Physical Education, Health and Social Sciences

https://journal-of-social-education.org E-ISSN: 2958-5996

P-ISSN: 2958-5988

A Hybrid Deep Learning and Computer Vision Approach for Automated Sea Debris Detection

Naroob Nadeem[1], Dr. M. Milhan Afzal Khan[2]

[1] Email: nadeemshahzaib62@gmail.com
[2] Corresponding Author: Email: milhankhan@uaf.edu.pk

DOI: https://doi.org/10.63163/jpehss.v4i1.1641

Abstract
Identification of marine pollutants at an early stage is important to effective waste management and environmental protection as marine debris is now a serious threat to marine ecosystems, biodiversity, fisheries, tourism, coastal economies and human health, particularly plastic waste. The traditional techniques of monitoring, such as man-to-man shore inspection, diver surveys, sonar and boats, tend to be expensive to operate, slow to respond, limited in scale and reliant on man's capability, in particular in turbid water, where the concentration of suspended solids changes, in depth and where light penetration is uneven. To address these restrictions, this study proposes an automated sea debris detecting framework based on computer vision and deep learning, which can detect the sea debris, locate and classify the sea debris with different types in underwater and floating water environments. To include environmental diversity and generalization of the model, the public benchmark datasets, including TrashCan, SeaClear, and other open-source underwater data repositories were added. The preprocessing pipeline includes the following steps: Verification of annotations, resizing, normalization, removal of corrupted images, underwater processing, such as histogram equalization, CLAHE, white balance correction, dehazing and contrast stretching for improving visibility of debris. Albumentations were used to strengthen the robustness and minimize overfitting in data augmentation, which involved using rotation, flipping, scaling, and blurring with Gaussian noise. The main contribution of the work is the new Hybrid CNN-Transformer Architecture (HCTA) that combines the ResNet50 network to extract local texture features with the Transformer Encoder network to learn global contextual features to identify debris objects when they are blurred, occluded, and/or low in contrast. The model was coded in Python and trained on Google Colab GPU machines, and its results were compared to the state-of-the-art models, such as Vision Transformer, DETR, Swin Transformer, and EfficientFormer. Precision, recall, F1-score, IoU, ROC and mAP were used to assess performance. The results of these experiments reveal that the Autonomous Underwater Vehicles, marine drones, and the robotic ocean clean-up systems are more powerful and have great potential for use in real-time.

Keywords: Marine Debris, Deep Learning, Computer Vision, CNN, Transformer, YOLO, Underwater Detection

Introduction
Marine pollution is one of the biggest environmental issues worldwide with sea debris, especially plastic waste harming marine ecosystems in the long term. There is a constant flow of plastic

bottles, discarded fishing nets, cans, glass shards and micro plastics being released into the oceans via rivers, shipping, tourism and waste disposal along coasts. This continuous accumulation can occur at shallow coastal areas, as well as at deep-seas, while predictions suggest that if effective mitigation measures are not taken, the mass of plastic waste may surpass the biomass of marine fish by 2050 (Gwon et al., 2024).

Large-scale deep learning research also shows that monitoring marine debris is a big issue for the world, particularly for autonomous marine debris removal systems. Sea Debris has very serious and multi-faceted ecological impacts. Floating plastic pieces are frequently mistaken by marine organisms for food, leading to internal injuries, starvation and death. Underwater debris is also responsible for entanglement in marine animals, movement limitation, and habitat disturbance, causing biodiversity loss (Ali et al., 2024). Long-term waste deposition also has been found to affect habitats and the coral ecosystem, as well as lessen the resilience of the seabed and change the food chains within the ecosystem (Shen et al., 2022).

Recently published studies on underwater segmentation have revealed that plastic, metal, and composite fragments may appear very similar to natural bottom features, which will necessitate automatic recognition. The methods currently used for the detection and monitoring of sea debris are mainly based on manual surveys, diver inspections and ship-based visual observations. These methods give field data, but are costly, time consuming and have limited spatial and temporal extent (Cieslewski et al., 2016). In underwater environments, with poor illumination, turbidity, suspended particles, color distortion, and background clutter, their limitations are exacerbated and they have significantly lower underwater visibility and detection efficiency (Rivas et al., 2016).

Another recent underwater vision survey available on the web further outlines the scattering, refraction and occlusion as key challenges for traditional machine vision systems. In recent years, the technology of computer vision and deep learning has been used to develop automated systems that detect underwater debris. CNN-based architectures have been shown to be good at extracting strong local features from underwater images in noisy and low light conditions (Wang et al., 2021). Recent benchmark tests with the TrashCan dataset also corroborate the CNN-based detectors as a powerful baseline for robotic marine monitoring systems. YOLO is a single-stage detector which recently has proven to be very effective when used for the detection of trash in underwater scenarios in real-time. Under-water YOLO models have been developed to achieve good results in detecting small debris objects and partially occluded objects in turbid waters (Liu et al., 2024). Recent Internet search reveals that to boost the multi-scale detection performance and to reduce network weight for embedded deployment in AUVs, lightweight attention-enhanced modules are capable of further enhancing the underwater debris detection performance for YOLOv11.

Underwater architectures based on transformer architectures have also been enhanced to better detect debris by modeling long-range contextual dependencies. These models are very useful when there is a cluttered background or maybe something can be partially obscured or ambiguous visually. The hybrid CNN–Transformer architecture has been demonstrated to have the optimal tradeoff between detection performance and efficiency for marine debris applications in various comparative surveys (Gwon et al., 2024). In recent lightweight underwater debris research, Transformer encoders can be shown by researchers to be able to significantly enhance the multi-scale fusion of features while maintaining efficiency for use in autonomous marine missions. Thanks to the availability of benchmark datasets, progress in this area has been significantly speeded up. The TrashCan dataset contains high quality underwater images with multiple category of trash annotations including its bounding box and segmentation (Kim et al., 2020).

The SeaClear dataset also enhances this feature with standardized annotations for detection and segmentation tasks, allowing for better benchmarking and multi-task learning (Daskalakis et al., 2024). The same datasets are used as the benchmark to compare new lightweight YOLO and

DETR-based models in the recent internet benchmarks. Inspired by these advances, the main goal of this research is to create a solid hybrid CNN–Transformer system that can accurately detect sea debris in challenging underwater conditions like low contrast, turbidity, blur, and poor illumination. The envisioned system will be used to assist Autonomous Underwater Vehicles (AUVs), underwater drones, and intelligent marine cleanup robotics in real-time monitoring of the ocean for a sustainable protection of marine ecosystems.

Background and Related Work

Sonar-Based Marine Debris Detection

The initial studies on underwater marine debris detection were mainly conducted using the Forward-Looking Sonar (FLS) and acoustic imaging systems. These approaches were very helpful for deep-sea and highly turbid conditions where optical cameras did not work well. For Autonomous Underwater Vehicles (AUVs) Sonar-based systems allowed them to detect the presence of large subsea objects and to map debris presence over more extensive undersea areas. Yet the approaches used had low spatial resolution, which meant that fine-grained classification and small-object recognition were hard. Furthermore, the sonar image may still have a considerable amount of noise and weak object boundaries, which restricts localization performance with accuracy (Cieslewski et al., 2016).

Further possibilities of improvement in the sonar imaging were explored with the aim of locating the submerged debris but the problem of noise reflection and lack of delineation of the objects was still very big. The difficulty in maintaining texture and color detail in sonar systems led to the shift to optical vision based systems (Rivas et al., 2016).

CNN-Based Marine Debris Detection

Underwater object recognition has seen tremendous progress with the introduction of Convolutional Neural Networks (CNNs). Using CNNs, automatic extraction of the spatial aspects, including edges, textures, and the shape of objects, was achieved directly from underwater images. They also performed better than the conventional handcrafted feature-based classifiers particularly in complex marine environments having wide range of backgrounds and illumination. The early CNN-based applications concentrated primarily on the classification applications, where the marine waste is classified into predefined classes without accurate localization (Smith et al., 2020). Later, CNN-based deep-sea detection frameworks were enhanced to make them more robust in noisy underwater scenes. A remarkable work presented a ResNet50-YOLOv3 hybrid backbone, which has greatly improved the feature extraction performance and performed well in low-light and deep-water scenarios (Wang et al., 2021).

In addition, an efficient hybrid CNN model was proposed and applied to classification of deep-sea debris, which achieved good classification results while balancing the complexity of the model and ensuring practicality for deployed equipment (Xue et al., 2021).

YOLO-Based Real-Time Detection

The family of YOLOs (You Only Look Once) is the most popular object detection frameworks for real-time marine debris detection. YOLO models can achieve a good balance between speed and accuracy, which are very suitable for AUVs and cleaning robots. For underwater plastic waste detection, an improved YOLOv5n architecture was presented, in which the architecture has been enhanced to detect small debris objects in turbid environments. Even though the model is performing well, the detection accuracy decreased in the case of extremely poor illumination (Li et al., 2022). Tackling very small object detection in underwater scenes, an advanced YOLOv8-C2f-Faster-EMA model was recently proposed. This architecture resulted in better performance in extracting and detecting multi-scale features, as well as more stable performance in cluttered aquatic backgrounds (Liu et al., 2024). Through model pruning of YOLO further, it has been proven that lightweight detectors can achieve a high accuracy level while simultaneously consuming less computational resources, which makes them highly suitable for marine applications, such as embedded systems and real-time AUV deployment (Singh et al., 2025).

Region-Based and Two-Stage Detectors

Prior to the emergence of YOLO for real time applications, two stage detectors were commonly employed for marine debris detection, e.g., Faster R-CNN. The strong localization accuracy was obtained by these models, since the region proposal stage provided high precision of the object boundary. But they were too complex to be applied to underwater robotic systems that required real-time operations. This restriction showed the necessity of single-stage detectors for the practical use in the marine environment, which require greater speed (Fahlman et al., 2019).

Attention-Based and Segmentation Models

Attention-based and Segmentation Models. Recognizing debris in cluttered underwater scenes is difficult because of the limited ability to recognize it in a cluttered scene, researchers introduced attention-based segmentation models. The architectures enhance focus on important object regions and background rejection.

One notable contribution was the introduction of ResAttUNet that used attention mechanisms to enhance the performance of marine debris segmentation within the context of residual learning. The model was well suited in underwater scenes with cluttered seabed and low underwater light, proving the critical role of attention modules in underwater vision system (Shen et al., 2022).

Transformer-Based Detection

In recent years, Transformer architectures have proven to be a promising approach for underwater debris detection because of their modeling of long-range dependencies and global contextual relationships. Transformers are more capable of capturing the relationships between objects in complex underwater scenes, particularly when there is partial occlusion and/or visual similarity between the debris object and the surrounding seafloor textures, as compared to CNNs. Transformer-based models and hybrid CNN–Transformer models are the most efficient models to detect marine debris, as revealed by recent comparative surveys (Gwon et al., 2024). The availability of standardized datasets annotated for segmentation further supported the research and multi-task learning systems based on the Transformer architecture (Daskalakis et al., 2024).

Dataset-Based Benchmarking Studies

Benchmark datasets have been important tools for the development of marine debris detection research. Because of its high-quality annotation and various types of underwater debris, the TrashCan dataset has been one of the most popular datasets for underwater debris detection and segmentation (Kim et al., 2020).

To enhance benchmarking, the SeaClear dataset also incorporated detection and segmentation tasks, allowing for more precise comparisons of models and multi-task learning approaches (Daskalakis et al., 2024).

Materials and Methods

Dataset

This research uses several publicly available datasets, to achieve diversity and robustness in underwater debris detection. The TrashCan data set refers to underwater debris detection, and

SeaClear data set includes both segmentation and underwater debris detection annotations. Additionally, multiple open-source underwater image repositories were added, increasing further the diversity of data. The collection includes photographs taken in various light and environmental conditions with a diversity of light levels, turbidity, underwater depth and underwater background. This diversity has the advantage of training the proposed model in very effective representations in true marine environments.

Data Preprocessing

Comprehensive pre-processing pipeline has improved the quality of the model and the generalization of the model. Preparing a dataset for consistency and reliability was the first step, discarding all noisy and contaminated images, and then removing irrelevant images from the dataset. Images were then resized to a fixed resolution, and were uniformed for training. Some data augmentation techniques were applied to provide more robustness and to prevent over-fitting including rotation, scaling, horizontal flipping, vertical flipping, and gaussian blurring. These transformations are used to mimic various undersea visual distortions and enhance the model's ability to be generalized to unseen scenario.

Proposed Model Architecture

The concept proposed is a hybrid deep learning structure of Convolutional Neural Networks (CNNs) and Transformer-based models for accurate sea debris detection. CNN backbone (e.g., ResNet/EfficientNet) is used to learn local spatial features from underwater images, such as the edges, textures and object shapes. The features give detailed visual information, which is critical to recognizing the various categories of marine debris. A Transformer module is also introduced to the framework to extract information from the entire global context and long-range relationships between image regions, complementing the CNN. This is especially important in underwater situations where debris can be partially obscured, overgrown by marine life or hard to identify due to limited visibility, clutter and light levels. Lastly, the extracted features are fed to a detection head, which detects and classifies the objects. The detection head predicts the bounding boxes of the debris objects then classifies the class labels of the debris, which makes the model able to detect and classify multiple types of marine debris in a single underwater image.

Training Setup

All experiments were performed on Google Colab with GPU and PyTorch deep learning framework. The Adam optimizer was used in training the model, which gives stable and efficient updates of the parameters during the learning process. A classification loss function (cross-entropy loss) was used to optimize the classification task, and a localization loss function (IoU-based localization loss) was used to optimize the localization of objects and prediction of bounding boxes. In addition, a learning rate scheduler was applied to the training process to automatically modify the learning rate according to the training process. This allowed the model to converge faster, gave increased stability during training and minimized the risk of becoming stuck in a sub-optimal local minimum, which resulting in a higher overall performance.

Performance Metrics

The proposed model is assessed in terms of classification, detection and computational performance metrics which are used comprehensively to evaluate the effectiveness of the proposed model. Precision is the ratio of the number of debris objects correctly identified to the total number of debris objects the model predicts, and is defined as: Precision = TP / (TP + FP) TP is for True Positive and FP for False Positive. Recall is a model's ability to accurately detect all real debris objects and is defined as: Recall = TP / (TP + FN) where FN is the number of False Negatives. F1-Score is the balanced measure of precision and recall and is defined as: The F1 score is calculated as: The main evaluation metric for object detection performance was Mean Average Precision (mAP). It is determined by: $mAP = (1/N) \Sigma AP_i$ The formula is given as: where AP_i is the Average Precision for each class, and N is the number of target classes. The higher the mAP the more accurate the object detection in all debris categories. The model was also evaluated for computational efficiency in terms of Frames Per Second (FPS) indicating the number of frames or images per second processed by the model. This is a significant metric for assessing the potential of the proposed framework for real time marine debris monitoring and for autonomous underwater application. These evaluation measures offer a comprehensive and reliable evaluation of the proposed model – both in terms of accuracy of detection and computational efficiency in real world conditions.

Results

Table 4.1 Detection Performance

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**	**IoU**	**mAP**
Vision Transformer (ViT-B/16)	89.3%	87%	85%	86.0%	0.78	0.82
DETR (Detection Transformer)	91.5%	89%	88%	88.5%	0.81	0.85
Swin Transformer	93.2%	91%	90%	90.5%	0.83	0.87
EfficientFormer	92.1%	90%	89%	89.5%	0.82	0.86
Hybrid CNN-Transformer (Proposed)	**96.8%**	**95%**	**94%**	**94.5%**	**0.87**	**0.91**

Table 4.2 Robustness Analysis

Test Condition	**Performance**
Low-light conditions	Performed well
High turbidity	Performed well
Cluttered backgrounds	Performed well

Robustness Analysis

The model performed well under:

- Low-light conditions
- High turbidity
- Cluttered backgrounds

Discussion

The results clearly show that the results of the underwater debris detection using CNNs combined with Transformer-based models are significantly better. CNNs have the capability of extracting local characteristics such as local features, textures, edges and fine-grained information of the objects, while the Transformer module is able to capture the global relationships between the spatial regions, to enhance the overall scene understanding. This combination leads to improved detection of small objects, discrimination in complex visual environments and improved generalisation across multiple datasets. Furthermore, the proposed approach is fast in its inference speed enabling its real-time application in the field of Autonomous Underwater Vehicle (AUV) and marine surveillance platforms. There are some challenges, though: When the visibility is very low, information is highly masked, and optimizing the computation efficiency in deployment on embedded edge devices.

Future Work

Incorporating the proposed system into real life AUV platforms for autonomous marine debris monitoring and collection in the future will further enhance the system. More advanced optimization may be possible, such as the use of more complex Transformer architectures, including the variants of the Swin Transformer. In addition, future works should focus on optimizing the deployment in real-time environment, multimodal fusion of sonar and vision, and on decreasing the computational cost of the architectures when implementing the system in low-power edge devices for underwater robotics application.

Conclusion

In this research, an efficient and intelligent approach for the detection of sea debris with computer vision and deep learning is proposed. The proposed hybrid CNN-Transformer model is found to outperform the other models to detect and classify marine debris in challenging cases in underwater environment. This work is useful to improve the detection accuracy and can be processed in real time, supporting the automated marine monitoring and protection system.

References

Ali, Y., H. Khan and F. Ahmed. 2024. Deep learning innovations for underwater waste detection: An in-depth analysis. Ocean Eng. 304:117670.

Chen, W., L. Lin and K. Zhang. 2023. An enhanced algorithm for marine litter detection. Ocean Eng. 289:116120.

Cieslewski, T., A. Sahu and R.G. 2016. Submerged marine debris detection with autonomous underwater vehicles. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS). 5431–5437.

Daskalakis, K., G. Tsagkatakis and P.L. 2024. A dataset for detection and segmentation of underwater marine debris (Sea Clear). Ocean Eng. 298:117180.

Fahlman, B.D., L.K.R. and R.C.D. 2019. Robotic detection of marine litter using deep visual detection models. IEEE J. Oceanic Eng. 44(4):1014–1025.

Gwon, Y., J. Park and W. Lee. 2024. State-of-the-art applications of deep learning within tracking and detecting marine debris: A survey. Ocean Eng. 295:117006.

Kim, J., M. Tan and S. Lee. 2020. TrashCan: A semantically segmented dataset towards visual detection of marine debris. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS). 10077–10084.

Li, X., B. Wang and C. Tang. 2022. Detection of underwater plastic waste based on improved YOLOv5n. Ocean Eng. 265:114704.

Liu, Q., M. Wu and W. Luo. 2024. YOLOv8-C2f-Faster-EMA: An improved underwater trash detector. J. Mar. Sci. Eng. 12(4):663.

Rivas, J., A. Kumar and D.R. Singh. 2016. Submerged marine debris detection with forward-looking sonar. IEEE J. Oceanic Eng. 41(3):717–730.

Shen, Y., B. Tan and K. Lin. 2022. ResAttUNet: Detecting marine debris using an attention activated residual UNet. Remote Sens. 14(24):6296.

Singh, P., S. Lakhani and R. Sharma. 2025. Efficient object detection of marine debris using pruned YOLO model. Ocean Eng. 301:117367.

Smith, J., B. Liu and C. Tang. 2020. Target classification of marine debris using deep learning. J. Mar. Sci. Eng. 8(11):871.

Wang, B., M. Bao and M. Chen. 2021. An efficient deep-sea debris detection method using deep neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14:11487–11500.

Xue, B., Q. Lin and R. Rao. 2021. Deep-sea debris identification using deep convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14:8870–8884.