Help Protect The Great Barrier Reef

Kavya Jaganathan
MS in Artificial Intelligence
Northwestern University
Evanston, USA
kavyajaganathan2022@u.northwestern.edu

Clarissa Cheam
MS in Artificial Intelligence
Northwestern University
Evanston, USA
clarissacheam2022@u.northwestern.edu

Srik Gorthy
MS in Artificial Intelligence
Northwestern University
Evanston, USA
srik.gorthy@u.northwestern.edu

***Abstract*—This project attempts to utilize different methods to perform underwater multi-object detection in real time to help mitigate COTS outbreaks and prevent the further degradation of the Great Barrier Reef.**

***Index Terms*—multi-object detection, yolov4, underwater images**

I. Introduction

Underwater object detection is crucial to a lot of different fields, such as homeland security, the repairing of undersea structures, maintenance, and of course, marine science [1]. This project in particular aims to maintain the health and mitigate the damage of the Great Barrier Beef by identifying crown of thorns starfish using multi-object detection during populations outbreaks (also known as COTS outbreaks) and eliminating a number of them. While crown of thorns starfish are a natural part of reef life, COTS outbreaks are one of the greatest threats to the Great Barrier Reef: during outbreaks, the starfish can number in millions, and each starfish consumes about 10 square meters of coral per year, which adds up to a significant amount of coral loss [2]. We will be performing multi-object detection on our dataset using various methods such as transfer learning using pre-trained models, multiple anchor box methods and the multiple density based object detector.

II. Data

In this project, we will recognize/predict the presence and position of crown-of-thorns starfish in sequences of underwater images taken at various times and locations around the Great Barrier Reef [3]. Predictions take the form of a bounding box together with a confidence score for each identified starfish. An image may contain zero or more starfish.

Data contains training set photos of the form video_video_id/video_frame_number.jpg. [train/test].csv contains metadata for the images.

- video_id - ID number of the video the image was part of. The video ids are not meaningfully ordered
- video_frame - The frame number of the image within the video. Occasional gaps in the frame number from when the diver surfaced
- sequence - ID of a gap-free subset of a given video. The sequence ids are not meaningfully ordered
- sequence_frame - The frame number within a given sequence
- image_id - ID code for the image, in the format 'video_id-video_frame'
- annotations - The bounding boxes of any starfish detections in a string format. The bounding box is described by the pixel coordinate (x_min, y_min) of its upper left corner within the image together with its width and height in pixels

III. Related Works

A. Underwater multi-object detection

Underwater object detection comes with many challenges. Some of the main challenges of underwater perception identified by Rizzini et al. 2015 include complex setup, distortion in signals and light propagation introduced by the water medium, and higher device costs. Light propagation in underwater environments is affected by absorption and scattering, which strongly affects visual perception [4]. A number of researchers have used various methods to perform underwater object detection. Chen et al. 2020 approached the problem with Invert Multi-Class Adaboost with deep learning to reduce the influence of noise on their proposed SWIPENet (Sample-WeIghted hyPEr Network) [5]. Galceran et al 2012 proposed a method that dynamically takes into account the environmental characteristics of sensed sonar data for underwater object detection with forward-looking sonar imaging [6]. Han et al. 2020 utilized a deep CNN model with two schemes to modify the CNN structure: 1) a 1 x 1 convolution kernel on a 26 x26 feature map that is then layered with a downsampling layer that resizes the output to 13 x 13, and 2) downsampling layer is added first, then the convolution, where the result is added to the last output to achieve detection [7]. Wang et al 2019 employed YOLOv3 to detect and identify the type of fish in an aquarium in the lab [8].

B. Multiple Anchor Box Method

Anchor boxes have long been used in object detection. Zhong et al. 2020 proposed an optimized anchor box for object detection, where the anchor shapes are dynamically learned, allowing for automatic adaptability to data distribution and the network learning capability. According to their paper, the accuracy with optimized anchor shapes consistently outperforms the baseline by around 0.5 1.2% with different numbers of

anchor shapes [9]. Ke et al. 2020 proposed a Multiple Instance Learning approach, referred to as Multiple Anchor Learning, to learn a detection model. They construct anchor bags and select the most representative anchors from the bag [10].

C. Mixture density-based object detector

Yoo et al 2019 proposed the use of a mixture density object detector for multi-object detection by estimating bounding box distributions for the input image. Their comparison of the MDOD with other object detection methods such as GTA-based models and EfficientDet showed that MDOD has an improved performance without structural changes, or heuristic and complex processing [10].

IV. Exploratory Data Analysis

Our data analysis on the train data is shown in the table 1.

Name	**Value**
Total Number of Videos	3
Total Number of Sequences	20
Total number of Images	23501
Number of Images in Video 0	6708
Number of Images in Video 1	8232
Number of Images in Video 2	8561
Total Number of Images with no annotations	18582 (79.07% of images)
Total Number of Annotations	11898
Number of Total Annotations in Video 0	3065
Number of Total Annotations in Video 1	6384
Number of Total Annotations in Video 2	2449

TABLE I
EDA OF THE TRAIN DATA

An example of images with high number(15) of annotations is shown with the boxes drawn in Figure 1.

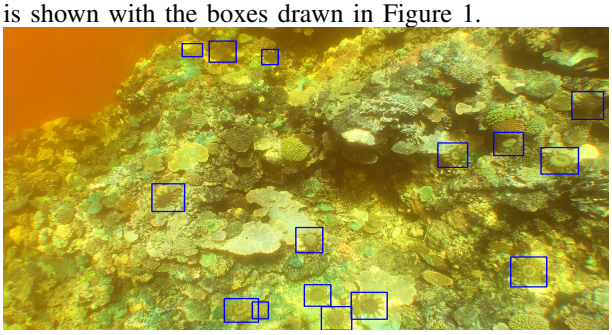

Fig 1: Annotated image 5778 in video 2

The given path of the COTS in the video sequences can be seen in the example in Figure 2. These paths trace the center of the boxes of the annotations.

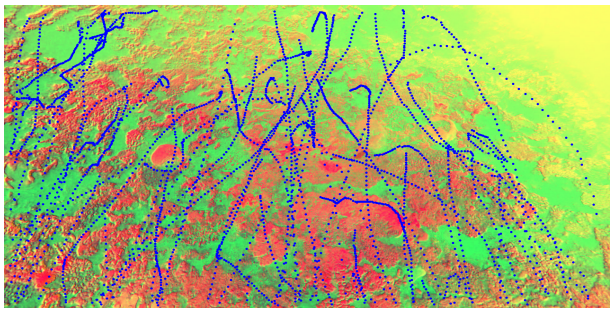

Fig 2: Starfish paths in video 0

V. Methodology

The basic aim is to utilise transfer learning to perform multi-object detection on the data. The network is trained to learn a variable number of bounding boxes around the crown-of-thorns starfish detected in the underwater images. The implementations chosen are the Yolov4 network proposed by Alexey Bochkovskiy et al. [11], Yolov3 proposed by Joseph Redmon et. al.[13] and FasterRCNN with anchor boxes proposed by Shaoqing Ren et al.[14]

A. Data Preparation

To feed our data to the Yolov4 and Yolov3 network the following data preparation was carried out:

- The JSON format annotations present in the CSV data file were parsed and converted to Yolov4 format. The Yolov4 format is as follows: class_id, xmin, ymin, width, height. In our scenario, we have only one class - COTS (Starfish) which we have mapped to the id 0.
- We perform an 80-20 train-validation split on the data to feed to the model.
- We resize the images to 416 x 416 to facilitate training the yolov4 network on the dataset.

1) Data Augmentation: We performed a few data augmentation techniques such as distortion, occlusion and augmentation with the following parameters.

- Saturation: 1.5, Exposure: 1.5, Hue: 0.1
- Mosaic: 1
- Jitter: 0.3 added to training data to prevent overfitting

B. Network Architecture for Yolov4

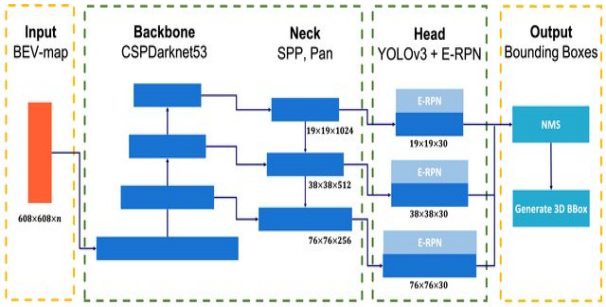

Fig 3: YoloV4 Network Architecture

C. Network Architecture for Yolov3

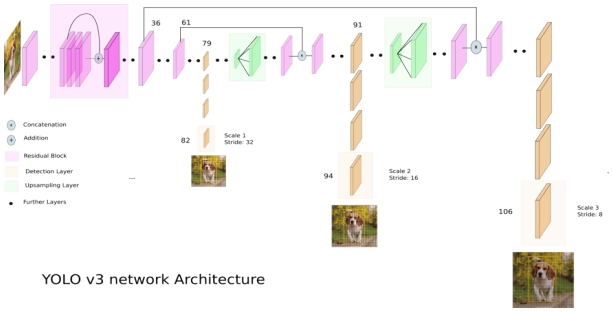

Fig 3: YoloV3 Network Architecture

D. Network Architecture for Faster RCNN

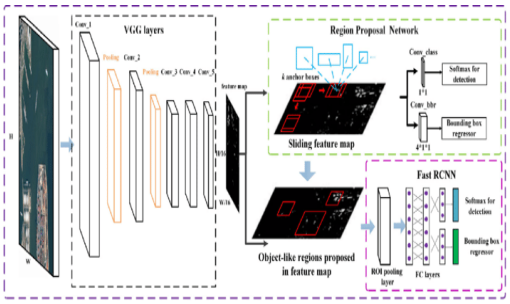

Fig 3: Faster RCNN Network Architecture

E. Minimum Average Precision - mAP

Precision and Recall is calculated using IoU value for a given IoU threshold. For example; If IoU threshold is 0.5, and the IoU value for a prediction is 0.7, then we classify the prediction as True Positive (TF). On the other hand, if IoU is 0.3, we classify it as False Positive (FP). Average Precision (AP) is the area under the precision-recall curve generated. The mean Average Precision or mAP score is calculated by taking the mean AP over all classes.

F. Network Training and Parameters for Yolov4

Since we are employing transfer learning we do not train from scratch. Instead we use pre-trained weights which have been trained upto 137 convolutional layers. This forms our baseline model.

Name	Value
batch	64
mini-batch/subdivision	32
filters	18
learning rate	0.005
decay	0.0005
momentum	0.949

TABLE II
PARAMETERS OF THE MODEL

1) Training Parameters:: The model is trained until average loss reaches 0.05 and the mAP stagnates at its highest value.

G. Network Training and Parameters for Yolov3

Since we are employing transfer learning we do not train from scratch. Instead we use pre-trained weights which have been trained upto 137 convolutional layers. This forms our baseline model.

Name	Value
batch	64
mini-batch/subdivision	64
width/height	704
learning rate	0.01
decay	0.0005
momentum	0.949

TABLE III
PARAMETERS OF THE MODEL

1) Training Parameters:: The model is trained until average loss reaches 0.05 and the mAP stagnates at its highest value.

H. Network Training and Parameters for FasterRCNN

Since we are employing transfer learning we do not train from scratch. Instead we use pre-trained weights of faster-rcnn_resnet50_fpn. The network is trained with the following parameters.

1) Training Parameters:: The model is trained until average loss reaches 0.06 and the mAP stagnates at its highest value.

Name	Value
batch	8
learning rate	0.005
decay	0.0005
momentum	0.9

TABLE IV
PARAMETERS OF THE MODEL

VI. GitHub Code

The code repository can be found in the following link.

https://github.com/srikg-msai22/DL-Group-11/

VII. Results and Future Work

The mAP and average loss have significantly better values than yolo models when using the FasterRCNN with anchor boxes. The next steps are focused on improving the accuracy of the Faster RCNN and the Yolov3/Yolov4 model. The yolo predictors are a 100x faster than the FasterRCNN and thus better suited for real time object detection for this specific problem statement. Thus, the next steps involve further improvement and running the model on a test dataset which is provided in the form of a video to check for prediction latency vs accuracy tradeoff.

Model	Run	mAP	Average_Loss
Yolov4	1	21.215%	0.00046
Yolov4	2	27.432%	0.00032
Yolov4	3	51.551%	0.00025
Yolov3	1	16.334%	0.00054
Yolov3	2	32.832%	0.00034
FasterRCNN	1	46.328%	0.06000
FasterRCNN*	2	54.820%	0.03400

A. Results

B. Future Work

- Analyzing the finding the prediction speeds for the YOLOv4 and Faster R-CNN models
- Building an ensemble of YOLOv4 and Faster R-CNN models
- Handling Low-Annotated Images
  - Two-Class Sets Approach: First building a model to classify an image to say if there is a COTS or if there isn't. Then use one of our models/build new models on the annotated images on the ones with possible annotations.
  - Annotated Image Semi-Supervised Approach: Build a model on the annotated images. Build another model to automatically annotate any new images and penalize the model towards checking if there are annotations or not.

REFERENCES

[1] S. Xie, J. Chen, J. Luo, P. Xie, and W. Tang, "Detection and tracking of underwater object based on forward-scan sonar," Intelligent Robotics and Applications, pp. 341–347, 2012.

[2] M. Hall and S. Cummins, "How scaring starfish could help to save the Great Barrier Reef," The Conversation, 03-Feb-2022. [Online]. Available: https://theconversation.com/how-scaring-starfish-could-help-to-save-the-great-barrier-reef-36759. [Accessed: 15-Feb-2022].

[3] J. Liu et al., "The CSIRO Crown-of-Thorn Starfish Detection Dataset", arXiv [cs.CV]. 2021.

[4] D. L. Rizzini, F. Kallasi, F. Oleari, and S. Caselli, "Investigation of vision-based underwater object detection with multiple datasets," International Journal of Advanced Robotic Systems, vol. 12, no. 6, p. 77, 2015.

[5] L. Chen, Z. Liu, L. Tong, Z. Jiang, S. Wang, J. Dong, and H. Zhou, "Underwater object detection using invert multi-class Adaboost with deep learning," 2020 International Joint Conference on Neural Networks (IJCNN), 2020.

[6] E. Galceran, V. Djapic, M. Carreras, and D. P. Williams, "A real-time underwater object detection algorithm for multi-beam forward looking sonar," IFAC Proceedings Volumes, vol. 45, no. 5, pp. 306–311, 2012.

[7] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater Image Processing and object detection based on deep CNN method," Journal of Sensors, vol. 2020, pp. 1–20, 2020.

[8] C.-C. Wang, H. Samani, and C.-Y. Yang, "Object detection with deep learning for underwater environment," 2019 4th International Conference on Information Technology Research (ICITR), 2019.

[9] Y. Zhong, J. Wang, J. Peng, and L. Zhang, "Anchor Box Optimization for Object Detection," 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 2020.

[10] W. Ke, T. Zhang, Z. Huang, Q. Ye, J. Liu, and D. Huang, "Multiple anchor learning for visual object detection," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[11] J. Yoo, H. Lee, I. Chung, G. Seo, and N. Kwak, "Training Multi-Object Detector by Estimating Bounding Box Distribution for Input Image," Computing Research Repository, vol abs/1911.12721, 2019.

[12] A. Bochkovskiy, C.-Y. Wang, en H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection", Computing Research Repository, vol abs/2004.10934, 2020.

[13] Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", Computer Vision and Pattern Recognition

[14] Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Computer Vision and Pattern Recognition