Underwater Object Detection using Faster RCNN

Supervised by Prof. Prithwijit Guha

Vignesh P(214156010) vigneshp@iitg.ac.in
Indian Institute of Technology Guwahati

Abstract

The Brackish dataset, an underwater dataset with 6 different categories of videos is taken and images are extracted and divided into train test and validation splits. A pretrained model of Faster RCNN with ResNeXT as RPN is taken from detectron2 and fine tuned on the brackish dataset for limited number of epochs considering computation power available. The results are evaluated on the test set and different Average Precision Values are noted. Using the results, in this project we conclude that pretraining on ordinary non - underwater dataset does not help on Underwater Object Detection Task.

1 Introduction

Underwater Object detection is one of the difficult object detection task because underwater circumstances results in poor image quality. The presence of suspended particles, blurry background, scattering and absorption of light makes the detection of small objects difficult.The present state of the art object detection models use Convolutional Neural Network(CNN). There are two different categories of CNN based object detectors namely single stage detectors and two stage detectors.In two stage object detectors, one model is used to extract region of objects, and a second model is used to classify and further refine the localization of objects. In this project I tried to implement a two stage object detector using the detectron2 library introduced by Meta.

2 Model

The models used in the project are **Faster RCNN**(Ren et al., 2015) with **ResNeXt** (Xie et al., 2016) models. The faster RCNN is the upgraded version of RCNN which is very slow multistage object detector. The faster RCNN has region proposal network (RPN) which is a fully convolutional network that generates proposals with various scales and aspect ratios. Faster RCNN is the combination of Region Proposal network(RPN) and Fast RCNN. ResNeXT is used as Region Proposal Network(RPN) in the model. The Fast R-CNN detector consists of a CNN backbone, an ROI pooling layer and fully connected layers followed by two sibling branches for classification and bounding box regression. ResNeXt is a network inspired by Inception's split-transform-merge scheme that aims to overcome the latter's impediments by having matching branches and blocks throughout the network.

Figure 1: Example image of Brackish Dataset with annotations and label

3 DataSet

The dataset used in the project is **Brackish Dataset** (Pedersen et al., 2019). A good deep learning model or object detection model requires large amount of data. The Brackish Dataset is a publicly available underwater dataset containing bounding box annotated sequences of images containing big fish, small fish, starfish, shrimps, jellyfish, and crabs captured in a brackish strait with varying visibility. The videos were categorized based on the main activity of the respective video and subsequently manually annotated with a bounding box annotation tool resulting in a total of 14,518 frames with 25,613 annotations.

4 Method

The pretrained Faster RCNN model with weights obtained from COCO Dataset is used in the project. The model is fine tuned with Brackish Dataset and evaluated on test data and the results are noted. The batch size taken is 64 frames for each pass. Out of 14518 frames, 11603 frames are used for training and 1467 frames are used for validation and 1468 frames used for test. The model is finetuned for 750 epochs. The model is evaluated for each 250 iterations and Average Precision scores are calculated for 6 different classes and different scales. The model weights are updated without freezing the bottom layers.The detectron2 library is used for loading and training the Faster RCNN Model.

5 Evaluation Metrics

The evaluation metrics used in the project are mean average precision (Hu et al., 2017), Average precision 50, Average precision 75, Average Precision Large(APL), Average Precision Medium(APM), Average Precision Small(APS). Average Precision is Area under the precision recall curve. The most common method used for calculation of the area of the curve is using 11 point interpolation (i.e) Taking 0:0.1:10 values of recall and calculating precision at each point and taking the mean of all the precisions. AP50 means the threshold value used for IOU is 0.5. The IOU means Intersection Over Union (Rezatofighi et al., 2019) which is calculated by taking ratio of the intersection and the union of predicted and actual bounding boxes. The AP 75 means threshold for IOU is 0.75. The APL, APM, and APS are calculated by average precision values at different scales.

6 Results

Different Average Precision values for the test data

Metric	Classes	Score
AP	crab	16.82
	fish	49.03
	jellyfish	39.09
AP50	All	30.425
AP75	All	18.53
APL	All	14.97
APM	All	16.28
APS	All	9.74

The Average Precision Scores are low because the model is trained only for less number of iterations and the pretrained model is trained on completely different dataset which has less noise. This result partially confirms that pretraining on other general dataset does not help for underwater object detection(Jesus et al., 2022).

References

Ronghang Hu, Piotr Dollár, Kaiming He, Trevor Darrell, and Ross B. Girshick. 2017. Learning to segment every thing. *CoRR*, abs/1711.10370.

Andre Jesus, Claudio Zito, Claudio Tortorici, Eloy Roura, and Giulia De Masi. 2022. Underwater object classification and detection: first results and open challenges. *CoRR*, abs/2201.00977.

Malte Pedersen, Joakim Haurum, Rikke Gade, Thomas Moeslund, and Niels Madsen. 2019. Detection of marine animals in a new underwater dataset with varying visibility.

Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2015. Faster R-CNN: towards real-time object detection with region proposal networks. volume abs/1506.01497.

Seyed Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian D. Reid, and Silvio Savarese. 2019. Generalized intersection over union: A metric and A loss for bounding box regression. *CoRR*, abs/1902.09630.

Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2016. Aggregated residual transformations for deep neural networks. volume abs/1611.05431.