Help Protect the Great Barrier Reef

Clément Apavou[1], Guillaume Serieys[1]

[1] MVA Master's student at École Normale Supérieure Paris-Saclay, France.

Introduction

The Australia's Great Barrier Reef is the world's largest coral reef and home to 1,500 species of fish, 400 species of corals, 130 species of sharks, rays, and a massive variety of other sea life. Unfortunately, this environment of magnificent beauty is threatened by the overpopulation of a coral-eating starfish species named the crown-of-thorns starfish (COTS).

Figure 1: Greet Barrier Reef

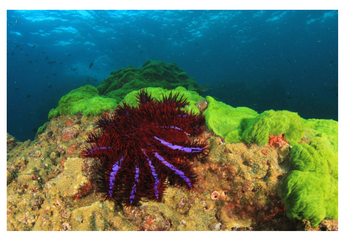


Figure 2: Crown-of-thorns

Scientists, tourism operators and reef managers established a large-scale intervention program to control COTS outbreaks to ecologically sustainable levels. To know where the COTS are, a traditional reef survey method, called "Manta Tow", is performed by a snorkel diver. While towed by a boat, they visually assess the reef, stopping to record variables observed every 200m. While generally effective, this method faces clear limitations, including operational scalability, data resolution, reliability, and traceability.

Application

In order to develop new survey and intervention methods in COTS control, the Great Barrier Reef Foundation established an innovation program. The new survey and intervention methods will consist in collecting thousands of reef images using underwater cameras and then apply AI technology with a near real-time object detection model trained on underwater videos of coral reefs to improve the efficiency and scale at which reef managers detect and control COTS outbreaks. Thus, our work will help researchers identify species that are threatening Australia's Great Barrier Reef and take well-informed action to protect the reef for future generations.

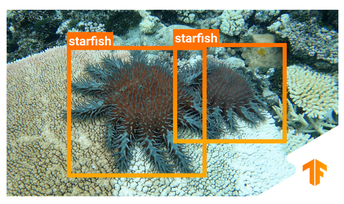
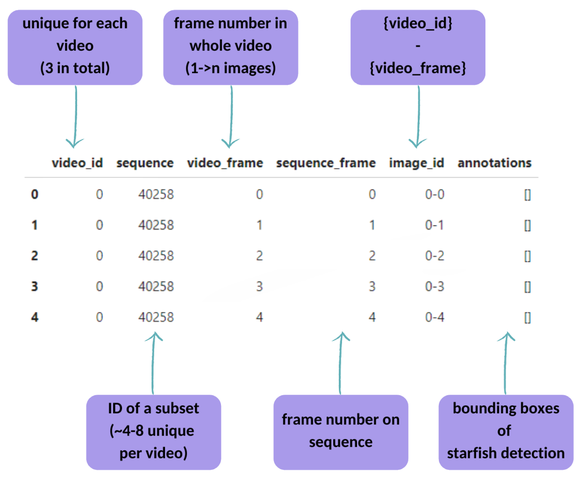


Figure 3: Crown-of-thorns.

Figure 4: Dataset image.

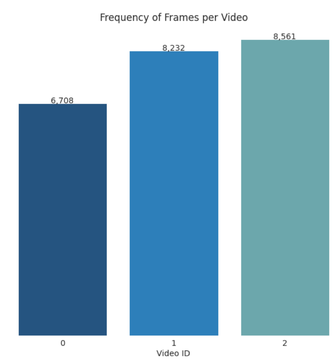
Dataset

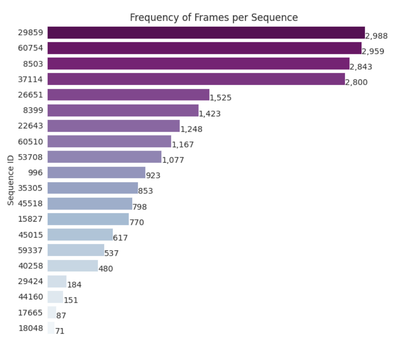
The dataset contains a training set of 23,501 annotated images and a test set of 13,000 images whose annotations are not accessible. Images are size 1280x720 (WxH). Each image has the following metadata:



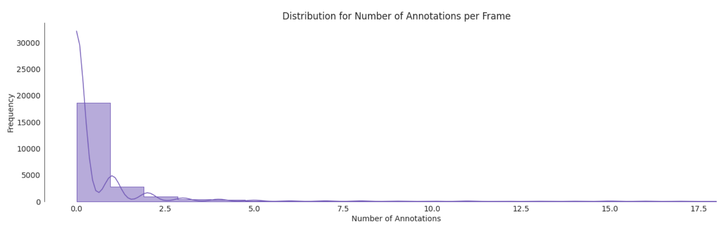
Dataset Analysis

There are a total of 3 videos. Each video is split into sequences. 1 video (id 0) is split into 4 sequences, while the other 2 videos (id 1 & 2) are split into 8 sequences. Each sequence does not have the same number of frames.





All images do not contain starfish. Indeed, there are approximately 79% frames with no annotation and only approximately 21% frames with at least 1 annotation. Specifically, of 23,501 images, only 4,919 have at least one annotation.



Difficulties

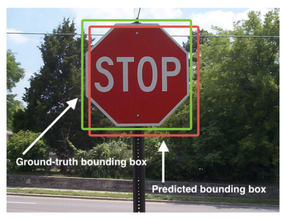
1. The crown-of-thorns starfish are difficult to distinguish because of their coral-like appearance. Also, images are taken from a distance which makes it even more difficult (Figure 4).
2. The training set contains images from only three videos. This details, can be a very challenging detail to train a model so that it generalizes well on data that it has never seen or with contexts different from those seen during the training.
3. There are many more images without annotation and less images with at least one annotation. If all dataset images are given during the training, the model will see few positive examples and therefore it will not detect well crown-of-thorns starfish.

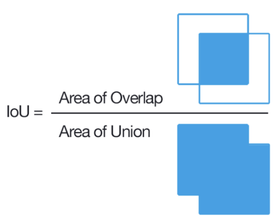
Evaluation

The competition is evaluated on the F2-Score at different intersection over union (IoU) thresholds. The F2 metric weights recall more heavily than precision, as in this case it makes sense to tolerate some false positives in order to ensure very few starfish are missed.

$$F_\beta = (1+\beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$$

The metric sweeps over IoU thresholds in the range of 0.3 to 0.8 with a step size of 0.05, calculating an F2-score at each threshold. For example, at a threshold of 0.5, a predicted object is considered a "hit" if its IoU with a ground truth object is at least 0.5.

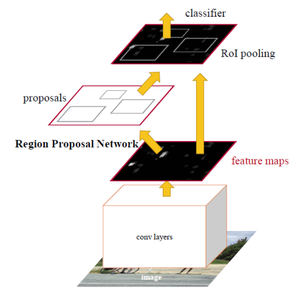




A true positive is the first (in confidence order) submission box in a sample with an IoU greater than the threshold against an unmatched solution box.

Method : Faster R-CNN (R-CNN)

Faster R-CNN (Ren & al. 2016) is a two-stage detector. The first stage is to generate region proposals by selective search (sliding window) with a Region Proposal Network (RPN) and the second stage is to detect and classify the object of each proposal.



During the training, there are two losses for the region proposal network (objectness and box regression) and two losses for the second stage (classification and box regression).

Method : RetinaNet (SSD)

RetinaNet (Lin & al. 2018) is a Single Shot Detector (SSD) using the Focal Loss for the classification which allow to tackle extreme imbalance between foreground and background classes during training. A SSD is faster (during inference) than a R-CNN based RPN like Faster R-CNN but less accurate. During the training there are only two losses one for the classification (Focal loss) and the second for the box regression (L_1 loss).

Experiment

For the experiments, we split the training set into two sets, the train set with frames of the videos 0 & 1 and the validation the video 2. We used data augmentation with albumentations and applied RandomBrightnessContrast, HorizontalFlip, RandomSizedBBoxSafeCrop (840x360). We used the implementation of Faster R-CNN and RetinaNet of the library torchvision models. It was pre-trained on COCO train2017 and the backbone ResNet-50-FPN was pre-trained on ImageNet. For the optimization, we used the stochastic gradient descent (SGD) with 16 images per minibatch an initial learning rate of 0.005 with a momentum of 0.9 and a scheduler that reduces learning rate when the F2-score of validation has stopped improving. Models were trained on Kaggle GPU Tesla P100-PCIE-16GB with 2 CPUs and the training visualization was made with Weights & Biases (wandb).

Results

Model	Split	F2 val.	F2 test
FasterRCNN	positive samples	0.4027	0.359
	all images	0.03622	0.322
RetinaNet	positive samples	0.3235	0.368

Best F2-score on the public leaderboard : 0.673.

Future Work

1. Try other splits (e.g add some negative samples to the training, split by sequences).
2. Try Vision Transformers in backbone and Yolo, one of the best single shot detector.
3. Pre-train model with the dataset and external underwater datasets.