

Underwater Object Detection using Tensorflow

Avinash Mahavarkar^{1,3,*}, Ritika Kadwadkar^{2,**}, Sneha Maurya^{3,***}, and Mrs. Smitha Raveendran^{4,****}

¹avinashmahavarkar@gmail.com

²ritika.kadwadkar1525@gmail.com

³snehamaurya1610@gmail.com

⁴smitha.raveendran@rait.ac.in

Abstract. Object Detection is a popular technology that detects instances within an image. In order to eliminate the barriers in Computer Vision technology due to the dissolution of the BGR(Blue-Green-Red) constituents with the increase in depth, it has been a necessity that the accuracy and efficiency of detecting any object underwater is optimum. In this article, we conduct Underwater Object Detection using Machine Learning through Tensorflow and Image Processing along with Faster R-CNN (Regions with Convolution Neural Network) as an algorithm for implementation. A suitable environment will be created so that Machine Learning algorithm will be used to train different images of the object. Open source Computer Vision has various functions which can be used for the image processing needs when an image is captured.

1 Introduction

Image processing is used to operate digital images using various algorithms. In image processing the image is in the form of two dimensional matrix on which certain procedures or algorithms are performed in order to get a favourable result. Some of these algorithms are contrast enhancement, dithering and half toning, feature detection etc. Object Detection is one of the primary applications of image processing. It is a procedure in which any stationary object or a moving object can be detected in any real time instance, an image or a video .For training the data sets machine learning plays a very important role. Machine Learning is a field within Artificial Intelligence in which the machine trains itself while adapting it to the changing environment. The trained dataset are the ones which consists of the images of objects or material which we have to detect and the tested datasets are different from the trained ones and are used for checking the accuracy of the system. 80 percent of images are used to train whereas 20 percent are used for testing. In this project, we have used Tensorflow which is an Object Detection API which uses Faster R-CNN as the machine learning algorithm . Within Tensorflow the pixels of images are in the form of matrix and operations are performed on them in order to obtain required results. There are various pre-trained models for detecting objects they are R-CNN, Mask R-CNN, YOLO (You Only Live Once) etc. However these pre-trained models have their own specified data sets, in order to make use of our self trained data we have to create a suitable environment after fine-tuning the pre-trained model. There

are various applications of Tensorflow which are focused on deep learning and training. Tensorflow makes use of Python programming, hence there is utilization of respective Python version.

2 Related Work

Exploration for the need of object detection began during the 1960's. There are various traditional methods which were used for object detection such as using haar classifiers, Support Vector Machine (SVM) along with Histogram of Oriented Gradients (HOG), Oriented FAST and rotated BRIEF (ORB) feature matching etc. However these techniques did not give accurate results in case of underwater object detection. Thus the following object detection and recognition techniques proposed by different authors gave a tremendous boost within this field. For an image to be recognized correctly its edges are taken into consideration so that a correct result is obtained. Guobo Xie and Wen Lu [18] explains as to why is edge detection important in case of image processing. They have used opencv functions to detect edges of an image by first converting it to gray scale, then by adding a filter, thresholding the obtained result thereby the output only displays the edges of an image. Apart from edge matching feature matching is another important aspect of detecting an object in any real time instance or an image. Lie et al. [15] have proposed the SURF method for object detection in which using the algorithm which is based on feature matching technique the edge features, key points and the color resemblance of an object is been detected. However the SIFT AND SURF algorithms are patented and to use them a proper permission and license has to be issued. An alternative for these meth-

*e-mail: Mailaddressforfirstauthor

**e-mail: Mailaddressforsecondauthorifnecessary

***e-mail: Mailaddressforlastauthorifnecessary

****e-mail: Mailaddressforlastauthorifnecessary

ods is the ORB method. Ethan Rublee et al.[16] proposes in his paper the ORB (Oriented fast and Rotated Brief) feature which is an open source algorithm and two magnitudes faster as compared to SIFT and SURF however the efficiency of is compromised in this case.

Rahul Dutt Sharma et al. [14] explains in his work about the subtraction of background within a real time object so as to find the correct result. It uses a three frame differencing which uses a AND operator and an OR operator this is because only the AND operator cannot detect the objects with a moving frame as a result an OR operator is used so as to detect the moving objects unlike the two frame differencing which uses only AND operator.

Xiu Li et al[8] proposed in their research paper about the detection of various fish species underwater. LifeCLEF Fish dataset along with a video dataset from Fish4knowledge were used in training the data and the machine learning algorithm used was Fast R-CNN.

J. Kim and S. Yu[9] designed a SONAR based ROV which used a convolution based neural network in order to detect object at a distance. It comprised of a main AUV with a small ROV to travel to distant places underwater.

Hai Huang et al [10] proposed underwater marine detection using the Faster R-CNN algorithm which has a greater mAP (Mean Average Precision) as compared to other CNN algorithms they took into account data augmentation as Faster R-CNN requires mass labelled samples of images.

R. Phadnis et al[10] have introduced a custom designed algorithm which would keep a track of objects within a specific diameter so as to prevent the wastage of time if any object is lost.

Zhe Chan et al.[13] have explained the use of an artificial collimated light source so as to avoid various problems of object detection underwater. A collimated light source have parallel rays of electromagnetic radiation and thus has lowest degree of dispersion. Its aim is to focus the artificial light on the object directly such that it collimates directly on the object.

Jonathan Huang et al. [17] in the paper proposed mentions about all the tensorflow models or machine learning algorithms which are available for object detection underwater. It explains about SSD, YOLO, CNN, RCNN Faster RCNN etc. The detailed explanation of all these algorithms their advantages and disadvantages are mentioned, their accuracy percentages, their working, etc. are all included in their publication.

Our project involves underwater object detection using Tensorflow with Faster R-CNN algorithm model as our building block. The survey based on the Faster RCNN algorithm is from the Shaoqing Ren et al [6] publication in which a detailed explanation of the working of Faster RCNN has been mentioned along with the important aspects of the algorithm and is proven to be the most accurate among the different convolution neural networks.

Sai, B Sasikala, T.[11] have proposed a paper based on object detection and their count within an image using the tensorflow object detection API. It used the Faster R-CNN algorithm same as proposed in this paper. However in this

paper we have taken into account the underwater object detection concept.

3 Object Recognition

3.1 Phases of object recognition

The object detection and recognition algorithms are basically used in order to find the presence of an object, its movement and orientation within an image or real time instance. For an object to be detected and recognised the algorithm should be able to determine the presence of an object or multiple objects. Object recognition includes the following phases: preprocessing, feature extraction, feature selection, modeling, matching, and positioning [1]. The various algorithms based on feature matching are ORB, Speeded Up Robust Features (SIFT), Scale Invariant Feature Transform (SURF) etc. However these techniques have low efficiency in characterising various objects since it uses the features of an object to detect its presence. Therefore many object detection algorithms were introduced in order to overcome the drawbacks of these techniques and they are CNN, R-CNN, YOLO etc. Every algorithm within itself has its pros and cons. The most commonly used object detection and recognition algorithms are YOLO and Faster R-CNN.

3.2 Tensorflow and faster R-CNN

In this paper, underwater object detection using Tensorflow in order to train the system and Faster R-CNN as a machine learning algorithm for detection and implementation has been proposed.

3.2.1 Faster R-CNN

The basis of Faster R-CNN starts from understanding as to what is CNN (Convolution Neural Network). CNN has flourished within the applications of deep learning in case of image or video processing. A CNN consists of 4 layers - Convolution layer, Relu layer, Pooling layer and Flattening layer. The convolution layer consists of a filter used for navigation, this filter keeps fluttering over an image and makes calculation of every pixel. Relu layer has a Relu activation function used for eliminating the negative values and rounding them off to zero. Within the pooling layer, using feature map size reduction, only important parameters are considered. The flattening layer is used for the conversion of matrix into a single vector. The Convolutional and the Pooling Layer, together form the i-th layer of a CNN. The layers could be increased depending upon as to how much efficiency we need and how complex the systems are however the cost of computation increases accordingly. However the CNN algorithm is an image classification algorithm and in order to detect objects and recognize them a few modifications have to be done such as drawing the box around certain objects detected in front of a background etc. We simply cannot use the CNN layers followed by an FC layer to do the same as the number

of objects at the output are not bounded to a specific number. Girshick et al. [3] introduced a region-based CNN (R-CNN) for object detection. In this method the image is divided in a certain number of recommended regions. The ubiquitous procedure of the working of an R-CNN algorithm is from the gathered input images, region proposals are extracted. Every proposal is thus elapsed through the CNN to compute the features so that the regions can be correctly classified. However the training time in this case goes until 83 hours. In order to overcome this problem, Fast R-CNN was introduced. The only difference between R-CNN and Fast R-CNN is that instead of dividing the image into regions initially it is divided after applying CNN. The training time required for this algorithm is 9 hours. The faster R-CNN requires less training time that is about 4 hours, hence it is more efficient and accurate as compared to the previous two generations of CNN based algorithms. In figure 1, the steps which are followed as per the faster R-CNN algorithm to find objects in an image are mentioned. The basic building block of a faster R-CNN

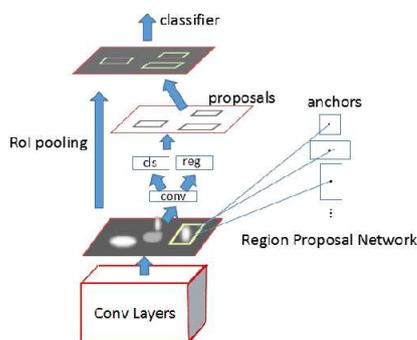


Figure 1. R-CNN Algorithm[6]

network is a Regional proposal Network (RPN) unlike R-CNN and Fast R-CNN which uses a selective search algorithm which makes the algorithm slower. An RPN can be defined as a platform that proposes the convolution features of any image or an instance within a video. The core of RPN is to detect different sizes of objects with different sizes of anchors. [7] It takes into account the boundary features of an object and by custom training the images RPN could be used in order to increase the efficiency of the output as it generates a supreme quality region of proposal with the increasing number of images. There are two steps which are in the Faster R-CNN they are: shared bottom convolutional layers, a region proposal network (RPN) and a region-of-interest (ROI) based [4] classifier. In this case the complete image is fed to the Convolution neural network. This network generates a convolution feature map which is comprised of the bottom convolution layers. The region proposal is predicted within an image. Based on this, RPN creates a respective object proposal, after which the Region of interest classifier estimates the label from a feature point which is gathered by ROI-pooling. Thus the classes of the object is recognized with respect to the trained images in the dataset. As you can see from the results table it is clear that the Faster R-CNN algorithm has the lowest test time i.e 0.2 seconds this is because of

the creation of RPN instead of using the selective search algorithm. The accuracy of the system is also increased.

3.2.2 Tensorflow

Tensorflow is an Object detection API which has an open source library for machine learning. It is used to detect multiple objects in real-time video streams.[5] The system is borne from real-world experience in conducting research more than one hundred machine learning projects throughout a wide range of Google products. [2] In this project, we have utilized the Tensorflow 1.14 version. TensorFlow uses dataflow graphs to represent computation, shared state, and the operations that mutate that state. [2] It can be used as an interface between the training models and the machine learning algorithms used for object detection in our case. There are various models available in tensorflow they are keras, resnet, mobilenet, inception and delf. These pre-trained models utilizing algorithms like R-CNN, Faster R-CNN, Fast R-CNN, YOLO etc. We can also create our own model and train them for object detection. The preliminary dataset which tensorflow uses is COCO (Common Objects in Context) which is compatible with all types of GPU's, handsets, CPU's, TPU's. Along with these tensorflow also supports transfer learning which helps us define our own classes in pretrained models thus eliminating the use of large datasets. However we can train a custom based database also, Tensorflow avails the feature of a custom trained model. The below flowchart in figure 2 represents the step by procedure of using a pre-trained models for detection of objects underwater

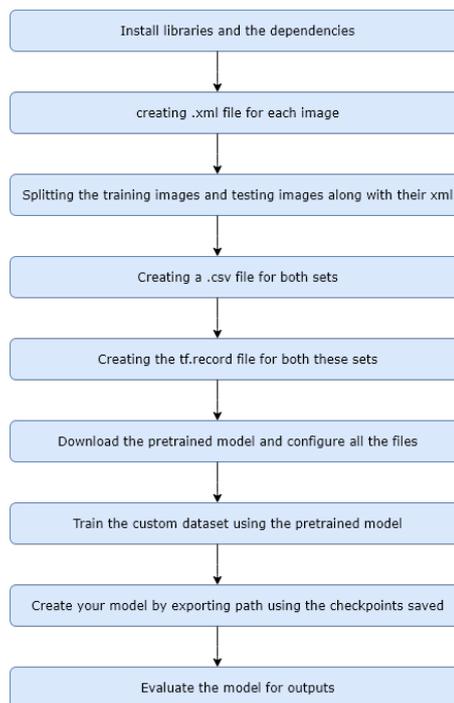


Figure 2. Tensorflow object detection flowchart to train models[6]

4 Experimental Results

4.1 Underwater Setup

We have used a predefined dataset to detect an object underwater. In Tensorflow we initiate a call regarding whichever dataset is required. We have used COCO dataset for our observations. This particular dataset consists of over 80 classes, 80000 training images and 40000 validation images which keep getting updated automatically. At Ramrao Adik Institute of Technology, we have an underwater tank setup which is used for research purposes. We have used this setup for performing our experiment. Figure 3 shows the image of the tank setup. In the tank, a camera is connected to the system. The camera used is a Super Hi Vision 2 Million Pixels CMOS camera. The Camera is flexible and is waterproof with a 8.5mm diameter. Due to its small size we can inspect areas under water, gaps and holes. The camera captures real time images with objects detected in it. The system used has Intel i5-8265U 1.8GHz processor with an 8GB installed RAM. One of the important aspect of our project is using the correct GPU. Our system has NVIDIA GEFORCE MX250, with the help of this version of GPU object detection was efficiently carried out.



Figure 3. Tank setup

Further, we put objects like Toothbrush, Cup, Bottle, Clock, etc in the tank. To test the efficiency of this experiment, we performed this in unclean water and where the source of light was less. As the water was impure and illumination was minimum, objects were still recognized efficiently. Whenever an object is present in the proximity of the camera, a bounding box appears on the screen and it also indicates the extent of feature that is matched with the trained images in percentage.

4.2 Results

So as we can see the system is able to execute the task of object detection and recognition using tensorflow and machine learning. Faster R-CNN was used in this project which in return increased the accuracy of the system. These are the results which we procured as shown in figure 4. Even in filthy water the camera is able to detect objects with an accuracy of percentages shown in the image result. The comparison of the results with other CNN algorithm are mentioned in table 1 which relates the accuracy of the system with accuracy of the training images.

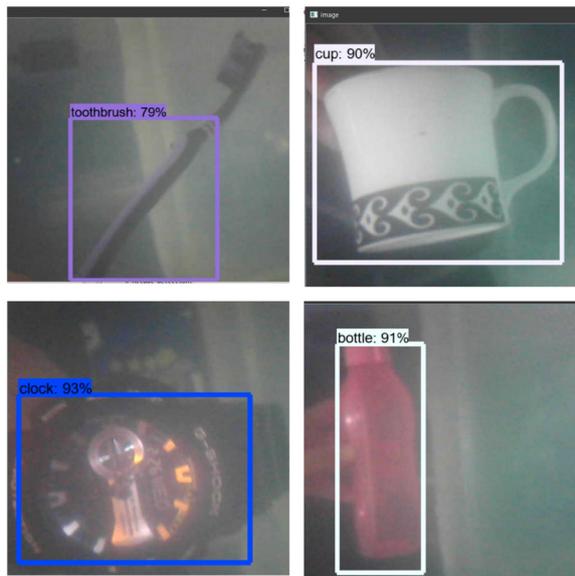


Figure 4. Objects detected under water

5 Conclusion

In this paper, we proposed the Faster R-CNN algorithm with Tensorflow for the underwater object detection. The algorithm used in our case was the Faster R-CNN algorithm which uses Regional proposal regions as the basic building block because of which the test time of detection is reduced since an RPN takes into account the boundary features of an object. By using a trained dataset it was clear that the accuracy of the system was optimal. However by using a custom trained dataset the accuracy and efficiency can be increased further more as the RPN is capable of generating a supreme quality region of proposal. Because of various effects such as dehazing, scattering of light and RGB color disappearance it becomes more and more difficult to detect objects. However by using such machine learning algorithms and proper training of images within the datasets these problems can be resolved to great extent.

Table 1. Results

Method	Test time	Training accuracy	Op accuracy
CNN	0.452 s	98.19 %	96.54 %
R-CNN	49 s	96.72 %	94.67 %
Fast R-CNN	2.3 s	97.98 %	95.68 %
Faster R-CNN	0.2 s	97.65 %	96.31 %

6 Future Work

The improvement we are trying to achieve in this project is that we are testing to add our own custom trained dataset to the model rather than using the default one. We are facing difficulties related to compatibility of the system and the various packages installed for tensorflow. After successfully attaining the same we will be able to detect the

objects which we train leading to the increased efficiency of the system.

References

- [1] Research on underwater object recognition based on YOLOv3 Hongbo Yang¹ • Ping Liu¹ • YuZhen Hu¹ • JingNan Fu¹
- [2] TensorFlow: A System for Large-Scale Machine Learning Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng, Google Brain
- [3] R. B. Girshick, Donahue, T. Darrel, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In CVPR, pages 580–587, 2014
- [4] Domain Adaptive Faster R-CNN for Object Detection in the Wild Yuhua Chen¹ Wen Li¹ Christos Sakaridis¹ Dengxin Dai¹ Luc Van Gool^{1,2} ¹Computer Vision Lab, ETH Zurich ²VISICS, ESAT/PSI, KU Leuven
- [5] R. Phadnis, J. Mishra and S. Bendale, "Objects Talk - Object Detection and Pattern Tracking Using TensorFlow," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1216-1219.
- [6] Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks Shaoqing Ren Kaiming He Ross Girshick Jian Sun.
- [7] X. Chen, Qiang Guo, Shuohao Li and Jun Zhang, "Holistic Vertical Regional Proposal Network for scene text detection," 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, 2017, pp. 72-77.
- [8] Fast Accurate Fish Detection and Recognition of Underwater Images with Fast R-CNN Xiu Li, Min Shang, Hongwei Qin, Liansheng Chen.
- [9] J. Kim and S. Yu, "Convolutional neural network-based real-time ROV detection using forward-looking sonar image," 2016 IEEE/OES Autonomous Underwater Vehicles (AUV), Tokyo, 2016, pp. 396-400.
- [10] Faster R-CNN for marine organisms detection and recognition using data augmentation Hai Huang, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi, Ai-Yun Zang.
- [11] Sai, B Sasikala, T.. (2019). Object Detection and Count of Objects in Image using Tensor Flow Object Detection API, 542-546.
- [12] R. Phadnis, J. Mishra and S. Bendale, "Objects Talk - Object Detection and Pattern Tracking Using TensorFlow," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1216-1219.
- [13] Underwater Object Segmentation Based on Optical Features Zhe Chen¹ ID, Zhen Zhang¹, Yang Bu², Fengzhao Dai², Tanghuai Fan³ and Huibin Wang^{1,*}
- [14] Optimized Dynamic Background Subtraction Technique for Moving Object Detection and Tracking 1 Rahul Dutt Sharma, 2 Shubh Lakshmi Agrwal, 3 Sandeep K. Gupta, 4 Anil Prajapati .
- [15] Learning SURF Cascade for Fast and Accurate Object Detection Jianguo Li, Yimin Zhang
- [16] ORB: an efficient alternative to SIFT or SURF Ethan Rublee Vincent Rabaud Kurt Konolige Gary Bradski
- [17] Speed/accuracy trade-offs for modern convolutional object detectors Jonathan Huang Vivek Rathod Chen Sun Menglong Zhu Anoop Korattikara Alireza Fathi Ian Fischer Zbigniew Wojna Yang Song Sergio Guadarrama
- [18] TensorFlow: A system for large-scale machine learning, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng