Design and Evaluation of Efficient Underwater Object Detection Systems Using AI

Seelam Ruchitha Mahi[1], Siddamsetti Vamsi Krishna[2],
Annam Komal Sai Mani Naga Vaibhav[3], Prayaga Vathsalya Sri Vijaya[4] and
Brij Kishor Tiwari[5]
{ruchithamahi@gmail.com[1], vamsikrishnasiddamsetti@gmail.com[2], vaibhav.ksmn@gmail.com[3], vijayaprayaga83@gmail.com[4], aeromantiwari@gmail.com[5]}

Department of Advanced Computer Science and Engineering, Vignan's Foundation for Science Technology and Research, Guntur-522213, Andhra Pradesh, India[1, 2, 3, 4, 5]

Abstract. Underwater object detection plays a crucial role in marine applications, where accurate detection is often prevented by obstacles like poor visibility, image distortion, and environmental noise. To address these issues, this study examines the effectiveness of the various deep learning models in detecting underwater objects. In this study YOLOv8, YOLOv10, YOLOv11, and YOLOv12 models were used. These model's performance on a wide range of underwater images is assessed based on their accuracy (mean average precision, or mAP), speed, and capacity to operate in challenging circumstances. Each model was trained and tested using the same data to provide a fair comparison. The results indicate that the more recent YOLO models, particularly YOLOv11 and YOLOv12, achieve high accuracy and speed marks, with an average mean precision of 91.5%. This study proves that using the most recent YOLO models can improve underwater object detection and support real-time marine applications like underwater exploration and autonomous underwater vehicles.

Keywords: Underwater Object Detection, Deep Learning, YOLOv8, YOLOv10, YOLOv11, YOLOv12, mean Average Precision (mAP).

1 Introduction

The undersea world is simultaneously one of the most strange and least investigated ecosystems on the planet. Automatic underwater object detection and recognition is a very important task in marine biology, underwater exploration, autonomous vehicle navigation, ocean preservation, or fisheries management to various regions. As human activity in underwater regions becomes increasingly extensive, the need for intelligent systems that are capable of performing real time detection and classification of undersea creatures or objects has been greatly encouraged. However, training accurate and robust object detectors for underwater images poses a set of difficult challenges unlike those in land-based image analysis.

Underwater images are fundamentally different from the images taken above the surface or from the air. Because light is scattered and absorbed by water, such images are usually characterized by low contrast, color deviation, noise, blur, and finite visibility. These challenges are exacerbated by the dynamic nature of underwater scenes, including the stationary movement of underwater organisms, the varying illumination, and the proximity to tagged particles and water flow. The Marine life we are supposed to find and identify even if it refers to the same

ICITSM-Part II 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2358015

textures and shapes, often overlaps or merges with the substrate so finding them even is a bigger challenge to identify them. For that reason, traditional computer vision techniques and even univariate convolutional neural networks by themselves consistently fail to work under such situations or achieve the same results as perceived by the end-user.

In recent years, the emergence of deep learning has revolutionized object detection with power-hungry models, such as the YOLO (You Only Look Once) family, demonstrating remarkable accuracy and speed performance in real-time scenario. The advancement of YOLO from its predecessors to its current editions (like YOLOv8, YOLOv10, YOLOv11, and YOLOv12) in terms of architecture development, optimization techniques and generation power is significant. These promotions render YOLO models' strong candidates in underwater object detection, where real-time filtering (at high speeds) is as important as correct predictions.

In this paper, four generations of YOLO architectures (i.e., YOLOv8, YOLOv10, YOLOv11, and YOLOv12) were thoroughly compared according to a well-designed underwater dataset. The dataset used (referred to as UWOD) is sourced from and curated and hosted on Kaggle, comprising 3,200 labeled images, and displays 3 marine species (Sea Cucumber, Sea Urchin, Scallop). The dataset is split into 2,560 training, 128 validation, and 506 test images to enable robust training and evaluation.

The dataset has been pre-processed for usage by deep learning applications, including image enhancement techniques that are used to improve contrast and visibility, data augmentation that may be used to increase malignancy model robustness, and manual (very careful) proofing to ensure labels are accurate.

All YOLO variant were trained by the same training pipeline and hyper-parameters (when it is appropriate) to have fair comparison in terms of relative difference between their performance. The results are imposed on the commonly used evaluation metrics of Industrial which are provided: Precision-Recall, mAP@50, mAP@50-95. These metrics indicate a thorough performance understanding of each model over different IoU thresholds levels of both detection accuracy and localization quality. While from the tested models, YOLOv11 and YOLOv12 achieved the high mAP@50 except that YOLOv10 saw better performances in mAP@50-95, i.e., all levels of detection difficulty.

By conducting an in-depth analysis of the performance of each YOLO variant, this research provides invaluable statistics on the efficiency of state-of-the-art deep learning models for underwater object detection. Results of the contribution will foster research and development and overall work development of the marine scientists engaged in developing real-time intelligent systems in underwater space. Moreover, to this comparison emphasizes the influence of architectural evolution of the YOLO framework, and demonstrates the great potential of deep learning in the solution of real-world problems in challenging visual conditions such as underwater.

2 Literature Review

Recently, underwater object detection and classification has received quite a bit of attention because it has many potential applications in marine surveillance, archaeology and environmental monitoring that draw interest of the researchers. Performance in underwater vision tasks has improved dramatically with the use of deep learning advancements, in particular CNNs and real-time models such as YOLO. Unlocking the potential of R-CNNs for object detection in aquatic environments: addressing image bias, with RetinaRPN, and probabilistic inference and adapting boosting 0250 reweighting to capture uncertaintyWe improve detection confidence on ambiguous items across heterogeneous source data sets, via detecting AWS_CAPTURES using a set of acquired and generated underwater annotations. Besides, Enhanced YOLOv7-Tiny model [2] is proposed for advanced feature extraction for better accuracy and speed, which can be a better solution for the real time underwater.

A YOLOv8 segmentation-based pipeline [3] employs image enhancement, pixel-wise masking, and sharpened bounding boxes to cater for noise, distortion, and low visibility, to enhance detection in complex underwater environments. CEH-YOLO [4] is a lightweight YOLOv8-based model that combines the modules of deformable attention, multiscale fusion and composite detection box proposal for the efficient detection of small and blurred underwater objects. DB-UODN [5] introduces a dual-branch feature extraction network on the basis of ECDB and DSPAC SPC modules, which can improve the robustness and accuracy of multi-scale targets in complex underwater conditions. CEH YOLO [6] further enhances quality using multiscale pooling and deformable at2 tention to obtain high accuracy and real-time performance with no extra computational burden.

A work [7] compared the four YOLOv3 variants, and demonstrated YOLOv3-SPP outperformed all the variants in accuracy, and YOLOv3-Tiny-PRN possessed the fastest processing speed, which verified YOLOv3 becomes an effective solution under turbid, low ILL conditions. A new enhancement method was used [8] to make clearer and more stable images for underwater biological observation and locate targets with better efficiency.

The YOLOv7-AC [9] combines ACmix, ResNet ACmix, GAM, K-means clustering to improve the feature extraction and detection accuracy, and achieves the best performance on URPC and Brackish datasets. A survey paper [10] investigates traditional and deep learning UOD methods, and introduces the potential APPLICATIONS OF UOD UOD has many practical applications for solving real-world problems.

issues such as degraded images and class imbalance and presents diagnostics like Diagnosis and TIDE to assess models. An enhanced EfficientDet-based method [11] with BiSkFPN and adversarial learning enhances detection performance of noise underwater images. Also hinges explainability to GradCAM and compares YOLO variants on Brackish dataset.

Finally, a deep learning-based coral detection scheme in [12] employs YOLO for reef monitoring, which can result in less manpower in coral data collection ecology with high detection performance. DeepSeaNet enhances detection in the presence of noise in an underwater environment with EfficientDet and various YOLO variants [13-14] with the block module to enhance this for improved visibility performance. UTDF-YOLOv5 integrates attention to YOLOv5 for real-time detection [15].

3 Methodology

In order to enhance the performance of underwater object detection, in this paper we deploy a structured approach using image enhancement and deep learning for detection. Some pre-processing such as contrast enhancement, colour correction and sharpness enhancement are performed to compensate the degraded underwater image. After the enhancement, the cutting-edge YOLO models are used to train and test the UWOD dataset to achieve an interactive marine object detection and identification procedure.

Dataset details: In this work, underwater object detection models were trained and tested on the UWOD dataset from Kaggle. The dataset contains overall 3,200 labeled images of various underwater settings. It is comprised of 3 classes, namely scallop, sea urchin, and sea cucumber. The data is divided into 2,560 for training, 128 for validation, and 506 for testing, a setup suitable for model training and evaluation as in Fig. 1. Each of the images are bounded, and therefore serves as valid data for object detection algorithms. With its high quality and diversity, this dataset is suitable to be used as a benchmark for the development of deep learning model on underwater object detection.

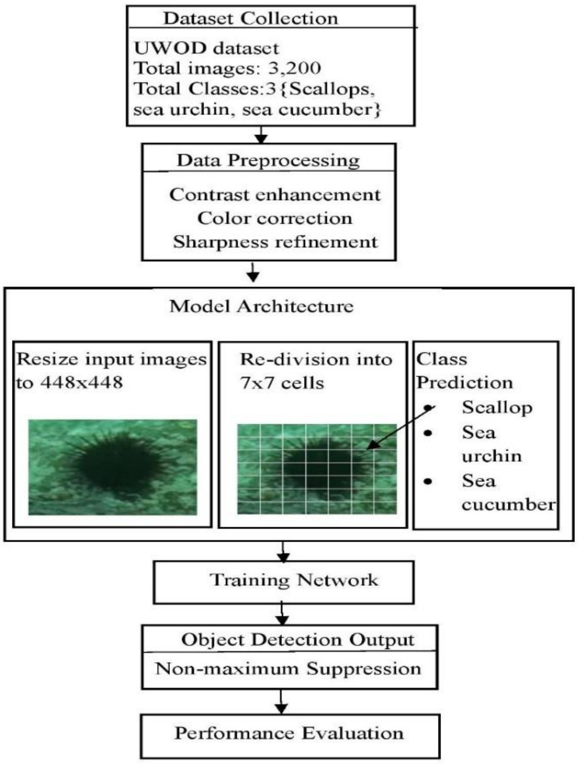


Fig. 1. Block diagram of the proposed method.

Data Preprocessing: Water lost and the characteristics of defects in underwater images are mainly caused by light scatters and absorbers in the water body, resulting in contrast weak, responsive low and so on. To mitigate these drawbacks, we have developed a pre-processing pipeline of contrast enhancing, colour correcting, and sharpness improving. Such strange facing approaches enhance the overall quality of the dataset and result in better object detection results.

The underlying principle of the image enhancement is to transform original image (RGB) to LAB color space. This further provides the ability to use Contrast Limited Adaptive Histogram Equalization (CLAHE) on the L-channel to increase local contrast while maintaining the natural color equilibrium in the image. Then, the image undergoes further sharpening using PIL’s sharpness enhancer. The change is described by means of the following mathematical relation:

$$\mathrm{I}''(\mathrm{x},\mathrm{y}) = \mathrm{I}'(X,Y) + \beta(\mathrm{I}'(X,Y) - blur(\mathrm{I}')(x,y)) \tag{1}$$

where $\mathrm{blur}(\mathrm{I}')$ denotes the Gaussian-blurred version of I', and β is the sharpness factor.

Next, contrast adjustment is applied using the transformation

$$I(X,Y) = \mu + \gamma\,(\mathrm{I}'(X,Y) - \mu) \tag{2}$$

here μ is the mean pixel intensity of the image and γ is the contrast factor (set to 1.0).to ensure uniformity and boost model performance, the entire enhancement pipeline is systematically applied to each image in the training, validation, and testing datasets, thereby maintaining consistency and improving object detection accuracy across all data splits.

(a) Raw Image (b)Preprocessed Image

Fig. 2. Comparison of Raw and Preprocessed Images.

Fig. 2 shows that the improvement of underwater visuals through preprocessing. Fig. 2(a) shows the original scene as it is shown, characterized by low contrast, a strong greenish tint, and reduced visibility. On the right, the enhanced image reflects the application of techniques like LAB color space transformation, CLAHE for better contrast, and sharpness adjustment, resulting in clearer visuals with improved object distinction and feature visibility. These enhancements make the image more suitable for object detection tasks.

Models used: The underwater object detection framework inte- grates multiple deep learning models, specifically tai- lored for the challenges of underwater imaging such as poor visibility, low contrast, and color distortion. The architecture incorporates the YOLO (You Only Look Once) family of models YOLOv8, YOLOv10, YOLOv11, and YOLOv12 each bringing enhancements in speed and accuracy for real-time object detection. YOLOv8 serves as the baseline model, featuring an optimized backbone and neck for efficient feature extraction from underwater images. It employs a CSPDarknet backbone with SPP and PAN layers, offering a balance between model complexity and detection precision.

YOLOv10 makes other improvements based on the previously described spatial pyramid pooling and anchor-free detection, which tend to be more suitable to detect objects of different scales (like sea cucumber, scallop and sea urchin). YOLOv11 and YOLOv12 further enhance these functionalities to include an attention mechanism, a feature fusion based on transformers and improved decoders to improve the localization and classification of marine organisms in the presence of noise in underwater scenarios. We use a same pre-processing pipeline of contrast adjustment, sharpness enhancement and color correction for each model to be trained or validated on improved underwater datasets.

The framework provides strong robustness of object detection in various underwater scenes by using these latest YOLO versions. Results compare detection accuracy and efficiency yield better results with newer versions of YOLO demonstrating their applicability to underwater with complex scenarios and subtle variations of objects.

4 Model Architecture

The YOLO (You Only Look Once) framework's fundamental ideas are merged into the recommended architecture for underwater object detection, which views object detection as a single regression problem, as opposed to a region proposal and classification combination. By predicting bounding box coordinates and class probabilities simultaneously and processing the entire image in a single forward pass, this single-stage detector ensures real-time performance.

Overall Architecture: The architecture consists of three main components: the backbone, neck, and head as shown in Fig. 3.

Backbone: The backbone extracts features from the input underwater image by utilizing both high level semantic cues and low-level textures. These features are then transferred through several attention modules, which are essential in underwater environments where visual quality is often degraded due to turbidity, color distortion, and scattering.

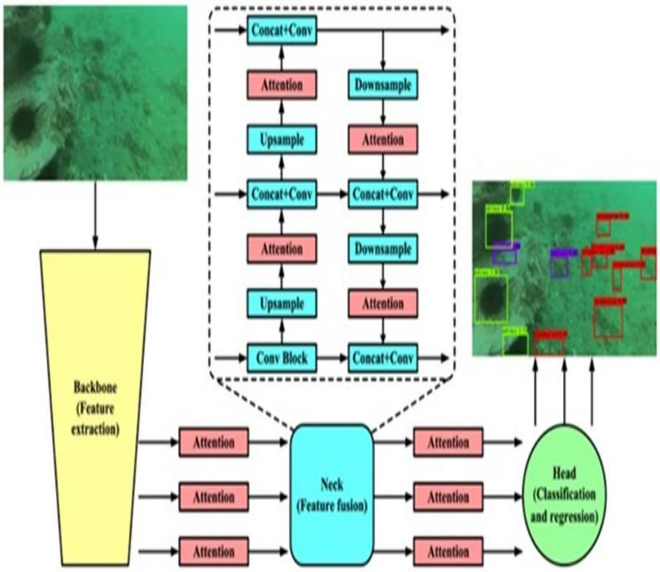


Fig.3. YOLOv8 Architecture adapted from Ganesan et.al.

Neck: The neck produces multi-scale feature fusion by merging convolution, concatenation, down sampling, and up sampling. A hierarchical attention developed structure enhances contextual understanding by focusing on important regions in both spatial and channel dimensions.

Head: The head is responsible for the classification and regression tasks, generating the final predictions, including object classes and bounding boxes.

Grid-Based Object Detection: The input image is resized to 448×448 pixels and divided into a grid of size $S \times S$, typically 7×7 Each grid cell is responsible for detecting objects whose center lies inside the grid cell. Each grid cell predicts B bounding boxes, each containing the following five components:

- Center coordinates (X, Y) relative to the grid cell.
- Width and height (W, H) normalized to the image dimensions, Confidence score C.

The confidence score C is computed as:

$$C = P_{object} \times IoU_{pred,truth} \tag{3}$$

Where P_{object} is the probability of an object being present in the grid cell, and $IoU_{pred,truth}$ is the Intersection over Union between the predicted and ground truth bounding boxes?

Each grid cell also predicts class probabilities for C object categories, resulting in $5B + C$ output values. For instance, if $B = 2$ and $C = 20$, each grid cell generates 30 values.

Bounding Box Normalization: The bounding box parameters are normalized for consistency. Let (X, Y) be the object's absolute center, and (W, H) be its width and height. If the upperleft corner of the corresponding grid cell is (X_a, Y_a) the normalized coordinates are as

$$x = \frac{X - X_a}{Cell\ size}\ and\ y = \frac{Y - Y_a}{Cell\ size} \tag{4}$$

$$\Delta W = \frac{W}{448}\ and\ \Delta H = \frac{H}{448} \tag{5}$$

Even though w and h are relative to the full image size, this normalization ensures that x and y stay within the [0, 1] range. After predictions are made, Non-Maximum Suppression (NMS) removes unnecessary overlapping boxes by retaining only the most confident box. Bounding boxes with an IoU greater than 0.5 are suppressed to ensure each object is detected only once.

5 Results and Discussion

In this study, four distinct YOLO-based models YOLOv8, YOLOv10, YOLOv11, and YOLOv12 were used to detect Underwater objects. The UWOD dataset was used to train and evaluate the models. Each image was preprocessed utilising Colour Space Transformation, contrast adjustment, and sharpness enhancement techniques to improve feature representation and reduce visual degradation of underwater images.

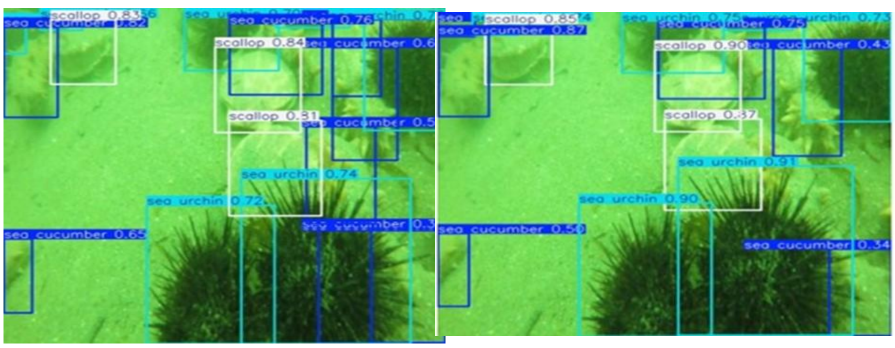
Fig 4 shows the results of detection of YOLOv8 and YOLOv10, while Fig. 5 shows the results of detection of YOLOv11 and YOLOv12.It can be observed that YOLOv8 (Fig 4(a)) detects most of the objects but with slightly decreased confidence scores. YOLOv10 Fig.4(b) builds upon this by generating more precise bounding boxes and confidence levels. YOLOv11 Fig.5(a) refines the predictions further, showing cleaner object localization and less overlap of boxes.YOLOv12 Fig.5(b) performs the most accurate detections, with consistently high confidence scores and clearly defined object borders.

Fig.6(a) and Fig.6(b) shows the performance of the YOLOv8 and YOLOv10 models and found around 90% of accuracy.Similarly Fig.7(a) and Fig.7(b) represents the performance of YOLOv11 and YOLOv12 models.From this results we can observe that YOLOv11 and YOLOv12 models are getting of higher accuracy around 91% .Among all YOLOv12 model gave the best performance with higher accuracy.This means YOLOv12 was able to learn well from the data and make correct predictions without overfitting or underfitting.

Table 1 shows that YOLOv11 and YOLOv12 achieved the highest mAP@50 (0.915) and precision (0.886) among all YOLO variants. YOLOv8 demonstrated notable speed and lightweight deployment, making it suitable for real-time applications despite a slight drop in accuracy. YOLOv12 also recorded the highest recall (0.880), effectively balancing true positive detection and minimizing false negatives. YOLOv10 delivered competitive performance, particularly in detecting small or partially obscured underwater objects due to its efficient feature aggregation design.

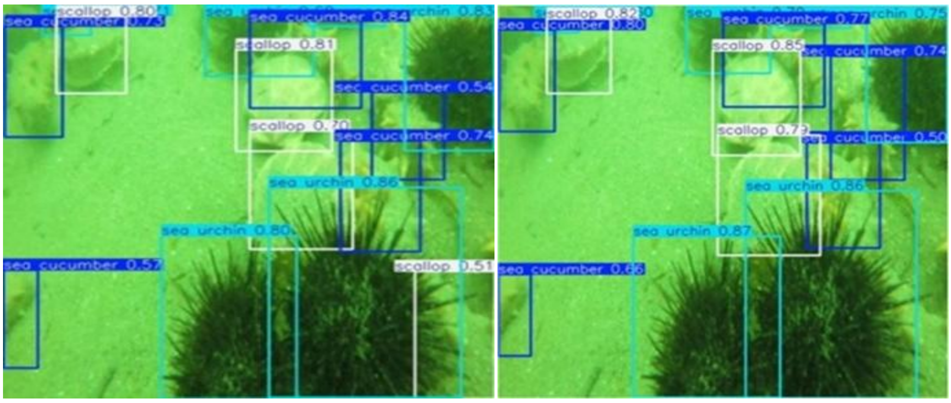
Table 2 displays the class-wise mAP@50 accuracy of all models for three types of underwater objects: scallop, sea cucumber, and sea urchin. While all YOLO variants performed well on larger and more visually different objects like Sea Urchin and Scallop, YOLOv11 had the highest accuracy (0.964 for Sea Urchin and 0.939 for Scallop). However, Sea Cucumber exhibited a relatively lower identification accuracy due to its size, shape, and higher visual similarity to the background.

Overall, the experimental findings demonstrate that the combination of effective image preprocessing and complex YOLO structures greatly enhances underwater item recognition. Out of all the models, YOLOv11 and YOLOv12 offered the best trade-off between precision, recall, and mAP, whereas YOLOv8 performed exceptionally well in scenarios requiring rapid and lightweight inference. Therefore, the specific requirements of the application whether it be ecological monitoring or real-time underwater surveillance should dictate the model selection.



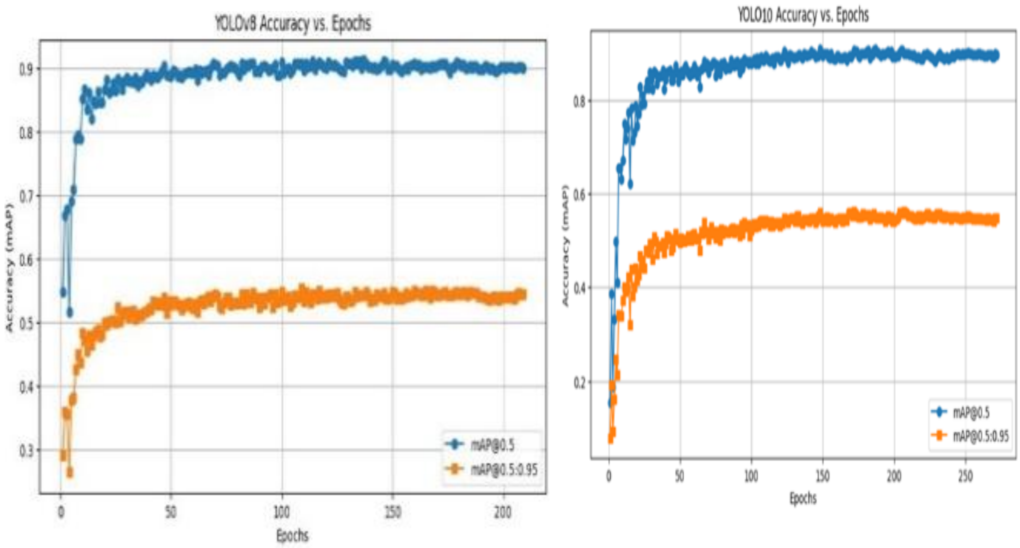
(a) Result of YOLOv8 (b)Result of YOLOv10

Fig. 4. Comparison of Detection Results from YOLOv8 and YOLOv10.



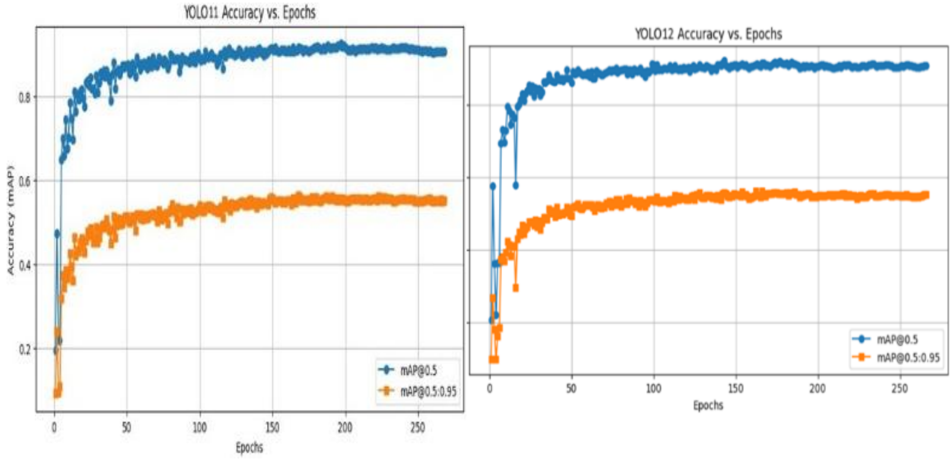
(a) Result of YOLOv11 (b) Result of YOLOv12

Fig. 5. Comparison of Detection Results from YOLOv11 and YOLOv12.



(a) YOLOv8 Accuracy vs Epochs (b) YOLOv10 Accuracy Epochs

Fig. 6. Training performance and accuracy of YOLOv8 and YOLOv10.



(a) YOLOv11 Accuracy vs Epochs (b) YOLOv12 Accuracy vs Epochs

Fig. 7. Training performance and accuracy of YOLOv11 and YOLOv12.

Table. 1. Performance Comparison of YOLO Variants on Underwater Dataset.

Metric	YOLOv8	YOLOv10	YOLOv11	YOLOv12
Precision(B)	0.873	0.845	0.886	0.866
Recall (B)	0.867	0.845	0.845	0.880
mAP@50(B)	0.906	0.906	0.915	0.915
mAP@50-95 (B)	0.551	0.562	0.562	0.560

Table. 2. Class-wise mAP@50 Accuracy for YOLO Variants.

Class	YOLOv 8	YOLOv 10	YOLO v11	YOLOv1 2
Sea Cucumber	0.839	0.843	0.843	0.843
Sea Urchin	0.954	0.956	0.964	0.957
Scallop	0.927	0.921	0.939	0.946

6 Conclusions

In this study we have done a comparative analysis using various deep learning models like YOLOv8, YOLOv10, YOLO v11, YOLOv12 for underwater object detection under challenging situations including poor visibity and noisy environments. By using annotated datasets and implementing preprocessing techniques segmentation and also morphological operations the proposed model achieved a better performance for different versions of YOLO for real time underwater applications and also it can be further used for the advancements in marine monitoring.

References

[1] P. Song, P. Li, L. Dai, T. Wang, and Z. Chen, “Boosting R-CNN: Reweighting R-CNN samples by RPN’s error for underwater object detection,” Neurocomputing, vol. 530, pp. 150–164, Elsevier.

[2] J. Ling, C. Zhang, and D. Du, “Underwater object detection algorithm based on improved YOLOv7-tiny,” Proceedings of the ICCSNT, 2023, pp. 28–31.

[3] A. Thakur, A. Dubey, R. Vashisth, and S. Tomar, “An improved underwater object detection based on YOLOv8 segmentation,” 2023.

[4] K. Ganesan, A. John, B. Saritha, and M. Manoj, “A deep learning approach to detecting objects in underwater images,” Cybernetics and Systems, 2023, pp. 1–16.

[5] X. Chen, M. Yuan, C. Fan, X. Chen, Y. Li, and H. Wang, “Research on an underwater object detection network based on dual-branch feature extraction,” Electronics, vol. 12, 3413, 2023.

[6] J. Feng and T. Jin,“CEH-YOLO: A composite enhanced YOLO-based model for underwater object detection,” Ecological Informatics,vol.82,102758, 2024.

[7] M. S. Asyraf, I. S. Isa, M. I. F. Marzuki, S. N. Sulaiman, and C. C. Hung, “CNN-based YOLOv3 comparison for underwater object detection,” 2021.

[8] Z. Liu, Y. Zhuang, P. Jia, C. Wu, H. Xu, and Z. Liu, "A novel underwater image enhancement and improved underwater biological detection pipeline," Journal of Underwater Imaging Techniques, 2022.

[9] K. Liu, Q. Sun, D. Sun, M. Yang, and N. Wang, "Underwater target detection based on improved YOLOv7," Proceedings of the ACM Multimedia Conference, 2023.

[10] L. Chen, Y. Huang, J. Dong, Q. Xu, S. Kwong, H. Lu, H. Lu, and C. Li, "Underwater object detection in the era of artificial intelligence: Current, challenge, and future," IEEE on Artificial Intelligence, 2024.

[11] M. Jain, "DeepSeaNet: Improving underwater object detection using EfficientDet and YOLO variants," Proceedings of the Deep Learning for Oceanography Workshop, 2024.
[12] Y. Ouassine, J. Zahir,Conruyt, M. Kayal, P. A. Martin, E. Chenin, L. Bigot, and R. Vignes Lebbe, "Automatic coral detection with YOLO: A deep learning approach for efficient and accurate coral reef monitoring," Conference on Marine Biological Detection, 2024.
[13] S. Jain, "DeepSeaNet: Improving underwater object detection using EfficientDet and YOLO variants," Proceedings of the Deep Learning for Oceanography Workshop, 2023.
[14] Z. Liu, B. Wang, Y. Li, J. He, and Y. Li, "UnitModule: A lightweight joint image enhancement module for underwater object detection," arXiv preprint arXiv:2309.04708, 2023.
[15] J. Wang and N. Yu, "UTD-YOLOv5: A real-time underwater targets detection method based on attention improved YOLOv5," arXiv preprint arXiv:2207.00837, 2022.