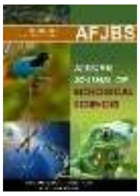
https://doi.org/10.48047/AFJBS.6.13.2024. 3141-3149

African Journal of Biological Sciences



Optimizing Underwater Fish Detection: A Comparative Study of YOLOv5 and FasterR-CNN Performance

1.D S S Lakshmi Kumari P
**Information Technology (Assistant Professor)SRKR Engineering College*
Bhimavaram, India pdsslkumari@srkrec.ac.in 0000-0001-7211-3301

2. Suresh Kumar Samarla
Computer Science Engineering (Assistant Professor)SRKR Engineering College
Bhimavaram, India ssk@srkrec.ac.in 0000-0003-1051-6812

3. Dr. Phani kumar Solleti
CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhrapradesh, India, phanikumar.solleti@gmail.com;

4. Dr.Rajababu M
Information Technology, Aditya University, Surampalem., Andhra pradesh, India, rajababu.makineedi@adityauniversity.in

Article History
Volume 6, Issue 13, 2024
Received: 18June 2024
Accepted: 02July 2024
doi:***10.48047/AFJBS.6.13.2024.*** 3141-3149

Abstract

Objective: This study uses the improved capabilities of two powerful Convolutional Neural Net- work models for real-time object detection, YOLOv5 and Faster R-CNN, to improve the accuracy and efficiency of underwater fish detection in a variety of complicated aquatic situations.

Methodology: The research involves a comparative analysis of YOLOv5 and Faster R-CNN models. These models are evaluated based on their detection speed, accuracy, and reliability in underwater settings. YOLOv5is known for its speed and lightweight architecture, while Faster R-CNN emphasizes detection accuracy through region proposal mechanisms. Various underwater scenes, including those with blurred images, are used to assess the performance of both models.

Results: The findings indicate that YOLOv5 excels in scenarios requiring real-time detection due to its speed. Conversely, Faster R-CNN demonstrates superior precision in complex image contexts, achieving an accuracy of 89.8%. This level of accuracy ensures reliable and swift fish detection even in blurred underwater scenes.

Conclusion: This study provides valuable insights into the strengths and weaknesses of YOLOv5 and Faster R-CNN in fish detection tasks. It contributes to the advancement of marine biology studies, sustainable fishing practices, and the preservation of aquatic ecosystems by offering robust tools for accurate and efficient fish detection.

***Index Terms*—Object detection,FasterRCNN,YOLOv5, fish detection**

1. Introduction

Fish target detection is important in many areas, including aquatic ecosystem management, fisheries research, and environmental monitoring. Traditional methods of fish detection often rely on manual observation or labor-intensive techniques, which can be time- consuming and prone to errors [3]. The field of automated fish detection systems has seen a paradigm shift with the introduction of computer vision and deep learning techniques, which present the possibility of improved efficiency and accuracy.

This study, in this context, focuses on the creation and application of a deep learning-based fish detection framework. Specifically, we employ the Faster R-CNN (Region-based Convolutional Neural Network) model, known for its effectiveness in object detection tasks. By leveraging annotated underwater images sourced from Kaggle,our objective is to train and evaluate the Faster R- CNN model to accurately detect fish targets within these images. This study embarks on an exploration of cutting-edge object detection technologies, specifically YOLOv5 and Faster R-CNN, to revolutionize the field of underwater fish detection. The focus is on leveraging these powerful convolutional neural network (CNN) models to tackle the inherent challenges of aquatic environments, such as variable visibility, back- ground noise, and the dynamic nature of marine life. The main goal of this study is to create and improve a very precise and effective method for identifying fish in submerged environments. Utilize advanced object detection models, specifically YOLOv5 and Faster R-CNN. Compare the performance of YOLOv5 and Faster R-CNN in terms of: Detection speed, Accuracy.

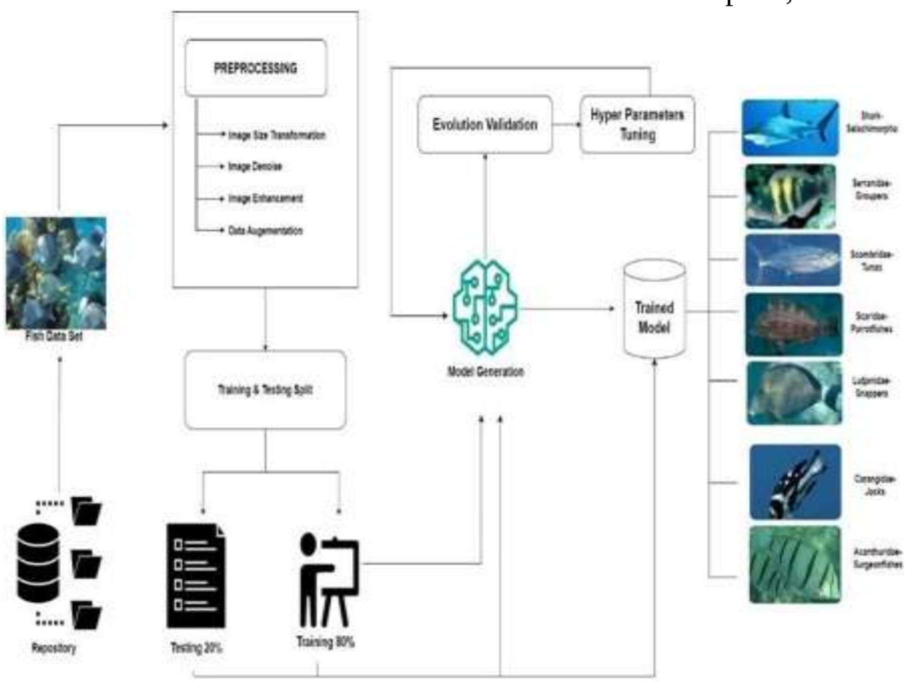


Figure. 1. Framework for proposed system

2. Literature Review

In this research the author Hu et.al [3] examines the principles of subaquatic imaging and the factors leading to quality degradation, providing a brief classification of existing methods and then highlights current general deep learning techniques for underwater image enhancement, video enhancement technologies along with standard datasets.

On another hand the author Wanghua Li et.al [6] proposed YOLOv5s for F4K dataset. They have used image preprocessing with a Gamma transform for better grayscale and contrast. The backbone is lightweight ShuffleNetv2, incorporating the SE channel attention methods to enhance model size and computation, enhancing detection speed. The BiFPN-Short network used for improved feature fusion and detection accu- racy.

The algorithm, a modification of Faster R-CNN, uses DenseNet for feature extraction, leveraging connections between layers to improve shellfish detection accuracy was proposed by Yiran Feng et.al. [2]. Additionally, it optimizes the proposal merging strategy by using Soft-NMS, enhancing the precision of the proposals.

The author Lingcai Zeng et.al [10] introduces the Faster R-CNN-AON network on VOCO7 dataset, which integrates an adversarial occlusion network (AON) with the standard Faster R-CNN detection algorithm. The AON competes with the Faster R-CNN by learning to obscure targets, making it harder to classify them correctly.

In this approach, the author Mehdi Fatan et al. [1] retrieved image edges and used texture information to classify images using a MLP neural network and SVM. The most effective technique for edge classification was found to be the combination of the MLP network with the 2D Fourier transform.

In order to improve the classification accuracy, transfer learning and data augmentation were used in conjunction with the DCCN models that Guan Wei Thum et al. [9]suggested to categorize underwater cable images that were taken from a series of underwater photographs. After these models were compared and their performance was discussed, MobileNetV2 performed better than the other models, produced reduced computational times, and had the highest accuracy 93.5.

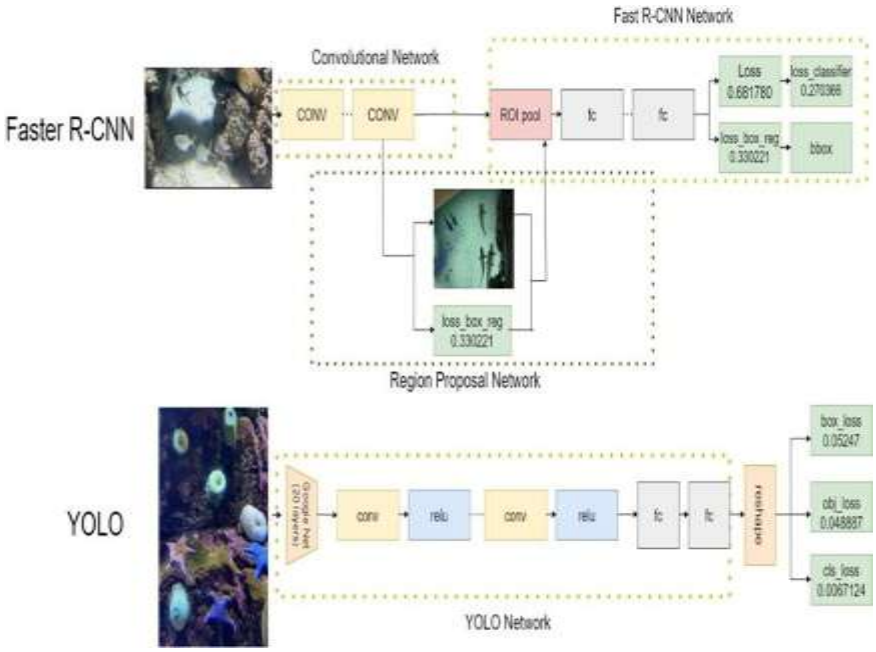


Figure. 2. FRCNN & YOLO Model Generation

On another hand Pengfei Shi et.al [8] proposed an approach for detecting and classifying underwater dam cracks. They use a dodging algorithm to correct uneven illumination in underwater images. Then, it detects cracks by analysing both local image block characteristics and global connected domain characteristics.

In this study Li G et.al [4], collected fish images using an infrared camera on a deep-sea truss-structure net cage, and a fish dataset was created. The results indicated that Faster R-CNN with the EfficientNetB0 backbone and FPN module was the most effective, achieving a high AP50 value of 0.85 and significantly shorter detection time, closely rivaling the best AP50 value of 0.86 obtained with VGG16 and all improvement modules plus data augmentation.

The author Li J et .al [5] proposed improved CME- YOLOv5 network aims to enhance fish detection in dense groups and small targets. They introduced the C3CA module, combining coordinate attention (CA) mechanism and cross-stage partial networks with 3 convolutions (C3) structure, replacing the C3 module in YOLOv5. This improves target feature extraction and detection accuracy. The efficient intersection over union (EIOU) loss function to optimize convergence rate and location accuracy.

In this study the author M. A. Rosales et.al [7], employed Faster R-CNN to develop a fish detector for locating fish occurrences within a frame. The marine biological item detection architecture proposed by Zhang J et al. [11] is built upon an improved YOLOv5 framework. To more successfully collect contextual in- formation, they have employed the RTM Det backbone architecture and the CSPLayer module with a large convolution kernel. Additionally, the YOLOv5 neck module incorporates the BoT3 module with multi-head self-attention (MHSA) mechanism, which improves detection accuracy and the model's performance insettings with dense targets.

In this study the author Zhang X et.al [12] pro- posed end-to-end aquatic animal detection framework that integrates image enrichment and object detection. The two-stage detection network with a dynamic IoU threshold as the backbone. The underwater image enhancement module (UIEM), which consists of de- noising, color correction, and deblurring sub-modules.

3. Framework For Proposed Methodology

In the proposed methodology the FasterRCNN and YOLOv5 models are shown in Figure 2.

A. *Faster R-CNN (Faster Region-based ConvolutionalNeural Network)*

In response to the limitations identified in the existing Faster R-CNN approach for underwater fish detection, this paper proposes an enhanced system thatintegrates MobileNet as the support architecture withinthe Faster R-CNN framework.

This study aims to address the key disadvantages of sensitivity to underwater conditions, computational intensity, and adapt- ability to varied species, among others. MobileNet,known for its lightweight and efficient architecture, is specifically designed for mobile and embedded vision applications. By leveraging depthwise separable convolutions, MobileNet drastically lowers both the computing cost and the number of parameters without sacrificing performance. Integrating MobileNet with Faster R-CNN allows for faster processing speeds and lower resource consumption, making real-time underwater fish detection more feasible. Moreover,the use of MobileNet as the backbone enables theproposed system to better handle the diversity of fish species. The efficient feature extraction capability of MobileNet, combined with Faster R-CNN's region proposal and classification mechanisms, permits for more precise identification and localization of a varied variety of fish species across various underwa ter scenes. To overcome the restrictions connected with anchor boxes and handling occlusions and over- laps, the proposed system incorporates adaptive anchor strategies and enhanced algorithms for detecting partially obscured or overlapping fish. This ensures more robust detection in densely populated underwaterenvironments.

Overall, the proposed system represents a significant advancement in underwater fish detection technology. By merging the strengths of MobileNet and Faster R-CNN, and addressing the specific challenges of underwater environments, this system offers a promising solution for accurate, efficient, and reliable underwater fish detection, with potential applications in marine biology, environmental monitoring, and sus tainable fishing practices. In Faster R-CNN, the first phase is to generate region proposals using a technique called Selective Search. This produces a set of budding bounding boxes that contains the objects. Then, these bounding boxes are passed through a Convolutional Neural Network (CNN) to extract features. These features is then fed into networks: the region proposal network (RPN) and the object detection network. The RPN is responsible for classifying the region proposals as either foreground (containing an object) or background (not containing an object). The object detection network takes the refined region proposals and classifies them into specific object categories. It also further refines the bounding box coordinates to accurately localize the objects. Overall, Faster R-CNN combines the processes of region proposal generation and object classification into a single network, making it more efficient and accurate compared to previous methods

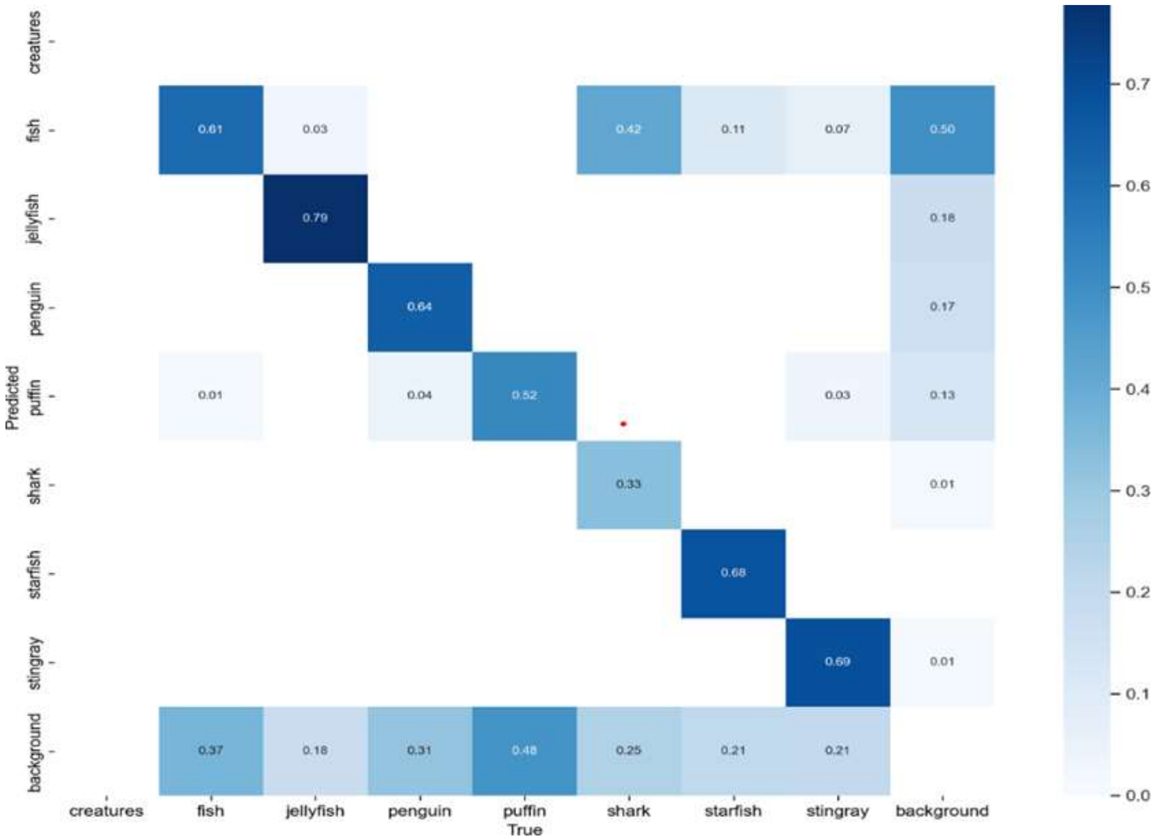


Figure. 3. Confusion matrix for the Yolov5 model

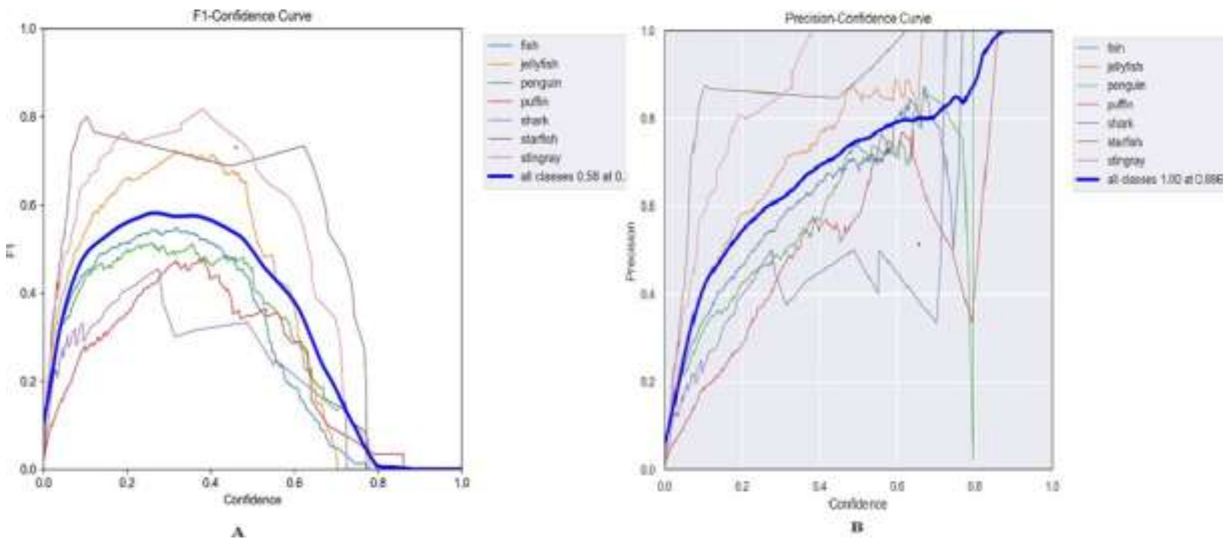


Figure. 4. F1-Confidence Curve & P-Confidence Curve for proposed model

.Data Collection

The dataset consists of a collection of images depicting various species of fish, catagorized into 7 distinct classes such as Fish, Jelly Fish, Stingray, Puffin, Pen- guin, Starfish,and Shark. The Proposed Framework is shown in Figure 1.

DATA SPLIT

The dataset has been segregated into three subsets to facilitate model training, validation, and evaluation:

1) **Test Set:** This subset contains 68 images, which are reserved for estimating the performance of trained models.
2) **Training Set:** Comprising 476 images, this sub- set forms the primary data. These images are utilized by algorithms to learn patterns and features representative of each

fish class, enabling them to make predictions during deployment.

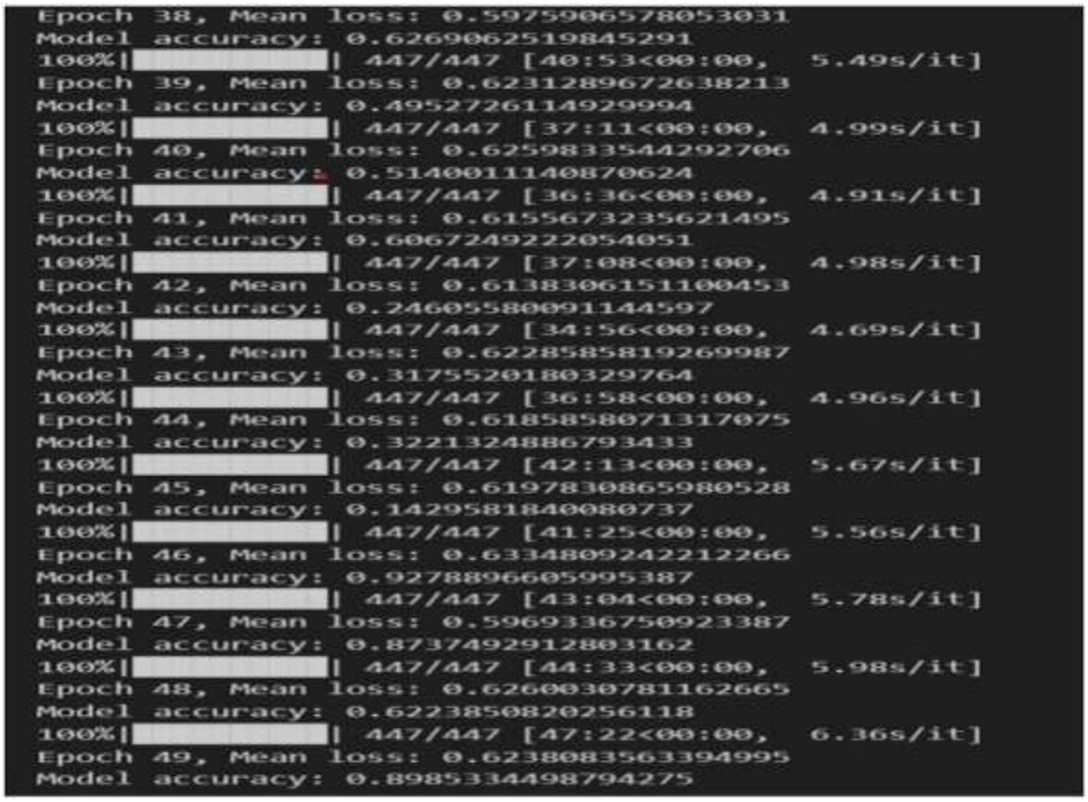
3) **Validation Set:** With 136 images, the validation set is used to assess the performance of the model during training, allowing for adjustments to hyper parameters or model architecture to improve overall performance.

Dataset Link: Fish Dataset

B. *YOLOv5*

Architecture: Yolov5, a one-step object detection algorithm, is well-known for its ability to forecast object speed, bounding box, and class probabilities. The architecture is divided into three main parts: a backbone for feature extraction (typically CSPDark- net53), a neck (using PANet for feature aggregation), and a head for bounding box prediction.

Detection Process: YOLO creates a grid out of the image and



```
Epoch 38, Mean loss: 0.5975906578053031
Model accuracy: 0.6269062519845291
100%|████████████| 447/447 [40:53<00:00,  5.49s/it]
Epoch 39, Mean loss: 0.6231289672638213
Model accuracy: 0.4952726114929994
100%|████████████| 447/447 [37:11<00:00,  4.99s/it]
Epoch 40, Mean loss: 0.6259833544292706
Model accuracy: 0.5140011140870624
100%|████████████| 447/447 [36:36<00:00,  4.91s/it]
Epoch 41, Mean loss: 0.6155673235621495
Model accuracy: 0.6067249222054051
100%|████████████| 447/447 [37:08<00:00,  4.98s/it]
Epoch 42, Mean loss: 0.6138306151100453
Model accuracy: 0.24605580091144597
100%|████████████| 447/447 [34:56<00:00,  4.69s/it]
Epoch 43, Mean loss: 0.6228585819269987
Model accuracy: 0.3175520180329764
100%|████████████| 447/447 [36:58<00:00,  4.96s/it]
Epoch 44, Mean loss: 0.6185858071317075
Model accuracy: 0.3221324886793433
100%|████████████| 447/447 [42:13<00:00,  5.67s/it]
Epoch 45, Mean loss: 0.6197830865980528
Model accuracy: 0.1429581840080737
100%|████████████| 447/447 [41:25<00:00,  5.56s/it]
Epoch 46, Mean loss: 0.6334809242212266
Model accuracy: 0.9278896605995387
100%|████████████| 447/447 [43:04<00:00,  5.78s/it]
Epoch 47, Mean loss: 0.5969336750923387
Model accuracy: 0.8737492912803162
100%|████████████| 447/447 [44:33<00:00,  5.98s/it]
Epoch 48, Mean loss: 0.6260030781162665
Model accuracy: 0.6223850820256118
100%|████████████| 447/447 [47:22<00:00,  6.36s/it]
Epoch 49, Mean loss: 0.6238083563394995
Model accuracy: 0.8985334498794275
```

Figure. 5. Accuracy of Faster RCNN model at different epoch’s

forecasts the bounding boxes and class probabilities for every grid cell. It employs non-maximum suppression (NMS) to fine-tune the predictions by removing over- lapping boxes according to their confidence scores, and it applies anchor boxes to enhance detection in a range of sizes. **Training:** model is trained on the dataset with annotated images, using a loss function that combines classification loss, objectness loss, and box regression loss. It employs various data augmentation procedures like flipping, scaling, and cropping to improve robust- ness and generalization.

C. *FasterR-CNN*

Architecture: Faster R-CNN is a 2-stage object detection algorithm, consisting of a RPN for generating object proposals and a Fast R-CNN detector to classify the proposals and refine their bounding boxes. It often uses feature extraction backbones like CSPDarknet53 Region Proposal Network (RPN): The RPN scans the images through a sliding window, proposing candidate object regions (anchors) and scoring them based on Objectness. It uses a set of pre-defined anchor boxes of various scales and aspect ratios to capture different object sizes.

Detection Process: For each proposal from the RPN, the Fast R-CNN detector extracts features using RoI (Region of Interest) Pooling, classifies the features into object categories, and refines the bound- ing box locations. The final step involves applying NMS to remove duplicate detections.

Training: Faster R-CNN is trained in a multi-stage process, initially focusing on the RPN and then fine-tuning the Fast R- CNN detector. The training uses a shared loss function that includes terms for classification, bounding box regression for the RPN, and similar terms for the Fast R-CNN detector.

Table 1. Evaluation Metrics for All Classes Using YOLOV5.

Class	Precision	Recall	mAP50	mAP50-95
all	0.612	0.555	0.543	0.191
fish	0.553	0.517	0.474	0.133
jellyfish	0.667	0.735	0.691	0.204
penguin	0.491	0.522	0.477	0.115
puffin	0.430	0.464	0.348	0.0943
shark	0.438	0.325	0.342	0.167
starfish	0.857	0.630	0.732	0.316
stingray	0.845	0.690	0.733	0.304

D. *Results Discussions*

The research contributes to the advancement of ma- rine biology studies, sustainable fishing practices, and the preservation of aquatic ecosystems by offering a robust tool for accurate and efficient fish detection. The results show the Faster RCNN algorithm achieves an accuracy of 89.8 which can achieve accurate and fast detection for fish targets in underwater blurred scenes. In the Table 1 shows the evaluation metrics precision, Recall, mAP50 and mAP95 for the seven classes in which the model is tested. It was observed mAP50 for yolov5 is 54% on this dataset. The confusion matrix is shown in Figure 3. The Figure 4 shows the F1-confidence and precision confidence of all the classes in the given dataset.

We have implemented YOLOv5 and faster-CNN to detect and classify the images. Epoch 50 is the best one compared to other epoch's as we have achieved maximum accuracy with 89.8% for Faster R-CNN which is shown in Fig5.The model hyper parameters is shown in Table 2. Comparison of models with existing works is shown in Table 3.

Table 2. Comparison between Faster-RCNN & YOLOV5

Phases	RPN + Faster-CNN detector	Concurrent bounding box regression & classification
Network Type	Fully Convolutional	Fully Convolutional
Backbone	Mobilenet	Darknet-53
Location	Anchor based	Anchor based
Detection	One (at 0.5)	Two (at 0.5 &
IOU Thresholds	Softmax loss for	0.95)
	classification, smooth L1 for regression	**Binary cross entropy**
Loss Function	Multi-stage	Single stage
Processing	object detection	object detection
mAP50	0.898	0.537
mAP50-95	0.145	0.182

Table 3. Comparison of Models with Accuracies.

Ref.ID	Dataset	Model	Accuracy
[8]	Custom dataset	Faster RCNN	85
[11]	URPC2020	YOLOv5	79.4
[12]	URPC2020	ResNet	69.4
Proposed Model	Roboflow	Faster RCNN	89.8

E. Conclusion & Future Enhancement

The study has successfully demonstrated the application of advanced convolutional neural networks, specifically YOLOv5 and an enhanced Faster R-CNN with a MobileNet backbone, in the field of subaquatic fish detection. The integration of MobileNet into Faster R-CNN, in particular, has offered a compelling balance between computational efficiency and detection precision, making it suitable for deployment in resource-constrained underwater settings. Implementing transfer learning and semi- supervised learning techniques could also enhance learning efficiency and model adaptability, particularly for rare or hard-to- detect species. Finally, increasing the collaboration between marine biologists, environmentalists, and AI researchers will ensure that the development of these technologies remains aligned with conservation goals and practical research needs. By emphasizing these areas, future improvements will further contribute to the sustainability and comprehension of marine ecosystems, as well as expand the technical capabilities of underwater fish detecting systems.

References

[1] M. Fatan, M. R. Daliri, and A. M. Shahri, "Underwater cable detection in the images using edge classification based on texture information," *Measurement*, vol. 91, pp. 309–317, Sep. 2016.
[2] Y. Feng, X. Tao, and E. Lee, "Classification of shellfish recog- nition based on improved Faster R-CNN framework of deep learning," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–10, Jun. 2021.
[3] K. Hu, C. Weng, Y. Zhang, J. Jin, and Q. Xia, "An overview of underwater vision enhancement: From traditional methods to recent deep learning," *Journal of Marine Science and Engineering*, vol. 10, no. 2, p. 241, Feb. 2022.
[4] G. Li, Z. Yao, Y. Hu, A. Lian, T. Yuan, G. Pang, and X. Huang, "Deep learning-based fish detection using above- water infrared camera for deep-sea aquaculture: A comparison study," Sensors, vol. 24, no. 8, p. 2430, Apr. 2024.
[5] J. Li, C. Liu, X. Lu, and B. Wu, "CME-YOLOv5: An efficient object detection network for densely spaced fish and small targets," *Water*, vol. 14, no. 15, p. 2412, Aug. 2022.
[6] W. Li, Z. Zhang, B. Jin, and W. Yu, "A real-time fish target detection algorithm based on improved YOLOv5," *Journal of Marine Science and Engineering*, vol. 11, no. 3, p. 572, Mar. 2023.
[7] M. A. Rosales, M. G. Palconit, V. J. Almero, R. S. Concepcion, J.-A. Magsumbol, E. Sybingco, A. A. Bandala, and E. P. Da- dios, "Faster R-CNN based fish detector for smart aquaculture system," in *Proceedings of the IEEE*, Nov. 2021, pp. 1–6.
[8] P. Shi, X. Fan, J. Ni, and G. Wang, "A detection and classifica- tion approach for underwater dam cracks," *Structural Health Monitoring*, vol. 15, no. 5, pp. 541–554, Sep. 2016.
[9] G. W. Thum, S. H. Tang, S. A. Ahmad, and M. Alrifaey, "Toward a highly accurate classification of underwater cable images via deep convolutional neural network," *Journal of Marine Science and Engineering*, vol. 8, no. 11, p. 924, Nov. 2020.

[10] L. Zeng, B. Sun, and D. Zhu, "Underwater target detection based on Faster R-CNN and adversarial occlusion network," *Engineering Applications of Artificial Intelligence,* vol. 100, p. 104190, Apr. 2021.

[11] J. Zhang, J. Zhang, K. Zhou, Y. Zhang, H. Chen, and X. Yan, "An improved YOLOv5-based underwater object-detection framework," *Sensors*, vol. 23, no. 7, p. 3693, Apr. 2023.

Zhang, X. Fang, M. Pan, L. Yuan, Y. Zhang, M. Yuan, Lv, and H. Yu, "A marine organism detection framework d on the joint optimization of image enhancement and ct detection," *Sensors*, vol. 21, no. 21, p. 7205, Oct. 2021.