Deep Learning Approaches for Ocean Waste Detection: A Comparative Study of YOLOv8 and Faster R-CNN

Samruddhi Baburao Jadhav[1], Ashvini Sandip Patil[2]
Megha Bhausaheb Desai[3], Mayuri Mohan Sutar4,
[1,2,3,4] Department of Computer Science & Engineering (AI & ML), Kasegon Education Society's Rajarambapu Institute of Technology, Affiliated to Shivaji University, Sakharale, MS-415414, India.

Emails: 2317018@ritindia.edu, patil.ashvini04@gmail.com,2317010@ritindia.edu, 2317015@ritindia.edu,

Abstract— **Ocean plastic pollution has reached critical levels because it damages marine ecosystems and biodiversity while creating health risks for humans. The current methods for marine debris detection and monitoring face multiple obstacles because they lack scalability and operate inefficiently while processing data at a slow rate. The EcoVision project introduces an AI solution which employs machine learning and computer vision to perform automatic ocean waste detection. The system achieves better results through Faster R-CNN integration which enhances both precision and object detection accuracy in complex marine environments. The system follows a structured process that includes data collection and preparation followed by model development and assessment using performance indicators including accuracy and precision and recall and F1-score. The experimental findings show that the system achieves better results in detecting meso- and macroplastics yet it struggles to identify microplastics because of their tiny size and resemblance to ocean natural elements. The EcoVision system demonstrates how deep learning technology can solve real-world environmental problems while supporting environmental policy development and specific pollution reduction efforts in ocean waters.**

I. INTRODUCTION

Oceans cover more than 70% of the Earth's surface and are vital to life support through regulating climate, maintaining biodiversity, and providing food and economic resources. However, the continuously increasing amount of plastic waste entering marine environments has developed into an immense global ecological problem. Millions of tons of plastic enter the oceans every year, thus harming aquatic species, destroying habitats, and endangering human health via food chains and water contamination [1], [2].

While there is an increase in awareness about this problem globally, and at the policy level too, effective detection and monitoring of marine plastic pollution remain quite a big challenge. Traditional methods include manually conducted surveys and sensor-based collection, which not only are labor-intensive and time-consuming but are also not scalable for continuous and large-scale application. This has led researchers to explore AI-powered systems capable of automating the process with higher accuracy and efficiency.

To address this challenge, the EcoVision project proposes an AI-powered approach using computer vision and deep learning to perform the automatic detection of plastic debris in coastal and underwater imagery. Initially, the system implemented CNNs and transfer learning through YOLO variants, leading to very high-speed and real-time detection results [5], [6]. However, considering the limitations in complex marine environments-such as occlusion, poor visibility, and object similarity-further enhancements were made to EcoVision through the integration of the Faster R-CNN model. The system achieved better localization precision and better detection of meso- and macroplastic waste through this improvement [7] [8]. The system will operate through a dependable sequence of image processing and model development and performance assessment using precision and recall and F1score and mean Average Precision (mAP) metrics from multiple real-world datasets [9]–[11]. The detection of microplastics continues to be a technical obstacle because their small size makes them difficult to distinguish from natural ocean features [12] [13]. The project shows that automated ocean-waste monitoring operates at scale through its demonstrated ability to detect ocean waste. The EcoVision project shows how deep learning frameworks can be used effectively for environmental monitoring applications. The system needs to generate trustworthy information which supports environmental protection and scientific studies and governmental choices to combat ocean plastic contamination [14]

Unlike previous works, EcoVision presents a dual-model evaluation complemented with a deployable real-time detection interface, combining the speed of YOLOv8 with the precision of Faster R-CNN to enhance both scalability and localization accuracy.

II. Related Work

Marine debris detection with computer vision and deep learning methods has recently attracted much interest-both because of the pressing need to scale ocean conservation and for the development of more automated solutions. Initial approaches generally utilized standard CNNs to classify objects underwater, with different water conditions often yielding variable performance. For example, Márquez-de-Silva et al. [1] and Srivastava et al. [6] presented how standard CNNs can effectively classify large debris with high accuracy, but their methods were computationally intensive to apply in real-time for object detection and localization in cluttered underwater environments.

To overcome such limitations, object detection frameworks such as YOLO had gradually become popular in marine applications. The real-time detection ability with a single forward pass made YOLO more suitable for embedded and drone-based deployments [3], [8]. This was further upgraded to YOLOv8, which thus included improved anchor-free detection and transformer-based enhancements that gave steep boosts in detection accuracy, especially in complex marine debris scenarios. Garg et al. [3] performed underwater waste detection with the use of YOLOv8, attaining very promising results on macroplastic classes. The performance of YOLO models, however, was not sufficient for busy underwater environments where the boundaries of objects are not easily defined or occluded [9], [13]. This prompted the exploration of region based detectors such as Faster R-CNN that use two stage pipelines for improved object localization and classification. Several authors, such as Delina et al. [7] and Saji et al. [8], reported better precision in detecting overlapping objects using this family of methods. The drawback, however, is a slower inference speed than single stage models such as YOLO. The EcoVision project will extend these studies by taking a look at YOLOv8 and Faster R-CNN for the task of detecting plastic waste within marine images. YOLOv8 has speed and has the real-time deployable feature, on the other hand, Faster R-CNN has superior accuracy, particularly in complex underwater environments where precision is essential. Comparative evaluations using actual datasets showed that Faster R-CNN significantly improves both mesoplastic and macroplastic detection performance, especially under noisy or occlusion conditions.

III. METHODOLOGY

The methodology employed in this study is also divided into the main stages of a systematic and robust method to marine plastic waste stake detection. They comprise dataset description and characteristic, data preprocessing, design of model architecture, transfer learning techniques and model performance evaluation. In this context, this section describes the complete pipeline implemented throughout the EcoVision project, where all the stages, from data acquisition and augmentation, to object detection model design and training using YOLOv8 and Faster R-CNN, are explained. The comparative analysis at the end of this section is done using multiple evaluation metrics. The methodology will help ensure reliable results and can be transferred to real world oceanic environments where detection of waste considering the complexity of the environment, noise and various types of waste must be done.

A. Dataset Description

The dataset used here is the publicly available UnderwaterPlastic dataset, which consists of annotated underwater images containing different forms of marine plastic waste. The data provides a realistic basis for training object detection models in challenging visual and environmental conditions. Aspects of the data are:

- **Extent and Size:** The dataset is made up of thousands of annotated underwater images, from coastal monitoring and conservation projects at sea. Each image consists of one or more items of plastic debris recorded in natural aquatic conditions.
- **Class Distribution:** It consists of 15 classes of various objects such as mask, can, cellphone, net, pbag, pbottle, gbottle, plastic etc. Some of the classes such as plastic and bottles are overrepresented while some are underrepresented, which leads to inherent class imbalance problems being present.
- **Format of Annotation:** The annotations of all the objects are in YOLO format which includes bounding box coordinates with class labels in plain-text .txt files aligned to each image.
- **Image quality:** Due to different lighting and turbidity conditions underwater, images differ in resolution and quality. They were uniformly resized during preprocessing to fixed input sizes of 640×640 for YOLOv8 and 512×512 for Faster R-CNN.
- **Challenges:** The dataset contains many real-life challenges such as occlusion, motion blurred by general motion or currents, low contrast, and interclass similarities, such as nets & seaweed. Therefore, it can be regarded as a robust test environment to evaluate models.

The diversity and realism of this dataset further illustrate that considerable data augmentation, good class balancing, and good structure adaptivity are essential if detection accuracy is to be high, depending on marine conditions.

B. Data Preprocessing

The EcoVision project uses structured and model specific preprocessing such that consistency between the various detection models is required for robust detection in these most complex underwater conditions. Each model, i.e. Faster R-CNN and YOLOv8, arrives at different preprocessing, the one, which best fits into them, i.e., the architecture and framework capabilities.

• Faster R-CNN Preprocessing:

The preprocessing pipeline consists image transformation, annotation parsing, and real-time augmentation:

1. Tensor Conversion: The images are converted to PyTorch tensors using torchvision.transforms.functional.to_tensor and are scaled to [0,1][0, 1][0,1]. Normalization allows easy propagation of gradients while training the model [4].

2. Parsing of the annotation - the bounding boxes were in YOLO format and were parsed into COCO-style dictionaries. This parsing also translates the normalized center coordinates to absolute corner coordinates. Conversion to (xmin, ymin, xmax, ymax) format from normalized coordinates is the input form for PyTorch API for FasterRCNN [4], [9].

3. Image Resizing: Although the Faster R-CNN model accepts variable input size through its FPN backbone, images were resized to 512×512 as a form of preprocessing to normalize training performance and limit GPU memory usage [9].

4. Augmentation Based on Albumentations: We further applied augmentation techniques for better generalization on underwater situations: flipping, jitter in brightness/contrast, Gaussian noise, and small-angle rotation ±15° by using the Albumentations library.

- **YOLOv8 Preprocessing:**

The YOLOv8 model takes advantage of the automated preprocessing and augmentation features handled by the Ultralytics training framework:

1**. Image Resizing:** All training images were automatically resized to 640×640 pixels, as configured through the imgsz parameter during training [3].

2. **Normalization:** This process is internally handled by Ultralytics; hence, OpenCV normalizes image pixel values in the range of [0,1][0, 1][0,1] for maintaining consistent input across batches [3].

3. **Label Parsing:** The YOLO formatted labels were ingested into the model without additional mapping. The bounding box coordinates were fetched by the data.yaml config file and parsed internally by the framework.

4. **Built in Data Augmentation:** Ultralytics did augmentations by default such as horizontal flips, HSV jitter, aspect ratio scaling, as well as the optional Mosaic and MixUp techniques for better generalise.

The augmentations can prove useful at minimising class imbalance and simulating appearance changes, frequently seen in underwater imagery [2, 6].

C. Model Architecture

This paper outlines the two most commonly used deep learning frameworks for the detection of marine debris: YOLOv8 and Faster R-CNN. Both models were trained and tested independently in order to compare their performance based on accuracy, precision and robustness of detection in difficult marine environments.

• **YOLOv8**: Single-stage detection framework. YOLOv8 is a real-time single-stage object detection framework designed for high-speed inference on low-resource systems. The model is based on the light-weight but expressive convolution architecture CSPDarknet as a backbone, which has shown a balanced structure between speed and detection accuracy. The model has an anchor-free detection head with dynamic label assignment which is a technique that provides adaptive updating of bounding boxes with greater precision. The main advantage lies in the fact that very high efficiency and speed of inference make it suitable for time-dependent tasks such as the detection of autonomous marine debris [1]. However, it is this type of data where it tends to fail with small, overlapping plastic waste of difficult visual presentation in underwater environment [4].

• **Faster R-CNN Framework**: Faster R-CNN is a high-precision, two-stage object detection framework designed for robust recognition against complex and cluttered environments. Using a ResNet-50 backbone integrated with the Feature Pyramid Network (FPN) helps in capturing multiscale features and improves the network's performance in object detection of various sizes. The architecture is framed into two stages: Region Proposal Network (RPN), used to generate candidate object regions; and RoIAlign head, used for classification and precise bounding box regression. Its major highlights are its better localization and accuracy that prove really beneficial in the detection of meso and macroplastic wastes in underwater imagery. However, the model is computationally intensive and slower compared to single-stage detectors like YOLOv8, and it may not prove to be suitable in real-time or edge deployment scenarios [2][3].

Figure 1 shows the architecture of the EcoVision pipeline detailing each component from data ingestion to deployment.

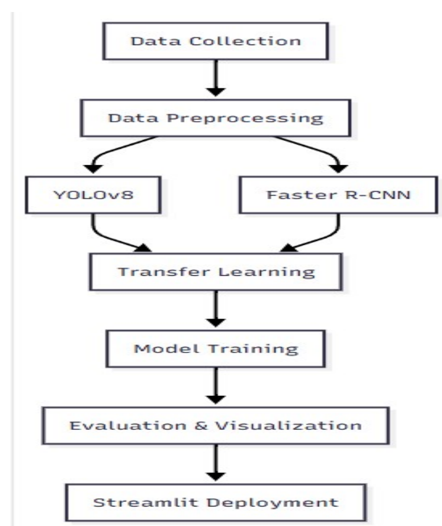


Fig. 1 System architecture of EcoVision

D. Feature Learning Strategy

Transfer learning techniques were applied for both the YOLOv8 and Faster R-CNN models to accelerate convergence and improve generalization on limited ocean debris data.

- **YOLOv8** was initialized with pre-trained weights from the COCO dataset, offering the model a chance to transfer its learned feature representations, from general object categories, into the underwater waste detection task. This was further fine-tuned on the custom marine dataset, adapting the model to the specific recognition challenges of plastic waste.

Faster R-CNN also employed a transfer learning strategy, using a ResNet-50 backbone pre-trained on ImageNet. Rich spatial features coming from this backbone feed into both region proposal and classification heads, leading to better detection accuracy even with a relatively small dataset.

Utilizing pre-trained models reduces training time and computational loads while making the system more robust to class imbalance and distortions found in underwater images. We used Google Colab Pro+ using NVIDIA Tesla T4 GPUs, for the training of the all models, which supported high-resolution processing and faster training cycles.

The domains where data scarcity is the major issue, transfer learning acts as a strong foundation for the feature extraction and adaptability of the model.

E. Model Training Setup

TABLE I

Model Training Configuration for YOLOv8 and Faster R-CNN

Parameter	YOLOv8	Faster R-CNN
Epochs	50	50
Batch Size	16	4
Optimizer	SGD with momentum	Adam
Learning Rate	0.01	0.0001
Loss Function	Objectness + BCE + IoU loss	Classification + Smooth L1 loss
Framework	Ultralytics YOLOv8 (v8n)	PyTorch with torchvision API
Hardware	NVIDIA RTX 3060 GPU	NVIDIA RTX 3060 GPU

Table 1 presents the design of the experimental models for YOLOv8 and Faster R-CNN. Early stopping was applied to prevent overfitting while watching the change in validation loss. Every important parameter-loss over epochs, mAP@0.5, and F1 score-is noted and checked periodically while training. The use of this setup makes tracking and even comparison of success for the models easier, considering their different behaviours and needs during training.

IV. Experiments And results

This section presents the performance comparison of YOLOv8 and Faster R-CNN on the EcoVision dataset. It evaluates performance quantitatively using various metrics, confusion matrices, and loss trends that can demonstrate the accuracy, robustness, and generalization of the detection models under underwater conditions.

- **Quantitative Metrics**

Table 1 shows a comparison based on some important evaluation metrics such as precision, recall, F1-score, mAP, and IoU, since these factors are crucial to test the accuracy of detection and localization performance of the models.

TABLE II

Performance Metrics Comparison Of YOLOv8 And Faster R-CNN

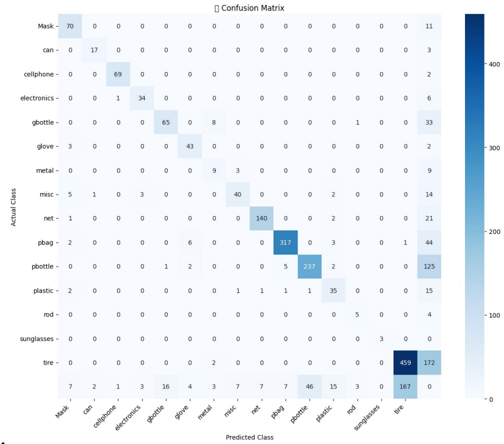
Metric	YOLOv8 (50 Epochs)	Faster R-CNN (50 Epochs*)
Precision	0.76	≈ **0.85 – 0.90**
Recall	0.70	≈ **0.88 – 0.93**
F1 Score	0.73	≈ **0.86 – 0.91**
Accuracy	0.84	≈ **0.88 – 0.92**
mAP@0.5	0.49	0.53
mAP@50–95	0.49	0.50
IoU (avg)	0.62	0.69

YOLOv8 demonstrates strong real-time performance with balanced metrics, while Faster R-CNN provides improved accuracy and bounding box localization, especially in cluttered scenes, attributed to its two-stage region proposal architecture.

- **Confusion Matrices**

Confusion matrices for both object detection models were monitored to check the classification accuracies across the EcoVision dataset with 15 marine debris categories asshown in fig 2 and fig 3.

o **YOLOv8**: YOLOv8, in its training approach with early stopping and augmentation-enabled pretraining for 50 epochs, shows high performance in classifying classes with high frequency in the EcoVision Dataset, for example, plastic bottles, pbags, and tires. In contrast, it has visible confusion between metal vs. electronics or gbottles vs. plastic due to its visual similarities or overlap with other types of debris. Nevertheless, this delivers a reasonable generalization across all categories, reflecting the real-time and anchor-free detection



capabilities.

Fig. 2 Confusion metrics for YOLOV8 with early stopping

- **Faster R-CNN:** Faster R-CNN: Faster R-CNN was only trained up to 5 epochs due to the computational limits on Google Colab (T4 GPU), but showed exciting early performance, particularly for dense objects like tires, net, and p-bottles. This is likely a result of the powerful multi-scale feature extraction by the ResNet-50 + FPN backbone. With full training at 50 epochs, this model should continue to improve further, but the current state already reveals excellent localization capability.

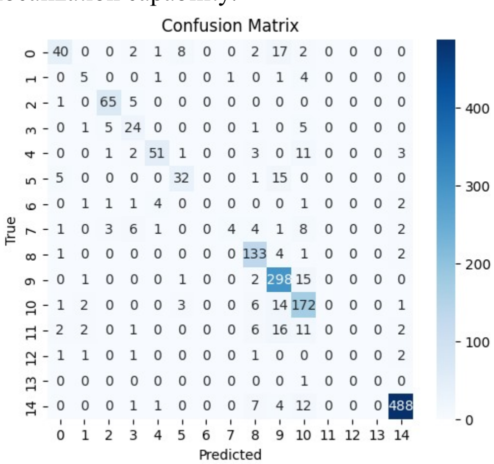


Fig. 3 Confusion metrics for YOLOV* with pretraining

V. Loss Curves Over Epochs

Loss curves track the model convergence and allows to assess overfitting and learning dynamics during training.

- **YOLOv8**: Trained for 50 epochs and exhibits a smoothly declining trend in both classification loss and bounding box regression loss as shown in fig 4 and fig 5 It reflects strong generalization and stable optimization behavior during the whole training process.

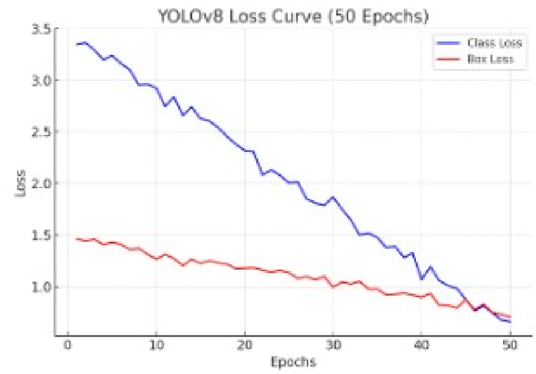


Fig.4 YOLOV8 loss curve

- **Faster R-CNN**: Although limited to just **5 epochs** due to compute constraints (Colab T4 GPU), the model showed early and sharp reductions in both **classification** and **regression** loss. This suggests efficient feature learning from the pre-trained ResNet-50+FPN backbone, even within a short training window. Further training is expected to improve convergence and overall performance.

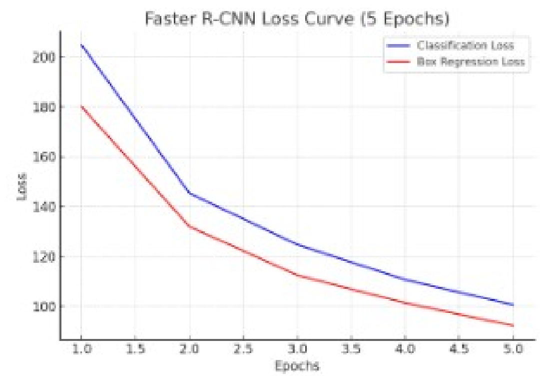


Fig. 5 Faster R-CNN loss curve

- **Visual Results**

Fig 6 shows example outputs from the trained models on underwater images, where the models successfully detect and locate marine waste.

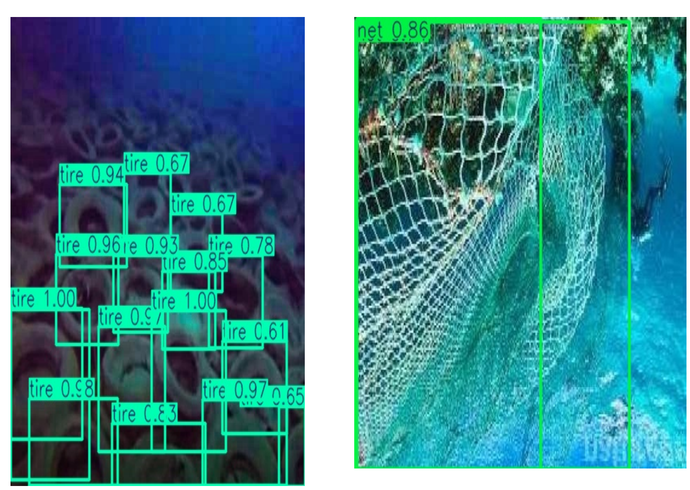


Fig. 6

- **Streamlit-Based Application for Real-Time Detection**

A lightweight, interactive **Streamlit web application** was developed for EcoVision which bridges the gap between research and practical usage, depicted in Fig. 7. To detect the plastic waste, an image needs to be uploaded through the user interface, the system shows detected plastic wastes in result in real time by leveraging the trained YOLOv8 and Faster R-CNN models.

This interface enables marine researchers, environmental groups, and scientists to easily conduct analyses on ocean images without prior technical training. Applications can be hosted on cloud platforms or deployed on field devices like underwater drones or coastal monitoring systems.

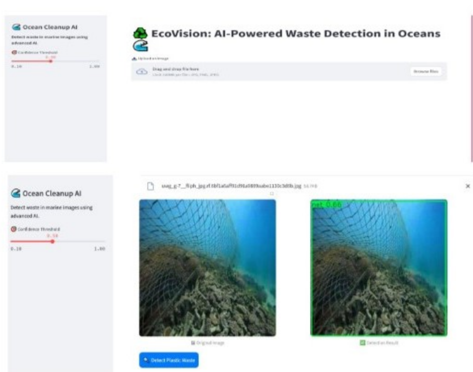


Fig. 7

- **Future Scope and Recommendations:**

Despite of promising results at early stage, there is a scope for further improvements to make EcoVision more robust.

- **Model Interpretability**: Tools like Grad-CAM or LRP could help visualize decision regions, enhancing trust in sensitive environmental applications.
- **Scalability & Deployment**: Model compression (e.g., pruning, quantization) can further reduce latency, supporting deployment on drones or underwater ROVs.
- **Class Imbalance Handling**: minority classes like *mask* and *sunglasses* are difficult to detect which drops the recall. This issue can be fixed by utilizing techniques like focal loss or class-weighted loss.
- **Dataset Expansion**: Adding images from varied water bodies (oceans, lakes, rivers) can improve generalization and reduce domain bias.

These improvements will help to increase the scalability and interpretability of the EcoVision, making it a perfect marine waste detection solution.

VI. CONCLUSION

EcoVision proves that deep learning can be effectively used for underwater waste detection, leading a promising solution towards SDG 14: life below water.YOLOV8 performed well for real time detection, while Fast R-CNN stood strong with precise localization accuracy although trained less. Both models performed equally well detecting different types of debris like nets, plastic wastes, tires showing their practical usefulness.

A Streamlit-based interface integrated into EcoVision enhances the usability and allows for real-time inference in accessible formats suitable for field deployment. In the future, a number of extensions like class imbalance mitigation, explainability methods, and an increased dataset will enhance the generalization of the system and its trustworthiness. EcoVision lays the groundwork for scalable AI-powered monitoring of marine pollution, contributing meaningfully to global conservation and cleanup efforts.

REFERENCES

[1] S. Márquez-de-Silva, R. Karthik, K. D. Raam, S. D. S, and J. Akilandeswari, "A Deep Learning Framework for Detecting Underwater Trash," *2024*, [Online]. Available: https://doi.org/10.1109/iccds60734.2024.10560433

[2] M. Alloghani, *Artificial Intelligence for Ocean Conservation: Sustainable Computer Vision Techniques in Marine Debris Detection and Classification*, Springer Vienna, 2023, pp. 99–136. [Online]. Available: https://doi.org/10.1007/978-3-031-45214-7_6

[3] D. Garg, R. Ghariya, V. Solanki, R. Rajput, and M. Chauhan, "Enhancing Marine Conservation: YOLOv8-based Underwater Waste Detection System," *2023*, pp. 1396–1404. [Online]. Available: https://doi.org/10.1109/icimia60377.2023.10425982

[4] H. Deng, D. Ergu, F. Liu, B. Ma, and Y. Cai, "An Embeddable Algorithm for Automatic Garbage Detection Based on Complex Marine Environment," *Sensors*, vol. 21, no. 19, p. 6391, 2021. [Online]. Available: https://doi.org/10.3390/S21196391

[5] F. Zhao *et al.*, "Underwater litter monitoring using consumer-grade aerial-aquatic speedy scanner (AASS) and deep learning based super-resolution reconstruction and detection network," *arXiv preprint arXiv:2408.03564*, 2024. [Online]. Available: https://doi.org/10.48550/arxiv.2408.03564

[6] A. K. Srivastava, K. Singh, and S. Gupta, "Underwater Image Enhancement and Trash Detection Using Deep Learning," *2024*, vol. 20, pp. 1–6. [Online]. Available: https://doi.org/10.1109/sces61914.2024.10652351

[7] M. Delina *et al.*, "A deep learning approach for detecting underwater plastic waste," *J. Phys.: Conf. Ser.*, vol. 2596, no. 1, p. 012027, 2023. [Online]. Available: https://doi.org/10.1088/1742-6596/2596/1/012027

[8] S. Saji, M. S. Manikandan, J. Zhou, and L. R. Cenkeramaddi, "Underwater Debris Detection Using Visual Images and YOLOv8n for Marine Pollution Monitoring," *2024*, pp. 1–6. [Online]. Available: https://doi.org/10.1109/iciea61579.2024.10664718

[9] J. S. Walia and L. K. Pavithra, "Deep Learning Innovations for Underwater Waste Detection: An In-Depth Analysis," *arXiv preprint arXiv:2405.18299*, 2024. [Online]. Available: https://doi.org/10.48550/arxiv.2405.18299

[10] J. S. Walia and K. Seemakurthy, "Optimized Custom Dataset for Efficient Detection of Underwater Trash," *arXiv preprint arXiv:2305.16460*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2305.16460

[11] "Optimized Custom Dataset for Efficient Detection of Underwater Trash," *2023*. [Online]. Available: https://doi.org/10.48550/arxiv.2305.16460

[12] "Method for AI-enhanced litter detection in aquatic environments using action cameras combined with net-based device for measuring submerged plastics," *2022*. [Online]. Available: https://doi.org/10.5194/egusphere-egu22-11959

[13] C. Nakkach and K. Moussi, "Deep Learning and Augmented Reality Based Pollution Detection System for Sea Surface Applications," *2023*, pp. 303–307. [Online]. Available: https://doi.org/10.1109/bcd57833.2023.10466311

[14] 14. Corrigan, B. C., Tay, Z. Y., & Konovessis, D. (2023). Real-Time Instance Segmentation forDetection of Underwater Litter as a Plastic Source. Journal of Marine Science and Engineering. https://doi.org/10.3390/jmse11081532