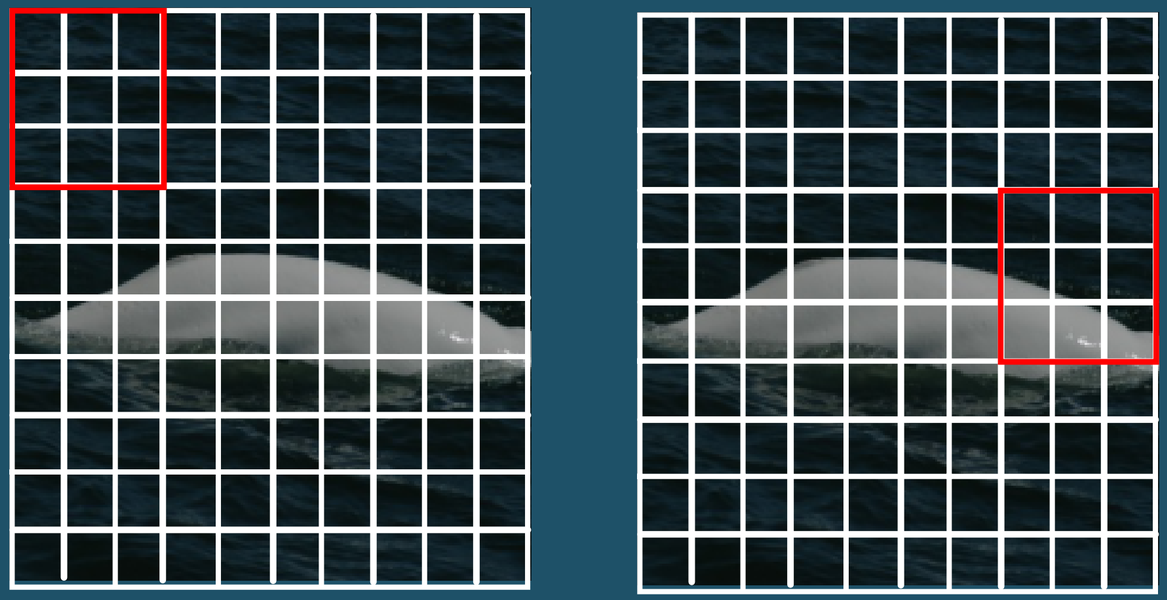
MEMORY ISSUES

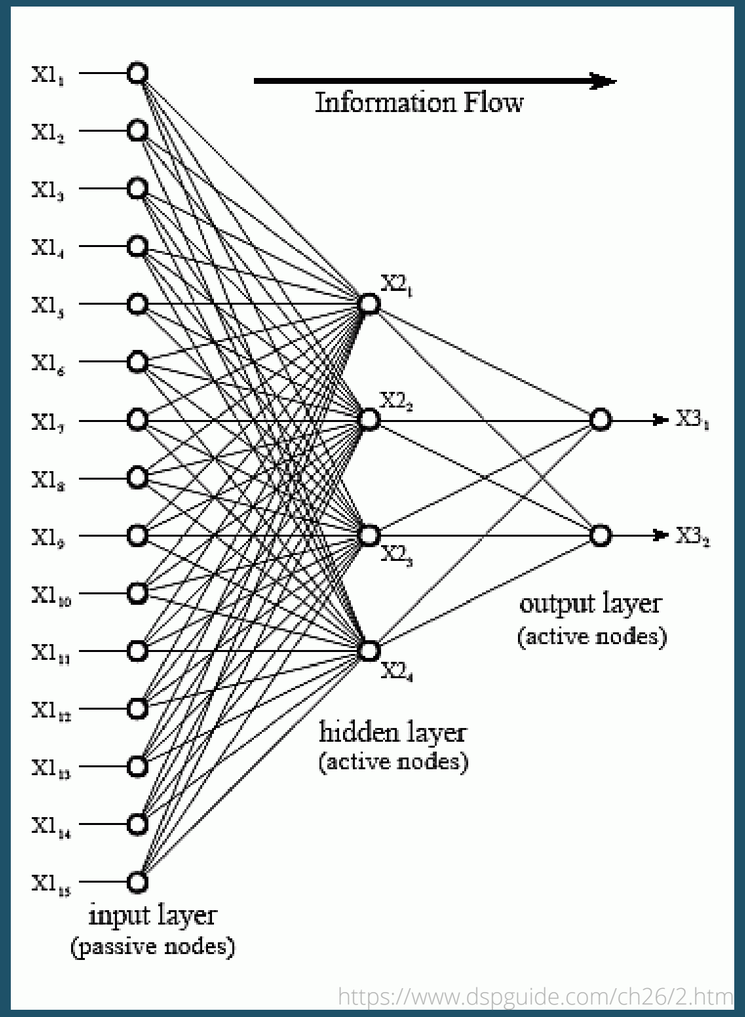
RAM overload

IMAGE DATA GENERATOR

- Real time feature engineering
- Validation split
- Shuffle data and feed to a model in batches, lowering memory usage
- Different versions of images each training epoch to reduce overfitting

CONVOLUTIONAL NEURAL NETWORKS (CNN)

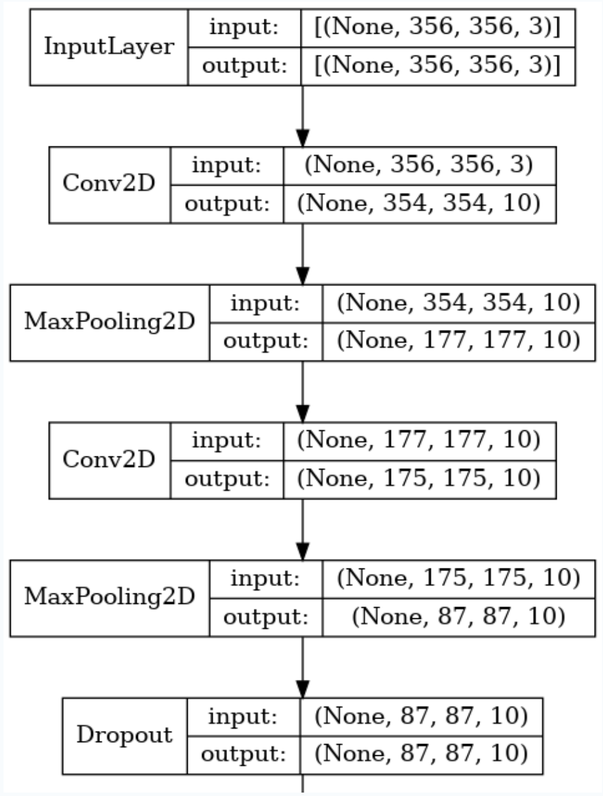


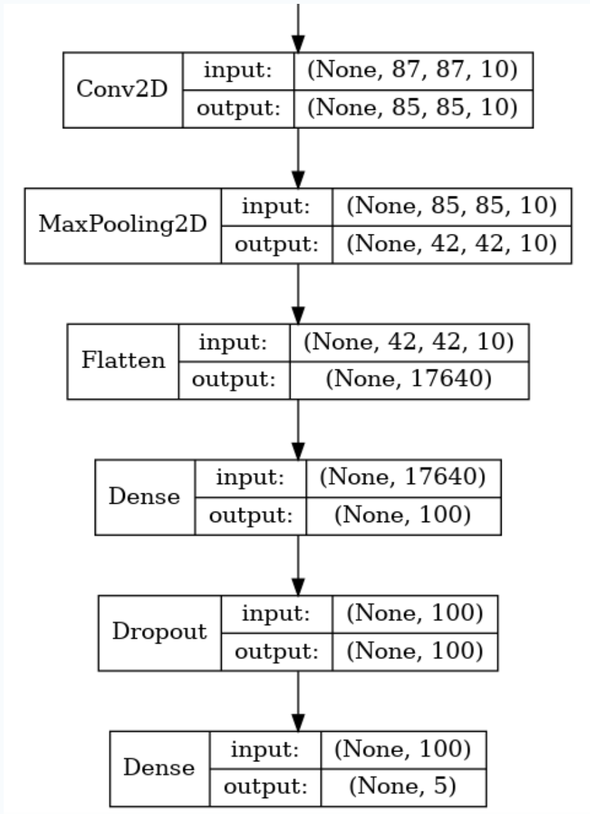


https://www.dspguide.com/ch26/2.htm

CNN

Architecture

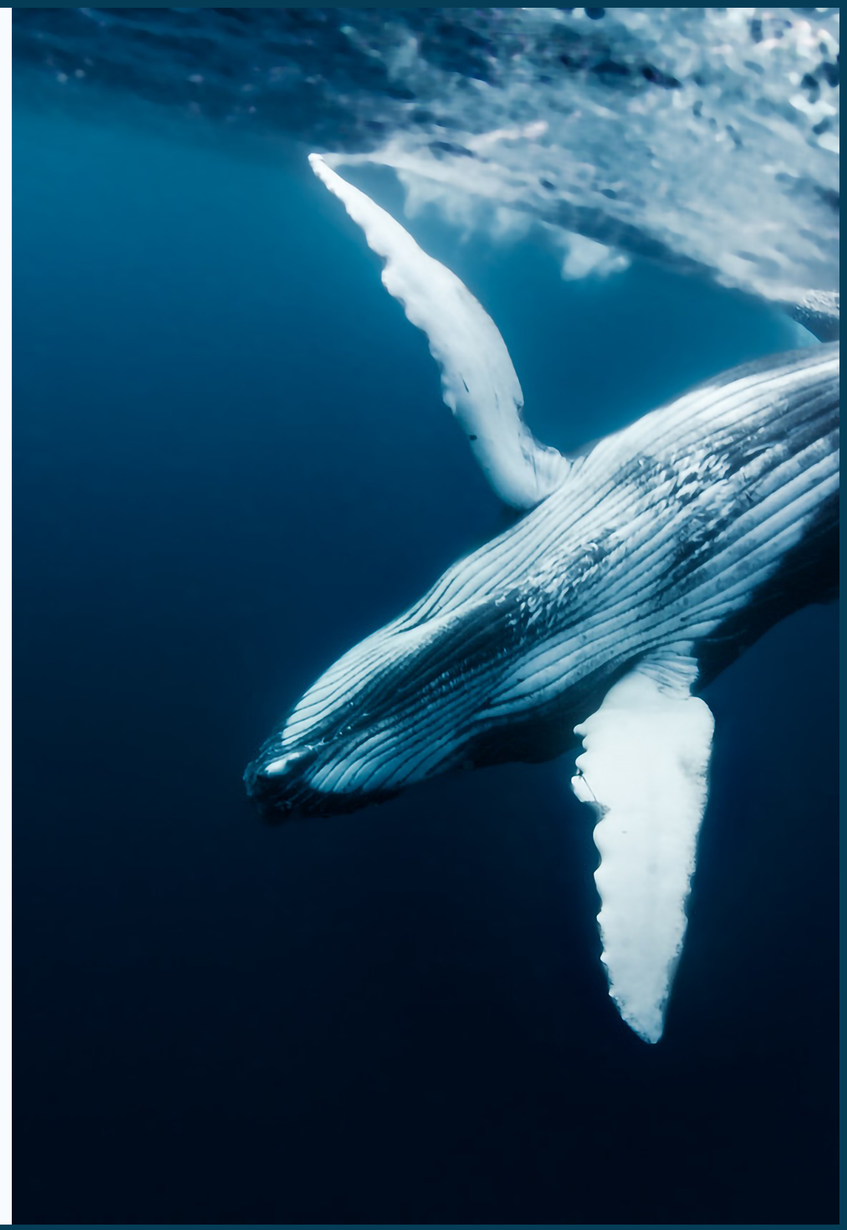




CNN

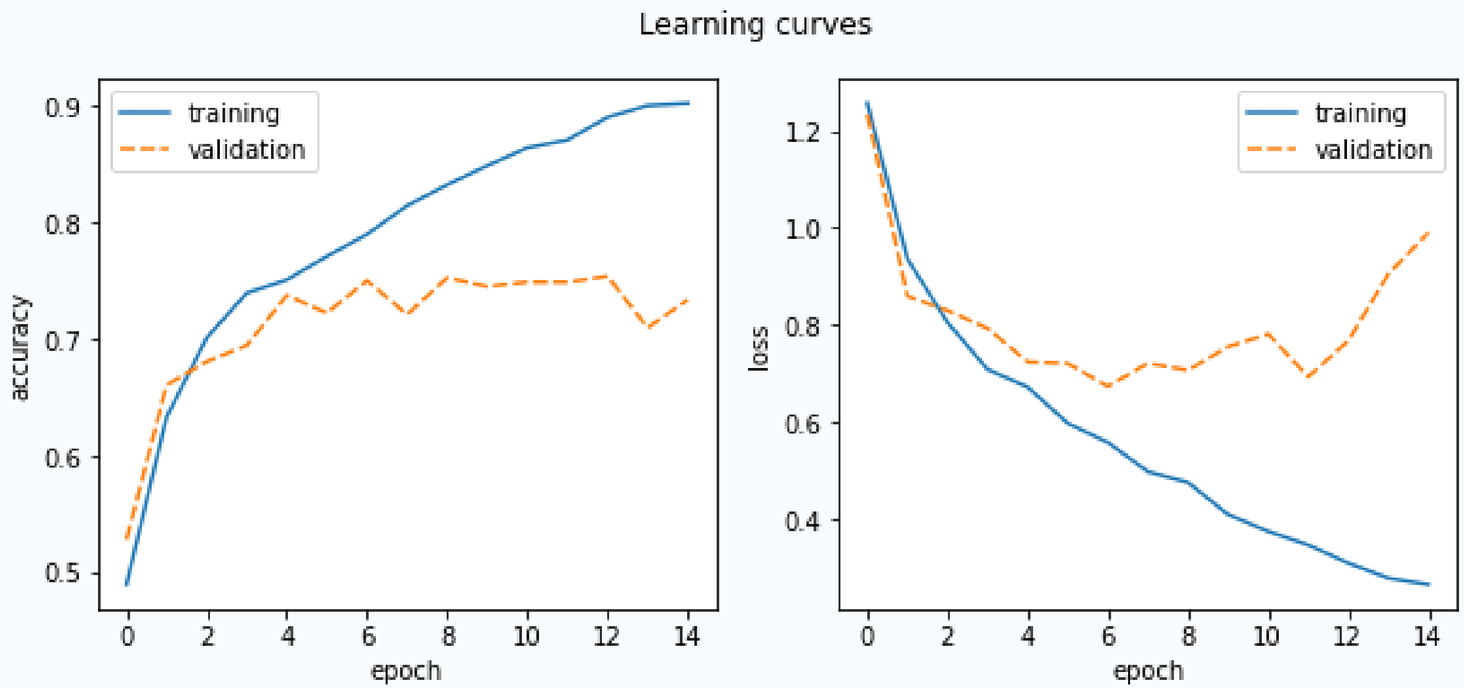
Hyperparameters

- Epochs 15
- Activation functions
  - Hidden: Rectified Linear Unit (ReLU)
  - Output: Softmax
- CNN layers
  - Batch size: 5
  - Kernel size: 3x3
  - Filters: 10
- Loss function: Sparse categorical cross-entropy
- Optimiser: Adam



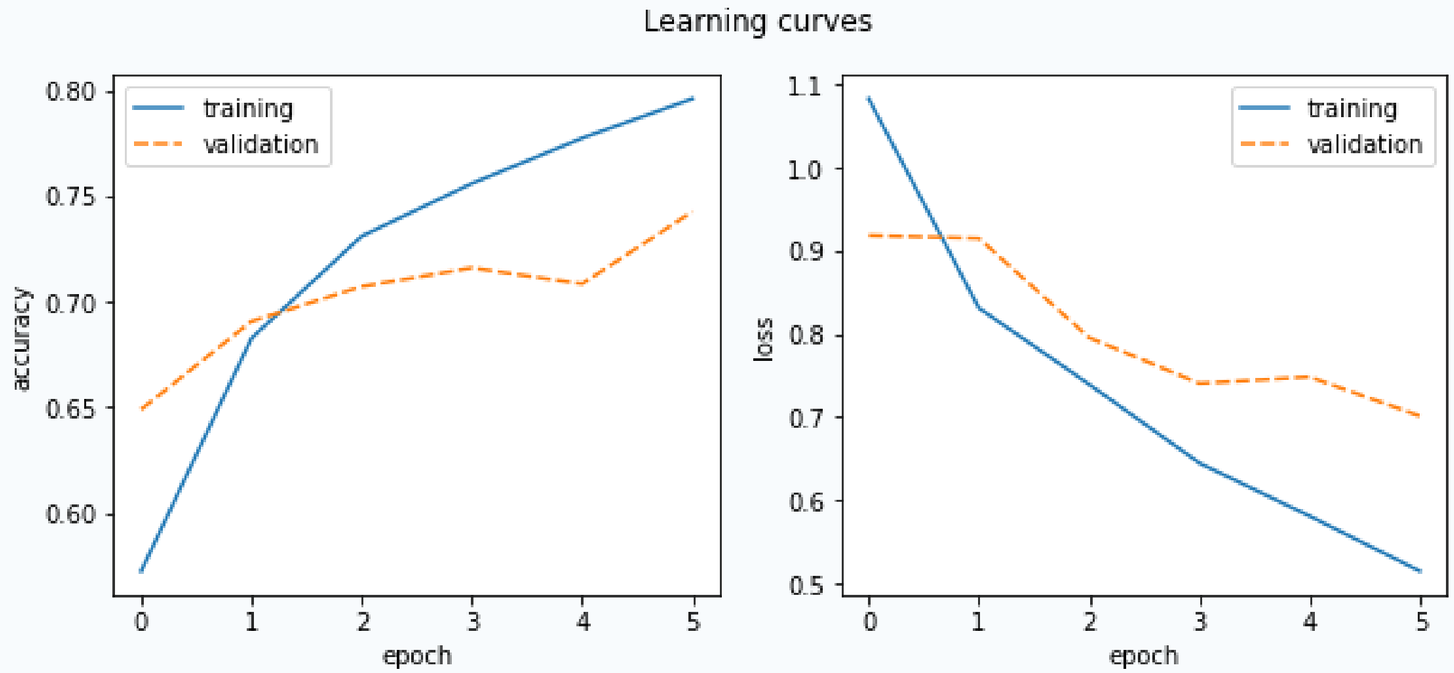
CNN

Results: training and validation sets



CNN

Results: training and validation sets



CNN

Results: testing set

Family	Precision	Recall	F1-score	Support
Balenidae	0.83	0.74	0.78	174
Balaenopteridae	0.63	0.63	0.63	250
Delphinidae	0.71	0.81	0.76	250
Monodontidae	0.83	0.87	0.85	250
Ziphiidae	0.59	0.32	0.42	68

Final testing set accuracy: 0.73

Actual: Balaenopteridae
Predicted: Delphinidae

Actual: Balaenopteridae
Predicted: Monodontidae

Actual: Balaenopteridae
Predicted: Balaenopteridae

Actual: Balaenopteridae
Predicted: Balaenopteridae

Actual: Delphinidae
Predicted: Delphinidae

Actual: Delphinidae
Predicted: Delphinidae

Actual: Balaenopteridae
Predicted: Balaenopteridae

Actual: Balaenopteridae
Predicted: Balaenopteridae

Actual: Balaenidae
Predicted: Balaenidae

Actual: Monodontidae
Predicted: Monodontidae

Actual: Delphinidae
Predicted: Delphinidae

Actual: Balaenopteridae
Predicted: Balaenopteridae

Actual: Delphinidae
Predicted: Monodontidae

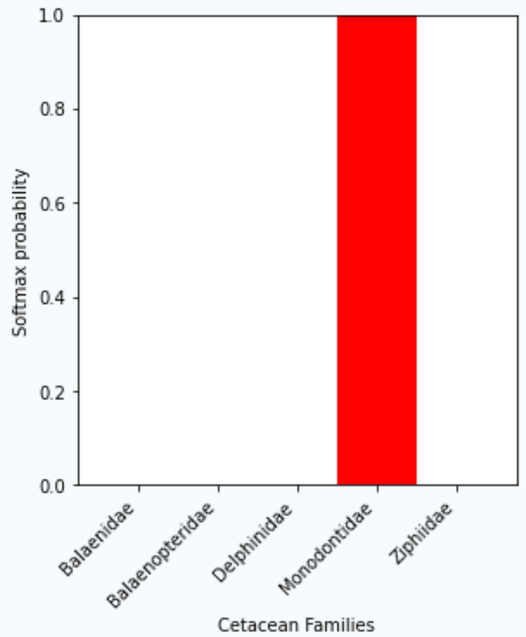
Actual: Delphinidae
Predicted: Delphinidae

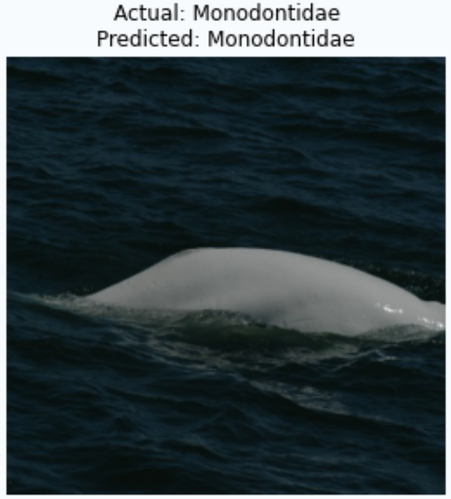
Actual: Monodontidae
Predicted: Monodontidae

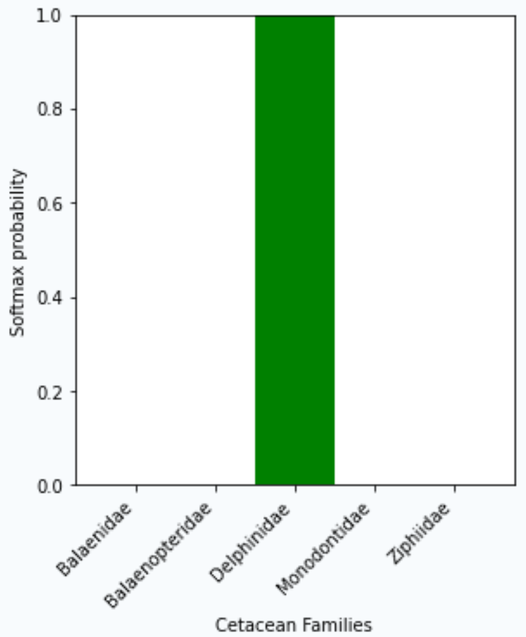
Actual: Balaenopteridae
Predicted: Balaenopteridae

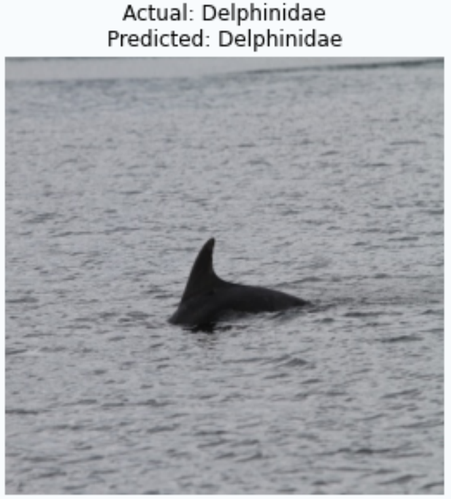
Actual: Balaenidae
Predicted: Balaenidae

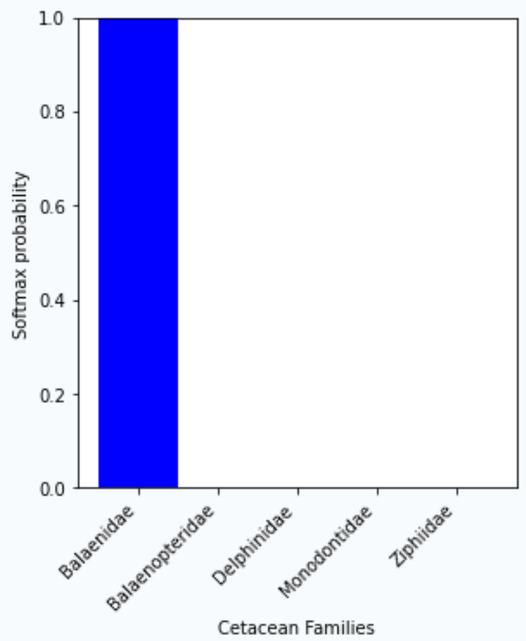
Actual: Delphinidae
Predicted: Delphinidae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families

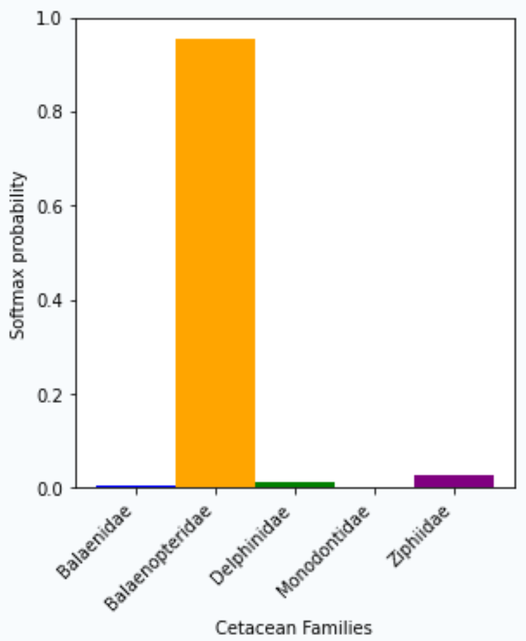

Actual: Monodontidae
Predicted: Monodontidae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families

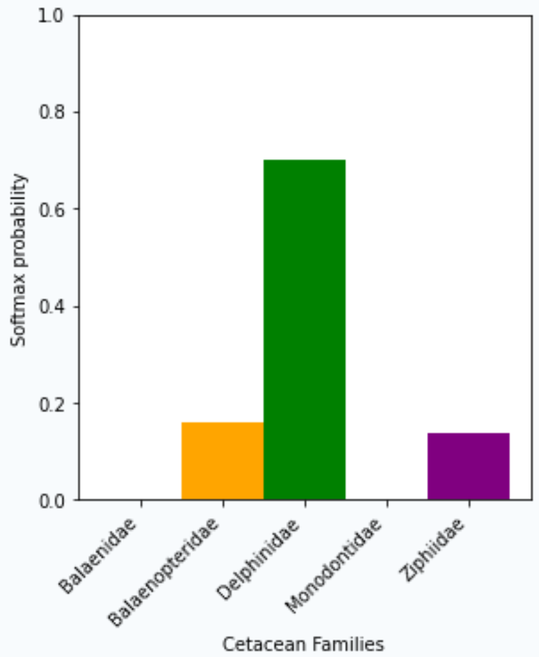

Actual: Delphinidae
Predicted: Delphinidae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families

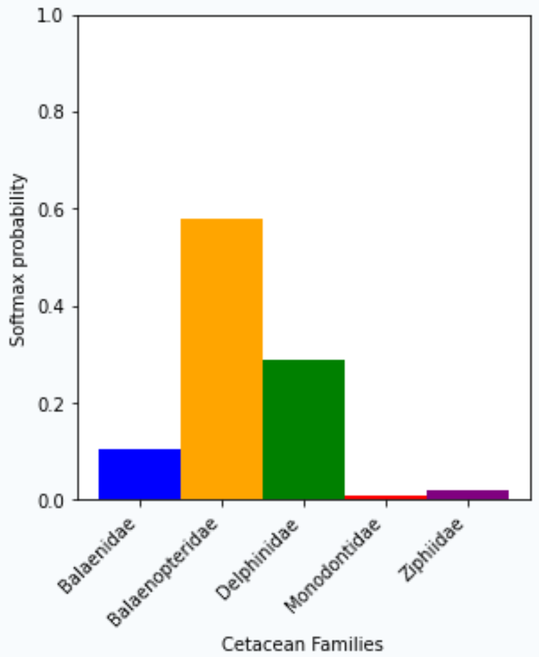

Actual: Balaenidae
Predicted: Balaenidae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families

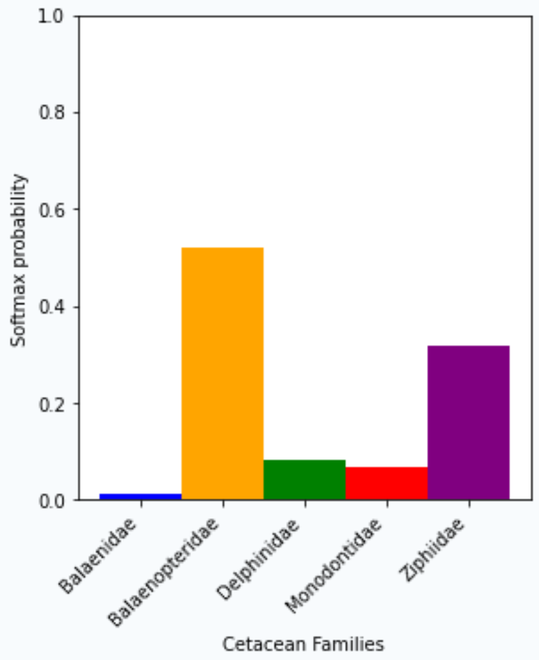

Actual: Ziphiidae
Predicted: Balaenopteridae

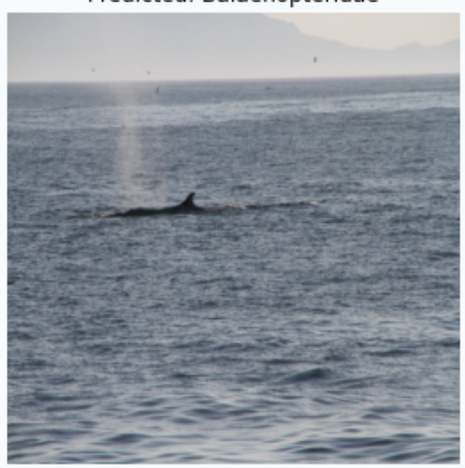

1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families

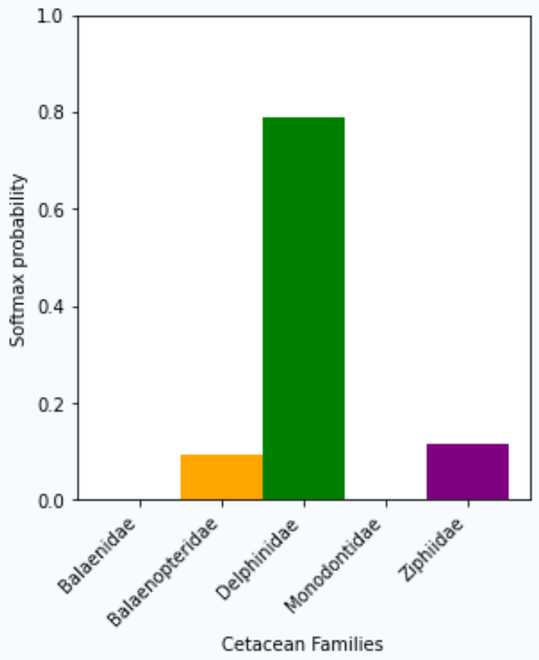

Actual: Delphinidae
Predicted: Delphinidae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families


Actual: Balaenidae
Predicted: Balaenopteridae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families


Actual: Balaenopteridae
Predicted: Balaenopteridae


1.0
0.8
0.6
0.4
0.2
0.0
Softmax probability
Balaenidae
Balaenopteridae
Delphinidae
Monodontidae
Ziphiidae
Cetacean Families


Actual: Ziphiidae
Predicted: Delphinidae

Final thoughts

Limitations

- Computational power and time
  - Limited modelling techniques
  - Limited regularisation
    - Image augmentation
    - Hyperparameter tuning
    - Gridsearch
  - Small sample of the dataset
  - K-fold cross validation

Future efforts

- Image resizing and according to original aspect ratio
- Reconsider target
  - Species-level
  - Individual-level
- Pipeline for modelling

THANK YOU

I WILL NOW TAKE QUESTIONS :)