Save the Whales Project
Data Science for the Common Good
August 20, 2021

Chhandak Bagchi

cbagchi@umass.edu

University of Massachusetts Amherst

Gizem Cicekli

gcicekli@umass.edu

University of Massachusetts Amherst

1   Introduction

When marine biologists encounter whales near the surface in the open ocean, they often want to rapidly assess the health of animals, including any of their calves. There are many reasons why they want to do this, ranging from routine assessments of population health, to urgent operational planning to disentangle whales that have been caught in fishing nets. This project developed an image processing pipeline that analyzes drone images of whales to automatically calculate their body conditions (mass, surface area, volume, etc.) hundred times faster than human experts.

2   Motivation

Historically, the true health of whales could only be assessed through whaling, which adversely impacts critically endangered whale species such as the North Atlantic right whale. Nowadays, the conventional method to assess the health of whales are by putting human divers into the water near the whales for direct inspection. But this method is time consuming, risky, and inaccurate. Advanced drone technology has made it possible for biologists to swiftly capture high resolution aerial images of whales and their calves at the surface. Along with these, biologists have developed a painstaking photogrammetric methodology for analyzing these aerial images to estimate whale shape, size, and body composition. But this analysis can take an hour or more per image and sometimes an accurate answer is needed in minutes. That's where this project makes its contribution. In partnership with Woods Hole Oceanographic Institution, Aarhus Institute of Advanced Studies, and the UMass Amherst Biology Department, our team tackled the problem of rapidly assessing the body condition of whales by applying state of the art computer vision techniques and advanced data science algorithms.

3   Data & Approach

Our partners furnished a corpus of over 12000 drone images of whales taken in the waters near Australia and Argentina. Of these, approximately 3000 came with usable photogrammetric metadata created by marine researchers marking the tip and tail of any visible whales.

A software prototype was created that performs the following tasks:

- **Object detection:** Draws bounding box around each whale in the image

- **Key point detection:** Marks key points on the detected whales, including rostrum (head) and fluke (tail)

- **Segmentation:** Differentiates each instance of the whale from the surrounding water

- **Body size estimation:** Calculates the size of all visible whales by estimating the overall length as well as widths at 5% intervals

4   Results

Although the prototype can make mistakes including cutting off the shape of the polygon or forming an incorrect shape, preliminary evaluation results still suggest that its performance is at least as accurate as human analysis for an overwhelming majority of images. But while human analysts may need an hour or more to analyze a single whale in an image, our image processing system can process the entire image in only four seconds.

5   Impact

The software prototype is now in the hands of researchers who plan to validate its results against other data. When successfully validated, this prototype will be able to be deployed on marine research ships, and there would be a repeatable method for applying data science to help other threatened species.

6   Acknowledgments