

Real-Time Underwater Trash Object Detection

Using Enhanced YOLOv8

Adimulam Raghuvira Pratap
Department of Computer Science and
Engineering
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada, India
pratapadimulam@vrsiddhartha.ac.in

Thumu Rampraveen Reddy
Department of Computer Science and
Engineering
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada, India
rampraveen2710@gmail.com

Lanke Nikhila
Department of Computer Science and
Engineering
Velagapudi Ramakrishna Siddhartha
Engineering College
Vijayawada, India
nikhilalanke04@gmail.com

Abstract— The accumulation of trash in aquatic and marine environments has severe and significant consequences on marine ecosystems, resulting in a persistent environmental hazard. While volunteer trash pickup efforts have grown in recent years, they still demand considerable time, effort, and resources. To effectively address these challenges, the implementation of automated solutions is essential. This project aims to develop an efficient object detection deep learning model which helps in effectively identifying the underwater trash objects. The key challenges of underwater image-based localization come from object detection under low light conditions, also the presence of numerous small and overlapping targets, which significantly increase the complexity. In this project we have used improved YOLOv8n object detection model achieving an impressive accuracy rate of mAP of 0.951. We have utilized diverse datasets covering a range of environmental conditions, including challenging low-light scenarios. Also, the integration of advanced preprocessing techniques and data augmentation has further enhanced our model's performance, ensuring reliable predictions in various underwater environments, thereby improving detection performance without introducing additional computational complexity.

Keywords— Trash detection, Marine pollution, Image enhancement, EfficientNet-B0, YOLOv8n.

I. INTRODUCTION

In recent times, there has been a drastic increase in the amount of trash being discarded into our water bodies. Shockingly, [1] approximately 70% of this trash sinks into the water, raising a significant threat to aquatic life and disrupting the delicate underwater ecosystem. The increasing amount of trash [27] being thrown into water bodies has turned our oceans into dumping grounds. Consequently, marine animals are struggling to survive in their own habitats. It is our responsibility to act as guardians of our oceans, keeping them clean and safe for all aquatic life. Manual cleaning methods, while noble in intent, are often labor-intensive, time-consuming, and impractical for large-scale ocean cleaning efforts. Relying solely on manual methods is not a sustainable or effective solution to address the widespread problem of underwater trash pollution.

To enhance cleaning activities in underwater environments, there is a growing need for technological advancements. Autonomous Underwater Vehicles (AUVs), typically used for surveillance, can be repurposed for cleaning activities. However, the primary focus should be on detecting trash objects in the underwater environment. Underwater

environments present significant challenges for trash detection, including low lighting conditions, turbidity, and water movements. Additionally, detecting a trash object itself poses challenges, as the object may not be in a fixed position. There is a need to develop a model capable of identifying trash objects from various angles and positions in diverse underwater environments.

This paper employs modified YOLOv8 (You Only Look Once version 8) for the detection and classification of trash objects. The dataset utilized is the Trash-ICRA19 dataset, consisting of images capturing various types of marine debris in different environmental conditions. Given the significant variability in water clarity and light quality across the dataset, we apply histogram equalization to enhance color contrast in potentially color-distorted images. Further, the dataset is augmented using mosaic augmentation technique within the YOLOv8 model [5] itself. The dataset comprises 9,576 images, each labeled with bounding. The model is trained and tested against a range of underwater scenarios to evaluate its performance.

The main goal of this paper is to efficiently detect the underwater trash objects. Underwater images captured using cameras typically exhibit blue or green shades. To address this, we employ histogram equalization to adjust the color contrast, enhancing the visibility of trash objects. Additionally, we utilize the mosaic data augmentation technique to improve detection from various positions. We enhance the feature extraction capabilities of the YOLOv8 architecture by replacing its feature extraction module with EfficientNet-B0, which enables better performance across different underwater scenarios.

This paper presents a unique approach to underwater trash object detection using an improved YOLOv8 model, which enhances detection accuracy and efficiency in challenging underwater environments.

II. LITERATURE REVIEW

The incorporation of various deep learning and computer vision techniques has resulted in major developments in the field of underwater trash detection.

[4] ChunMing Wu worked on improving the YOLOv5 algorithm for detecting underwater garbage. This research involved replacing the YOLOv5 feature extraction module with MobileNetV3 and incorporating the CBAM to improve feature extraction in both spatial and channel dimensions.

[3] YOLOv8, with its high accuracy (mAP 0.84) and rapid detection capabilities, outperforms YOLOv5 in underwater environments, addressing challenges of object shape variability and low visibility.

Wei Zhou's [6] work introduced a novel object detection network for identifying underwater marine debris. This network features CSPDarknet53 backbone, which combines an efficient channel attention module with depthwise over-parameterized convolution to extract deep semantic features of the debris. The inclusion of the PixelShuffle PANET module for enhanced feature fusion, which uses dilated parallel modules (DPMs) with multiscale dilated rates, improves detection accuracy for objects of different sizes.

In the realm of underwater vision enhancement and object detection, a significant study by researchers proposed a CNN approach [8] to address weak illumination in underwater images. This study utilised both the max-RGB and shades of gray methods to enhance images and modified the CNN structure using improved schemes. This features a downsampling layer followed by a convolution layer, showed superior results and achieved a mean Average Precision (mAP) of about 90%.

P. Srinivas Babu [7] introduced an improved configurable underwater object detection system using YOLO-based CNN. This system detects objects in real-time, sending captured images with their coordinates to a specified email address. This model is effective in various underwater scenarios, demonstrating the practical applicability of real-time detection systems.

Gangqi Chen [23] came up with a hybrid transformer-based framework for detecting small underwater objects. Using a lightweight transformer-based network and a fine-grained feature pyramid network, this framework efficiently detects small and feeble objects in degraded underwater images. It utilizes global contextual information to enhance the understanding of the underwater scene.

Further improvements in underwater object-detection were seen with an improved YOLOv5-based framework [29] proposed by researchers. This framework incorporated several key enhancements, including the RTMDet backbone framework with a Cross-Stage Partial Layer featuring a large convolution kernel for capturing comprehensive contextual information. In addition, a common convolution layer was added to efficiently extract valuable information from images, resulting in superior mean Average Precision (mAP@0.5) on underwater data.

Neha Singh [11] made advancements in detection, classification, and segmentation of underwater objects using advanced computer vision techniques. A notable study integrated YOLOv8 with TC-YOLO and the SAM algorithm to enhance performance in underwater environments. The TC-YOLO adds a Transformer module and CA modules for better image processing, while the SAM algorithm improves real-time segmentation, crucial for understanding complex underwater environments.

Corrigan [28] proposed neural network models, Mask R-CNN and YOLACT, for real-time segmentation of underwater trash litter in images. YOLACT's lightweight nature enables faster detection, making it suitable for litter collection tasks using AUVs, while Mask R-CNN is well-suited for surveying trash distribution.

These studies collectively highlight the significant progress in underwater trash detection, employing various innovative techniques and deep learning models to enhance accuracy and efficiency in challenging underwater environments. These insights and technological advancements provide a strong foundation for our research on utilizing improved YOLOv8 for detecting and classifying underwater trash objects.

III. METHODOLOGY

A. Dataset Description

The primary dataset used for training the trash detection model is "Trash-ICRA19". It is a bounding-box labelled dataset of underwater trash. This dataset, derived from the J-EDI marine debris dataset, includes videos captured in real-world underwater environments, featuring varied depths, lighting conditions, and object types. The dataset contains marine debris, remotely operated vehicles (ROVs), and biological entities such as plants and animals, from which 5,700 annotated images were extracted.

To further enhance the model's robustness, additional images from the TrashCAN 1.0 dataset were used, including those captured in various underwater environments. The cumulative dataset comprises 9,576 images, categorized into three classes: Bio (biological objects), ROV (remotely operated vehicles), and Trash (underwater debris) and their distribution is shown in Fig. 1. Images were divided into three sets: training (76%), validation (19%), and testing (5%), for model development and evaluation.

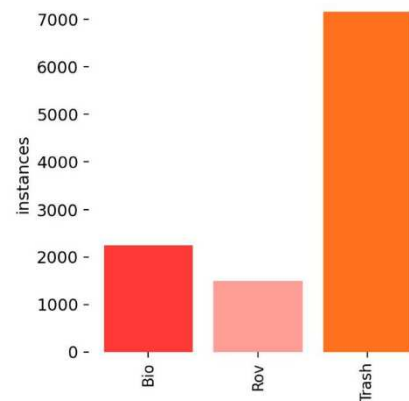


Fig 1: Class-wise Distribution of Images in Dataset

B. Dataset Pre-processing

To ensure optimal performance in underwater trash detection, a series of advanced preprocessing techniques [18] were applied to the dataset, addressing key challenges such as noise, color distortion, uneven illumination, and poor contrast, resulting in clearer and more accurate input images.

1) *Homomorphic Filtering*: To address the non-uniform illumination commonly found in underwater environments, homomorphic filtering was applied. This technique separates the illumination and reflectance components of the image, allowing the adjustment of illumination to normalize lighting and enhance contrast. Enhancing the contrast makes the

boundaries of trash objects more discernible, improving detection accuracy.

2) *Wavelet Denoising*: Wavelet denoising was utilized to mitigate the effects of Gaussian noise, which is prevalent in underwater images. This technique preserves essential image features, such as edges, while effectively removing noise. Reducing noise improves the model's accuracy in identifying and localising trash objects.

3) *Contrast-Limited Adaptive Histogram Equalization (CLAHE)* [16]: CLAHE was used for improving the local contrast without over-enhancing low-light underwater images. CLAHE adaptively clips the histogram, ensuring that the contrast is increased without amplifying noise. This technique improves the visibility of underwater objects, rendering them more distinguishable from the background.

4) *Anisotropic Filtering*: To smooth homogeneous regions while preserving and enhancing edges, anisotropic filtering was applied. This technique helps improve object segmentation accuracy, ensuring that trash objects are accurately delineated from the background. By emphasizing the edges, the model can identify and localize the trash more effectively.

5) *Color Channel Equalization*: To address the color distortions often present in underwater images, color channel equalization was performed. This technique balances RGB channels, corrects predominant color biases, and ensures that objects are more visually distinct. Reducing colour distortions helps in improving the model ability in accurately classifying the trash objects.

C. Improved YOLOv8 Architecture

To effectively address the challenges of underwater trash detection, we implemented architectural modifications to YOLOv8, focusing on computational efficiency, feature refinement, and sensitivity to small and irregularly shaped objects. The following sections detail these key enhancements. Fig. 2 represents the architecture diagram of proposed system.

1) *Backbone Optimization with EfficientNet-B0*: The original YOLOv8 backbone was replaced with EfficientNet-B0, chosen for its ability to optimize feature extraction while minimizing computational load. EfficientNet-B0 employs a compound scaling technique that proportionally scales network depth, width, and input resolution. This balanced scaling enables EfficientNet to maintain high representational power while achieving a significant reduction in floating-point operations (FLOPs). The backbone's use of depthwise separable convolutions reduces parameter overhead, allowing the model to perform efficiently in diverse underwater conditions with minimal computational complexity.

2) *Enhanced Multi-Scale Detection and Feature Fusion*: To improve detection across scales, particularly for small and overlapping objects, we introduced an additional multi-scale detection head. This head captures high-resolution, shallow-layer feature maps that retain fine-grained spatial details essential for small object detection. The multi-scale head's output is combined with other layers to form a comprehensive feature map, allowing the model to effectively adapt to objects of various sizes.

3) *Adaptive Attention Mechanisms for Region and Channel Focus*: Underwater environments introduce complex lighting and visibility issues, which necessitate precise attention mechanisms. We integrated Coordinate Attention to improve spatial localization and enhance the model's focus on key image regions. Coordinate Attention encodes positional information along two spatial directions, enabling the network to capture both short- and long-range dependencies without excessive computational overhead. The module generates separate 1D encodings along the height and width of the feature map. These encodings are passed through a shared MLP to yield adaptive attention weights, enhancing the ability of model to distinguish objects from background noise, particularly in low-contrast settings.

4) *Deformable Convolution and Adaptive Objectness Scoring*: To further adapt to the spatial variability of underwater trash objects, Deformable ConvNets v2 were integrated into the Neck and Bottleneck layers. Unlike traditional convolutional layers, which operate on a fixed grid, deformable convolutions dynamically adjust their sampling positions. This flexibility allows the model to better capture spatially diverse features, which is critical for detecting irregularly shaped trash objects. The addition of modulation scalars further refines the influence of each sampling point, resulting in more accurate feature extraction.

D. Model Training and Evaluation

Coming to the training of our modified YOLOv8 model, began with loading and preparing the dataset covering various types of environments, with each image annotated. To ensure balanced evaluation and minimize overfitting, we split this dataset into training and validation sets. We also implemented caching techniques, which sped up data retrieval during training by keeping images and labels readily accessible, ultimately making the training process more efficient.

To accelerate the model's learning and improve initial accuracy, we have initialized the YOLOv8 modified model with 'yolov8n.pt' pre-trained weights, allowing the model to build on prior knowledge from similar object detection tasks. This initialization helped the model recognize general patterns, textures, and structures often found in underwater environments.

For optimization, we used the AdamW optimizer, configured with momentum of 0.9 and a step size of 0.001429. AdamW's adaptive nature and weight decay feature enhanced the stability of model and helped in preventing the overfitting, allowing it to adapt effectively during training. To enhance the robustness of model in varying underwater environments, we have incorporated data augmentation methods like blurring, grayscale transformation, and CLAHE (contrast-limited adaptive histogram equalization). These augmentations introduced variations in lighting, perspective, and object appearance, making the model more resilient and capable of generalizing to new data.

Hyperparameter tuning was another essential step in this process. Through iterative testing, we fine-tuned the batch size, number of epochs, learning rate, and momentum to find the optimal combination for maximum accuracy and

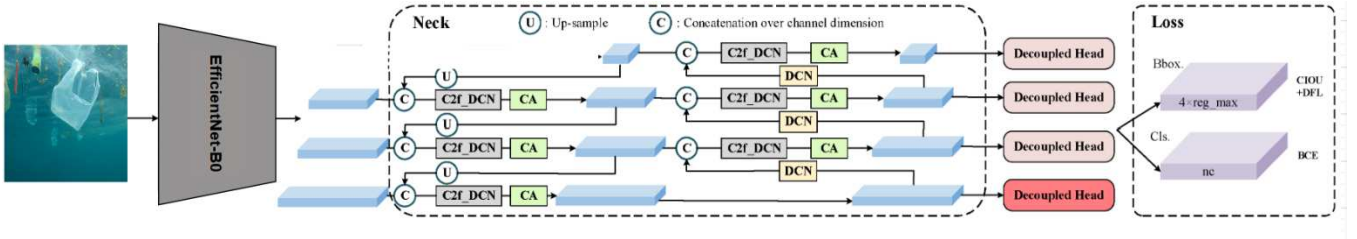


Fig 2: Architecture Diagram

convergence speed. This tailored approach ensured that our model was well-suited to the unique challenges of underwater detection.

E. Evaluation Metrics

We used a few evaluation metrics for assessing the performance.

1) *Mean Average Precision (mAP)*: This metric provides an overall measure of detection accuracy by calculating the average precision across all object classes and IoU thresholds. It balances the model's precision and recall, offering a comprehensive view of detection quality.

2) *Precision*: This measures accuracy of positive predictions by determining the ratio of true positive detections to the total number of predicted positives, reflecting the effectiveness model in reducing false alarms and irrelevant classifications.

3) *Recall*: Recall measures the model's ability to detect actual objects by determining the fraction of relevant instances it correctly identifies in an image.

4) *F1-Score*: F1-Score merges precision and recall into one measure using their harmonic mean, making it valuable for assessing performance when balancing false positives and false negatives.

5) *Intersection over Union (IoU)*: IoU evaluates the overlap between the model's predicted box and the actual object's boundary. A higher IoU threshold (e.g., 0.5 or more) signifies better localization by requiring greater alignment for a detection to be valid.

IV. RESULTS AND ANALYSIS

This Result and Analysis section presents the evaluation of the modified YOLOv8 model, highlighting its performance across key metrics and visualizations. Each metric and graph offer's valuable insights of the model performance in identifying underwater trash and its consistency across various evaluation parameters.

A. Results

The modified YOLOv8 model exhibited strong performance on key metrics, achieving a precision of 0.913, which indicates that 91.3% of detected objects were correctly classified. With a recall of 0.918, indicating model had identified 91.8% of actual objects present. The model's mAP at an IoU threshold 0.5 (mAP50) was 0.951, demonstrating its capability to accurately localize objects with an acceptable degree of overlap. Additionally, the mAP averaged across the IoU thresholds from 0.5 to 0.95 was 0.747, providing a wider view of detection performance across different levels of localization strictness. Table I summarizes these metrics for

each class, including "Bio", "Rov", and "Trash", illustrating consistent accuracy across different object types.

TABLE I: RESULTS OF THE PROPOSED SYSTEM

Class	Precision	Recall	mAP50	mAP50-95
All	0.913	0.918	0.951	0.747
Bio	0.925	0.939	0.965	0.799
Rov	0.874	0.934	0.945	0.786
Trash	0.942	0.880	0.942	0.656

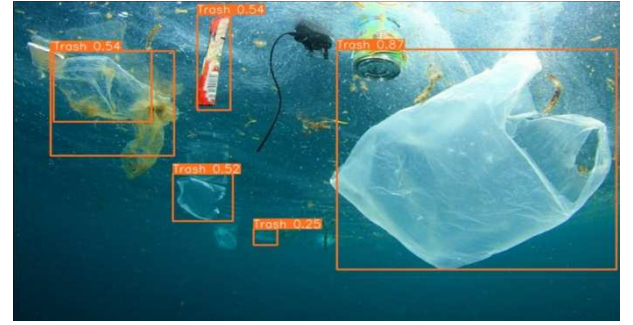


Fig 3: Test Image

Fig. 3 shows the testing of a new image with the model, the prediction shows that the model is able to predict the trash objects of even small size.

B. Precision-Recall Curve

The Precision-Recall (PR) curve, shown in Fig. 4 offers a visual depiction of the model's performance. High precision values for classes such as "Bio" (0.965), "Rov" (0.945), and "Trash" (0.942) indicate that the model consistently

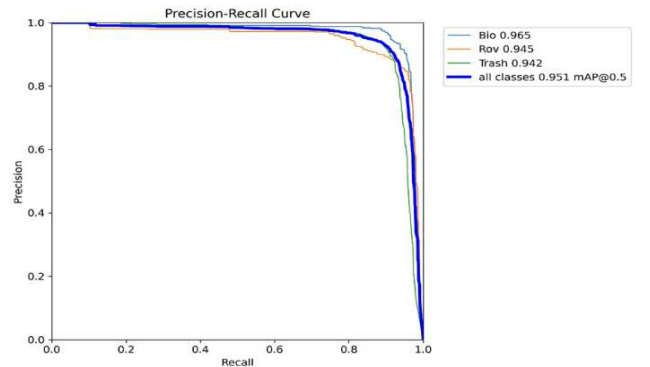


Fig 4: Precision-Recall Curve

minimizes false positives. This curve illustrates the model's capability to achieve a high detection rate while accurately classifying objects with minimal errors, making it a valuable tool for assessing performance on individual classes.

C. Train/Box Loss and Train/Obj Loss Curves

The Train/Box Loss and Train/Obj Loss curves in Fig. 5 visualize the training dynamics by tracking the loss incurred during bounding box regression (Train/Box Loss) and object classification (Train/Obj Loss). These curves indicate the model's convergence rate over training epochs, helping assess whether it reached a stable point and optimized effectively. The gradual reduction in loss values demonstrates effective learning, confirming that the model consistently improved its predictions as training progressed.

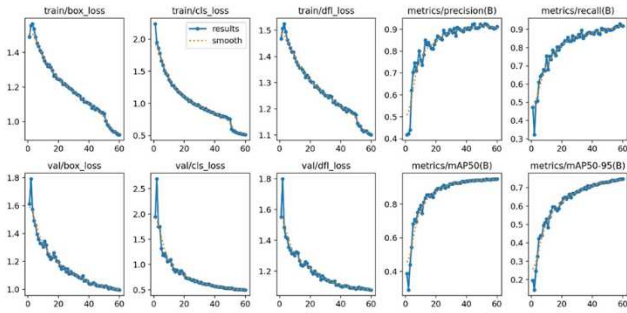


Fig 5: Train/Box Loss and Train/Obj Loss Curves

D. Confusion Matrix

The confusion matrix in the Fig. 6, offers a comprehensive overview of the model's predictions for each class, identifying correct classifications as well as misclassifications. For instance, the model correctly classified 472 instances as "Bio" but misclassified 68 instances into other categories. For the "Rov" class, it correctly identified 386 instances and misclassified 93. Similarly, in the "Trash" class, 1570 instances were correctly classified, while 188 instances were misclassified. This matrix highlights the model's strengths in accurate classification while identifying areas where misclassifications occur more frequently, guiding potential areas for further model refinement.

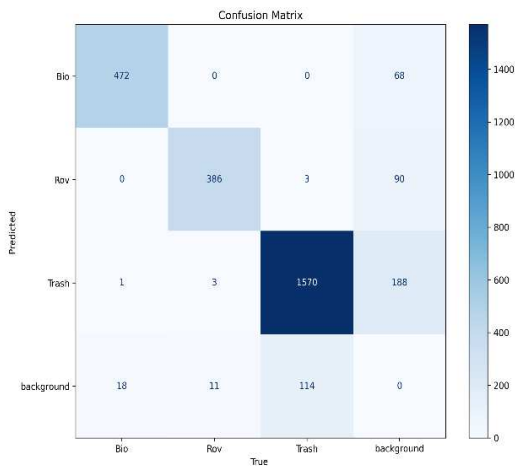


Fig 6: Confusion Matrix

E. Comparative Analysis

Table II provides a comprehensive comparison of this work with previous studies, highlighting advancements in underwater trash detection and the effectiveness of the proposed approach.

TABLE II: COMPARATIVE ANALYSIS

Work	Algorithm	Accuracy
[3]	YOLOv8	84%
[7]	YOLO based CNN	79%
[8]	CNN approach	90%
[11]	Integrated YOLOv8 with TC YOLO & SAM algorithm	95%
[23]	hybrid transformer network	80%
[28]	YOLACT and Mask R-CNN	90%
[29]	Improved YOLOv5	79.8%
This work	Improved YOLOv8	95.1%

V. CONCLUSION

In conclusion, this study introduces a robust underwater trash detection system developed using the YOLOv8 object detection algorithm, achieving an impressive accuracy rate of 95.2%. Our system's effectiveness is reinforced by the use of diverse datasets that account for challenging environmental conditions, including low-light scenarios, along with advanced preprocessing and data augmentation techniques that improve prediction reliability. The results demonstrate the potential of advanced machine learning technologies to address pressing environmental issues, emphasizing the role of continued research in marine conservation. Future advancements could involve integrating autonomous underwater vehicles (AUVs) for automated trash collection, enhancing the system's capability to operate in difficult-to-reach or hazardous environments. Furthermore, expanding the system to enable real-time monitoring and data analysis could offer valuable insights into patterns of marine debris, aiding in the development of targeted policies for more effective conservation efforts.

REFERENCES

- [1] Barnhart, Andrew, et al. "The Great Pacific Garbage Patch." (2017).
- [2] Derraik, J. G. B., "The pollution of the marine environment by plastic debris: a review," *Marine Pollut. Bull.*, vol. 44, no. 9, pp. 842-852, 2002.
- [3] D. Pavani et al., "YOLO object detection method for underwater trash collecting robot: Octacleaner," *IEE (TEMSCON-ASPAC)*, Bengaluru, India, 2023.
- [4] Wu, C., Sun, Y., Wang, T., et al., "Underwater trash detection algorithm based on improved YOLOv5s," *J. Real Time Image Proc.*, vol. 19, pp. 911-920, 2022.

- [5] Bawankule et al., "Visual detection of waste using YOLOv8," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, IEEE, 2023, pp. 869-873.
- [6] W. Zhou et al., "YOLOTrashCan: A deep learning marine debris detection network," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1-12, 2023, Art. no. 5002012, doi: 10.1109/TIM.2022.3225044.
- [7] P. S. Babu et al., "Underwater Object Detection Using Image Processing," *IJRESM*, vol. 4, no. 7, pp. 315-319, Jul. 2021.
- [8] Han et al., "Underwater Image Processing and Object Detection Based on Deep CNN Method," *J. Sensors*, vol. 2020, Article ID 6707328, 2020.
- [9] Chen, X. et al., "Research on an Underwater Object Detection Network Based on Dual-Branch Feature Extraction," *Electronics*, vol. 12, no. 16, p. 3413, 2023.
- [10] Liu, Q., Huang, W., Duan, X., et al., "DSW-YOLOv8n: A New Underwater Target Detection Algorithm Based on Improved YOLOv8n," *Electronics*, vol. 12, no. 18, p. 3892, 2023.
- [11] Underwater Marine Life Study using Yolo V8 - PALAK SAINI, DEEPALI YADAV, NEHA SINGH - IJFMR Volume 5, Issue 6, November-December 2023. DOI 10.36948/ijfmr.2023.v05i06.10565
- [12] Jia et al., "Underwater Object Detection in Marine Ranching Based on Improved YOLOv8," *J. Mar. Sci. Eng.*, vol. 12, no. 1, p. 55, 2024.
- [13] Zhang et al., "Efficient Small-Object Detection in Underwater Images Using the Enhanced YOLOv8 Network," *Appl. Sci.*, vol. 14, no. 3, p. 1095, 2024.
- [14] Fulton et al., "Trash-ICRA19: A Bounding Box Labeled Dataset of Underwater Trash," Data Repository for the University of Minnesota, 2020.
- [15] Hong et al., "TrashCan 1.0 An Instance-Segmentation Labeled Dataset of Trash Observations", 2020.
- [16] L. Lu et al., "Comparative study of histogram equalization algorithms for image enhancement," *Mobile Multimedia/Image Proc. Sec. Appl.*, vol. 7708, pp. 337-347, 2010.
- [17] Jian et al. "Underwater image processing and analysis: A review," *Signal Processing: Image Communication* 91 (2021): 116088.
- [18] Y. Qi et al., "A comprehensive overview of image enhancement techniques," *Arch. Comput. Methods Eng.*, pp. 1-25, 2021.
- [19] P. Kaur et al., "Data augmentation for object detection: A review," in *2021 IEEE Int. Midwest Symp. Circuits Syst. (MWSCAS)*, pp. 537-543, 2021.
- [20] A. Guo et al., "A lightweight YOLOv8 integrating FasterNet for real-time underwater object detection," *J. Real-Time Image Proc.*, vol. 21, no. 2, pp. 1-15, 2024.
- [21] Z. Li et al., "Underwater object detection based on improved transformer and attentional supervised fusion," *Inf. Technol. Control*, vol. 52, no. 2, pp. 397-415, 2023, doi: 10.5755/j01.itc.52.2.33214.
- [22] L. Ge et al., "Advanced deep learning framework for underwater object detection with multibeam forward-looking sonar," *Struct. Health Monit.*, vol. 0, no. 0, 2024, doi: 10.1177/14759217241235637.
- [23] G. Chen et al., "HTDet: A hybrid transformer-based approach for underwater small object detection," *Remote Sens.*, vol. 15, p. 1076, 2023, doi: 10.3390/rs15041076.
- [24] S. Shah and J. Tembhurne, "Object detection using convolutional neural networks and transformer-based models: A review," *J. Electr. Syst. Inf. Technol.*, vol. 10, p. 54, 2023, doi: 10.1186/s43067-023-00123-z.
- [25] G. Kalaierasi et al., "A deep learning approach to detecting objects in underwater images," *Cybernet. Syst.*, 2023, doi: 10.1080/01969722.2023.2166246.
- [26] D. L. Rizzini et al., "Investigation of vision-based underwater object detection with multiple datasets," *Int. J. Adv. Rob. Syst.*, vol. 12, no. 6, 2015, doi: 10.5772/60526.
- [27] J. I. Gerber, "Study: Up to 11 million tons of plastic waste on the sea bottom," *Trending Topics*, Apr. 8, 2024, accessed May 9, 2024.
- [28] B. C. Corrigan et al., "Real-time instance segmentation for detection of underwater litter as a plastic source," *J. Mar. Sci. Eng.*, vol. 11, p. 1532, 2023, doi: 10.3390/jmse11081532.
- [29] J. Zhang et al., "An improved YOLOv5-based underwater object-detection framework," *Sensors*, vol. 23, p. 3693, 2023, doi: 10.3390/s23073693.