ICS 637 Final Report

Jayden Tactay

December 20, 2024

1 Introduction

At a brief overview, this project uses deep learning to perform object detection. The task calls for the detection of cetaceans in ocean images, with an emphasis on detecting dorsal fins of dolphins. Given images taken at sea, the deep learning model should determine whether those images contain sea mammals and produce bounding boxes around any mammals present.

This object detection task is part of a larger project of automatic photo quality classification for cetacean photo-identification. Marine biologists use techniques like photo-identification to track population sizes and trends, like migration patterns, of various marine animals, aiding in conservation and protection efforts. Effective photo-identification requires high-quality images of distinctive animals; however, as social species like dolphins often travel in groups, manually filtering and cataloging images of cetaceans is time- and labor-intensive, often serving as the bottleneck of photo-identification tasks. Automatic photograding would drastically increase identification efforts, for which cetacean detection is a crucial first step.

2 Data

The dataset used by the model consists of various images taken at sea by marine biologists from the University of Hawaii at Manoa and National Oceanic and Atmospheric Administration (NOAA). These images are collected as part of field surveys and other research expeditions in various locations (e.g., the Mariana Islands). While unlabeled data is plentiful, the current dataset consists of 2414 hand-labeled images. After augmentation, the dataset consists of 5872 images, split into the following subsets: 5213 train, 442 validation, 217 test. The training set is used to train the model, while the validation set is used for early stopping, hyperparameter tuning, and model selection. The test set is used for a final, unbiased qualitative and quantitative evaluation of the model—on the Roboflow hub of the project, reported model performance metrics are calculated using the test set. The labels mostly come from two primary labelers, who have examined and modified the annotations of any external labelers.

2.1 Pre-processing and Augmentation

Before training, each image is automatically oriented to maintain a consistent orientation across all images in the dataset. A rotation of up to 15 degrees is also randomly applied to the training images, increasing the size of the training split from 1755 images to the 5213 used during training. During training, images are resized to 640x640 pixels. The dataset is also artificially expanded even further using the default training-time random augmentations defined by the Ultralytics YOLO model implementation:

- Adjust hue by 0.015
- Adjust saturation by 0.7
- Adjust value (brightness) by 0.4
- Translate by 10% of image size
- Scale by 50%
- Flip horizontally with 50% probability
- Erase 40% of the image

3 Model

The model is an Ultralytics YOLO11m object detection model that has been fine-tuned on the cetacean dataset. As the task involves detecting cetaceans in an image, the latest pretrained state-of-the-art YOLO11 object detector is an ideal backbone for the model due to its increased performance over previous YOLO versions. The model uses the medium version of YOLO11 to maintain a balance between performance and efficiency.

3.1 Hyperparameters

Hyperparameters were hand-optimized through iterative design and side-by-side model comparisons. Different values for the number of epochs, batch size, and early stopping patience were tested. Epoch count ranged from 50 to 100, and batch sizes of 16 and 32 were directly compared under similar training settings. Early stopping patience was increased by ranges of 5 from 5-15 epochs, with 10 epochs proving to be the best duration for observed improvement without excessively training. Augmentation strategies, specifically that of deterministic or probabilistic image rotation (i.e., physical or synthetic dataset growth), were compared with each other. After testing each combination of hyperparameter and augmentation of interest, the model with the highest mAP on the validation set was chosen: 32 batch size, 100 epochs, and deterministic rotation by up to 15 degrees. As the object detection task is part of an ongoing project, previous iterations of the model were also compared with the best new model using the validation set of the latest dataset version. This analysis affirmed that the current model achieves the best performance compared to older models and models with different hyperparameters.

3.2 Evaluation and Analysis

The final model was evaluated on the test set, obtaining 91.3% mAP, 87.9% precision, and 82.4% recall.

While better metrics do exist for the model's bounding box capabilities (i.e., intersection over union, or IoU), the calculated metrics still provide a helpful indication of model performance. Apart from how well the model draws boxes, it is also important that the model can properly identify cetaceans in an image, as false negatives from missing an animal entirely are costly. As such, the current performance of the model indicates that its detection capabilities will adequately generalize. Furthermore, the partner researchers conduct field surveys and data collection at consistent times during each expedition, so lighting conditions and color compositions of future images are not expected to deviate heavily from current training images.

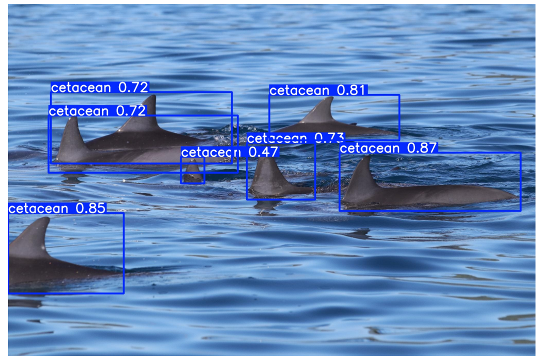

Figure 1: The model does well with complex scenes containing many cetaceans. This image has an appropriate number of bounding boxes, with each box sufficiently encapsulating each animal.

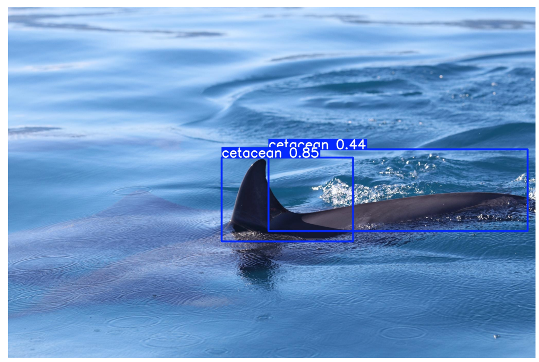

Figure 2: However, the model may also make errors. For example, this image received two bounding boxes, but only one animal is present.

4 Conclusion

The object detector serves an integral first step in the photograding pipeline. It even works well as a standalone model for detecting cetaceans, performing better than existing models (especially on the kinds of images used specifically for cetacean photo-identification). In fact, this current iteration of the model itself is being utilized for a separate, but related, experiment. Naturally, the model still has room for improvement, but given that it performs its task proficiently in its current state, efforts can now be focused on other areas of the photograding pipeline.