JEE Journal of Ecological Engineering

Journal of Ecological Engineering 2024, 25(7), 58–69
https://doi.org/10.12911/22998993/187970
ISSN 2299–8993, License CC-BY 4.0

Received: 2024.04.17
Accepted: 2024.05.14
Published: 2024.06.01

Exploring Deep Learning for Underwater Plastic Debris Detection and Monitoring

Abdelaadim Khriss[1*], Aissa Kerkour Elmiad[1], Mohammed Badaoui[2], Alae-Eddine Barkaoui[3], Yassine Zarhloule[3]

[1] Laboratory LARI, FSO, Mohammed Premier University, Oujda, Morocco
[2] Laboratory MSD, ESTO, Mohammed Premier University, Oujda, Morocco
[3] Laboratory 2GPMH Lab., FSO, Mohammed Premier University, Oujda, Morocco
* Corresponding author's e-mail: abdel.abdelkrs@gmail.com

ABSTRACT

In this paper, a comparative evaluation of state-of-the-art deep learning models for object detection in underwater environments focusing on marine debris detection was presented. The performance of four prominent object detection models was investigated, including: Faster R-CNN, SSD, YOLOv8, and YOLOv9, using two different datasets: TrashCAN and DeepTrash. Through quantitative analysis, the accuracy, precision, recall, and mean average precision (mAP) of each model across different object classes and environmental conditions were evaluated. The obtained results show that YOLOv9 consistently outperforms the other models, demonstrating superior precision, recall, and mAP values on both datasets. Furthermore, the stability and convergence behavior of the models during training were analyzed, highlighting the excellent stability and adaptability of YOLOv9. The obtained results underscore the effectiveness of deep learning-based approaches in marine debris detection and highlight the potential of YOLOv9 as a robust solution for environmental monitoring and intervention efforts in underwater ecosystems.

Keywords: marine debris monitoring, deep learning, YOLOv9, YOLOv8, faster rcnn, ssd.

INTRODUCTION

Recent research has highlighted the profound impact of plastic waste on marine pollution. This pollution is primarily caused by various types of plastic waste that are littered on land and can either float or sink into different layers of the ocean. These plastics pose a direct threat to marine life through ingestion or entanglement, resulting in metabolic disruption and, in severe cases, mortality in aquatic species (Anjana et al., 2020; Li et al., 2016). The impacts extend beyond aquatic ecosystems to humans and coastal economies (Alabi et al., 2019). Seafood consumers may be exposed to marine plastic pollution as small plastic particles find their way into the organs of marine organisms. Several efforts are underway to quantify floating marine plastics in the marine epipelagic layers, where organisms surface to obtain oxygen and sunlight (Markic et al., 2020). Quantifying floating marine debris is achieved through current monitoring methods that use tools such as manta trawls and marine vehicles such as interceptors that operate in open waterways to collect surface plastics. However, these strategies face challenges due to their labor-intensive nature, high cost, and potential harm to marine life (Belioka et al., 2023). To address these challenges, remotely operated vehicles (ROVs) can withstand harsh ocean conditions, making them suitable replacements for manual underwater visual inspections. Leveraging the advancements in computer vision technology, ROVs are now equipped to detect objects underwater, as demonstrated in (Krause et al., 2020; Nava et al., 2023). This technology aids ROVs in various tasks, including the detection and inspection of plastic debris, tracking its movement, scene reconstruction, and other related activities (McLean et al., 2020; Corrigan et al., 2023; Aguirre-Castro et al., 2019; K Jothikrishna et al., 2023). Conventional techniques for detecting plastic debris in environments have traditionally relied on edge information extracted from images. However, these methods frequently encounter diminished accuracy due to adverse underwater conditions. Recent advancements have led to a deeper exploration of convolutional neural networks (CNNs) as a superior approach for enhancing the detection of plastic debris in marine environments (Khriss et al., 2024).

However, detecting both floating and underwater debris poses distinct challenges. For floating debris detection, issues include detecting small objects with indistinct features and weather conditions affecting detection accuracy (Zhang et al., 2024; Qiao et al., 2022). In contrast, underwater debris detection faces challenges, such as water clarity, depth, and light availability, influencing factors like water turbidity, complex background conditions, and lighting variations. These factors contribute to noise, texture distortion, uneven illumination, low contrast, and limited visibility in underwater images (Jesus et al., 2022). For further details on these challenges, readers may refer to studies (Lavers et al., 2016; Khriss et al., 2024).

Several studies have investigated the effectiveness of different deep learning models and datasets in detecting and classifying marine debris, particularly plastic waste, in various oceanic environments. Many research efforts have aimed to improve the identification of marine debris and litter in underwater habitats using a variety of deep learning models and datasets. (Watanabe et al., 2019) used YOLOv3 to enhance garbage detection and identify debris floating on the ocean surface. Their work achieved accuracies of 69.6% and 77.2% in detecting undersea life and marine debris, respectively. Notably, their study shifted the focus of marine debris detection from the ocean surface to the deep ocean, broadening the scope of detection efforts. In Fulton et al. (2019), the performance of different object detection models, namely YOLOv2, Tiny-YOLO, Faster R-CNN, and SSD, was evaluated using the J-EDI dataset to detect marine debris. The assessment was based on mAP and Average Intersection over Union (Avg. IoU) scores. YOLOv2 achieved an mAP of 47.9% and an Avg. IoU of 54.7%, while Tiny-YOLO scored 31.6% in mAP and 49.8% in Avg. IoU. Faster R-CNN displayed superior performance with an mAP of 81.0% and an Avg. IoU of 60.6%. SSD attained an mAP of 67.4% with an Avg. IoU of 53.0%. (Tata et al., 2021) evaluated the effectiveness of YOLOv4-Tiny and YOLOv5-S on the Deep-Trash and JAMSTEC JEDI datasets. They demonstrated high precision and accuracy in identifying marine plastic debris, achieving a mAP of 85%. Their study highlighted the model's capability to distinguish plastic from non-plastic objects with similar appearances. (Aleem et al., 2022) investigated augmentation and preprocessing methods, resulting in similar accuracies in the latest experiments. Using the forward looking sonar image (FLS) Marine Debris Dataset, they implemented image preprocessing techniques, including median and CLAHE filters. Employing Faster R-CNN with VGG16 and ResNet architectures, they attained overall accuracies of 93% and 91%, respectively. These findings underscore the significance of augmentation and preprocessing approaches in achieving comparable accuracies. Another study in the paper (M Bhanumathi et al., 2022) investigated the performance of the YOLOv4 and YOLOv5 algorithms for detecting marine plastics in epipelagic water layers. Scrapped dataset images from the internet were employed for evaluation, with YOLOv5-S exhibiting significantly higher precision, mAP, F1-score, and inference speed than YOLOv4. The algorithm achieved an impressive 85% mAP for image inputs. (Singh et al., 2023) used diverse open-source datasets and videos representing various ocean environments from different countries. Their study compared the performance of YOLOv7, YOLOv5s, YOLOv6s, Faster R-CNN, and Mask R-CNN. YOLOv7 and YOLOv5s demonstrated superior performance with mAP scores of 96% and precision-recall scores of 96–93%, respectively, highlighting their effectiveness in marine debris detection tasks.

While these studies have made significant strides in leveraging deep learning models to detect plastic debris in underwater environments, several critical gaps in the current research landscape have emerged. Firstly, the absence of standardized benchmarking protocols and datasets poses a significant challenge in accurately assessing and comparing the performance of different detection models. Without a consistent benchmarking framework, it becomes difficult to discern the relative strengths and weaknesses of various approaches, hindering the progress towards more effective detection methodologies. Secondly, while some studies have conducted comparative analyses of different deep learning architectures on specific datasets, there remains a notable dearth of research comparing the performance of these architectures across diverse datasets. Such comparative analyses are essential for understanding the generalizability and robustness of detection models under varying environmental conditions. The performance of several state-of-the-art models was comprehensively assessed, including the YOLO variants incorporating the latest advancements such as YOLOv8 (Glenn et al., 2023) and YOLOv9 (Wang et al., 2024) architecture, alongside Faster RCNN (Ren et al., 2015) and single-shot detector (SSD) (Liu et al., 2016) object detection. The selection of models is based on several considerations. These include the proven performance of the models in object detection tasks, their established reputation within the computer vision community, their use of different methodologies, the inclusion of recent versions to show progress, and their encouraging results in similar tasks. The evaluation was conducted using two distinct datasets: the Trashcan Dataset (Hong et al., 2020) and the DeepTrash Dataset (JAMSTEC 2012). By leveraging these datasets, which offer diverse underwater environments and debris scenarios, the authors aimed to provide a thorough comparison of the detection capabilities of different models under varying conditions. Through meticulous analysis and evaluation, the conducted study sought to identify the most effective models for underwater debris detection, thus contributing to the advancement of solutions for preserving marine ecosystems.

The rest of the paper is structured as follows. The materials and methods section outlines the dataset characteristics, deep learning models employed, experimental procedures, and evaluation metrics. In Results, findings are presented alongside visual aids and compared with existing methods. Discussion interprets results in the context of plastic debris detection, acknowledging study limitations and suggesting future research directions. Finally, conclusion summarizes key findings and highlights the contribution of the paper and future work directions for environmental monitoring and deep learning.

MATERIALS AND METHODS

This section, presenting the methodology set up and elaboration on trained deep learning models, begins with the selection of YOLOv8, YOLOv9, Faster R-CNN, and SSD methods. These approaches were chosen for their established track record of achieving remarkable performance in object detection tasks across various domains (Tan et al., 2021). Each method brings unique strengths to the table, including real-time processing capabilities and high precision in detecting objects of diverse shapes and sizes. By employing these methodologies, the authors aimed to provide a comprehensive evaluation of their effectiveness in detecting and locating marine debris and other underwater objects, thereby contributing to the advancement of environmental monitoring and intervention strategies.

Network architecture

Faster R-CNN

The Faster Region-based Convolutional Neural Network (Faster R-CNN) architecture represents a significant advancement in object detection, and operates through a multi-stage pipeline, Figure 1 with remarkable accuracy. Faster R-CNN introduces the concept of region proposals, which are candidate bounding boxes generated by a selective search algorithm or similar methods. The faster R-CNN pipeline consists of the following key steps:

- region proposal – initially, the input image undergoes a selective search algorithm to propose a set of region proposals, which are potential bounding boxes containing objects. These proposals are typically generated based on various cues such as color, texture, size, and shape;
- feature extraction – each region proposal is then warped to a fixed size and fed into a pre-trained CNN, such as VGG, ResNet, or AlexNet, to extract feature representations. This step results in a feature vector for each region proposal;
- classification and localization – the feature vectors are used to classify the content of each proposed region and refine their bounding box coordinates. This is achieved through additional layers in the network, often involving separate branches for classification and bounding box regression;
- non-maximum suppression (NMS) – to address redundancy among region proposals, a non-maximum suppression algorithm is applied. This step filters out highly overlapping bounding boxes, retaining only the most confident detections;
- The loss function combines classification loss and bounding box regression loss. Represented as:

$$L_{total} = \lambda_{cls} L_{cls} + \lambda_{reg} L_{reg} \quad (1)$$

where: L_{cls} is the classification loss, which measures the discrepancy between the predicted class probabilities and the ground truth class labels; L_{reg} is the bounding box regression loss, which measures the difference between the predicted bounding box coordinates and the ground truth bounding box coordinates; λ_{cls}, λ_{reg} are hyperparameters used to balance the influence of the classification and regression terms.

YOLO

Unlike traditional methods that rely on multi-stage pipelines, YOLO adopts an innovative one-stage approach, revolutionizing object detection by conducting it in a single forward pass through the network. This eliminates the need for computationally intensive region proposal techniques and post-processing steps, resulting in faster inference speeds. Real-time detection is made possible with a single pass through the network, as depicted in Figure 2. By employing a unified CNN, this architecture directly predicts bounding boxes and class probabilities from the entire image, thereby simplifying the detection process. To calculate the loss function, several components are considered, encompassing object localization, objectness score, and class prediction. The total loss is a combination of these components, where each term contributes to the overall optimization objective of the YOLO model.

$$L = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \left[\begin{array}{c} (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + \\ \left(\sqrt{w_i} - \sqrt{\widehat{w}_i}\right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i}\right)^2 \end{array} \right] + \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} (c_i - \hat{c}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} (c_i - \hat{c}_i)^2 + \sum_{i=0}^{S^2} (p_i(c) - \hat{p}_i(c))^2 \quad (2)$$

where: S represents the number of grid cells; B represents the number of bounding boxes per grid cell; λ_{coord}, λ_{obj}, λ_{noobj} are constants used to balance the influence of the different components of the loss function.

These terms collectively contribute to the loss function, guiding the optimization process during training to improve object detection performance.

SSD

The SSD is another pioneering architecture in object detection, offering a unique blend of accuracy and speed. Unlike conventional methods relying on intricate multi-stage pipelines, SSD simplifies the process by integrating object detection into a single neural network, similarly to YOLO, as shown in Figure 2. However, SSD introduces a different approach to achieving real-time inference speeds while maintaining high detection accuracy. In SSD, the detection process involves the simultaneous prediction of bounding boxes and class probabilities at multiple scales within the network. This is achieved by incorporating multiple convolutional layers with different aspect ratios and sizes, allowing the model to effectively detect objects of varying scales and aspect ratios. To effectively train the SSD model, a carefully designed loss function is used to optimize the network parameters. The loss function consists of several components, each contributing to accurately localizing and classifying objects in images.

$$L_{SSD} = \frac{1}{N}(L_{loc} + L_{conf}) \quad (3)$$

where: L_{loc} represents the localization loss, which measures the discrepancy between the predicted bounding box coordinates and the ground truth bounding box coordinates; L_{conf} represents the confidence loss, which quantifies the accuracy of objectness scores and class predictions; N

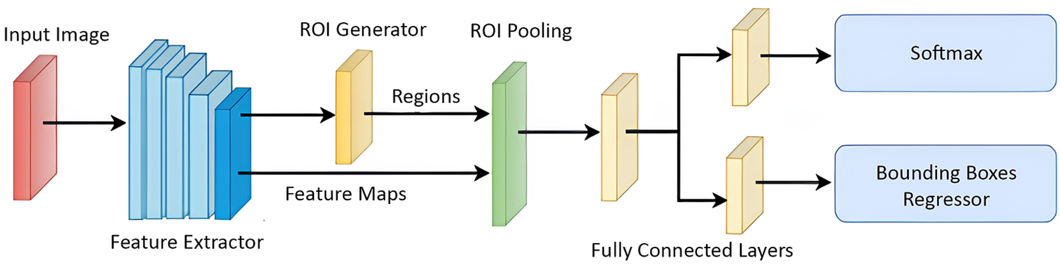

Figure 1. Two stage detector

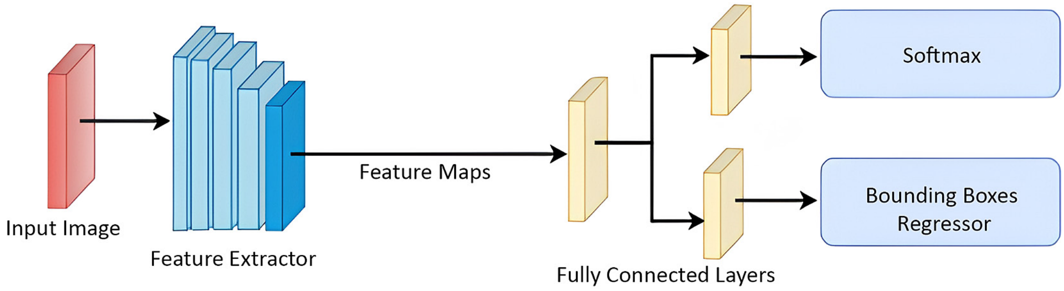

Figure 2. One stage detector

denotes the total number of matched default boxes across all training samples.

The localization loss L_{loc} is calculated using smooth loss, ensuring robustness to outliers and preventing exploding gradients during training. On the other hand, the confidence loss L_{conf} is computed using a softmax function over class scores, penalizing misclassifications and encouraging accurate object predictions.

Dataset

The evaluation of the considered models was conducted using two distinct datasets: the TrashCAN dataset (Hong et al,. 2020) and the DeepTrash dataset (JAMSTEC 2012). The TrashCAN dataset served as the primary training source, consisting of 7212 annotated images portraying underwater debris, ROVs, and marine life. These images were meticulously annotated with bitmaps, for instance, segmentation and bounding boxes, sourced primarily from the JAMSTEC E-Library of Deep Sea Images (J-EDI) dataset, curated by the Japan Agency of Marine-Earth Science and Technology, Figure 3 depicts a sample from the TrashCAN dataset.

In contrast, the DeepTrash dataset was curated from field videos capturing marine plastic across various locations in California, including South Lake Tahoe, Bodega Bay, and San Francisco Bay. This dataset reflects the diverse challenges encountered in real-world marine environments, with variations in quality, depth, and visibility deliberately included to emulate harsh conditions, Figure 4 showcases samples from the DeepTrash dataset. The datasets introduced specific difficulties, including low visibility, visual noise, and objects of different forms. These challenges underscored the importance of robust models capable of handling diverse environmental conditions and object detection scenarios in the context of marine plastic detection.

Experimental setup

In the adopted experimental configuration, the NVIDIA Tesla T4 graphics card paired with the NVIDIA driver version 525.105.17 was used, which ensures seamless compatibility with CUDA version 12.0. This choice was made to take advantage of the powerful computing capabilities of the Tesla T4, which is known for its ability to efficiently handle large datasets due to its substantial GPU memory of 15360 MiB. For hyperparameters, the Adam optimizer with a learning rate set to 0.001 and a batch size of 16 was used. Furthermore, the input images were standardized to a resolution of 640× 640 to maintain consistency across experiments and to ensure compatibility with the architecture.

Evaluation metrics

In evaluating the effectiveness of the model enhancements, a set of evaluation metrics was used to measure the performance of the model

Figure 3. Samples from TrashCAN dataset

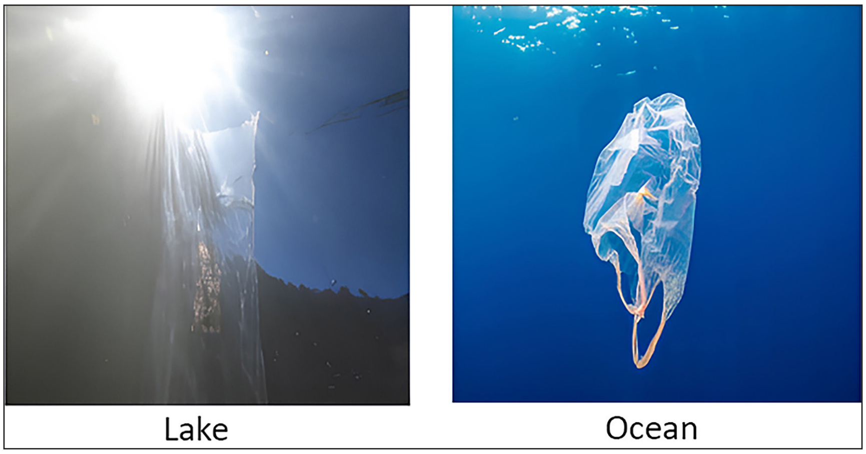

Figure 4. Samples from DeepTrash dataset

across several critical parameters. These metrics were selected to provide a comprehensive understanding of the detection accuracy, false alarm rates, and overall effectiveness of the model in capturing relevant information while minimizing exposure to irrelevant or erroneous data. Among the metrics used, precision, recall, F1 score, and mean average precision (mAP) emerged as central components of the evaluation framework.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (5)$$

$$F1_{score} = \frac{2 \times Precision_{\ class} \times Recall_{class}}{Precision_{\ class} + Recall_{class}} \quad (6)$$

$$mAP = \frac{1}{n} \sum_{i=1}^{n} AP(i) \quad (7)$$

RESULTS

Quantitative analysis

Table 1 presents a comparative evaluation of models across the two distinct datasets: TrashCAN and DeepTrash. YOLOv9 consistently demonstrates superior performance across both datasets, excelling in various metrics. In the TrashCAN dataset, YOLOv9 achieves the highest precision at 87.81% and the best recall at 85%. Additionally, it outperforms other models in mAP at IoU thresholds of both 0.5 90.13% and 0.95 70.22%. Similarly, on the DeepTrash dataset, YOLOv9 maintains its dominance with a precision of 77.76% and the highest recall at 82.14%. It also exhibits exceptional performance in mAP at both IoU thresholds of 0.5 85.33% and 0.95 72.48%. These results underscore the effectiveness and robustness of YOLOv9 in object detection tasks across diverse datasets, consistently surpassing its predecessors, such as YOLOv8, as well as competing models like Faster RCNN and SSD.

In addition, Tables 2 and 3 provide a granular analysis focused on the TrashCAN and DeepTrash datasets, presenting performance metrics for each object class. These detailed insights demonstrate the consistency and robustness of YOLOv9 across a wide range of object classes, reinforcing its position as the leading solution for

Table 1. Evaluating the accuracy of the model

Dataset	Model	Precision (%)	Recall (%)	$mAP_{0.5}$ (%)	$mAP_{0.95}$ (%)
Trash CAN dataset	Faster RCNN	86.9	84.52	88.88	69
	SSD	86.55	84	87.37	68.4
	YOLOv8	87.09	84.55	89	70.1
	YOLOv9	**87.81**	**85**	**90.13**	**70.22**
Deep trash dataset	Faster RCNN	75.85	79.77	84.23	69
	SSD	74.59	79.28	84.44	68.57
	YOLOv8	77.15	81.9	85.28	71.9
	YOLOv9	**77.76**	**82.14**	**85.33**	**72.48**

Table 2. Model performance in marine object detection using the DeepTrash samples

Model	Class	Instances	Precision (%)	Recall (%)	$mAP_{0.5}$ (%)	$mAP_{0.95}$ (%)
YOLOv9	All	2041	86.1	81.9	83.4	65.5
	Plastic	1215	91.4	98.4	99.0	78.9
	Animal and plant	824	88.3	97.2	98.1	85.7
	Shipwrecks	2	78.5	50	53.2	31.8
YOLOv8	All	2041	77.69	80.8	85.3	63
	Plastic	1215	90.54	95.1	98	76.2
	Animal and plant	824	85.45	96	97	83.5
	Shipwrecks	2	76.28	50	51.5	28.6
Faster R-CNN	All	2041	77.3	78.29	84.96	63.1
	Plastic	1215	90.2	94.6	97.35	76
	Animal and plant	824	84.88	95	96.1	82.25
	Shipwrecks	2	75.8	46.55	49.8	28.18
SSD	All	2041	82.6	79.9	80.3	60.93
	Plastic	1215	88.4	96.5	96	77.12
	Animal and plant	824	86.3	95.8	96	76.7
	Shipwrecks	2	77.8	47.58	50.2	30.8

Table 3. Performance metrics for each class in object selection using YOLOv9 on the TrashCAN dataset

Class	Instances	Precision (%)	Recall (%)	$mAP_{0.5}$ (%)	$mAP_{0.95}$ (%)
All	2426	87.8	85.0	90.2	70.3
Animal_crab	83	83.2	66.3	78.0	49.3
Animal_eel	59	86.1	66.1	83.9	54.2
Animal_etc	46	89.1	71.1	86.3	57.3
Animal_fish	149	85.4	80.5	89.2	67.1
Animal_shells	60	78.5	60.8	71.4	44.7
Animal_starfish	59	76.7	88.1	81.9	48.5
Plant	94	91.9	85.1	87.6	58.0
Rov	700	90.2	86.8	94.9	83.4
Trash_bag	174	90.4	85.1	87.6	58.0
Trash_bottle	29	87.5	96.9	96.2	84.2
Trash_branch	70	86.6	92.4	97.2	78.5
Trash_clothing	15	89.6	93.3	92.6	88.6
Trash_container	106	93.0	87.7	94.7	75.3
Trash_cup	17	99.3	100.0	99.5	77.3
Trash_net	23	84.7	78.3	91.4	67.8
Trash_pipe	28	88.5	96.4	96.8	84.4
Trash_rope	16	74.9	93.4	83.1	64.9
Trash_snack_ wrapper	14	95.4	92.9	93.1	83.2
Trash_tarp	24	95.6	83.3	95.5	85.0
Trash_unknown_ instance	550	88.5	82.2	89.4	65.3
Trash_wreckage	27	88.5	85.9	91.9	82.5
Trash_can	83	87.6	90.4	95.6	74.7

Table 4. Performance metrics for each class in object setection using YOLOv9 on the DeepTrash dataset

Class	Instances	Precision (%)	Recall (%)	$mAP_{0.5}$ (%)	$mAP_{0.95}$ (%)
All	2041	86.1	81.9	83.4	65.5
Plastic	1215	91.4	98.4	99.0	78.9
Animal and plant	824	88.3	97.2	98.1	85.7
Shipwrecks	2	78.5	50.0	53.2	31.8

marine debris detection. In contrast, Table 2 provides a comprehensive comparison of the models. YOLOv9 emerges as the best performer, consistently demonstrating the highest precision, recall, and mAP at both 0.5 and 0.95 Intersection over Union (IoU) thresholds across all classes and overall metrics. In particular, YOLOv9 achieves remarkable precision and recall rates, particularly excelling at detecting plastic objects with a precision of 91.4% and a recall of 98.4%. YOLOv8 follows closely behind YOLOv9, showing competitive results but with slightly lower performance metrics in most classes. The faster R-CNN and SSD, while still delivering respectable performance, lag behind YOLOv9 and YOLOv8 in precision, recall, and mAP scores. All models struggle to achieve high precision and recall for the Shipwrecks class, indicating a common challenge in detecting objects within this category. Overall, the results underscore the effectiveness of YOLOv9 in object detection tasks, particularly in scenarios involving multiple classes, and highlight its robustness and reliability across different detection challenges.

Table 3 provides a comprehensive breakdown of the performance metrics for individual object classes within the TrashCAN dataset when using the YOLOv9 object recognition model. Each row in the table corresponds to a specific object class and details the number of instances present and the Precision, Recall, and mAP scores at IoU thresholds of 0.5 and 0.95. Notably, the model demonstrates strong overall performance across all classes, achieving an aggregate precision of 87.8%, recall of 85.0%, mAP0.5 of 90.2%, and mAP0.95 of 70.3%. Analysis of specific classes reveals variations in detection accuracy, with some classes, such as “trash cup” and “trash bottle”, exhibiting higher precision and recall scores, while others, such as “animal crab” and “animal shells”, exhibit comparatively lower scores. In addition, the table highlights class imbalances, with certain classes having significantly more instances than others, potentially impacting overall model performance. The mAP scores provide further insight into the ability of the model to accurately localize objects across different IoU thresholds, indicating its effectiveness in detecting marine debris and other objects within the TrashCAN dataset. In further detail, Figure 6 shows sample results of the object detection process, providing a visual representation of the performance of the model in identifying marine debris and other objects within the TrashCAN dataset.

Model complexity and the stability

The effectiveness of object detection models depends not only on their accuracy and speed, but also on their stability throughout training. In this section, the performance and stability of various object detection models were examined to understand their convergence behavior and adaptability. By analyzing the loss function curves, the authors aimed to identify the strengths and weaknesses of each model in handling marine object detection tasks. In Figure 5, the performance and stability of the training models are compared. YOLOv9 stands out due to its groundbreaking features including programmable gradient information (PGI) and generalized efficient layer aggregation network (GELAN). This model demonstrates remarkable accuracy and speed in object detection tasks. Most significantly, its stability is exceptional, as evidenced by a smooth loss function curve, suggesting efficient convergence and adaptability during training. In contrast, YOLOv8 shows moderate stability compared to YOLOv9. Its loss function curve indicates challenges in convergence, suggesting potential difficulties in effectively training and adapting the model. The faster R-CNN, leveraging a region proposal network (RPN), represents another advancement in object detection. However, it shows lower stability compared to both YOLOv9 and YOLOv8, as evidenced by a fluctuating loss function

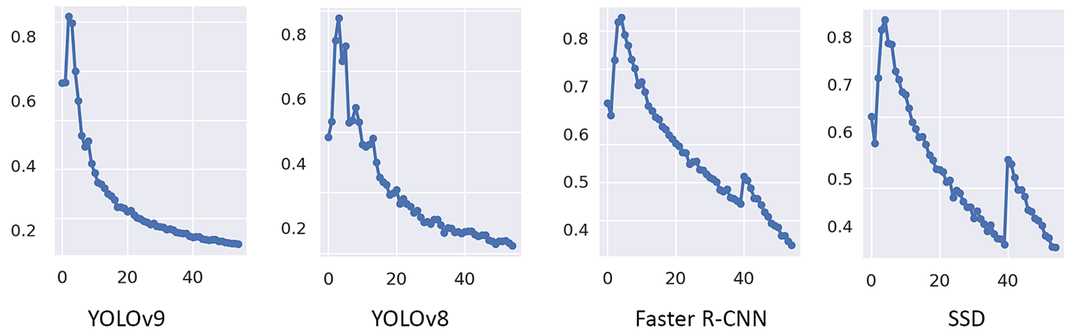

Figure 5. Training stability

Figure 6. Results obtained from employing YOLOv9 on the TrashCAN dataset samples

curve. This variability suggests less consistent convergence during training. Lastly, the SSD demonstrates the lowest stability among the models analyzed. Its erratic loss function curve and limitations in adaptability indicate notable challenges, which could affect its reliability and performance in marine object detection tasks.

DISCUSSION

The quantitative analysis provided valuable insights into the performance of different object detection models, highlighting their effectiveness in detecting marine debris across different datasets. To build on these findings, it is important to explore potential implications and directions for further research and development in this area, including the challenges posed by noisy scenes in marine environments. First, the comparison of object detection models across the TrashCAN and DeepTrash datasets revealed a consistent superiority of YOLOv9 over other models such as Faster RCNN and SSD. This advantage underscores the effectiveness and robustness of YOLOv9 for marine debris detection and positions it as a leading solution for environmental monitoring efforts. Delving deeper into the performance metrics for individual object classes within these datasets, variations in detection accuracy were observed across categories. While YOLOv9 demonstrated strong overall performance, challenges remained in accurately detecting less common objects, such as “animal crabs” and “shipwrecks”. This highlights the need for further optimization in the detection of rare and less common debris objects. In addition, analysis of model complexity and stability underscored the importance of efficient convergence and adaptability during training. The remarkable stability of YOLOv9, attributed to features such as PGI and GELAN, enhances its reliability in practical scenarios. However, other models such as YOLOv8, Faster RCNN, and SSD exhibited varying degrees of stability, suggesting potential implications for their real-world performance and usability. Further exploration of marine debris detection in noisy scenes is a promising area for future research. Noisy scenes, characterized by cluttered backgrounds and varying environmental conditions, pose significant challenges for accurate object detection. Innovative approaches, such as data augmentation, sensor fusion, and domain-specific knowledge integration offer opportunities to improve the robustness and adaptability of object detection models in noisy marine environments.

CONCLUSIONS

In conclusion, the comprehensive analysis of object detection models for marine debris detection provides valuable insights into their performance, stability, and potential for real-world applications. The superiority of YOLOv9 over other models, such as faster RCNN and SSD, across different datasets underscores its effectiveness and robustness in environmental monitoring efforts. While YOLOv9 demonstrates strong overall performance, challenges in accurately detecting less common debris objects still remain, highlighting the need for further optimization and refinement of object detection algorithms. In addition, analysis of model stability underscores the importance of efficient convergence and adaptability during training, with YOLOv9 exhibiting remarkable stability due to its innovative features. Looking forward, further exploration of marine debris detection in noisy scenes represents a promising avenue for future research. Innovative approaches, including data augmentation, sensor fusion, and integration of domain-specific knowledge, offer opportunities to improve the robustness and adaptability of object detection models in challenging marine environments.

Acknowledgements

We would like to express our heartfelt thanks to the contributors of the Remedies project for their generous funding and support provided through the Horizon Innovation Actions - Prevent and eliminate pollution of our ocean, seas, and waters (Horizon-Miss-2021-Ocean-03). We deeply appreciate the dedicated efforts of all project partners, researchers, and collaborators who have worked together to devise and implement effective strategies for the valorization and prevention of plastic litter. Our collective actions are aimed at addressing the pressing challenges confronting our oceans, and we acknowledge the invaluable commitment of everyone involved in advancing Remedies towards significant progress in realizing a sustainable and cleaner future for marine ecosystems.

REFERENCES

1. Khriss, A., Elmiad, K.A., Badaoui M., Barkaoui A. and Zarhloule, Y. 2024. Advances in machine learning and deep learning approaches for plastic litter detection in marine environments. Journal of Theoretical and Applied Information Technology, 102(5).
2. Markic A., Gaertner J.-C., Gaertner-Mazouni N. and Koelmans A.A. 2020. Plastic ingestion by marine fish in the wild. Critical Reviews in Environmental Science and Technology, 50(7), 657–697.
3. Aleem, A., Tehsin S., Kausar S. and Jameel

A. 2022. Target classification of marine debris using deep learning. Intelligent Automation & Soft Computing, 32(1).

4. Bhanumathi, M., Dhanya, S., Gugan, R. and Kirthika, K.G. 2022. Marine Plastic Detection Using Deep Learning. 11.
5. Corrigan, B.C., Tay, Z.Y. and Konovessis, D. 2023. Real-time instance seg- mentation for detection of underwater litter as a plastic source. Journal of Marine Science and Engineering, 11(8), 1532.
6. Wang, C.-Y., Yeh, I.-H. and Liao, H.-Y.M. 2024. Yolov9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616.
7. McLean, D.L., Parsons, M.J.G., Gates, A.R., Benfield, M.C., Bond, T., Booth, D.J., Bunce, M., Fowler, A.M, Harvey, E.S, Macreadie, P.I., Charitha, B. Pattiaratchi, Rouse, S., Thomson, P.G., Partridge, J.C., Todd, V.L.G., Jones, D.O.B. 2020. Enhancing the scientific value of industry remotely operated vehicles (rovs) in our oceans. Frontiers in Marine Science, 7, 220.
8. Tata, G., Royer, S.-J., Poirion, O. and Lowe, J. 2021. Deepplastic: A novel approach to detecting epipelagic bound plastic using deep visual models. arXiv preprint arXiv: 2105.01882.
9. Jocher, G., Chaurasia, A. and Qiu, J. 2023. Ultralytics yolov8.
10. Japan Agency for Marine-Earth Science and Technology (JAMSTEC). JAMSTEC deep-sea debris database, 2012.
11. Walia, J.S. and Seemakurthy, K. 2023. Optimized custom dataset for efficient detection of underwater trash. arXiv e-prints, pages arXiv–2305.
12. Jesus, A., Zito, C., Tortorici, C., Roura, E., De Masi, G. 2022. Underwater object classification and detection: first results and open challenges. In OCEANS 2022-Chennai. 1–6. IEEE.
13. Jothikrishna, K., Rithika, S.M., Swetha, S.V. and Kavitha, K. 2023. Remotely operated underwater vehicle (rov). In 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 1–4.
14. Hong, J., Fulton, M.S. and Sattar, J. 2020. Trashcan 1.0 an instance-segmentation labeled dataset of trash observations.
15. Watanabe, J.-I., Shao, Y. and Miura, N. 2019. Underwater and airborne monitoring of marine ecosystems and debris. Journal of Applied Remote Sensing, 13(4), 044509.
16. Anjana, K., Hinduja, M., Sujitha, K., Dharani, G. 2020. Review on plastic wastes in marine environment biodegradation and biotechnological solutions. Marine Pollution Bulletin, 150, 110733.
17. Lavers, J.L., Oppel, S., Bond, A.L. 2016. Factors influencing the detection of beach plastic debris. Marine Environmental Research, 119, 245–251. Elsevier.
18. Tan, L., Huangfu, T., Wu, L. and Chen, W. 2021. Comparison of yolo v3, faster r-cnn, and ssd for real-time pill identification.
19. Belioka, M.-P. and Achilias, D.S. 2023. Microplastic pollution and monitoring in seawater and harbor environments: A meta-analysis and review. Sustainability, 15(11), 9079.
20. Fulton, M., Hong J., Islam, Md J. and Sattar, J. 2019. Robotic detection of marine litter using deep visual detection models. In 2019 international conference on robotics and automation (ICRA), pages 5752–5758. IEEE.
21. Alabi, O.A., Ologbonjaye, K.I., Awosolu, O. and Alalade, O.E. 2019. Public and environmental health effects of plastic wastes disposal: a review. J Toxicol Risk Assess, 21, 1–13.
22. Aguirre-Castro, O.A., Inzunza-Gonz´alez, E., Garc´ıa-Guerrero, E.E., Tlelo-Cuautle, E., L´opez-Bonilla, O.R., Olgu´ın-Tiznado, J.E. and C´ardenas-Valdez, J.R. 2019. Design and construction of an rov for underwater exploration. Sensors, 19(24), 5387.
23. Qiao, G., Yang, M., Wang, H. 2022. A Detection Approach for Floating Debris Using Ground Images Based on Deep Learning. Remote Sensing, 14(17), 4161.
24. Ren S., He K., Girshick R. and Sun J. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28.
25. Krause, S., Molari, M., Gorb, E.V., Gorb, S.N., Kossel, E. and Haeckel, M. 2020. Per- sistence of plastic debris and its colonization by bacterial communities after two decades on the abyssal seafloor. Scientific reports, 10(1), 9484.
26. Nava, V., Chandra, S., Aherne, J., Alfonso, M.B., Ant˜ao-Geraldes, A.M., Attermeyer, K., Bao, R., Bartrons, M., Berger, S.A., Biernaczyk, M. 2023. Plastic debris in lakes and reservoirs. Nature, 619(7969), 317–322.

27. Li, W.C., Tse, H.F. and Fok, L. 2016. Plastic waste in the marine environment: A review of sources, occurrence and effects. Science of the total environment, 566, 333–349.

28. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu C.-Y. and Berg, A.C. 2016. Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, 21–37. Springer.

29. Zhang, H., Fu, W., Li, D., Wang, X., Xu, T. 2024. Improved small foreign object debris detection network based on YOLOv5. Journal of Real-Time Image Processing, 21(1), 21.