Whale and Dolphin Identification Using Machine Learning

Identify whales and dolphins by unique characteristics

Sitao Chen, Justin Wong, Martin Lim
April 13, 2022

Berkeley
UNIVERSITY OF CALIFORNIA

Introduction

"Facial Recognition" for whales and dolphins

- Current approach
  - Manual matching by human eye
  - Tails, dorsal fins, heads, etc…
- Challenges
  - Time intensive and tedious
  - Human error
  - Limited scalability

Berkeley
UNIVERSITY OF CALIFORNIA

Problem Statement

- **Input:** Digital images of distinguishing features of a dolphin or whale (eg. dorsal fin, body markings, etc.)
- **Main:** Identify the **species** of the dolphin or whale.
- **Sub:** Identify the **ID number** of the dolphin or whale.

Objective

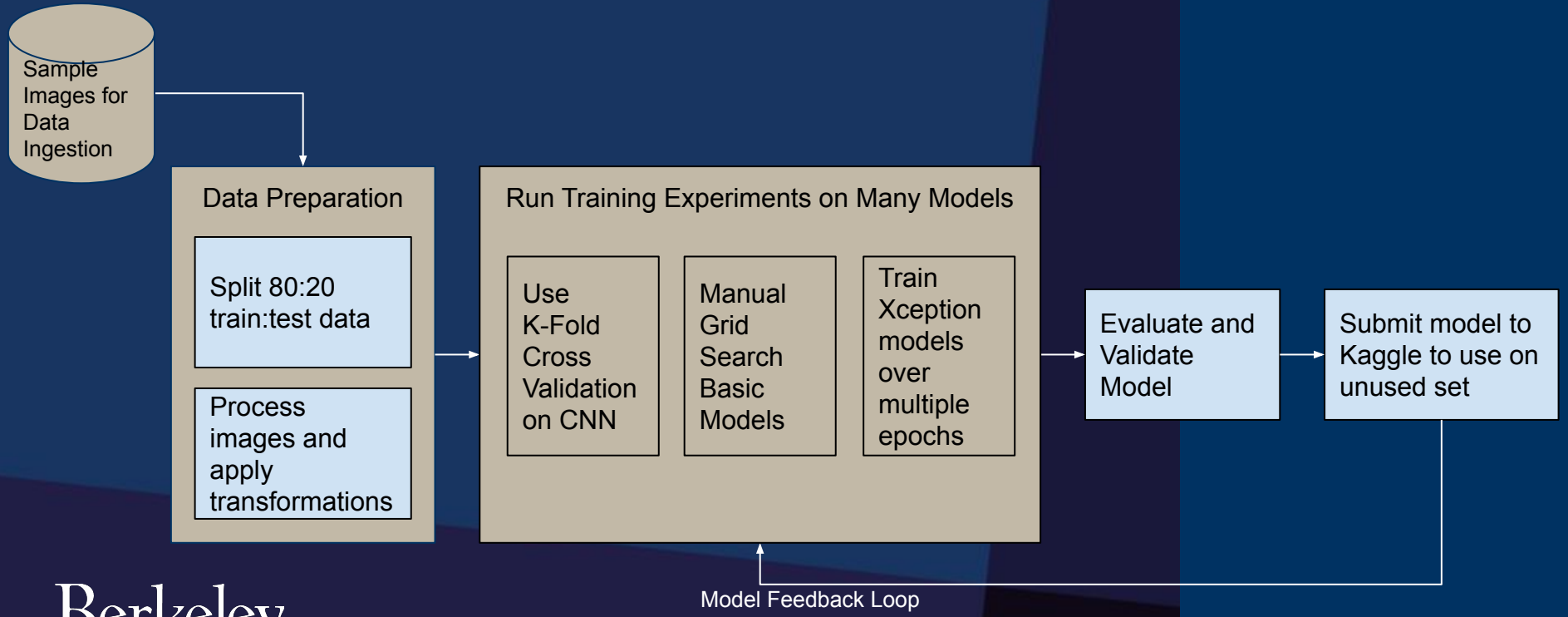
- To **accurately identify dolphin and whale from images of their unique**, distinguishing features (eg. fins, tails, body markings, etc.) using different machine learning techniques

Berkeley
UNIVERSITY OF CALIFORNIA

Approach/ Methodology

1. A subset of training images was created for data ingestion for model development
2. Transformations were applied on the images during the training process in order to standardize the input
3. A number of classifiers and clustering methods were used to predict both the species and the ID number of the individual images from the training sample
4. The different models were evaluated by accuracy and F1, the best model will be used to train with the entire training set

Block Diagram

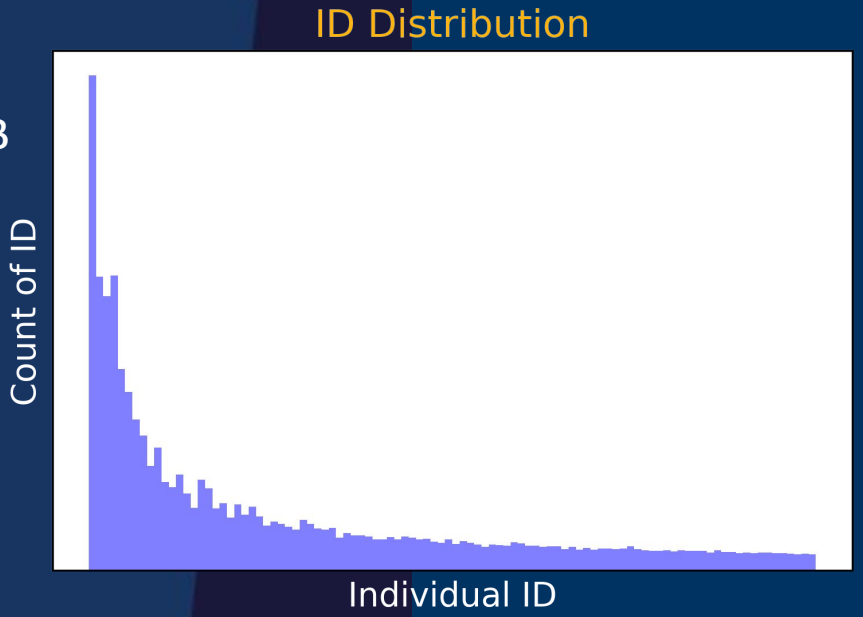


Berkeley
UNIVERSITY OF CALIFORNIA

The Data and Images

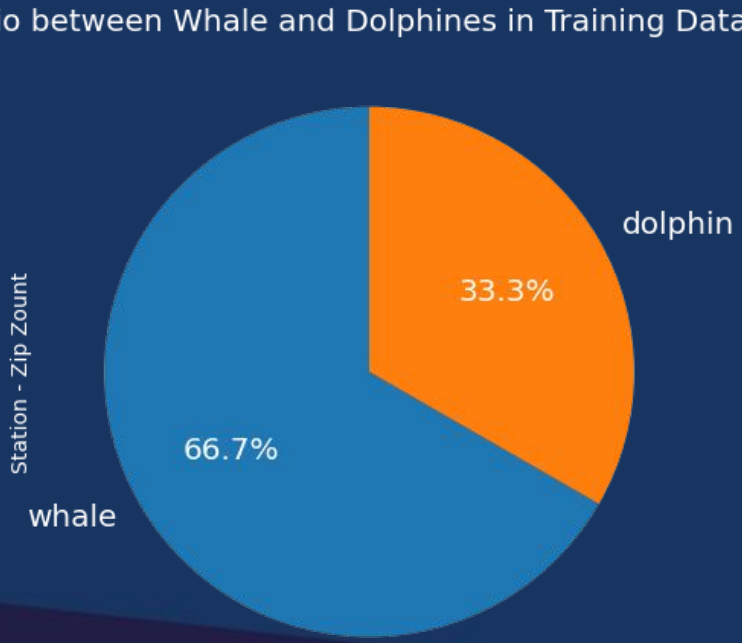
There are two datasets, training and test, containing over 78,000 images of whales and dolphins in JPEG format with total size of 62.06 GB

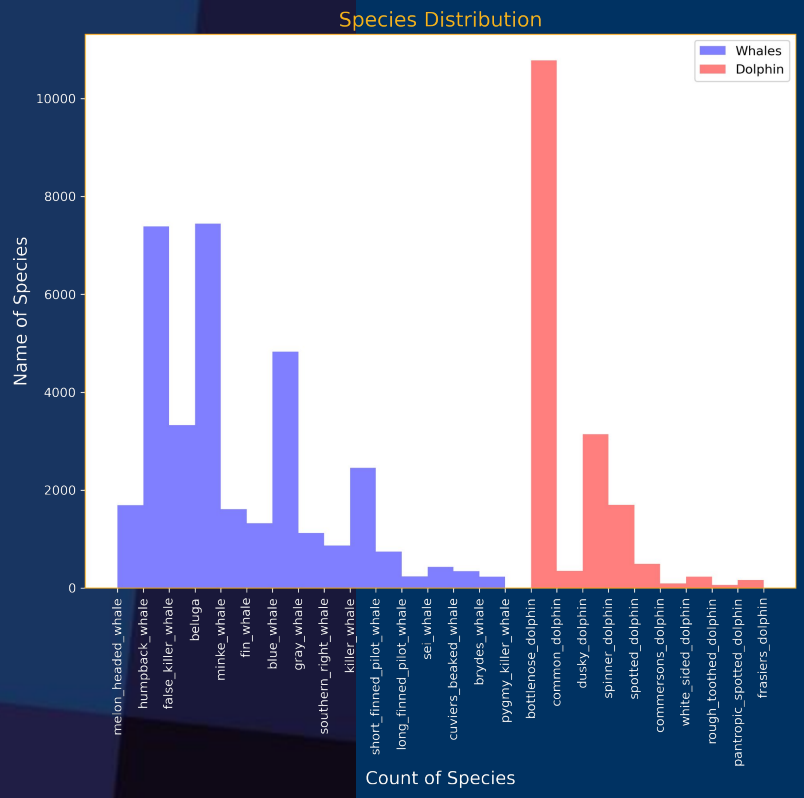
- Training set
  - 51033 images, file size between 19 kB ~ 2600 kB
  - 15587 unique individual
  - 16 species of Whales, 35155 images
  - 10 species of Dolphin, 15878 images
- Test set (unlabeled)
  - 27956 images



Berkeley
UNIVERSITY OF CALIFORNIA

The Data and Images Visualized





Berkeley
UNIVERSITY OF CALIFORNIA

What is considered success?

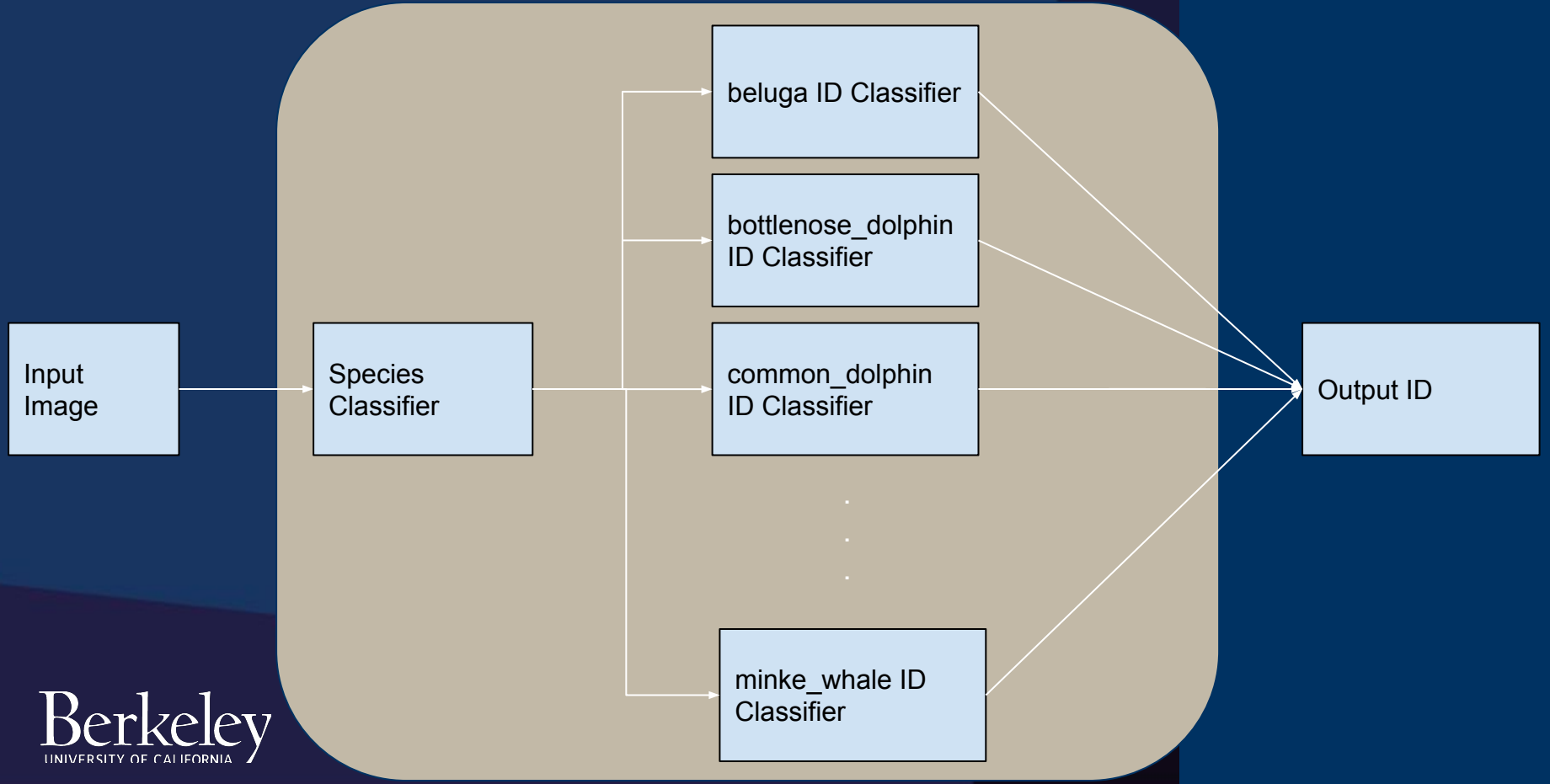
- Accuracy in predicting species > 50%
- Accuracy in predicting Individual ID > 6.42e-10
  - 6.42e-10 is 1 / 15587

Experiments

There are four experiments:

1. Basic Classifiers
   - k-Nearest Neighbors, K-Mean Clustering
   - Bernoulli Naive-Bayes, Gaussian Naive-Bayes
   - Logistic Regression
   - Decision Tree, Random Forest, Two-layer Random Forest Classifier
   - Neural Network
2. Basic Convolutional Neural Net
3. Xception (complex deep learning model with LSTM layers)
4. Hybrid model (if time is not the constraint)
   - 2 layered model, where first predicts species, and second predicts individual id

Hybrid Model - Two-Layer Model (POC)



Berkeley
UNIVERSITY OF CALIFORNIA

Results

Species Prediction

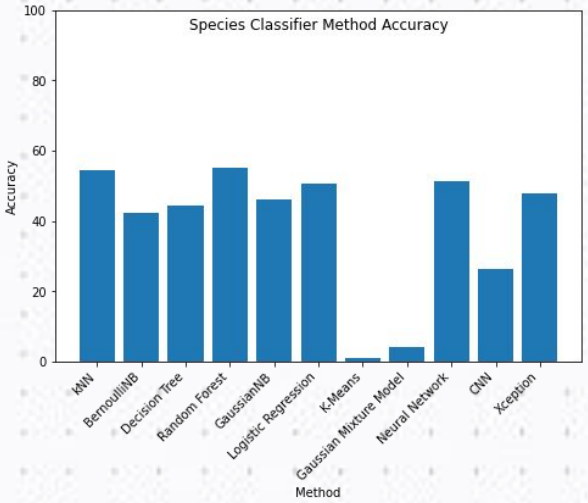
Method	Best Accuracy (%)	Best Parameter for Accuracy	Best F1 Score	Best Parameter for F1 Score
K-Nearest Neighbors	54.593	K = 10	0.5104	K = 10
Bernoulli Naive-Bayes	42.296	Alpha = 10	0.3813	Alpha = 2
Decision Tree	44.296	Max depth = 5	0.4235	Max depth = 10
Random Forest	55.111	Max depth = 50	0.5090	Max depth = 50
Gaussian Naive-Bayes	46.0	Var smoothing = 0.01	0.458	Var smoothing = 0.01
Logistic Regression	1.407	L2 = 0.1	0.0068	L2 = 0.5
K-Means	0.889	N/A	0.0064	N/A
Gaussian Mixture Model	4.0	N/A	0.0417	N/A
Neural Network	51.185	N/A	0.4565	N/A
Convolutional Neural Network	27.38	N/A	N/A	N/A
Hybrid Xception	**96.30**	**N/A**	**N/A**	**N/A**

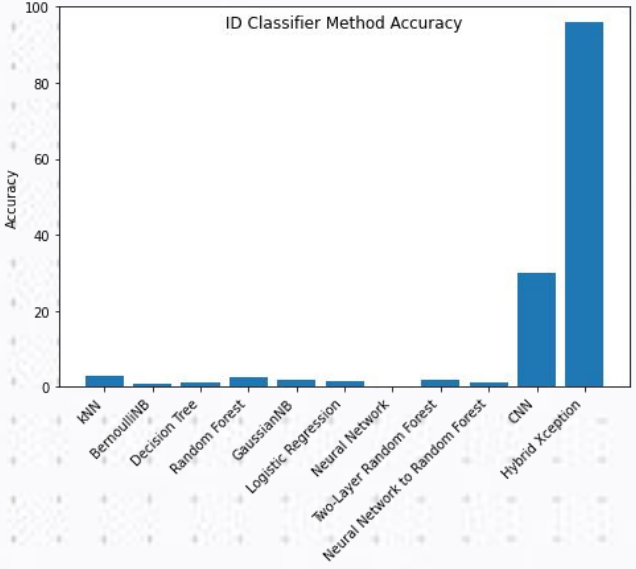
ID Prediction

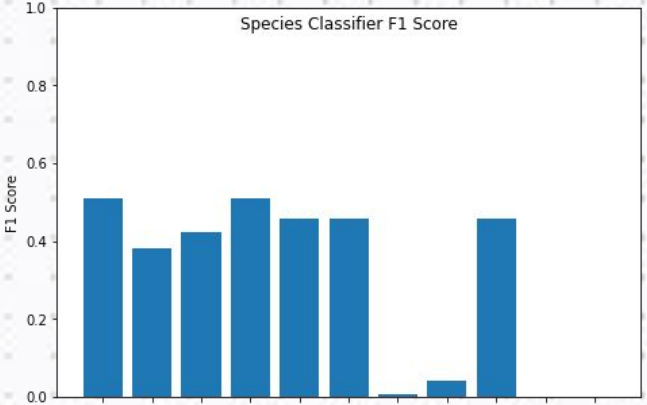
Method	Best Accuracy (%)	Best Parameter for Accuracy	Best F1 Score	Best Parameter for F1 Score
K-Nearest Neighbors	2.889	K = 1	0.0224	K = 1
Bernoulli Naive-Bayes	1.037	Alpha = 1e-10	0.0057	Alpha = 1e-10
Decision Tree	1.185	No max depth	0.0091	No max depth
Random Forest	2.667	Max depth = 1000	0.0186	Max depth = 1000
Gaussian Naive-Bayes	1.926	Var smoothing = 0.01	0.0120	Var smoothing = 0.01
Logistic Regression	50.667	L2 = 1.0	0.4564	L2 = 1.0
Neural Network	0.074	N/A	~0	N/A
Two-Layer Random Forest	1.926	N/A	0.0144	N/A
Neural Network to Random Forest	1.259	N/A	0.0076	N/A
Convolutional Neural Network	29.931	N/A	N/A	N/A
Xception	**48**	**N/A**	**N/A**	**N/A**

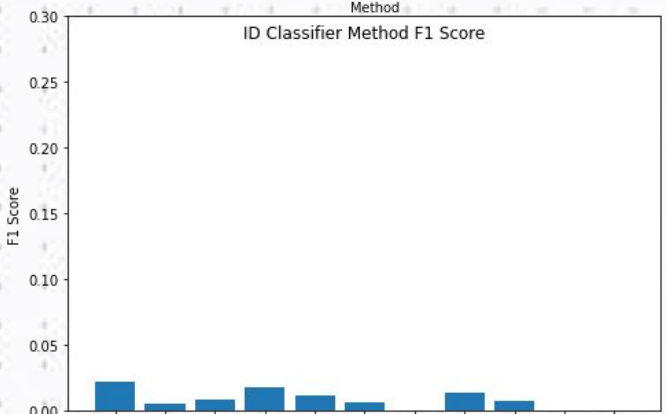
Berkeley
UNIVERSITY OF CALIFORNIA

Tests/ Graphs/ Discussions







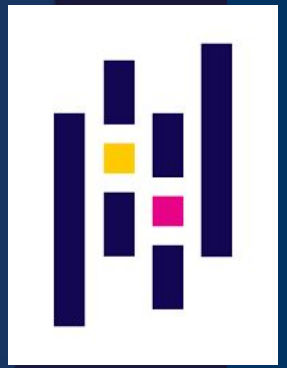


Constraints

- The sample dataset for model development:
  - 6750 data points
  - 80%/20% train/test split
  - 26 species
  - 1350 individuals (5 images each)
- Training using the full dataset / computing resources was lengthy and time-consuming:
  - 38,000 training images max due to the time constraints on Kaggle
  - 43,200.90 (12 hours until time out) seconds to run
- Model development was performed on the training test set only (no labels for the test set)
- All images except a few were RGB so gray-scaling was required for all images

Standards

- Python Packages used:
  - Math
  - Numpy
  - Pandas
  - Matplotlib
  - Tensorflow
  - Xception Model
  - h5py
  - Keras
  - SK Learn
  - Keras
  - SciPy
- Python Version: 3
- Jupyter Notebooks (local training)
- Kaggle (for larger-scale training)
- Disk Storage used: 68+ GB
- RAM Memory used: 32 GB



kaggle

Berkeley
UNIVERSITY OF CALIFORNIA

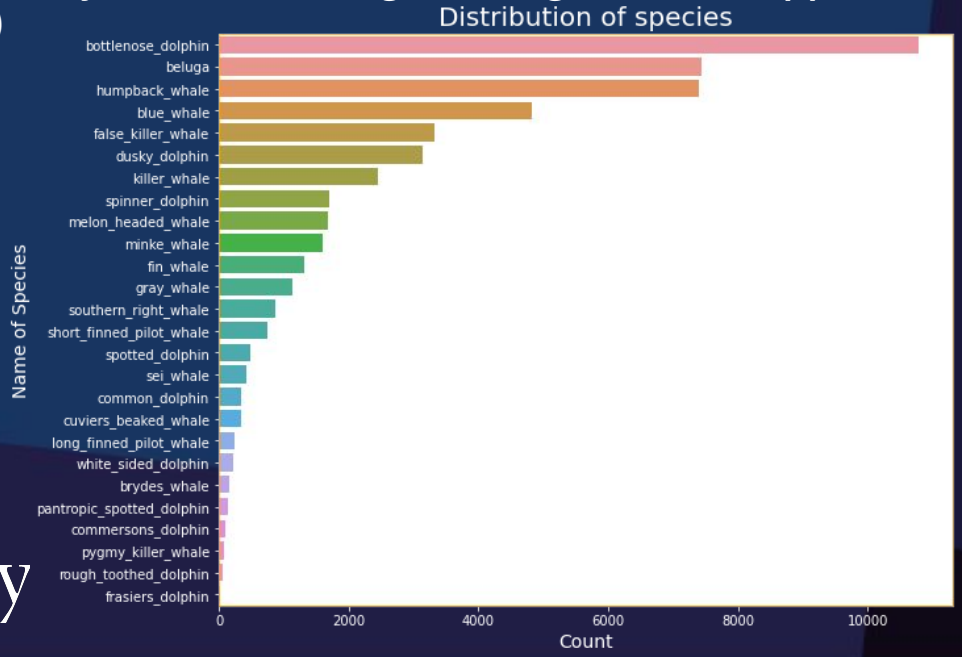
Comparison

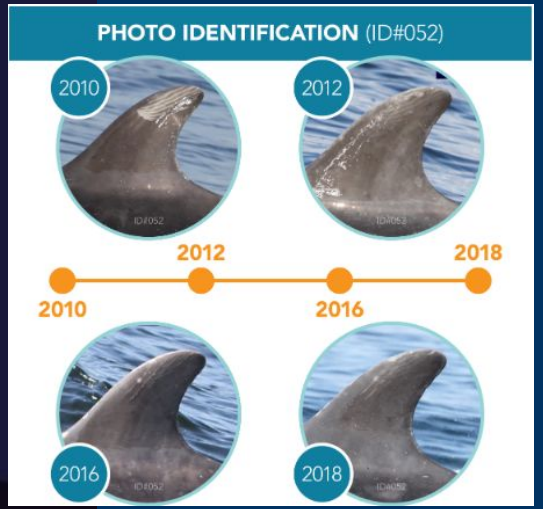
Accuracy and F1 are compared among each model

- Xception and basic CNN predicted ID well probably because the hidden layers extracted additional features in the images but poorly predicted species probably because of this
- Random Forest may have had a good blend of feature importance and mitigating overfitting.
- K-Nearest Neighbors memory/distance-based learning may have grouped similar species/individuals near each other
- The drawback for the Naive-bayes algorithms may have been that the features were not completely independent of each other

Limitations of the Study

- Lack of many images of the same individual → individual ID identification was very hard
- Trained on the 38,000 images max (due to timeout limitations on Kaggle)
- Aging and injuries caused slight changes in their appearance (unaccounted for in our models)





Berkeley
UNIVERSITY OF CALIFORNIA

Future Work

- Ensemble methods
- Siamese model (human facial ID)
- Bounding boxes to crop out surrounding objects and a mask to remove the pixel variation in the images due to background features
- Flip the images so that all pictures will face the same way
- Increase the contrast in images to improve neural network models

Q&A

Thanks for listening!

Berkeley
UNIVERSITY OF CALIFORNIA