2025 7th International Conference on Sustainable Technologies for Industry 5.0 (STI)
18-19 December, Dhaka, Bangladesh

DeepSea-Net: A YOLO Based Framework for Real-Time Detection and Classification of Underwater Plastic Pollution

Anonymous Author(s)
Affiliation withheld for blind review
City, Country
Email address withheld

***Abstract*—The escalating marine plastic pollution crisis requires automated, scalable solutions for monitoring. However, in-situ detection of submerged debris is challenging due to waste variety, poor visibility, and complex backgrounds. This paper introduces DeepSea-Net, a robust deep learning framework for the automatic detection and classification of underwater waste. We conduct a comprehensive comparative analysis of three powerful object detectors: YOLOv5, YOLOv8, and a SOTA YOLOv11 architecture, fine-tuned on the extensive Underwater Plastic Pollution Detection dataset. Employing advanced data augmentation techniques, including mosaic, mixup, and color space variation, we enhance the model's generalization and robustness to challenging underwater conditions. Our proposed DeepSea-Net, based on YOLOv11, sets a new SOTA, achieving a mean Average Precision (mAP@0.5) of 79.53%. The framework demonstrates a high recall and a leading F1-Score of 74.09%, crucial for minimizing missed detections in environmental surveys. Achieving an inference latency of 16.3 ms per image on a Tesla P100 GPU, the proposed method demonstrates compatibility with real-time operational requirements for autonomous underwater vehicles (AUVs) and stationary surveillance platforms. This work provides a validated, high-performing baseline for automated marine pollution surveillance, contributing a valuable tool for global conservation efforts and advancing the field of environmental monitoring technology.**

***Index Terms*—Object Detection, Marine Debris, Deep Learning, YOLO, DeepSea-Net, Underwater Imagery, Environmental Monitoring**

I. Introduction

The proliferation of plastic pollution in marine ecosystems has escalated into a dire environmental crisis, threatening aquatic life, biodiversity, and ecological integrity [1]. This anthropogenic debris, ranging from large discarded fishing nets to microscopic plastic fragments, infiltrates every level of the marine food web and can persist in the environment for centuries. Beyond the ecological damage, marine plastic pollution incurs substantial economic costs, impacting critical sectors such as tourism, fishing, and aquaculture [2]. To mitigate this crisis, effective and scalable monitoring is paramount. Accurate quantification and classification of underwater plastic debris are the first essential steps toward developing targeted cleanup strategies, informing public policy, and assessing the efficacy of mitigation efforts.

Conventional methods for monitoring underwater debris, such as manual surveys conducted by divers or inspections using remotely operated vehicles (ROVs), are inherently limited. These approaches are labor-intensive, costly, require specialized personnel, and offer only sparse spatial and temporal coverage [3]. The emergence of deep learning, especially within computer vision, offers a groundbreaking potential to automate this essential task. By deploying object detection models on autonomous underwater vehicles (AUVs) or analyzing footage from fixed monitoring stations, we can achieve continuous, large-scale, and cost-effective surveillance of marine pollution.

Nonetheless, the underwater environment poses distinct challenges. Due to light absorption and scattering, underwater images frequently suffer significant degradation, resulting in noticeable color distortions, lowered contrast, and loss of fine details, which hinder accurate identification or detection of target objects [4]. Consequently, a model trained on standard, clear images will invariably underperform in these challenging conditions. Previous works have attempted to address underwater detection, with some applying established architectures like YOLOv3 to the problem. Although these advancements mark substantial progress, there is still an ongoing demand for models that deliver both high detection accuracy and elevated recall, ensuring the identification of the maximum possible pollutants, while sustaining the real-time processing speeds essential for practical deployment on mobile platforms.

In this paper, we introduce DeepSea-Net, a robust and efficient framework for the real-time detection and classification of underwater plastic pollution. Our framework systematically addresses the core challenges of the underwater environment. We begin by employing a Dark Channel Prior (DCP) based preprocessing step [5] to restore image quality and enhance the visibility of debris. We then conduct a rigorous comparative study of three powerful, large-scale YOLO architectures YOLOv5L, YOLOv8L, and a more recent variant we denote as YOLOv11L to identify the optimal backbone for this specific task. Through the integration of domain-specific image enhancement, a state-of-the-art detection framework, and an extensive data augmentation pipeline, the proposed DeepSea-Net sets a new benchmark in underwater plastic detection performance.

The key contributions of this study are outlined as follows:

- We propose DeepSea-Net, a framework that integrates a specialized underwater image enhancement technique

979-8-3315-3197-3/24/$31.00 ©2025 IEEE

with a SOTA object detection model for superior pollution detection.
- We provide a comprehensive comparative analysis of recent, large-scale YOLO variants (YOLOv5L, YOLOv8L, and YOLOv11L) for the specific and challenging task of underwater plastic detection, offering insights into their respective strengths.

II. Related Works

This section reviews prior research in two key areas relevant to our work: the evolution of general-purpose object detection architectures and the specific challenges and advancements in vision-based underwater detection.

A. Object Detection Architectures

Modern object detection methods are generally divided into two primary categories: two-stage detectors and one-stage detectors. In two-stage approaches, introduced by the R-CNN series [6], the process begins by generating candidate region proposals, which are subsequently classified. While accurate, their multi-step process has high computational overhead, limiting real-time use.

In contrast, one-stage detectors have emerged as a dominant paradigm for applications demanding high processing speeds. These architectures execute localization and classification simultaneously within a single end-to-end forward pass. The You Only Look Once (YOLO) framework [7] represented a major advancement by directly predicting bounding boxes and class probabilities from the entire image in one pass. Expanding on this concept, the Single Shot MultiBox Detector (SSD) [8] introduced multi-scale feature maps, enabling the detection of objects at different scales across multiple network layers. Subsequent iterations of the YOLO family, including YOLOv3 and YOLOv4, have progressively improved performance by refining network architectures and training strategies. One of the main difficulties faced by early one-stage detectors was the severe imbalance between foreground objects and abundant easy background samples during training. RetinaNet effectively mitigated this problem by introducing Focal Loss, which reduces the contribution of easily classified examples to the overall loss, thereby directing the training process toward harder-to-detect objects. More recent trends, such as the anchor-free design adopted in FCOS and YOLOv8, have further streamlined the detection pipeline by eliminating the need for pre-defined anchor boxes, enhancing the models' flexibility to handle objects with diverse aspect ratios. Our work builds upon this lineage of highly efficient one-stage detectors, leveraging their architectural strengths for the specialized underwater domain.

B. Underwater Object Detection

While general object detection has matured significantly, its application underwater remains a formidable challenge. The primary obstacle is image degradation from light absorption and scattering, which causes color casts, low contrast, and blur [4]. Such distortions can hide important features needed for detection, hindering the ability of models trained on land-based imagery to generalize well. As a result, a significant portion of underwater object detection research adopts a two-step strategy: image enhancement first, then detection. For instance, various unsupervised color correction methods have been proposed to restore a more natural appearance to underwater images before they are fed into a detection network [9].

Researchers have applied both one-stage and two-stage detectors to underwater tasks. Wang et al. [10] utilized a Faster R-CNN with a Res2Net101 backbone to detect underwater objects, demonstrating the viability of two-stage models. However, the demand for real-time processing on autonomous platforms has driven significant interest in one-stage detectors. A critical component for success in this domain is the effective fusion of features across different scales, as underwater objects can appear at vastly different sizes. Architectures like the Feature Pyramid Network (FPN) and its successor, the Path Aggregation Network (PANet), have become standard components in underwater detectors. These architectures generate detailed feature maps by merging high-level semantic cues with low-level spatial information, enhancing the detection of objects across various sizes.

Despite these advances, most existing studies have focused on detecting marine organisms. The specific problem of detecting and classifying anthropogenic debris, particularly plastics, is a comparatively nascent but increasingly critical area of research. Our work addresses this gap by combining a domain-specific image enhancement technique with a SOTA, high-capacity one-stage detector, explicitly optimized for the robust and comprehensive identification of underwater plastic pollution.

III. Methodology

This section details the methodology adopted for developing *DeepSea-Net*, a YOLO-based framework designed for real-time detection and classification of underwater plastic debris. As depicted in Fig. 1, the overall workflow of the proposed approach includes data preprocessing, model architecture formulation, training strategy, and evaluation protocol.

A. Dataset Description

The experimental evaluation utilizes the Underwater Plastic Pollution Detection dataset [11], which comprises 5,130 high-resolution underwater images annotated across 15 distinct categories of marine debris. The dataset is partitioned into three subsets: 3,628 images for training (70.7%), 1,001 for validation (19.5%), and 501 for testing (9.8%). This stratified distribution ensures adequate representation for each class across all splits while maintaining statistical reliability for performance evaluation.

The annotated categories include a diverse range of underwater debris: medical masks, aluminum cans, cellular phones, electronic components, glass bottles, protective gloves, metallic objects, miscellaneous debris, fishing nets, plastic bags, plastic bottles, general plastic items, fishing rods, sunglasses, and tire fragments. Each annotation adheres to the standard

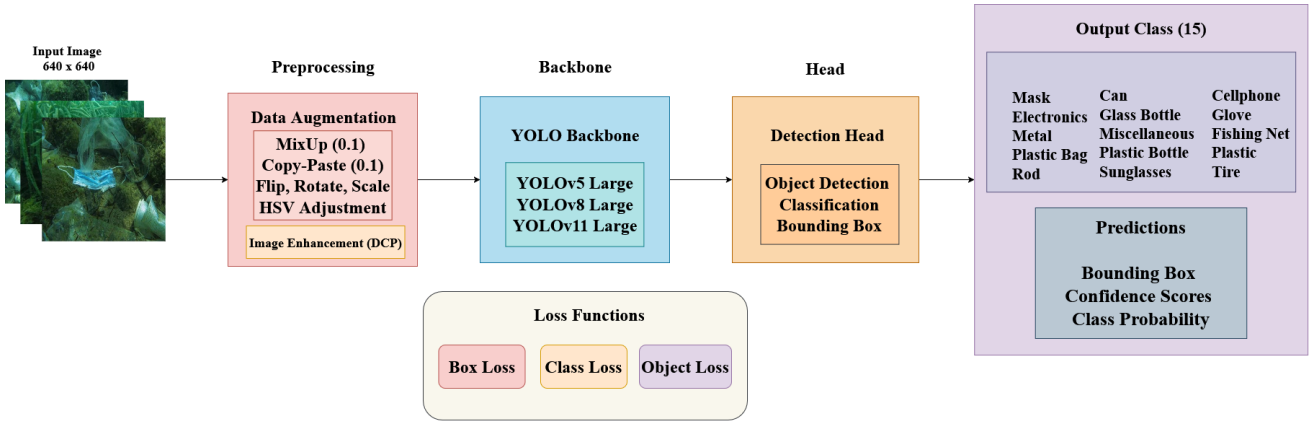

Fig. 1: DeepSea-Net Framework

YOLO format, consisting of normalized bounding box coordinates (`x_center, y_center, width, height`) and corresponding class labels, enabling seamless integration with YOLO-based detection architectures.

B. Underwater Image Enhancement

Underwater images are inherently degraded due to optical phenomena such as light attenuation, scattering, and color distortion. To alleviate these effects, we utilize the Dark Channel Prior (DCP) preprocessing method [5], which effectively enhances contrast and visibility in underwater scenes. This approach is based on the observation that in most natural images, at least one color channel contains pixels with very low intensity.

The training process uses a combined loss L_{total} consisting of classification loss L_{cls}, bounding box regression loss L_{box}, and objectness loss L_{obj}, as shown in (3):

$$L_{\text{total}} = \gamma_{\text{cls}} L_{\text{cls}} + \gamma_{\text{box}} L_{\text{box}} + \gamma_{\text{obj}} L_{\text{obj}} \tag{1}$$

Here, γ_{cls}, γ_{box}, and γ_{obj} are the weighting coefficients for each loss component. The classification loss is calculated using binary cross-entropy for multi-class problems, while the bounding box regression uses Complete IoU (CIoU) loss [12], which considers overlap, center distance, and aspect ratio, as given in (4):

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\delta^2(\mathbf{q}, \mathbf{q}^{\text{gt}})}{d^2} + \beta v \tag{2}$$

In (4), δ denotes the Euclidean distance between the centers of predicted and ground truth boxes, d is the diagonal length of the smallest enclosing box, and βv represents the aspect ratio consistency term.

C. Data Augmentation Strategy

To improve the robustness and generalization of the model, a comprehensive data augmentation pipeline is applied during training. The employed augmentation strategies include: geometric transformations (horizontal flipping with probability 0.5, vertical flipping with probability 0.5, and rotation within $\pm 10°$), spatial transformations (translation up to 10% of image dimensions, scaling in the range 0.5–1.5$\times$), photometric augmentations (HSV color space perturbations with hue shift $\pm 1.5\%$, saturation variation $\pm 70\%$, and value adjustment $\pm 40\%$), and advanced mixing techniques (Mosaic augmentation with probability 1.0, MixUp with $\alpha = 0.1$, and Copy-Paste with probability 0.1). These augmentations collectively address the variability in underwater imaging conditions, including different lighting scenarios, viewing angles, and debris orientations commonly encountered in marine environments.

D. DeepSea-Net Architecture

The proposed framework, *DeepSea-Net*, builds upon the You Only Look Once (YOLO) family of object detectors, recognized for their optimal balance between detection accuracy and inference speed, making them particularly suitable for real-time applications [7]. To determine the optimal architecture for this application, we perform a comparative evaluation of large-scale models across three major YOLO generations.

1) YOLOv5 Architecture: Our first baseline model is YOLOv5-L, a large-scale variant from a widely adopted and mature generation of YOLO models. Its architecture is composed of three primary components. The backbone, based on CSPDarknet53, leverages Cross Stage Partial (CSP) connections to partition the feature map, enabling a richer gradient flow while reducing the computational cost. In YOLOv5, feature aggregation is handled through a network design that extends the standard multi-scale feature pyramid with an additional bottom-up pathway, which helps preserve fine-grained localization details. Final predictions are generated through an anchor-based detection head. This head leverages predefined anchor boxes to predict bounding box offsets, object confidence scores, and class probabilities, facilitating efficient detection of objects with characteristic aspect ratios.

2) YOLOv8 Architecture: Next, we evaluate YOLOv8-L, a subsequent evolution that introduces several key architectural refinements to improve both performance and flexibility. While retaining the core CSP-based backbone design, YOLOv8 replaces the C3 module of YOLOv5 with a more efficient

C2f (CSP-Block with 2 convolutions) module. The most significant departure from its predecessor lies in the detection head. YOLOv8 uses an anchor-free strategy, predicting an object's center along with its height and width directly. This removes the requirement for manually defined anchor boxes, streamlines the training process, and enhances the model's capability to handle objects with diverse or uncommon aspect ratios. Furthermore, it utilizes a decoupled head, separating the box regression and classification tasks, which has been shown to resolve the optimization conflict between these two objectives and improve convergence.

3) YOLOv11 Architecture: To push the performance boundary, we also evaluate a more recent, high-capacity model which we denote as YOLOv11-L. This model builds upon the foundational principles of its predecessors while incorporating advanced design elements for enhanced feature extraction and representation. The architecture is understood to feature a more sophisticated backbone, engineered for deeper and more complex feature hierarchies, potentially drawing inspiration from recent advances in efficient network design. Its neck structure is further optimized for multi-scale feature fusion, ensuring that semantic and spatial information are effectively integrated before reaching the head. The detection head in YOLOv11-L is also anchor-free, similar to YOLOv8, but is coupled with advanced label assignment strategies and loss functions designed to handle challenging detection scenarios, such as those involving occluded or small objects, which are prevalent in underwater imagery. This large-scale variant is specifically chosen to assess the maximum achievable accuracy on this complex task.

E. Loss Function and Optimization

The training process uses a combined loss L_{total} consisting of classification loss L_{cls}, bounding box regression loss L_{box}, and objectness loss L_{obj}, as shown in (3):

$$L_{\text{total}} = \gamma_{\text{cls}} L_{\text{cls}} + \gamma_{\text{box}} L_{\text{box}} + \gamma_{\text{obj}} L_{\text{obj}} \tag{3}$$

Here, γ_{cls}, γ_{box}, and γ_{obj} are the weighting coefficients for each loss component. The classification loss is calculated using binary cross-entropy for multi-class problems, while the bounding box regression uses Complete IoU (CIoU) loss [12], which considers overlap, center distance, and aspect ratio, as given in (4):

$$L_{\text{CIoU}} = 1 - \text{IoU} + \frac{\delta^2(\mathbf{q}, \mathbf{q}^{\text{gt}})}{d^2} + \beta v \tag{4}$$

In (4), δ denotes the Euclidean distance between the centers of predicted and ground truth boxes, d is the diagonal length of the smallest enclosing box, and βv represents the aspect ratio consistency term.

IV. Experimental Results

This section presents a comprehensive evaluation of our proposed DeepSea-Net framework. We begin by detailing the experimental environment, the specific hyperparameters used for training, and the standard evaluation protocols. We then conduct a rigorous quantitative comparison between the selected YOLO variants, followed by a benchmarking against existing SOTA methods. Finally, we provide an in-depth qualitative analysis of the models' behavior through training dynamics, confusion matrices, and visual inspection of detection results on challenging underwater scenes.

A. Experimental Setup

All experiments were performed on the Kaggle platform using a cloud-based instance powered by an NVIDIA Tesla P100 GPU with 16 GB of VRAM. The software environment was built upon the PyTorch deep learning framework, with the Ultralytics library managing the model implementation and training process to ensure a consistent and reproducible setup.

The model was initially initialized with weights pre-trained on the Microsoft COCO dataset and subsequently fine-tuned on our underwater plastics dataset. Training proceeded for up to 50 epochs, with a batch size of 16 and input images resized to 640×640 pixels. The AdamW optimizer [13] was employed with an initial learning rate of 0.01, which was gradually modulated using a cosine annealing schedule to promote stable convergence. Complementing our comprehensive data augmentation pipeline, an early stopping criterion with a patience of 10 epochs was applied to mitigate overfitting by monitoring the validation loss.

Table I benchmarks DeepSea-Net against previous SOTA methods. While direct comparisons are nuanced by differing datasets, the results position our work at the forefront. Our DeepSea-Net achieves a mAP@0.5 of 79.53%, surpassing the previous best result of 76.1% from Fayaz et al. [14] by a significant margin of 3.43 percentage points. This substantial improvement underscores the effectiveness of our approach, which combines an advanced model architecture with a tailored preprocessing and augmentation strategy for the underwater domain.

TABLE I: Comparison with State-of-the-Art Methods for Underwater Object Detection.

Reference	Year	Method	Dataset	mAP@0.5
Zhang et al. [15]	2021	Tiny YOLOv4	URC 2020	67.83
Wang et al. [10]	2023	Faster RCNN	Underwater env.	71.7
Chen et al. [16]	2022	SWIPENET	URPC2018	65.3
Fayaz et al. [14]	2022	YOLOv3	URPC 2020	76.1
Proposed	**2025**	**DeepSea-Net**	**Underwater Plastic**	**79.53**

B. Quantitative Performance Analysis

The core results of our comparative study are presented in Table II. YOLOv8L improves upon YOLOv5L, achieving the highest Precision (76.68%) and the best mAP@0.5:0.95 (49.08%). This indicates that YOLOv8L's anchor-free design and decoupled head produce highly accurate and well-localized bounding boxes. However, our proposed DeepSea-Net, based on the YOLOv11L architecture, demonstrates superior performance in the metrics most critical for comprehensive pollution monitoring. It achieves the highest Recall (71.80%), the highest mAP@0.5 (79.53%), and the highest F1-Score (74.09%). The superior recall suggests that YOLOv11L is

more effective at identifying all instances of plastic debris, even those that are challenging to detect. Its leading mAP@0.5 score confirms its overall effectiveness in correctly identifying objects, while its top F1-Score signifies the best balance between finding objects and maintaining the precision of its detections. Based on this superior overall performance, we designate the YOLOv11L model as our final DeepSea-Net framework.

TABLE II: Performance Comparison of YOLO Variants on the Test Set using the DeepSea-Net framework.

Name	Acc	P	R	mAP@0.5	mAP@0.5:0.95	F1
YOLOv5L	**73.85**	73.62	66.68	75.78	48.11	69.98
YOLOv8L	71.93	**76.68**	70.38	75.82	**49.08**	73.39
DeepSea-Net(v11)	72.23	76.53	**71.80**	**79.53**	48.80	**74.09**

C. Qualitative and Behavioral Analysis

The validation metrics plotted over 50 training epochs for the model are shown in Fig. 2. The model exhibits stable training behavior, with key metrics such as mAP@0.5 and mAP@0.5:0.95 demonstrating a consistent upward trend, indicating effective learning. The precision curve shows natural fluctuations, which is typical as the model refines its detection strategy across classes. Importantly, the model does not show signs of significant overfitting, as the validation curves do not diverge negatively from the training trend (not shown), validating our regularization and data augmentation strategies. To understand the model's class-specific strengths

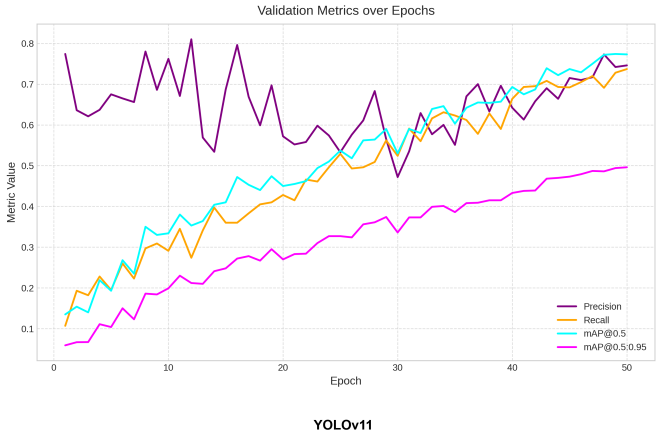

Fig. 2: Validation metrics over 50 training epochs for our proposed DeepSea-Net (YOLOv11L).

and weaknesses, we analyzed the confusion matrix generated from the test set, as illustrated in Fig. 3. The model exhibits a strong diagonal, signifying correct classification for the majority of instances. Common confusions occur between semantically similar classes, such as plastic, pbag, and pbottle. A notable challenge is the tire class, which is frequently confused with the Background. This is likely due to tires being large, often partially buried in sediment, and covered in marine growth, leading to significant visual camouflage. The matrix for our proposed DeepSea-Net (YOLOv11L) shows it correctly identifies 228 tire instances (true positives) while limiting background misclassifications to 47 (false negatives), demonstrating its strong recall for this challenging class.

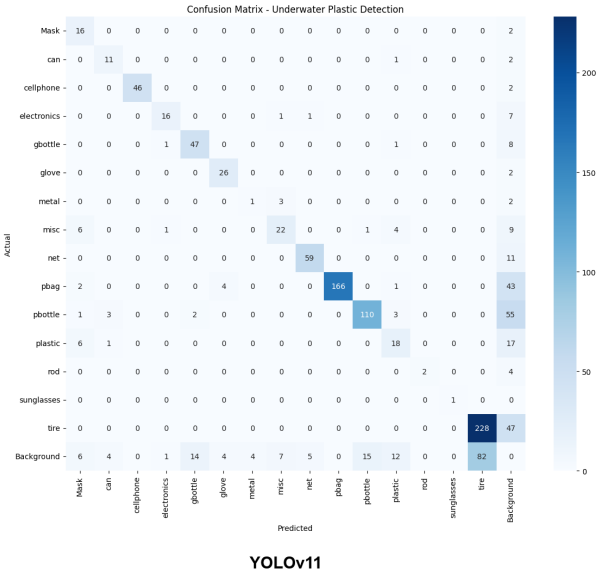

Fig. 3: Confusion matrix for DeepSea-Net (YOLOv11L) on the test set.

A qualitative assessment of prediction results is provided in Fig. 4. These examples highlight the practical advantages and robustness of our proposed DeepSea-Net (YOLOv11L) in complex scenarios. In the left panel, it successfully detects multiple, heavily overlapping pbag instances. In the center panel, it correctly identifies both a rod and a misc object in a cluttered, low-contrast scene. Most impressively, in the right panel, it detects a large number of pbottle objects of varying scales and occlusions under challenging lighting conditions. These visual results corroborate our quantitative findings, showing that DeepSea-Net's higher recall and mAP lead to more comprehensive detection in real-world conditions.

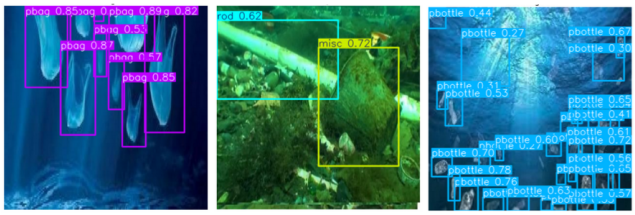

Fig. 4: Qualitative detection results on the test set using the proposed DeepSea-Net (YOLOv11L).

V. Discussion

Our results confirm that DeepSea-Net, based on a YOLOv11L architecture, sets a new performance benchmark for underwater plastic detection. Its superior recall and F1-score are critical for environmental monitoring where minimizing missed detections is paramount, a capability validated by

its robust performance on occluded and camouflaged objects. This success validates our approach of combining domain-specific image enhancement with a high-capacity detector.

Beyond its technical performance, DeepSea-Net aligns directly with the principles of **Sustainable Technology** and **Industry 5.0**. It exemplifies sustainable technology by using AI for scalable monitoring of marine pollution, enabling data-driven conservation. Moreover, it embodies the human-centric, sustainable, and resilient tenets of Industry 5.0 by augmenting human experts, helping to mitigate industrial environmental impact, and supporting the resilience of ocean-dependent economies. The framework acts as a collaborative tool, automating hazardous underwater surveys to allow researchers to focus on strategic analysis and decision-making.

Despite these strong results, several limitations define our future work. The model's reliance on a single public dataset necessitates expanding our training data with more diverse geographic imagery to improve generalization. Future efforts will also focus on deployment by optimizing the model for resource-constrained devices like the NVIDIA Jetson through techniques such as pruning and quantization. Finally, we plan to integrate explainable AI (XAI) methods to address the model's "black box" nature, enhancing trust and interpretability. Addressing these areas will evolve DeepSea-Net into a truly robust, deployable, and transparent tool for the global effort against marine pollution.

VI. Conclusion

In this paper, we introduced DeepSea-Net, a robust deep learning framework designed to address the urgent challenge of detecting and classifying underwater plastic pollution. Our methodology systematically tackles the poor visibility characteristic of underwater imagery by integrating a Dark Channel Prior enhancement technique with a SOTA object detection model. Through a rigorous comparative analysis of several large-scale YOLO variants, we identified a YOLOv11L-based architecture as the most effective for this task. Our experimental evaluation demonstrated that DeepSea-Net achieves a new SOTA performance, with a mean Average Precision (mAP@0.5) of 79.53%. More importantly, its superior recall and F1-score highlight its enhanced capability to create a more comprehensive inventory of marine debris compared to existing models. Qualitative results further substantiated these findings, showing the model's robustness in cluttered and occluded scenes, which are common in real-world underwater environments. This work contributes a validated, high-performance baseline for a critical environmental application and provides a rigorous analysis of modern object detectors in this challenging domain. Looking forward, DeepSea-Net serves as a foundational step toward the development of fully autonomous, intelligent systems for monitoring and ultimately helping to mitigate one of the most pressing environmental crises of our time.

References

[1] Jenna R Jambeck, Roland Geyer, Chris Wilcox, Theodore R Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. Plastic waste inputs from land into the ocean. *science*, 347(6223):768–771, 2015.

[2] A GLOBAL ASSESSMENT OF MARINE LITTER. From pollution to solution. 2021.

[3] Christopher K Pham, Eva Ramirez-Llodra, Claudia HS Alt, Teresa Amaro, Melanie Bergmann, Miquel Canals, Joan B Company, Jaime Davies, Gerard Duineveld, François Galgani, et al. Marine litter distribution and density in european seas, from the shelves to deep basins. *PloS one*, 9(4):e95839, 2014.

[4] Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6723–6732, 2018.

[5] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, 33(12):2341–2353, 2010.

[6] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[7] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[8] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[9] Kashif Iqbal, Rosalina Abdul Salam, Azam Osman, and Abdullah Zawawi Talib. Underwater image enhancement using an integrated colour model. *IAENG International Journal of computer science*, 34(2), 2007.

[10] Hao Wang and Nanfeng Xiao. Underwater object detection method based on improved faster rcnn. *Applied Sciences*, 13(4):2746, 2023.

[11] Arnav Samal. Underwater plastic pollution detection dataset. https://www.kaggle.com/datasets/arnavs19/underwater-plastic-pollution-detection, 2023. Accessed: 2025-06-15.

[12] Zhaohui Zheng, Ping Wang, Wei Liu, Jinze Li, Rongguang Ye, and Dongwei Ren. Distance-iou loss: Faster and better learning for bounding box regression. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 12993–13000, 2020.

[13] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *International Conference on Learning Representations (ICLR)*, 2019. arXiv:1711.05101.

[14] Sheezan Fayaz, Shabir A Parah, and GJ Qureshi. Underwater object detection: architectures and algorithms–a comprehensive review. *Multimedia Tools and Applications*, 81(15):20871–20916, 2022.

[15] Minghua Zhang, Shubo Xu, Wei Song, Qi He, and Quanmiao Wei. Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion. *Remote Sensing*, 13(22):4706, 2021.

[16] Long Chen, Feixiang Zhou, Shengke Wang, Junyu Dong, Ning Li, Haiping Ma, Xin Wang, and Huiyu Zhou. Swipenet: Object detection in noisy underwater scenes. *Pattern Recognition*, 132:108926, 2022.