CS230

North American Marine Mammal Detection for Enhancing Whale-Ship Collision Avoidance

Chris Barresi
Department of Computer Science
Stanford University
cbarresi@stanford.edu

Abstract

At sea collisions with whales are endangering the biggest animals to have ever roamed our planet. Current methods for maritime vessels to detect these animals are expensive, inaccurate, and inefficient. With detection failures leading to catastrophic injury of both the animal and the vessel. To improve the use of resources for government officials, ecologists, and other researchers, we are proposing an auto-detection algorithm using oceanic imagery as input. Using 1000 images of whales from online datasets, we labeled the images to train a modified version of ResNet 50. The resulting model—a Faster R-CNN—with a F1-Score = 0.76, Log Average Miss Rate = 0.4, and mean Average Precision = 0.876 is able to accurately detect cetaceans species in their environment.

1. Introduction

As maritime traffic increases, so have collisions with the biggest animals to roam our oceans. Just in 2021 more than 500 whale strandings were recorded (US only), most struck by a ship. As these species are in decline, there is an increased need to properly identify them to avoid colliding with them, preventing their death and economic impact to the vessel and the municipality that ends up with the whale carcass.

We propose visual identification of the marine species as the most economical and viable alternative, given that most navigational radars for big merchant vessels have a blind spot at the front of the vessel within 1000 yds are not good at discerning small objects on the surface of the water, and sonar systems are block by the ship's front wake.

This project has two main challenges: 1) Relatively small databases to choose from (compared to other animals such as cats or dogs)—given that whales, may present different aspects of their physiology. 2) Proper classification of the cetacean species might be difficult in some instances due to their physiological similarities; therefore it might be more viable to have a more accurate detection system of the different physiological aspects commonly presented above water to identify an animal close by than proper classification of the species.

2. Dataset

a. Raw Data

One thousand images from Open Image Dataset V6 were gathered, encompassing images of the most popular cetacean species. These images contain the most typical aspect of cetaceans above water, such as tail flukes, dorsal fins, flippers, and waterspouts. A minor section of these images includes whole body portion of the cetacean species.

Unfortunately, the dataset is skewed with cetaceans observed at theme parks—like Killer Whales and dolphins—and cetaceans observed during whale watching tours—such as baleen whales, specifically, humpback whales. The dataset is divided into 50 images for testing, 150 images for validation, and 800 images for Training.

b. Labeling

This dataset was manually labeled to reduce instances of errors. The labeling scheme includes the desired detection features, such as: tail flukes, flippers, dorsal fins, and general physiological shape. Manual labeling was accomplished using Matlab's Image Labeling app.

3. Methods

a. Proposed Model Architecture

A ResNet 50 neural network was modified to perform transfer learning and create a model that detects and classifies cetaceans in an image. This was accomplished by freezing all but the last three layers of the network and substituting these layers with new classification layers for our model, box regression layers, and ROI max pooling layers. Then defined anchor boxes and Region Proposal Layers to improve the algorithm performance and speed. This model transforms the pretrained ResNet 50 network into a Faster R-CNN, which favors efficient and accurate detection, over speed.

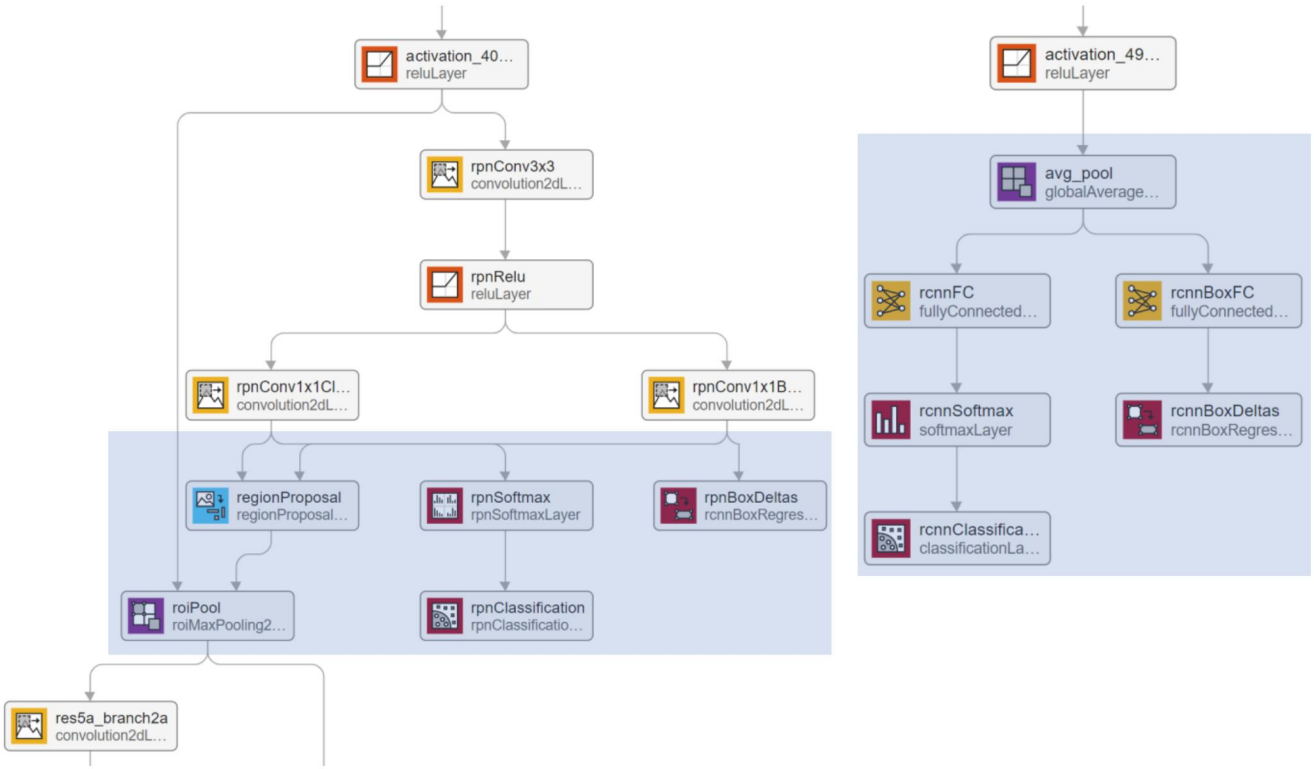


Figure 1. Faster R-CNN Modified Layers to ResNet 50 pretrained network. New Classification layers, Box regression Layers, and ROI max pooling layers

b. Anchor Boxes

For our model to accurately detect our desired features, new anchor boxes were defined using Matlab's estimateAnchorBoxes function. This function evaluates all the training

data detection boxes and returns the most optimal anchor boxes. A total of 7 anchor boxes were utilized to detect the different features of the cetacean species.

c. Training Parameters

For our model, Stochastic gradient descent with momentum was selected for training. Using 0.9 momentum parameter, an initial learning rate of 1E-3, and a learning rate drop factor of 0.1 per epoch of training for 10 epochs. Although not much change was observed after the sixth epoch. The training set was shuffled between every epoch. All other hyperparameters were set as default and allowed Matlab to optimize them during training. In addition, due to memory constraints of the hardware used (Nvidia RTX 1060), minibatches of 4 were used for training with image input size of [227 227 3].

4. Results

a. Training Metrics

During training, the training set and validation set accuracy, RMSE, and Loss were measured. The validation set was evaluated every 100 iterations of training. At the conclusion of training both sets had achieved accuracies greater than 90%, with minimal loss and relatively low RMSE.

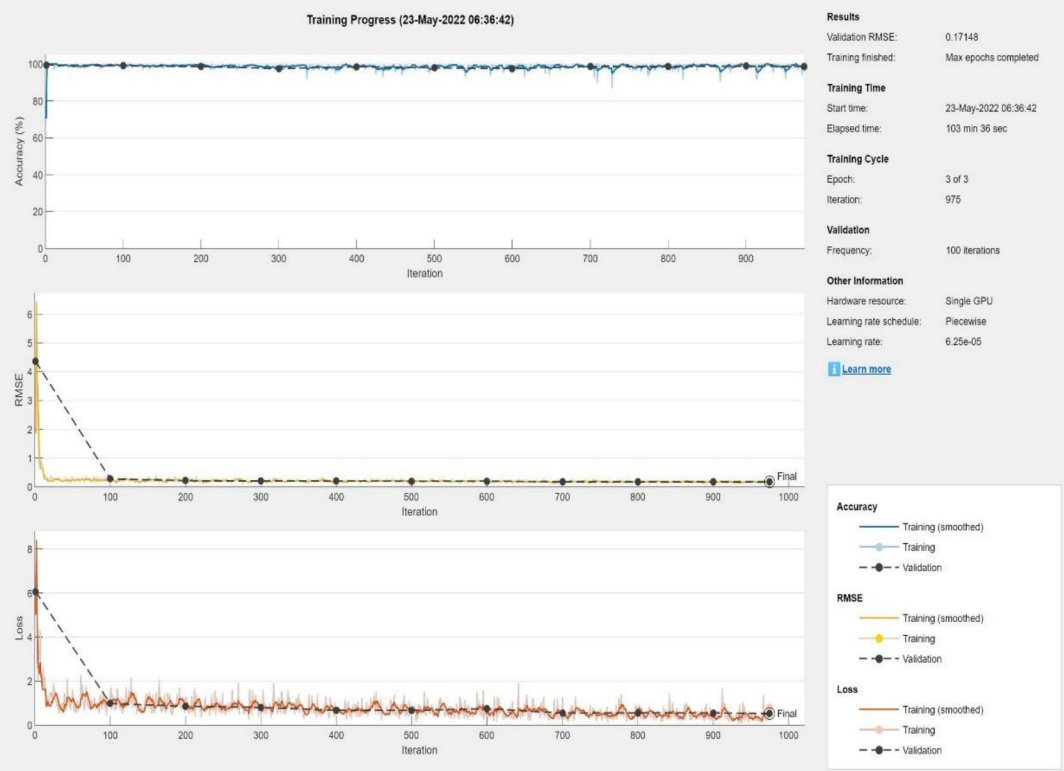


Figure 2. Training Metrics: Accuracy, RSME, and Loss for 3 epochs of training

b. Evaluation Metrics

Due to the tasking of the model, we have chosen F1-score, mAP-score, and Log Average Miss Rate as our main performance indicators. With these metrics, we assure to detect as many cetaceans as possible while keeping bounding boxes close to ground truth. From the two metrics, we have chosen mAP as our main performance indicator to be able to compare our results to those from the research community.

i. F1: Harmonic mean of precision and recall: $F1 = 2(P \cdot R)/(P + R)$

1. Precision (P): ratio of true positives and all positive predictions.
2. Recall (R): ratio of true positives and all positive ground truth.

ii. mAP (mean Average Precision): Mean area under the precision recall curve.

These parameters were determined using Matlab's evaluateDetectionMissRate, evaluateDetectionPrecision, and bboxPrecisionRecall functions.

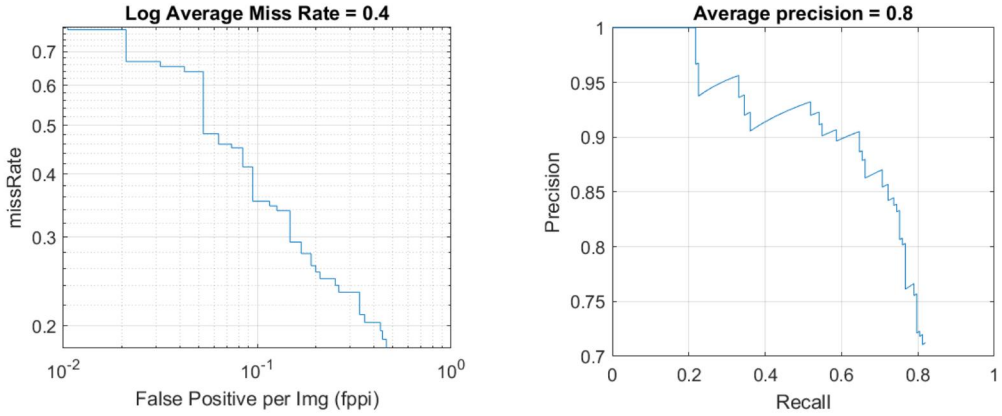
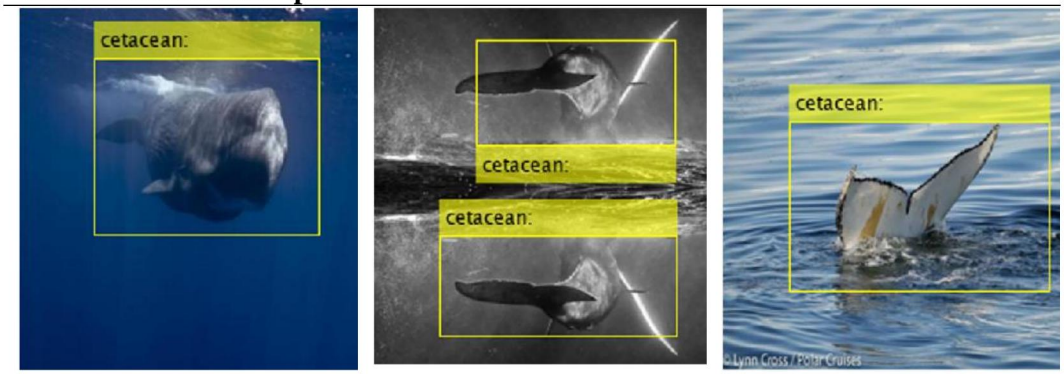


Figure 3. Evaluation metrics for our Model with and F1-Score = 0.76. Log Average Miss Rate = 0.4. mean Average Precision = 0.8

c. Detection Examples

Good Detection Examples



Poor Detection Examples

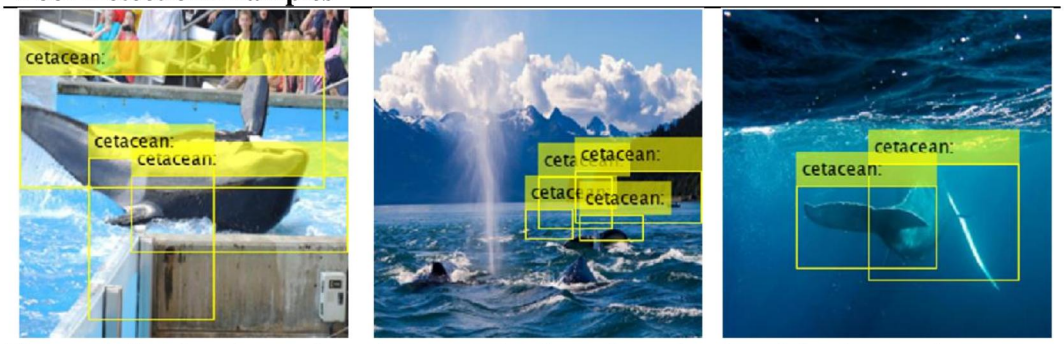


Figure 4. Good and poor detection examples from the trained model using random images from the internet, not contained in our test, training, or validation set.

5. **Discussion**

Looking at the results, the Faster R-CNN model—using pretrained weights from the ResNet 50 neural network—is effective and accurate at detecting cetacean species in their natural environment. Given the small dataset size and the variability of the images (different physiological aspects presented, and variations from species to species), we believe our results

are very promising. Although the model poor detections are mainly contributed not no misclassification or detection, but rather to detecting different physiological aspects without considering that holistically it belongs to the same animal. A feature that can be further improved by using simple algorithms that check which body parts were detected and stitch them together for the final detection box.

6. Conclusion and Future Work

In this work, we have presented a modified version of the ResNet 50 neural network that acts as a Faster R-CNN specifically crafted for cetacean detection in an oceanic environment. From a baseline ResNet 50 model, we have iterated to find a model with good performance metrics, obtaining successful results. However, we think that the characteristics of our dataset and labeling scheme hinder the performance of the presented model.

As future work, after modifying the labeling scheme, this model can be used to automatically label a bigger dataset to accurately detect and classify the different species of cetaceans and different body parts presented in their environment to the observer. Apart from the labeled images, multiple information could be extracted—such as population size, size of the cetacean, proximity relations to other species, and relative motion with respect to the observer.

7. References

P. C. Gray, K. C. Bierlich, S. A. Mantell, A. S. Friedlaender, J. A. Goldbogen, and D. W. Johnston, “Drones and convolutional neural networks facilitate automated and accurate cetacean species identification and photogrammetry,” in Methods in Ecology and Evolution, 2019.

M. Tan, R. Pang, and Q. V. Le, “Efficientdet: Scalable and efficient object detection,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 10781–10790, 2020.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779–788, 2016.

K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in Proceedings of the IEEE international conference on computer vision, pp. 2961–2969, 2017.