Shark and Human Tracking from Drone Footage

Nic Tekieli
Harvey Mudd College
ntekieli@hmc.edu

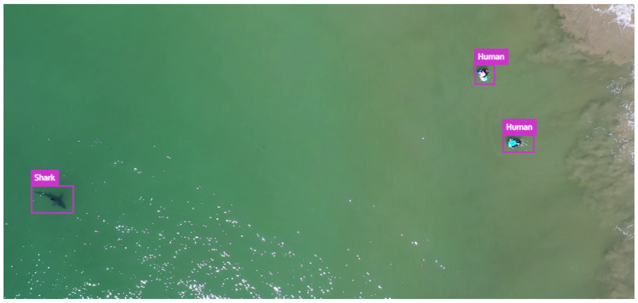


Figure 1: Manually labeled input image with separate labels for sharks and humans that we use to train our model

1. Introduction

1.1. Background

Marine animal behavior is a field that has heavily utilized drone footage to gather data. Studying the behavior of marine animals in the presence of disturbances in bodies of water is difficult with the availability of data, and drone footage has helped this field greatly. One of the subsets of this field is the study of shark locomotion data in the presence of surfers and bodyboarders near coasts. Sharks, like many other marine animals, have escape responses that are modulated by the size and distance of other bodies in water. Understanding how sharks react to disturbances in their habitat is incredibly important for the design and preservation of marine reserves in order to help conservation of these ecologically crucial animals.

1.2. Motivation

Drone footage is used as the main tool to gather locomotive data in order to study shark behavior. At the SharkLab at CSU Long Beach, researchers have been recently started using drones to gather data on white sharks along the Californian coast, but the large number of videos collected and no system to automate the extraction of data from these videos has created a bottleneck impeding research progress. This paper is meant to demonstrate a fully automatic method for labeling images extracted from drone footage of sharks and human in bodies of water using a model trained on manually labeled footage.

1.3. Context

Object detection is a very popular field in computer vision, and using drone footage has become increasingly easier to utilize due to advancements in drones and modeling algorithms. However, using drone footage for gathering data above bodies of water raises more challenges. Many of the most popular models have limited accommodations for tracking objects in water. This paper intends to describe an approach to using drone footage over bodies of water in a manner effective enough to be utilized to extract locomotive data of sharks in water.

2. Method

2.1. Data Handling

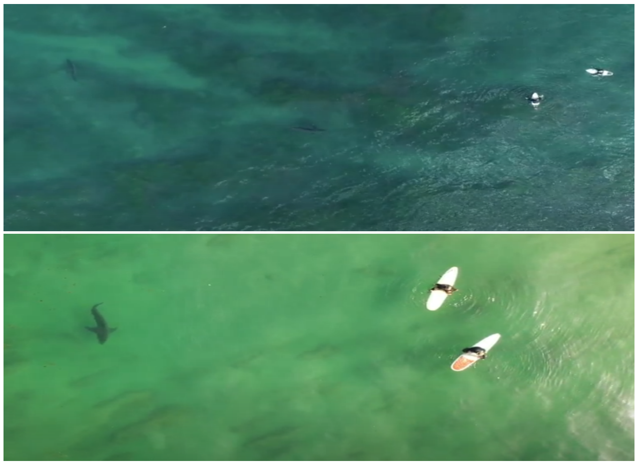


Figure 2: Above: Medium quality video showing two sharks and two humans. Below: Good quality video showing one shark and two humans

CSULB provided a number of videos grouped by quality for either ”medium” or ”good” videos, as seen in figure 2.

These drone footage files can contain any number of sharks or humans and some videos include the shoreline as well. We chose to not include any bystanders on the coast as they would not interact with the sharks in any meaningful way.

In order to extract images from the drone footage, we used a simple OpenCV scripts to extract JPEG images from the files. We then used these images with the Computer Vision Annotation Tool (CVAT), as seen in figure 3, to manually annotates frames from the videos with labels for sharks and humans. The process of manually labeling the images was important to have ground truth samples for our model.

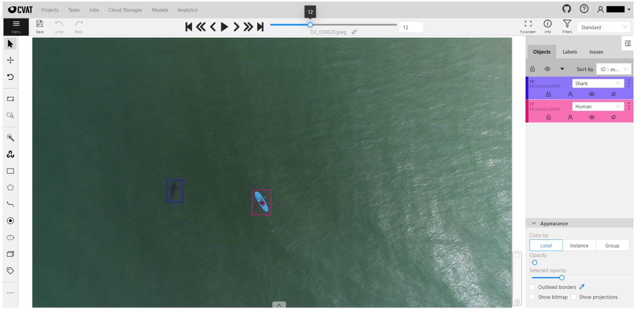
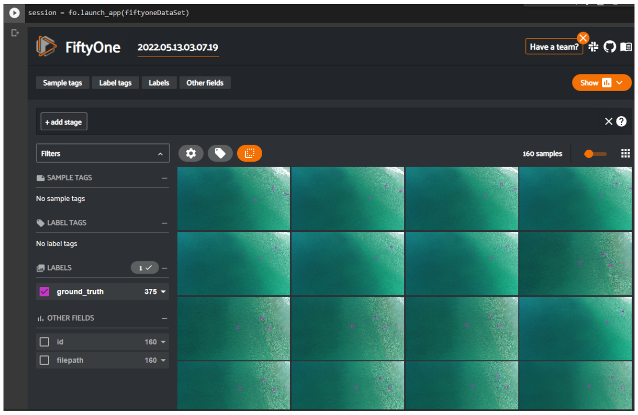


Figure 3: Utilizing CVAT to annotate images frame-by-frame

Using CVAT allowed us to extract the annotated images and the JSON file containing the label information in the COCO format. COCO is a large object detection dataset that saves annotation labels as a JSON and is commonly used in object detection models.

Once the images and the JSON containing the annotation data were extracted as a COCO file, we used FiftyOne, an open-source tool that allowed us to visualize our data in a more succinct window, as seen more clearly in figure 4.



```
session = fo.launch_app(fiftyoneDataSet)
```

Figure 4: FiftyOne open source tool for viewing ground truth labels

2.2. Model

For this dataset, we decided to use PyTorch's Faster R-CNN model with a ResNet-50-FPN backbone. This model has been pretrained on the COCO dataset, which contains over 200k labels images.

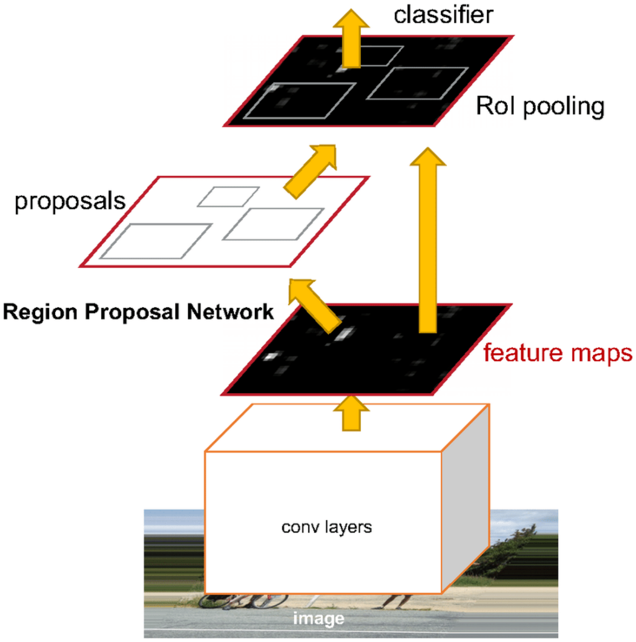


Figure 5: Faster R-CNN is a single, unified network for object detection

Faster R-CNN is an object detection model that utilizes a region proposal network with the convolutional neural network that is able to simultaneously predict object boundaries and prediction scores for each position. The model extracts feature maps from the images using the ResNet50 FPN and computes location and objectness scores using the region proposal network, as visualized in figure 5.

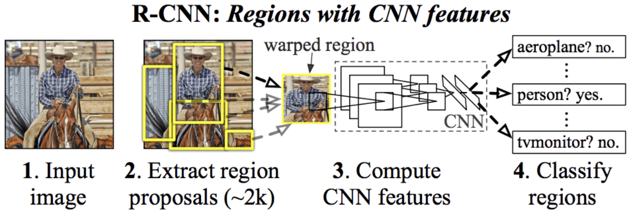


Figure 6: Faster R-CNN is a single, unified network for object detection

We used this model to build a network that could classify if an object was a shark or human and determine bounding box locations using inference.

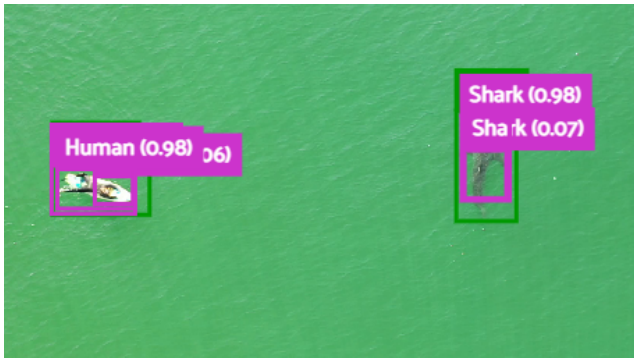


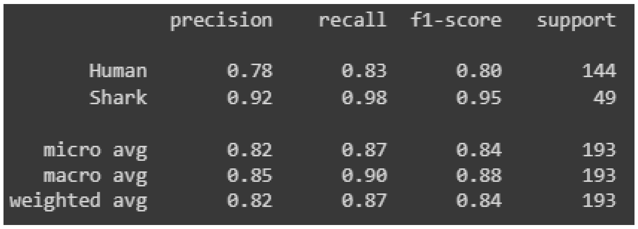
Figure 7: Evaluation images show predictions with confidence intervals

To train this model, we first had to create a new dataset class using PyTorch's Dataset class. This was necessary in order to organize the data in a manner that would make inputting our training data into the model more streamlined. Using our custom dataset class, we were able to train our Faster R-CNN with around 400 manually annotated images over 10 epochs. This gave us our trained model that we could then evaluate new images with. We gave our model 100 randomly selected drone footage images to test the accuracy of our model.

Using FiftyOne's filtering tools, we were able to filter our detections to get a set of high confidence predictions that we could use to evaluate our model.

3. Results and Evaluation

We used the COCO evaluation protocol to aggregate results. We were able to view a classification report for the two classes in our dataset and compute the mean average-precision (mAP) of our prediction model.



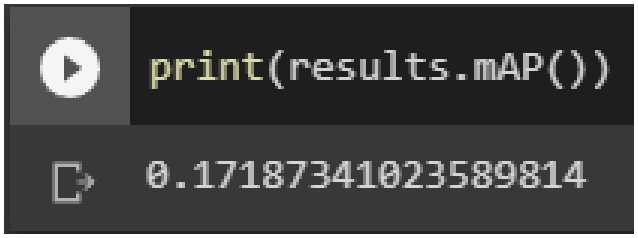
```
              precision    recall  f1-score   support

       Human       0.78      0.83      0.80       144
       Shark       0.92      0.98      0.95        49

   micro avg       0.82      0.87      0.84       193
   macro avg       0.85      0.90      0.88       193
weighted avg       0.82      0.87      0.84       193
```

Figure 8: Classification data for the precision and recall score of the two classes

To compare this mAP score of 0.172 which comes from our model that is trained on around 400 images, the pre-trained Faster R-CNN evaluated on the COCO 2017 Zoo dataset, which is trained on almost 120,000 images, evaluates with a mAP score of 0.394.



```
print(results.mAP())
0.17187341023589814
```

Figure 9: mAP score for our model

We used FiftyOne's plot function to view a graph of our results that show the precision-recall (PR) curves for the two classes in our model, as seen in figure 10. PR graphs are always monotonically decreasing as we trade-off precision and recall. For this model, the average precision for Sharks was 0.226 and the average precision for Humans was 0.117.

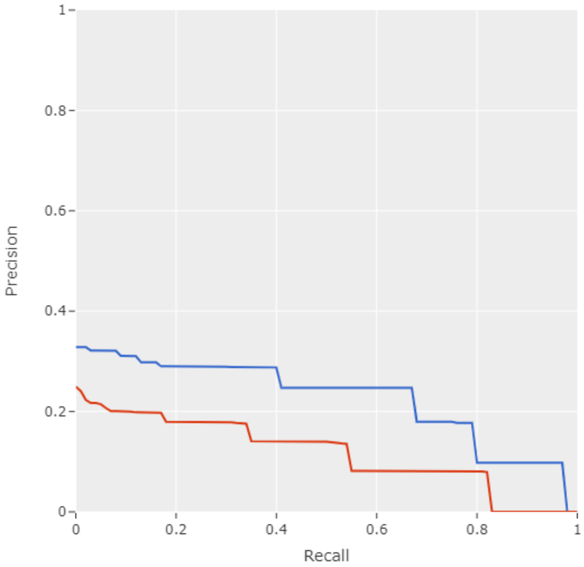


Figure 10: PR curves for specific classes in our model

3.1. Limitations

We were able to extract object labels and positions for Humans and Sharks using this model, although to varying degrees of success on lower quality videos. This can be seen most commonly when we come across images with poor lighting, where the shark is occluded by waves or shadows. Future projects dealing with similar problems may want to look at other CNNs that are more accommodating with drastic lighting changes from one dataset to another. The model also tends to create multiple bounding boxes around smaller portions of the sharks or humans, but these generally have low confidence values and can be discarded when observing the prediction data.

References

[1] Brownleem, Jason (2019), A Gentle Introduction to Object Recognition With Deep Learning, https://machinelearningmastery.com/object-recognition-with-deep-learning/

[2] Seamone, S. et al. (2014) Sharks modulate their escape behavior in response to predator size, speed, and approach orientation. Zoology, 117, 377-382

[3] S. Sambolek and M. Ivašić-Kos, ”Detecting objects in drone imagery: a brief overview of recent progress,” 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2020, pp. 1052-1057, doi: 10.23919/MIPRO48935.2020.9245321.

[4] Tucker, J.P. et al. (2021) White shark behaviour altered by stranded whale carcasses: Insights from drones and implications for beach management. Ocean and Coastal Management, 200, 1-8