Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multimodal Deep Learning Framework for Detection and Autonomous Removal of Submerged Ocean Plastic Waste

Shahen Ramnath

54081882@mylife.unisa.ac.za

University of South Africa

Sreedevi Vallabapurapu

University of South Africa

Suvendu Chandan Nayak

Silicon University

Research Article

Keywords: Deep learning, Submerged plastic, Multimodal sensing, Semantic segmentation, Marine robotics, Ocean sustainability

Posted Date: November 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11197394/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Multimodal Deep Learning Framework for Detection and Autonomous Removal of Submerged Ocean Plastic Waste

Shahen Ramnath, Sreedevi Vallabapurapu , and Suvendu Chandan Nayak

Abstract

Marine plastic pollution represents a persistent environmental challenge, particularly when plastic waste becomes submerged, trapped in marine habitats, or deposited on the seabed where conventional monitoring and clean-up methods become difficult. This chapter presents a multimodal deep learning framework for the detection, localisation, classification, and autonomous removal of submerged plastic waste. The framework integrates optical imaging, sonar and acoustic sensing, depth information, deep learning, and autonomous marine robotics. Convolutional neural networks, object-detection models including YOLO and Faster R-CNN, semantic segmentation approaches including U-Net and DeepLab, and reinforcement learning are incorporated into a perception-to-action pipeline for AUV and ROV deployment. The research design addresses underwater visibility, multimodal data fusion, limited annotated data, three-dimensional localisation, lightweight onboard inference, autonomous navigation, grasp planning, and environmentally sensitive retrieval. The framework provides a structured pathway towards scalable AI-supported monitoring and removal of submerged plastic waste while supporting long-term marine environmental sustainability. The proposed framework supports Sustainable Development Goal 14 (Life Below Water), particularly Target 14.1, by advancing AI-driven monitoring and potential removal of submerged plastic waste.

Keywords

Deep learning · Submerged plastic · Multimodal sensing · Semantic segmentation · Marine robotics · Ocean sustainability

Shahen Ramnath (corresponding author)

School of Computing, University of South Africa, Johannesburg, South Africa

e-mail: 54081882@mylife.unisa.ac.za

Sreedevi Vallabapurapu

School of Computing, University of South Africa, Johannesburg, South Africa

Suvendu Chandan Nayak

Department of Computer Science & Engineering, Silicon University, Bhubaneswar, India

1 Introduction

Marine plastic pollution significantly affects marine environments and can persist within marine ecosystems for long periods (United Nations Environment Programme 2026). Visible surface plastic can be addressed through beach cleaning, surface skimming, and mechanical collection, whereas submerged plastic presents a greater challenge. Plastic waste may sink, become trapped in coral or kelp habitats, remain suspended in the water column, or become (Islam 2026). These conditions make detection, localisation, classification, and removal more difficult. Submerged plastic can harm marine ecosystems and biodiversity, while smaller fragments become increasingly difficult to detect and retrieve. Conventional monitoring and removal methods are limited by safety, cost, depth, habitat disturbance, and poor underwater visibility. These challenges support the need for automated systems that integrate multiple sensing technologies, intelligent analysis, and robotic intervention. Artificial intelligence (AI) deep learning, can support this need by analysing underwater images, sonar, acoustic data, and video for detection, classification, segmentation, and autonomous decision-making. CNNs, Vision Transformers, object-detection and segmentation models, and reinforcement learning can therefore be integrated into intelligent ocean-monitoring and clean-up systems.

Underwater image enhancement and pixel-level marine-debris segmentation represent different components of the underwater perception pipeline (Wang et al. 2025). However, improved visual quality does not necessarily translate into improved debris detection or segmentation performance. This creates a need to evaluate image enhancement in relation to downstream perception tasks rather than visual appearance alone. The proposed framework therefore evaluates enhancement and segmentation together, while also investigating domain-adaptation methods to address differences between training data and images acquired under varying underwater conditions (Walker et al. 2024). Multimodal sensing and robotic systems have been investigated for underwater object detection and marine-debris removal. However, autonomous submerged-plastic removal still requires the integration of perception, localisation, navigation, onboard computation, and environmentally sensitive retrieval. AUVs and ROVs also operate under energy and computational constraints, while underwater image variability can reduce AI model generalisation. Differences in depth, turbidity, illumination, camera characteristics, and biofouling create challenges when AI models are transferred across environments. These limitations motivate a framework that combines multimodal perception with efficient onboard processing and cross-environment evaluation.

The proposed framework addresses key challenges related to underwater visibility, limited annotated data, domain shift, three-dimensional localisation, onboard computational

constraints, autonomous navigation, robotic retrieval, and environmental safety. Accordingly, this chapter presents an integrated multimodal deep learning framework for the detection and autonomous removal of submerged ocean plastic waste. The framework combines underwater sensing, image and sonar processing, multimodal perception, three-dimensional localisation, autonomous navigation, robotic manipulation, and environmental performance assessment. This research supports SDG 14 (Life Below Water), particularly Targets 14.1 and 14.2, through AI-driven approaches to marine plastic pollution and ecosystem protection.

2 Related Work

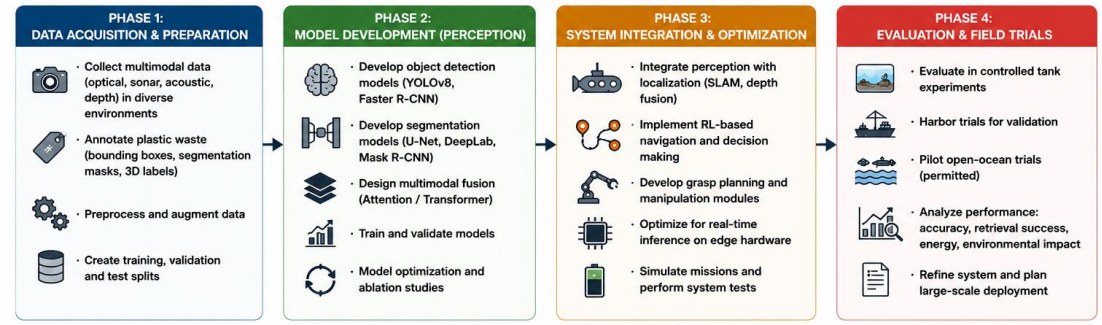
Deep learning has been increasingly applied to underwater debris detection, classification, and segmentation. CNN-based approaches extract spatial features from underwater imagery, while object detectors provide target localisation and segmentation models identify object boundaries at pixel level. EnYOLO combines underwater image enhancement and object detection using a lightweight backbone and domain-adaptation strategy, demonstrating the importance of addressing both image quality and deployment efficiency (Wen et al. 2024). Pixel-level segmentation is particularly important for submerged debris because object boundaries can support size estimation, separation from the background, and subsequent robotic interaction. OceanWaveNet developed for underwater plastic-waste segmentation (Ramnath et al. 2026), informs the research for the detection & segmentation component, while the present chapter considers a broader integration framework. WaterBiSeg-Net further addresses marine-debris segmentation through multi-scale feature enhancement, gated aggregation, and boundary-related processing (Zhang et al. 2024). These studies demonstrate progress in underwater perception, but accurate segmentation alone does not address localisation, navigation, or physical retrieval.

Vision Transformers, transfer learning, augmentation, and domain adaptation can support underwater perception where labelled data are limited. However, variations in turbidity, illumination, camera characteristics, and biofouling can reduce generalisation across environments, as highlighted by the SeaClear Marine Debris Dataset (uraš et al. 2024). Optical sensing provides detailed visual information but degrades under low illumination and turbidity. Acoustic sensing is less dependent on visible light but introduces challenges such as low contrast, imaging ambiguity, and variable spatial resolution (Zhou and Mizuno 2026). Combining these modalities can therefore improve underwater perception, although sensor alignment remains a challenge. UAMFDet addresses acoustic–optical misalignment through feature fusion and cross-attention (Chen et al. 2025). Other modalities, including depth sensing, LiDAR, and hyperspectral imaging, provide complementary information, although their suitability depends on operating conditions (Goddijn-Murphy et al. 2024). No single sensing modality is therefore suitable for all underwater environments.

Autonomous marine robotics extends AI-based perception towards physical debris removal. AQUA-SLAM integrates acoustic, visual, and inertial information for underwater localisation (Xu et al. 2025), while Han et al. demonstrate tactile sensing for underwater grasping (Han et al. 2025). At the system level, SeaClear integrates robotic platforms with litter detection, navigation, mapping, and seabed collection (Ilioudi et al. 2026). Together, these studies demonstrate progress in localisation, manipulation, and autonomous debris removal. However, a key challenge remains the integration of these capabilities into a computationally efficient framework that can detect, localise, and retrieve submerged plastic across changing underwater conditions while minimising disturbance to marine habitats.

3 Proposed Framework

The proposed research design follows an iterative and modular pipeline consisting of four principal phases depicted in Figure 1 data collection and curation, perception-model development, robotic integration and control, and evaluation and field validation. Each phase feeds into the next while allowing repeated model refinement, domain adaptation, and performance assessment.



Research methodology and system-development phases for the proposed autonomous submerged-plastic detection and retrieval framework.

Summary of the four-phase research methodology.

Phase	Core Activities	Output
Phase I: Data Acquisition	Collect multimodal underwater data from optical, sonar, acoustic, depth, and positioning sensors; annotate plastic debris; apply augmentation and domain adaptation; and create training, validation, and test sets.	Prepared and annotated multimodal dataset for model development.
Phase II: Model Development	Develop and compare preprocessing, object detection, segmentation, multimodal fusion, and depth or pose-	Trained perception models and three-dimensional target information for

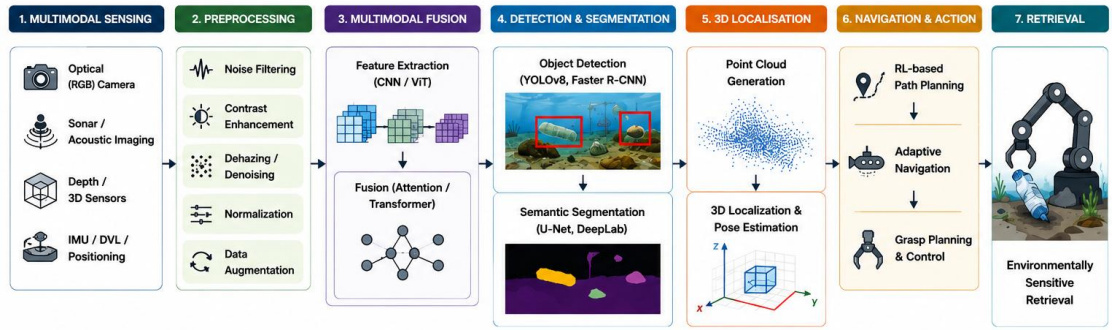
Phase	Core Activities	Output
	estimation approaches.	robotic interaction.
Phase III: System Integration	Optimise models for AUV and ROV hardware and integrate perception with localisation, navigation, grasp planning, and robotic retrieval.	Integrated perception-to-action system for autonomous submerged-plastic retrieval.
Phase IV: Evaluation	Evaluate perception, localisation, computational efficiency, robotic performance, environmental safety, and system robustness using the metrics in Table 2 and simulation, tank, harbour, and open-ocean trials.	Performance assessment and validation of the proposed framework under progressively realistic underwater conditions.

Phase IV evaluation metrics for the proposed framework.

Evaluation Area	Metric	Purpose
Detection	mAP, Precision, Recall, F1-score	Measure object-detection accuracy, false detections, missed targets, and the balance between precision and recall.
Segmentation	IoU, Dice	Measure the overlap between predicted and ground-truth debris segmentation masks.
Depth Estimation	RMSE	Measure the error between predicted and true target depth.
Computational Performance	Inference latency, FPS, power consumption	Assess whether the perception models are suitable for real-time onboard AUV and ROV operation.
Robotic Retrieval	Retrieval success rate	Measure successful grasps relative to total grasp attempts: $R_{\text{success}} = \frac{N_{\text{successful}}}{N_{\text{attempts}}} \times 100.$
Retrieval Efficiency	Throughput	Measure the quantity of debris removed over mission time: $T = \frac{M_{\text{removed}}}{t_{\text{mission}}}.$
Energy Efficiency	Energy per kilogram	Measure the energy required to remove debris: $E_{\text{kg}} = \frac{E_{\text{consumed}}}{M_{\text{removed}}}.$
Environmental Safety	Habitat disturbance	Assess observable effects such as seabed contact, disturbance to marine organisms, and seafloor disruption.

The proposed model displayed in Figure 2 with multimodal underwater sensing. Optical cameras, sonar or acoustic sensors, and depth information provide complementary observations of the underwater environment. Sensor data are preprocessed before being supplied to deep learning models. Detection and segmentation outputs are fused with

depth information to produce a three-dimensional representation of potential plastic targets. These outputs guide autonomous navigation, grasp planning, and retrieval.



Proposed multimodal deep learning framework for detection, localisation, and autonomous removal of submerged ocean plastic waste.

4 Simulation

Building on the OceanWaveNet segmentation stage, a synthetic proof-of-concept simulation was developed to evaluate three-dimensional localisation, navigation, and retrieval. The simulation assumptions are summarised in Table 3.

Synthetic parameters used in the proof-of-concept simulation.

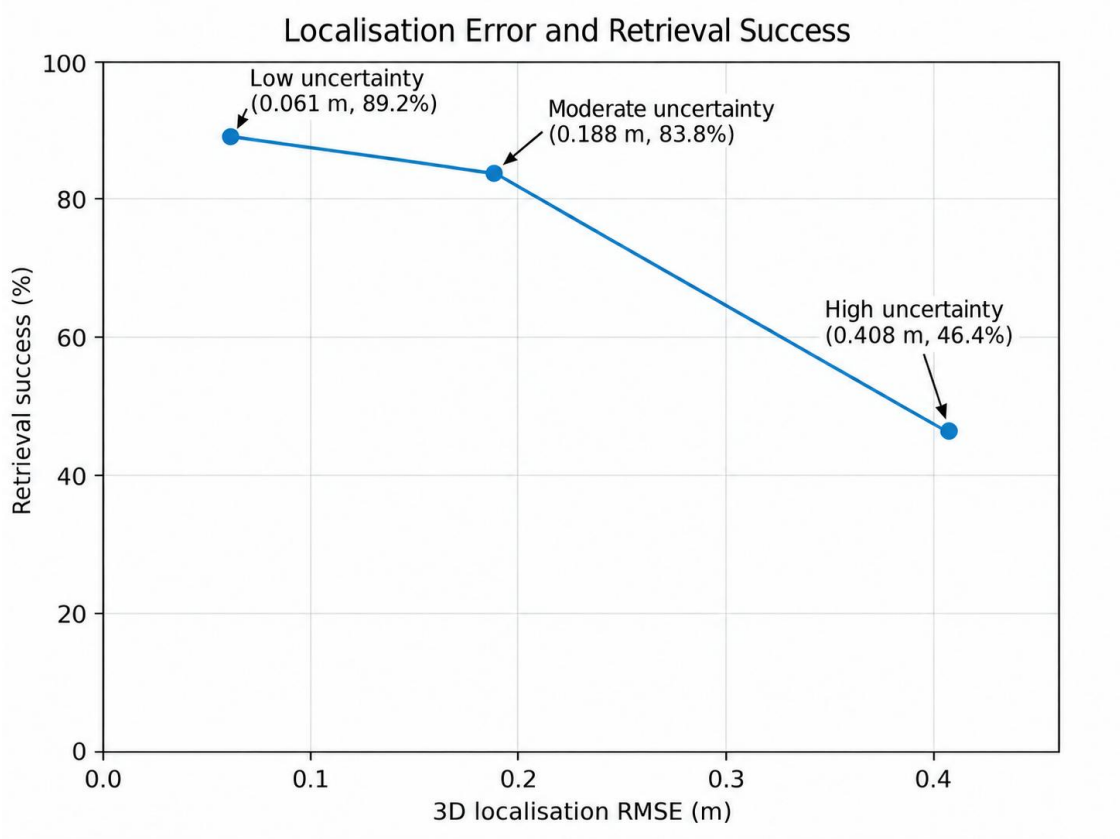
Simulation Parameter	**Synthetic Value / Assumption**
Segmentation input resolution	512×512 pixels
Absolute AUV/ROV operating depth	2000–3000 m
Local sensor-to-target range	1–10 m
Simulation trials	500 Monte Carlo trials per uncertainty condition
Camera model	Pinhole camera; $f_x = f_y = 500$ pixels, $c_x = c_y = 256$ pixels
Centroid uncertainty	2, 6, and 12 pixels standard deviation for low, moderate, and high uncertainty
Local range uncertainty	0.05, 0.15, and 0.35 m standard deviation for low, moderate, and high uncertainty
Orientation uncertainty	5°, 10°, and 18° standard deviation for low, moderate, and high uncertainty
Grasp localisation tolerance	0.40 m
Grasp orientation tolerance	25°
Minimum habitat-safety	0.30 m

Simulation Parameter	Synthetic Value / Assumption
buffer	
Navigation method	2.5D A* path planning
Local navigation workspace	12×12 m synthetic environment
Navigation grid	40×40 cells
Synthetic obstacle density	4%, 8%, and 12% for low, moderate, and high uncertainty

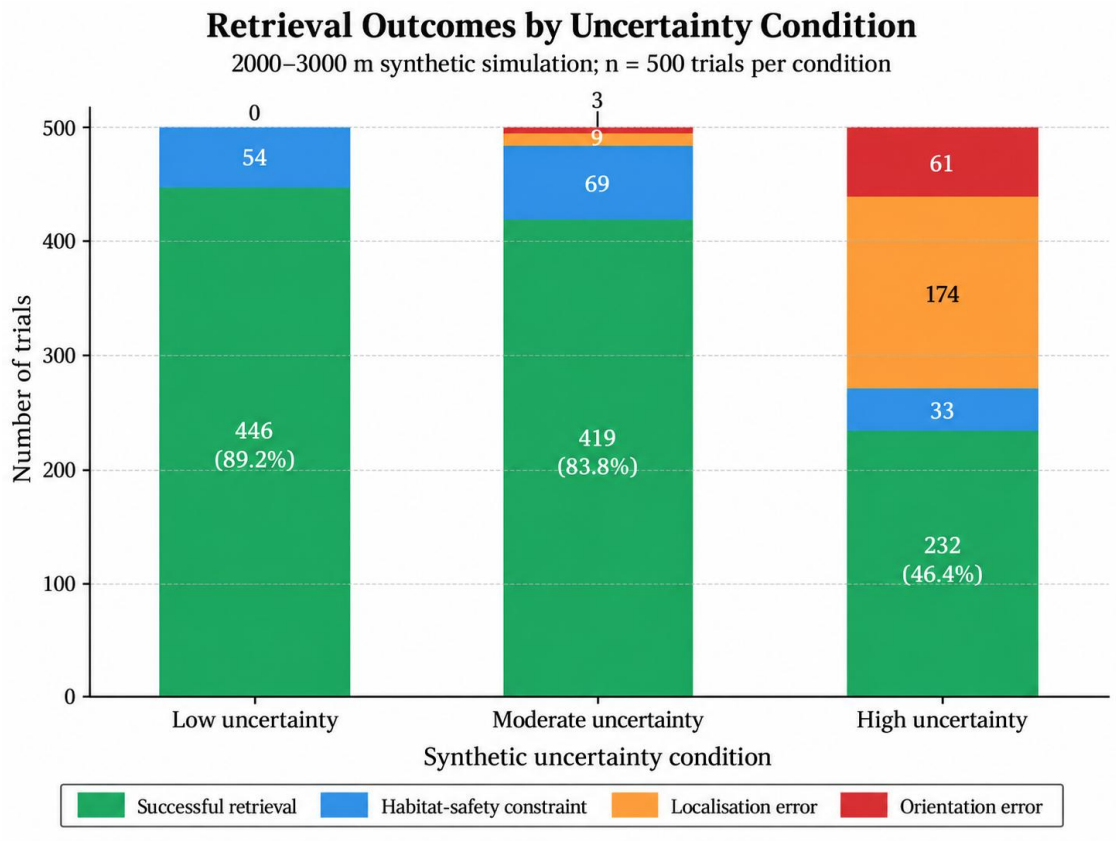
As shown in Table 4, localisation error increased with increasing uncertainty, with 3D RMSE rising from 0.062 m under low uncertainty to 0.408 m under high uncertainty. The A* planner identified a feasible path in all simulated trials under the selected synthetic obstacle conditions. Retrieval success decreased from 89.2% under low uncertainty to 46.4% under high uncertainty, indicating that increasing localisation and orientation uncertainty can propagate into downstream retrieval performance.

Proof-of-concept simulation results across 500 trials per uncertainty condition.

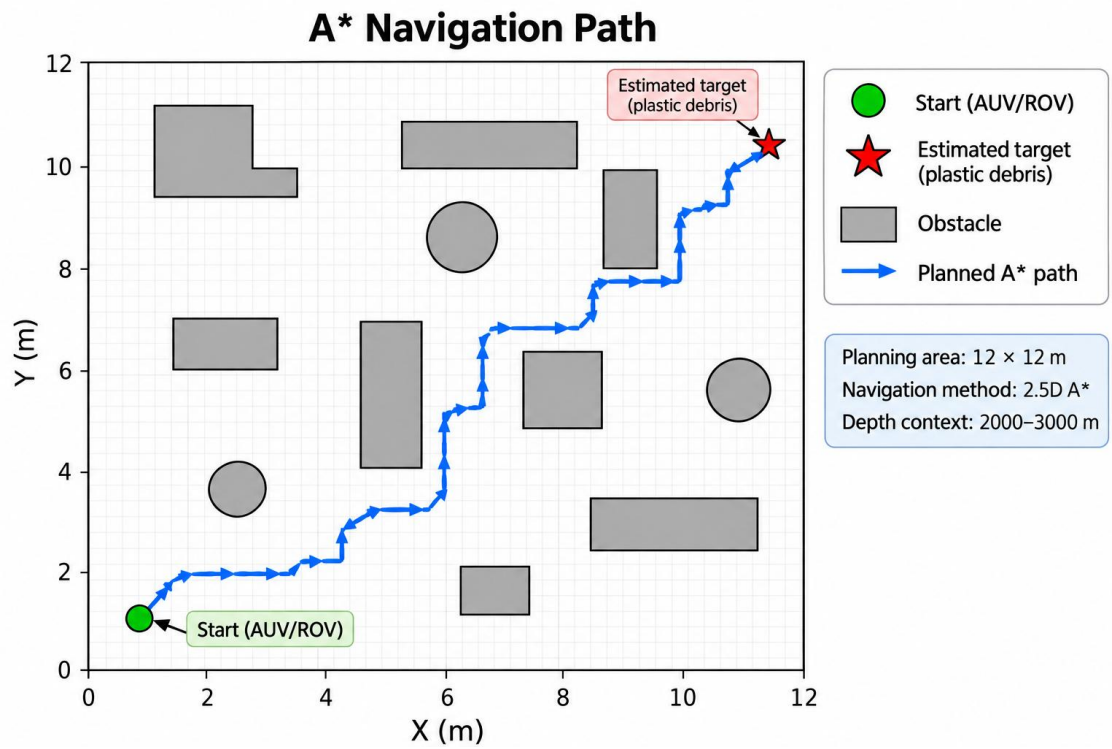
Condition	Depth (m)	3D RMSE (m)	Navigation (%)	Path (m)	Retrieval (%)
Low uncertainty	2482.2	0.062	100.0	7.28	89.2
Moderate uncertainty	2493.0	0.188	100.0	7.33	83.8
High uncertainty	2501.3	0.408	100.0	7.31	46.4



Relationship between three-dimensional localisation error and retrieval success across low, moderate, and high uncertainty conditions in the synthetic deep-water simulation.



Retrieval outcomes under low, moderate, and high synthetic uncertainty conditions.



A navigation path within the synthetic local workspace at an assumed AUV/ROV operating depth of 2000–3000 m. The planner identifies a collision-free path from the vehicle start position to the estimated target while avoiding synthetic obstacles.*

4.1 Discussion

The simulation indicates that three-dimensional localisation accuracy is a critical link between perception and autonomous retrieval. Retrieval displayed in figure 4 success decreased from 89.2% under low uncertainty to 46.4% under high uncertainty as localisation RMSE increased from 0.062 m to 0.408 m. The failure analysis strengthens this argument: under high uncertainty, 174 of the 500 trials failed because of localisation error and a further 61 because of orientation error, compared with no localisation or orientation failures under low uncertainty. This suggests that successful debris detection alone is insufficient for autonomous removal; the estimated target position and orientation must also remain within the tolerance required for robotic interaction. The results in Table [tab:comparison] therefore support evaluating the proposed framework as a connected perception-to-action system rather than assessing individual components in isolation.

The navigation results in Figure 5 should, however, be interpreted separately from retrieval performance. The A* planner identified in a feasible path in simulated trials, while retrieval success declined substantially as uncertainty increased. This distinction shows that reaching an estimated target location does not necessarily result in successful retrieval. Previous research similarly treats localisation, navigation, and underwater manipulation as distinct but connected problems. The present simulation complements AQUA-SLAM, SeaClear, learning-based-manipulation directions by examining how localisation and

orientation uncertainty can propagate into retrieval outcomes, rather than demonstrating superiority over these existing systems.

The results remain proof-of-concept and should not be interpreted as evidence of physical deep-sea performance. The simulated AUV/ROV operating depth was restricted to 2000–3000 m, while target sensing was modelled within a local range of 1–10 m. The simulation did not model depth-dependent pressure effects, deep-water vehicle dynamics, sensor degradation, communication constraints, or physical manipulation forces. In addition, the grasp tolerance, orientation tolerance, habitat-safety buffer, and obstacle distributions were synthetic. The results therefore demonstrate the feasibility of evaluating error propagation across localisation, navigation, and retrieval under controlled conditions, while physical AUV/ROV experiments and more realistic deep-water modelling remain necessary before conclusions about operational effectiveness can be made. The proposed framework supports SDG 14 through submerged plastic-waste monitoring and environmentally sensitive retrieval. However, further experimental validation is required to establish its effectiveness in reducing marine pollution and protecting marine ecosystems.

Study	**Primary Focus**	**Reported Numerical Result(s)**
AQUA-SLAM (Xu et al. 2025)	Underwater multisensor localisation and SLAM	Translation RMSE: 0.06–0.31 m. Rotation RMSE: 1.51–4.10°. Evaluation: 8 WaveTank sequences using 10-run averages.
Liu et al. (Liu et al. 2025)	Learning-based underwater manipulation	Grasp success: $20/20$ (100%). Mean grasp time: 13.52 s. Self-learning improvement: 41%. Rescue retrieval time: $41.23 \rightarrow 33.54$ s (22.9% improvement).
SeaClear (Ilioudi et al. 2026)	Integrated debris detection, localisation, navigation, and collection	mAP@0.5: 0.935. mAP@0.5:0.95: 0.711. ROV pose RMSE: 1.3 m. Bistrina autonomous collection: $3/4$ (75%). Lokrum collection: $4/4$ targets over five attempts.
Proposed simulation	3D localisation, A* navigation, and synthetic retrieval	3D RMSE (L/M/H): $0.062/0.188/0.408$ m. Navigation success (L/M/H): 100%. Retrieval success (L/M/H): $89.2/83.8/46.4\%$. Simulation trials: 500 per uncertainty condition.

Note: L/M/H denotes low, moderate, and high uncertainty conditions, respectively.

4.2 Limitations

Several limitations should be considered when interpreting the proposed framework and proof-of-concept simulation. Limited labelled underwater datasets may restrict model generalisation across locations, depths, sensor types, visibility conditions, and debris characteristics. Although augmentation, transfer learning, and domain adaptation can reduce data limitations, representative real underwater observations remain necessary for validation. Environmental conditions also constrain sensing, as severe turbidity can reduce optical performance while sonar may contain noise or insufficient resolution for small debris. In addition, AUV and ROV platforms must balance the computational demands of deep learning with limited energy available for sensing, propulsion, communication, and manipulation. Robotic retrieval introduces further uncertainty because debris may be buried, entangled, biofouled, or located near sensitive habitats. Finally, the proof-of-concept results are synthetic an operate at a depth of 2000–3000 m and do not model the full effects of deep-water pressure, vehicle dynamics, sensor degradation, communication constraints, or physical manipulation. Consequently, the simulation demonstrates framework behaviour under controlled conditions rather than operational performance in the real ocean.

5 Conclusions

This chapter presented a multimodal deep learning framework for the detection and autonomous removal of submerged ocean plastic waste, integrating underwater sensing, segmentation, three-dimensional localisation, navigation, and robotic retrieval. Building on the segmentation stage, the proof-of-concept simulation examined how localisation and orientation uncertainty can affect downstream system performance. As uncertainty increased, three-dimensional localisation RMSE increased from 0.062 m to 0.408 m, while retrieval success decreased from 89.2% to 46.4%, demonstrating that accurate perception and localisation are important for reliable autonomous retrieval. These results are based synthetically and therefore demonstrate framework behaviour rather than physical deep-sea performance. Future work should replace synthetic assumptions with real underwater perception outputs, and progress from controlled tank experiments to harbour and open-ocean trials. Further development should also improve multimodal sensing robustness, onboard computational and energy efficiency, and environmentally sensitive robotic manipulation. The framework therefore provides a structured basis for developing and evaluating AI-supported marine robotic systems for submerged plastic waste removal.

Ethics Ethics, Consent to Participate, and Consent to Publish declarations: not applicable

Declarations This research received no funding.

Data availability The synthetic datasets generated and analysed during the proof-of-concept simulations are available from the corresponding author on reasonable request.

Chen, Haojie, Zhuo Wang, Hongde Qin, and Xiaokai Mu. 2025. “Uamfdet: Acoustic-Optical Fusion for Underwater Multi-Modal Object Detection.” *Journal of Field Robotics* 42 (4): 970–83.

Goddijn-Murphy, Lonneke, Victor Martinez-Vicente, Heidi M Dierssen, et al. 2024. “Emerging Technologies for Remote Sensing of Floating and Submerged Plastic Litter.” *Remote Sensing* 16 (10): 1770.

Han, Yuxin, Chunyu Li, Deqing Mei, Pan Chen, and Yancheng Wang. 2025. “Development of Four-Fingered Dexterous Robotic Hand with Integrated Tactile Sensors for Underwater Objects Grasping.” *Sensors and Actuators A: Physical*, 117315.

Ilioudi, Athina, Stefan Sosnowski, Elisabeth Banken, et al. 2026. “The SeaClear System: An Intelligent Multi-Robot Solution for Autonomous Cleanup of Marine Debris on the Seabed.” *Engineering Applications of Artificial Intelligence* 170: 114094.

Islam, FA Samiul. 2026. “Marine Plastic Pollution: Ecological Risks, Environmental Distribution, and Sustainable Mitigation Perspectives.” *EQA-International Journal of Environmental Quality* 75: 1–24.

Liu, Ruoshi, Huy Ha, Mengxue Hou, Shuran Song, and Carl Vondrick. 2025. “Self-Improving Autonomous Underwater Manipulation.” *2025 IEEE International Conference on Robotics and Automation (ICRA)*, 16915–22.

Ramnath, S., S. Vallabapurapu, and C. Bester. 2026. “AI Model for Submerged Plastic Waste Removal in Ocean for Environmental Sustainability.” *Intelligent Computing and Sustainable Innovation in Technology (IC-SIT), Pp. 990 IEEE*.

United Nations Environment Programme. 2026. “Plastic Pollution and Marine Litter.” https://www.unep.org/topics/ocean-seas-and-coasts/ecosystem-degradation-pollution/plastic-pollution-marine-litter.

uraš, Antun, Ben J Wolf, Athina Ilioudi, Ivana Palunko, and Bart De Schutter. 2024. “A Dataset for Detection and Segmentation of Underwater Marine Debris in Shallow Waters.” *Scientific Data* 11 (1): 921.

Walker, Joseph L, Zheng Zeng, Chengchen L Wu, Jules S Jaffe, Kaitlin E Frasier, and Stuart S Sandin. 2024. “Underwater Object Detection Under Domain Shift.” *IEEE Journal of Oceanic Engineering* 49 (4): 1209–19.

Wang, Hao, Kevin Köser, and Peng Ren. 2025. “Large Foundation Model Empowered Discriminative Underwater Image Enhancement.” *IEEE Transactions on Geoscience and Remote Sensing* 63: 1–17.

Wen, Junjie, Jinqiang Cui, Benyun Zhao, et al. 2024. “EnYOLO: A Real-Time Framework for Domain-Adaptive Underwater Object Detection with Image Enhancement.” *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 12613–19.

Xu, Shida, Kaicheng Zhang, and Sen Wang. 2025. “AQUA-SLAM: Tightly Coupled Underwater Acoustic-Visual-Inertial SLAM with Sensor Calibration.” *IEEE Transactions on Robotics* 41: 2785–803.

Zhang, Wenming, Bofeng Wei, Yaqian Li, Haibin Li, and Tao Song. 2024. “WaterBiSeg-Net: An Underwater Bilateral Segmentation Network for Marine Debris Segmentation.” *Marine Pollution Bulletin* 205: 116644.

Zhou, Xiaoteng, and Katsunori Mizuno. 2026. “Benchmarking AI-Driven Acoustic Monitoring for Floating Marine Debris: Challenges in Deep Learning-Based Debris Extraction.” *Marine Pollution Bulletin* 222: 118655.