Global Journal of Engineering and Technology [GJET].
ISSN: 2583-3359 (Online)
Frequency: Monthly
Published By GSAR Publishers
Journal Homepage Link- https://gsarpublishers.com/journal-gjet-home/

Comparative analysis of YOLOv5 and YOLOv7 forUnderwater Object Detection Cases

BY

Moayed D. Daneshyari[1], Sannihita N Yalamanchili[2]

[1,2]Department of Computer Science, California State University, East Bay 25800 Carlos Bee Blvd, Hayward, CA 94542, USA

Article History

Received: 01/12/2024
Accepted: 05/12/2024
Published: 07/12/2024

Vol – 3 Issue – 12

PP: - 07-11

Abstract

This research study intends to evaluate and compare the performance of YOLOv5 and YOLOv7 utilizing the trashcan and brackish datasets in order to design an optimal underwater object recognition system for the ROV BlueRov2. Experimental settings were created to assess how well these algorithms performed with various equipment arrangements, particularly in murky environments. Using YOLOv5 and YOLOv7, a lightweight object identification approach was presented to overcome the difficulties in underwater object detection, such as low visibility, color bias, and small targets. This approach attained a high mean average precision (mAP). Moreover, it shows how to locate objects in murky waters quickly and accurately. Overall, this study offers information on how to create underwater object identification algorithms that are optimized, which can increase the effectiveness and efficiency of ROV systems and help to lessen the environmental impact of marine garbage or help in the research of the marine environment.

1. INTRODUCTION

There are many different types of aquatic life in the marine environment, which is a dynamic and complex ecosystem. But as marine debris pollution increases, it has become a serious threat to this environment. Marine debris includes fishing gear, clothing, wood, and plastic. We need better ways of managing this issue.

The marine sector has already made extensive use of the technologies like computer vision and machine learning [2]. These technologies make it feasible for AUVs to effectively locate and remove marine debris by detecting and recognizing things from digital photos or movies acquired bycameras. The paucity of data in the maritime environment, the constantly changing environment, and the difficulty in identifying decaying items like garbage are only a few of the difficulties with employing computer vision for AUVs [2]. The volume and variety of datasets can be increased using picture improvement and augmentation techniques, and neural network scientists must create new and improvednetworks for computer vision. One of the most sophisticated neural networks now in use is YOLOv5, which works in real-time and can recognize objects much more quickly than other networks like Region Convolutional Neural Network (R-CNN) [3]. This makes it perfect for AUVs. However, YOLOv5's accuracy in classifying marine trash and marine life has not been assessed in the literature as of yet, thereforeit is still unclear whether it is suitable for deployment on AUVs [3].

AUVs using computer vision technology have a wide range of possible uses in the marine environment, from pelagic fisheries and underwater archaeology to the researchof marine ecosystems and the conservation of marine organisms [2]. However, because of their sluggish processing speed and enormous model sizes, deep convolutional neural networks (DCNN)-based object detectors currently in use areineffective for use in underwater contexts. There are additional difficulties with underwater imaging, such as highnoise, poor visibility, blurred edges, poor contrast, and color cast, which can lower the quality of the photographs [4]. Also making detection more difficult are the frequent clustering and tiny size of underwater targets [4]. To overcome these difficulties and enable efficient underwater object identification, lightweight detectors that are accurate and tiny in size are needed [4].

This study examines the YOLOv5 neural network model's performance and accuracy in classifying marine detritus and marine life [3]. Our goal is to assess whether YOLOv5 is appropriate for deployment on AUVs by contrasting our findings with prior research. We'll also look into the advancement of portable detectors for finding objects underwater and evaluate how well they work in the marine environment [2, 4]. In the end, this paper advances our knowledge of the marine ecosystem and aids in thecreation of more practical and efficient solutions to the growing issue of marine trash [1].

*Corresponding Author: Moayed D. Daneshyari

II. METHODOLOGY

A. Determining the Network Architecture

The marine sector has already made extensive use of the technologies known as computer vision and machinelearning. These technologies make it feasible for AUVs to effectively locate and remove marine debris by detecting andrecognizing things from digital photos or movies acquired bycameras. The paucity of data in the maritime environment, the constantly changing environment, and the difficulty in identifying decaying items like garbage are only a few of the difficulties with employing computer vision for AUVs. The volume and variety of datasets can be increased using pictureimprovement and augmentation techniques, and neural network scientists must create new and improved networks for computer vision.

The You Only Look Once (YOLO) family of object detection models has released its fifth version, known as YOLOv5. It was made available by Ultralytics in June 2020 and is an enhancement over YOLOv4[8]. In contrast to multi-stage detectors like Faster R-CNN, which require many passes, YOLOv5 is based on a single-shot detector (SSD) architecture, which implies that it only has to make one pass over an image to detect objects.

To extract features from the input image, YOLOv5 uses a backbone network called CSPDarknet. Based on the Darknet architecture used in earlier iterations of YOLO, CSPDarknet enhances the network's information flow by including a cross-stage partial network (CSP) module. Scaled-YOLOv4 is a brand-new anchor-free bounding box prediction technique that YOLOv5 also makes use of to boost object recognition precision.

It has been demonstrated that YOLOv5 performs at the cutting edge on a variety of object detection benchmarks, including COCO and Open Images. Additionally, it has been used in various applications, including autonomous driving, drone-based surveillance, and traffic monitoring. The quickness of YOLOv5 is one of its main advantages. YOLOv5s is appropriate for real-time applications. Furthermore, YOLOv5 has a minimal memory footprint, allowing for deployment on edge devices with constrained processing power.

Due to its speed, accuracy, and usability, YOLOv5 has grown to be a popular option for object identification tasks and marks an important improvement over earlier versions of the YOLO model.

B. Datasets

The two main datasets we train on are the TrashCan Dataset and the Brackish Dataset where the Trashcan Dates consists of 7212 photos and the Brackish Underwater Dataset contains 14,674 photos.

- ***Trashcan Dataset:***

The TrashCan dataset[6] and the Brackish dataset[6]wereboth used in the study report.

There are now 7,212 photos in the TrashCan dataset[6], which is a collection of tagged images. These pictures show rubbish, remotely operated vehicles (ROVs), and a diverse range of underwater plants and animals. Instance segmentation annotations, which are bitmaps with a mask identifying which pixels in the image contain each object, arethe format used for the annotations in this dataset. The J-EDI (JAMSTEC E-Library of Deep-sea Images) collection, which is managed by the Japan Agency of Marine-Earth Science and Technology (JAMSTEC), is the source of the imagery used in the TrashCan dataset[10, 6]. Videos from ROVs run by JAMSTEC since 1982, primarily in the Sea of Japan, are included in this dataset. The table shows all the classes and annotations for the classes in Fig 1.

The dataset comes in two variations called TrashCan-Material and TrashCan-Instance, which represent various configurations of the object class. The TrashCan dataset's[6] ultimate purpose is to create reliable and effective trash identification techniques that can be used by onboard robots. The TrashCan dataset[6] is, to the best of our knowledge, the first instance-segmentation annotated dataset of underwater rubbish. Previous datasets have been developed that feature bounding box-level annotations of trash in marine habitats.

By making this information publicly available, the marine robotics community can advance research on this difficult issue and get a step closer to finding a solution to the pressing issue of autonomous trash detection and removal. As a result, the TrashCan dataset[6] is a useful tool for scientists and programmers studying marine robotics andautonomous waste detection. The inference on images isshown in Fig. 3.

- ***Brackish Underwater Dataset:***

For scientists studying underwater robotics and computer vision, the Brackish Underwater Dataset[7] is an invaluable tool. The underwater objects in this collection were photographed in brackish water, which is a combination of freshwater and saltwater. The collection consists of 14,674 photos with annotations at the object level for 10 object classifications. The table shows all the classes and annotations for the classes in Fig 2.

The photos in the Brackish Underwater Dataset[7] were taken in the American Chesapeake Bay estuary with an underwater camera attached to a remotely operated vehicle (ROV). The high-definition camera and positioning system on the ROV allowed for precise object labeling.

The Brackish Underwater Dataset[7] contains classifications of objects such as fish, crabs, oysters, seagrass, and other kinds of trash. Bounding boxes are used as annotations in the dataset since they show the position and dimensions of each object in the image. Researchers developing item identification and recognition systems for aquatic situations will find this dataset to be especially helpful. The Brackish Underwater Dataset[7] is a useful tool for developing and testing computer vision models because of the wide range of item classes it contains and the sheer volume of annotated photos.

The Brackish Underwater Dataset[7] is an important addition to the pool of publicly accessible underwater datasets and can

*Corresponding Author: Moayed D. Daneshyari

help with the creation of underwater robotic systems that are more precise and effective. The inference on images is shown in Fig. 4.

Class	***Annotations***
rov	3317
trash_unknown_instance	2756
trash_bag	908
animal_fish	764
trash_container	510
plant	507
trash_can	459
animal_starfish	398
animal_eel	343
trash_branch	336
animal_crab	309
animal_shells	249
animal_etc	235
trash_wreckage	165
trash_pipe	156
trash_net	127
trash_bottle	126
trash_tarp	121
trash_rope	117
trash_snack_wrapper	84
trash_clothing	82
trash_cup	59

1. TrashCan Dataset

Class	Annotations
Big fish	3241
Crab	6538
Jellyfish	637
Shrimp	548
Small fish	9556
Starfish	5093

2. Brackish Underwater Dataset

C. Model training and testing:

The YOLOv5s[8] neural network architecture was obtained from the YOLOv5 repository[8] and trained using the NVIDIA Tesla P100-PCIE-16GB hardware from Google Colab[9], which provides efficient computing power and easy accessibility. The training process involved 150 epochs with a batch size of 16 on the Yolov5 with the TrashCan Dataset and 10 epochs for the Brackish Dataset. The model leverages the Trashcan dataset[6] and uses Stochastic Gradient Descent as the default learning rate during training. After completion of training, the best weights were taken and deployed on the testing set to evaluate the model's performance using validation and testing sets.

The evaluation metrics utilized in this study include Mean Average Precision (mAP), Precision(P), and Recall(R). The Intersection Over Union measures the performance of the predicted bounding boxes. Meanwhile, mAP evaluates the network's ability to recognize four classes of images based on the average precision at different recall values. True Positive (T.P.) is an outcome where the model correctly predicts the positive class, while False Positive (F.P.) is an outcome where the model incorrectly predicts the positive class. Recall reflects the ratio between True Positive and the actual Positive, while False Negative (F.N.) is an outcome where the model incorrectly predicts the negative class. The mAP is calculated by generating a precision-recall curve using the I.O.U. and integrating this curve.

Overall, the use of YOLOv5s architecture, coupled with Google Colab's hardware, allows for efficient training and evaluation of the model's performance.

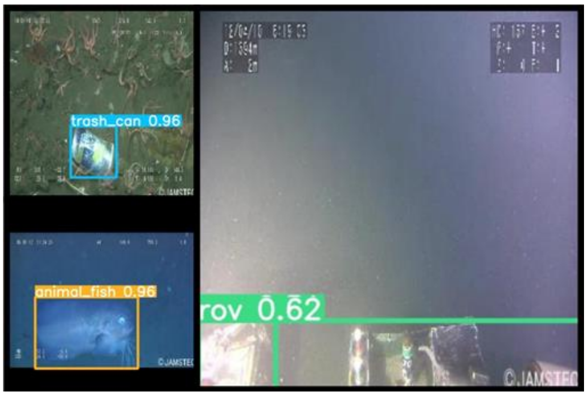

Fig 3.Inference on Images from TrashCan Dataset

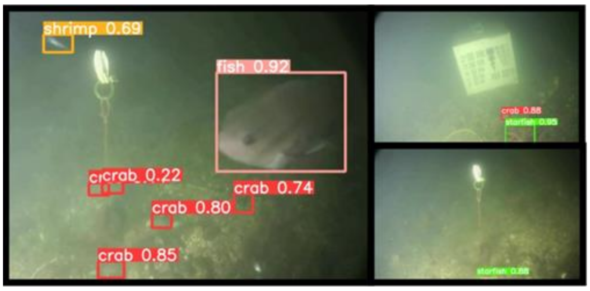

Fig 4. Inference on Images from Brackish Dataset

III. EVALUATION METRICS

The AP is the most popular metric used to gauge the precision of the detections among the various annotated datasets utilized by object detection competitions and the scientific community.

Precision, Recall, and mAP@0.5 are the assessment metrics utilized in this study to evaluate the model's performance [1].

To understand what Precision, Recall, and the mAP@0.5 are

*Corresponding Author: Moayed D. Daneshyari

we need to know of True Positive(TP), False Positive(FP), and False Negative(FN).

- True Positive: The given model predicts the bounding box correctly.
- False Positive: The given model incorrectly predicts the bounding box.
- False Negative: Incorrectly predicts the value as negative when its actual value is positive.

Precision (P) evaluates the model's capacity to recognize negative samples as the percentage of all positive examples that are genuinely true. Precision is calculated as,

Recall (R), which assesses the model's capacity to recognize positive samples, is the precision rate of eachcategory. P is the formula for precision, where TP is the number of positive samples that the model predicts will be positive and FP is the number of negative samples that the model predicts will be positive.

Recall is calculated as R = TP / (TP + FN), where FN is the number of positive samples that the model projected would be negative [1].

The AP (Average Precision) value is correlated with the mAP (mean Average Precision). Each additional positive case will correspond to a precision rate value (Pi), and the average of the n Pi for this class is known as AP@0.5. This istrue for a class of samples with n positive cases when the threshold of confusion matrix IOU is 0.5.

The mean value of AP@0.5 for all classes is known as mAP@0.5. A higher mAP@0.5 indicates a better likelihood that the model will maintain high precision with high recall. mAP@0.5 gauges the trend of model precision with recall.

In conclusion, the performance of the model is assessed in this study using the precision, recall, and mAP@0.5 criteria. While recall assesses the model's capacity to identifypositive samples, precision assesses the model's capacity to identify negative samples. A greater value suggests better performance according to mAP@0.5, which tracks the relationship between model precision and recall.

Below we see in Fig 5, and 6 the precision curve, recall curve, and precision-recall curve for the YOLOv5 on the TrashCan Dataset and we see that it has a Precision, Recall, and mAP of 97.2%, 96%, and 91.5% respectively. Whereas, for the YOLOv5 on the Brackish Underwater Dataset the Precision, Recall, and mAP values are 91.5%, 88.9%, and 94.3% respectively.

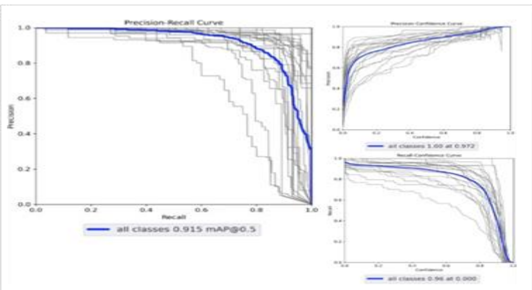

Fig 5. Precision-Recall curves for YOLOv5 on TrashCan

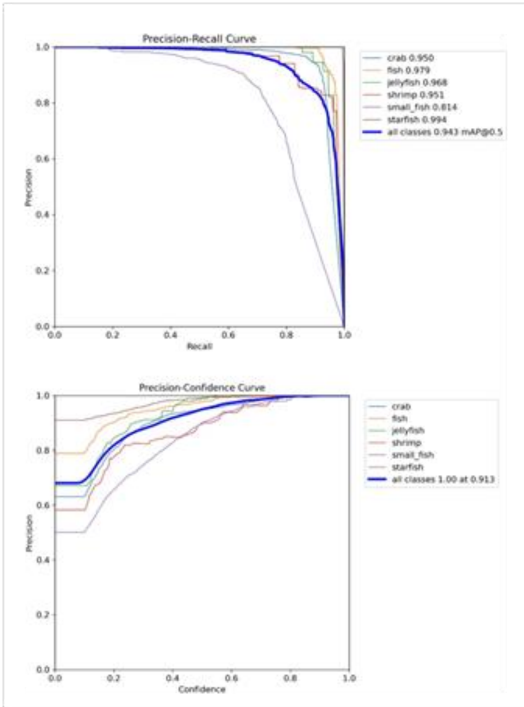

Fig 6. Precision-Recall curves for YOLOv5 on Brackish

IV. RESULTS

The results are evaluated for object detectors based on three main metrics, the Recall, Precision, and mean Average Precision(mAP) where this is the average score over all the categories. Object detection performance is assessed using intersection over union (IoU), which compares the predicted bounding box to the ground truth bounding box. The basic metric is the mAP@0.5 where the mean Average Precision values are calculated using an IoU threshold of 0.5. The thresholds range from [0.5, 0.55, 0.6, ..., 0.90, 0.95] but we are only interested in the 0.5 since it's the most accurate and is denoted as mAP@0.5 here. Out of the training and testing we have done, the YOLOv5 model on the Brackish Dataset has obtained the highest mAP@0.5, despite being trained with only 10 epochs, making it a better choice. The following are the results in Fig. 7.

Model	*Dataset*	*Precision*	*Recall*	*mAP@0.5*
Yolov5	TrashCan	97.2%	96%	91.5%
Yolov5	Brackish	91.5%	88.9%	94.3%
Yolov7	TrashCan	94.5%	98%	71.6%
Yolov7	Brackish	30.1%	80%	79.2%

Fig 7. Results for YOLOv5 and YOLOv7

REFERENCES

1. Lambertini, E., & Kozin, I. (2019). Marine Debris as a Global Environmental Problem: Introducing a Solutions-based Framework Aimed at Balancing Ocean Health and Economic Development.

*Corresponding Author: Moayed D. Daneshyari

Frontiers in Marine Science, 6, 1-13. https://doi.org/10.3389/fmars.2019.00555

2. Mendoza-Camacho, F., & Pérez-Corona, J. (2021). Underwater object detection: a survey. Journal of Ambient Intelligence and Humanized Computing, 12(4), 3381-3392.
3. Reddy, N. P. C., S. V. S. R. Ayyagari, and G. R. Reddy. "Deep learning based underwater object detection: A survey." Journal of King Saud University-Computer and Information Sciences 33, no. 3 (2021): 279-291.
4. Huang, Yeqing, Zhenyu Liu, Zhifeng Xu, and Miao Zhang. "Underwater Object Detection: A Survey." arXiv preprint arXiv:2102.09951 (2021).
5. Mendoza-Camacho, F., & Pérez-Corona, J. (2021). Underwater object detection: a survey. Journal of Ambient Intelligence and Reddy, N. P. C.,
6. S. V. S. R. Ayyagari, and G. R. Reddy. "Deep learning based underwater object detection: A survey." Journal of King Saud University-Computer and Information Sciences 33, no. 3 (2021): 279-291.
7. Roboflow. Trashcan Dataset. Available: https://universe.roboflow.com/ applied-machine-learning/trashcan-dataset/dataset/3
8. Roboflow. Brackish Underwater Dataset. Available: https:// universe.roboflow.com/brad-dwyer/brackish-underwater/dataset/2
9. Yolov5 by Ultralytics. Available: https://github.com/ultralytics/yolov5 Google Colab. Yolov5. Available: https://colab.research.google.com/ github/ultralytics/yolov5/blob/master/tutorial.ipynb
10. Yolov7. Available: https://github.com/WongKinYiu/yolov7 [10]Trashcan Dataset Available:https://conservancy.umn.edu/handle/ 11299/214865
11. M. H. Zahid, N. N. Siddique, and H. Farooq, "A Survey on Performance Metrics for Object-Detection Algorithms," 2020 27th International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 1-6, doi: 10.1109/IWSSIP48289.2020.9145130, Sept. 2020.
12. Yuan, F., Jiang, J., Zhang, Y., & Yang, Y. (2021). Marine debris detection model based on the improved YOLOv5. Measurement, 109994. https://doi.org/10.1016/j.measurement.2021.109994
13. M. Javed, J. Kim, and M. Baig, "Visual Marine Debris Detection using YOLOv5s for Autonomous Underwater Vehicle," Remote Sensing, vol. 13, no. 22, article no. 4706, doi: 10.3390/rs13224706, Nov. 2021.
14. H. Zhou, Q. Fang, X. Zhu, and J. Xiong, "Marine debris detection model based on the improved YOLOv5," Journal of Physics: Conference Series, vol. 1993, article no. 022090, doi: 10.1088/1742-6596/1993/2/022090, Oct. 2021.
15. B. Fan, Z. Guo, and H. Zhang, "An Improved Method of Marine Debris Detection Based on YOLOv5," Sensors, vol. 22, no. 23, article no.9897, doi: 10.3390/s22239897, Nov. 2022.
16. V. S. Varshini and S. Anusha, "Marine Debris Detection using Deep Learning Techniques," International Research Journal of Engineering and Technology, vol. 8, no. 7, pp. 1235-1239, doi: 10.23956/ijermt.v8i7.591, Jul. 2021.
17. [17] S. Pedersen, S. Tripathy, J. Olsen, J. Steffensen, and M. L. Jørgensen,"Detection of Marine Animals in a New Underwater Dataset with Deep Learning," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0-0, doi: 10.1109/ CVPRW.2019.00172, Jun. 2019

*Corresponding Author: Moayed D. Daneshyari