

Deep Learning-Based Visual Approaches for Underwater Garbage Detection

Dario Perhat*, Fausto Ferreira*,†

* University of Zagreb Faculty of Electrotechnical Engineering and Computing
Laboratory for Underwater Systems and Technologies (LABUST)
10000 Zagreb, Croatia

† CoE MARBLE - Centre of Excellence in Maritime Robotics
and Technologies for Sustainable Blue Economy
10000 Zagreb, Croatia
dario.perhat@fer.unizg.hr

Abstract—Marine debris severely impacts aquatic ecosystems and poses long-term environmental and economic risks. Manual underwater garbage removal is costly and inefficient, making visual detection by Autonomous Underwater Vehicles (AUVs) a promising alternative. This paper surveys deep learning approaches for camera-based underwater debris detection, focusing on real-time methods suitable for AUV deployment. Most reviewed studies employ large annotated datasets and convolutional neural networks, primarily YOLO-based. Their performance is compared with other techniques and datasets to evaluate detection effectiveness. The analysis highlights key trade-offs and aims to support the development of robust and accurate frameworks for underwater garbage detection.

Keywords—underwater debris detection, deep learning, convolutional neural networks, marine environment, object detection, autonomous underwater vehicles

I. INTRODUCTION

Due to growing amounts of solid waste, underwater ecosystems are increasingly polluted, threatening marine life and coastal areas [1]. Improper disposal causes underwater pollution, which harms marine flora and fauna, demanding proper monitoring and removal solutions [2]. However, detecting debris underwater is challenging due to low light, visual noise, and poor contrast, as shown in Fig. 1 (plastic bottle example).

To address these challenges, computer vision and machine learning offer effective automated detection solutions. These are not the only options and while vision remains popular, sonar [3], [4] and LiDAR [5], [6] have been used in low-light environments. Moreover, while for better detection, typically, the camera is positioned on-board AUVs or Remotely Operated Vehicles (ROVs), UAVs have been used in coastal settings [7]. In such cases, the camera can only detect objects at the surface or on land but not underwater and thus that type of works are not included in this paper. Likewise, approaches combining more than one sensor, e.g. hyperspectral Raman

This work has been funded by the European Union under the Horizon Europe Program, Grant No: 101087162. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

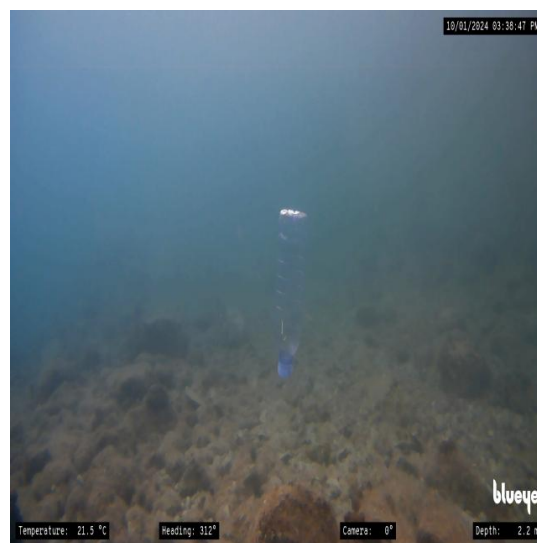


Fig. 1. A transparent plastic bottle underwater hardly identifiable using machine learning.

spectroscopy combined with LiDAR [8] are excluded from the survey as well. This paper focuses on camera-based methods for their superior resolution and visual information which are useful not only for detection but also for classification or segmentation. Moreover, for a focused contribution that can be self-contained in a conference paper and given their popularity, the article surveys deep learning (DL) techniques only. The goal is to provide a short review of the state of the art focusing on the most advanced techniques that have also shown promising results in a diversity of fields and thus are representative of the state of the art in object detection and in particular in underwater garbage detection.

Among DL techniques, YOLO (You Only Look Once) [9] is one of the most widely used object detection algorithms both in general and in underwater garbage detection and classification. Indeed, several versions of YOLO have been used in this outperforming other models like Tiny-YOLO [10],

SSD [11], and Faster R-CNN [12] as it shall be described in Section III.

Enhanced variants such as YOLOv5s-GTB [13] and YOLO-TrashCan [14] integrate techniques like GhostBottleneck [15] and Transformer-based feature extractors for better performance and efficiency. Moreover, several lightweight models such as YOLOX-S and ShuffleNetv2 [16] are evaluated for real-time detection onboard AUV/ROV.

Among the surveyed approaches, some use multiple cameras in [17], combining multiple viewpoints to improve accuracy. Although results show improved detection, delays in data transmission affect performance, indicating the need for more robust system designs.

This paper briefly compares state-of-the-art deep learning methods for underwater debris detection, evaluating accuracy, efficiency, and real-time applicability. Section II outlines the methodology, Section III presents results, and Section IV concludes with future work directions.

II. METHODOLOGY

The literature review was conducted primarily via IEEE Xplore and other scientific databases such as Scopus and IOPscience. Keywords such as underwater/marine garbage/litter/trash/debris detection/recognition, AI-based detection, AUV/ROV/underwater robot, vision/camera/video, deep learning/CNN/YOLO were used individually and in combination. In terms of inclusion and exclusion criteria, for the environment, only underwater studies were considered and no surface/coastal or above-water garbage detection was considered. For the sensor modality, only works using camera/video were included with sonar/LiDAR approaches mentioned only for context but not analyzed in detail. As the title indicates, the focus is on deep learning (mostly YOLO, Faster R-CNN and SSD) although hybrid setups where CNN features feed a classical classifier are included as well. For better comparison, only papers reporting quantitative metrics such as mAP, precision, recall, Frames per Second (FPS) are included. Finally, only peer-reviewed journal articles and full conference papers written in English and published between 2015 and 2025 were considered. In total, from 48 identified papers, around 20 relevant works were selected for analysis given the page limit of a conference paper.

This section categorizes the reviewed works by machine learning approach and dataset used. Fig. 2 shows the division into YOLO-based and non-YOLO-based methods, as well as dataset grouping. This structure is maintained in Section III.

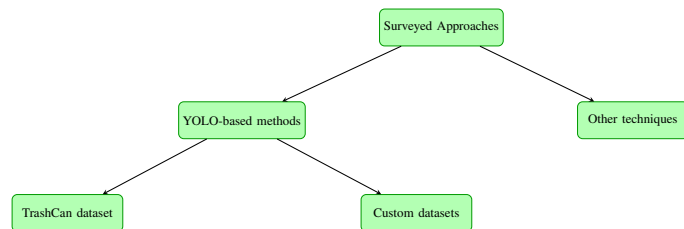


Fig. 2. Overview of machine learning methods and datasets used in underwater garbage detection.

III. RESULTS AND DISCUSSIONS

The reviewed works aim to improve model efficiency by balancing detection accuracy and processing speed, which is crucial for real-time systems like AUVs. This section examines key methods, their evolution, and main performance indicators such as precision, mAP, recall, and inference time.

A. YOLO Techniques

Most selected studies focus on YOLO-based approaches, so these are discussed first. Within them, several use the same dataset (TrashCan), which is analyzed firstly, followed by methods using other YOLO variants.

Additionally, although not used in the reviewed works, the DeepTrash dataset is worth noting. It contains over 30,000 labeled underwater images categorized by waste type and has been applied in other studies for deep learning evaluation.

1) *TrashCan Dataset*: The TrashCan dataset [18] is one of the most prominent datasets for underwater garbage detection. It consists of 7,212 images with various underwater objects such as plastics, nets, and other debris. Two recent studies used this dataset: one with a modified YOLOv4 [14] (YOLO-TrashCan) and another with YOLOv5s [19]. YOLOTrashCan achieves a mAP of 65.01%, 5.23% higher than YOLOv4, with 62% recall and 75% precision. Its model size is reduced to 214MB and inference speed is 36.17 FPS, making it suitable for real-time use.

In [19], YOLOv5s was chosen for its compact size (7MB), 7.5M parameters, and 140 layers. It achieved a mAP of 68.1% and an inference speed of 153.9 FPS. Precision and recall are not reported, preventing direct comparison, but the results suggest YOLOv5s is better suited for fast, efficient applications than YOLOTrashCan.

YOLOv5s-GTB [13], an improved version of YOLOv5s, includes architectural enhancements like GhostBottleneckV2, Transformer modules, and BiFPN. These contribute to a 0.9% mAP@0.5:0.95 improvement, 11.0% smaller model, and 10.7% lower complexity. However, it has lower inference speed than YOLOv5s, making it less suitable for time-critical use, despite its higher accuracy.

YOLOv8 [20] introduced detection and segmentation, achieving a box mAP of 45% and mask mAP of 36.2%. Although less accurate for detection than YOLOv5s-GTB, its segmentation capability suits tasks requiring detailed object shapes.

Another study [21] evaluated YOLOv8n, achieving 86.01% mAP, 88% precision, and 94.6% recall, with a compact 21MB size. However, its inference time (133.6 ms or 7.5 FPS) limits real-time usability. Interestingly, smaller models like YOLOv5s-GTB often outperform larger ones like YOLO-TrashCan in accuracy, showing that compactness can correlate with better performance.

A recent comparative study [22] introduced YOLOv9, tested alongside YOLOv8, SSD, and Faster R-CNN on TrashCan and DeepTrash. YOLOv9 achieved mAP@0.5 values of 90.13%

and 85.33%, respectively, outperforming YOLOv8n and affirming YOLOv9 as a current state-of-the-art solution for underwater debris detection.

To provide a clearer comparison of these methods, Table I summarizes the key metrics, including mAP, inference time, and model size.

TABLE I
PERFORMANCE COMPARISON OF DEBRIS DETECTION METHODS USING THE TRASHCAN DATASET

Method	Model	mAP	Inference Time(FPS)	Model Size(MB)
[14]	YOLOTrashCan	65.01%	36.18	214
[19]	YOLOv5s	68.1%	153.9	7
[13]	YOLOv5s-GTB	90.3%	117.64	12.9
[20]	YOLOv8	45%	NS	NS
[21]	YOLOv8n	86.01%	7.5	21
[22]	YOLOv9	90.13%	NS	NS

Apart from the models compared in Table I, several recent studies have introduced enhanced YOLO-based architectures tailored for underwater environments. For instance, SFD-YOLO [23] builds upon YOLOv8 by incorporating high-resolution inputs via a Residual Dense Network (RDN), improving the detection of small and partially occluded objects. This resulted in a performance boost from 84.5% to 91.2% mAP@0.5. Similarly, AIT-YOLOv7 [24] integrates attention mechanisms (CBAM), Swin Transformer blocks for resolution enhancement, and a diffusion-based image restoration module. Evaluated on the TrashCan dataset, it achieved 81.4% mAP@0.5, outperforming most baseline models used in earlier studies. These approaches illustrate how hybrid architectures that combine image enhancement and detection can significantly improve performance in challenging underwater conditions.

2) *Custom Dataset*: Custom datasets play a key role in advancing underwater waste detection, allowing focus on specific debris types and conditions not covered in general datasets like TrashCan. They enable tailoring to actual underwater conditions, including visibility and specific waste types.

Several reviewed studies developed custom datasets for targeted detection. [16] used 4,549 real-world underwater images with diverse debris types. [25] collected 5,720 annotated images of plastic waste from J-EDI videos. [26] created hybrid datasets by overlaying synthetic waste objects on real underwater backgrounds. [27] compiled over 50,000 frames with 8,000+ annotated objects under different lighting and visibility conditions. [28] captured 6,600 images of plastic bags, nets, and stones under varied illumination. [29] built a dataset from online images and YouTube, focusing on gloves and masks in marine settings.

In [25], YOLOv2 and Tiny YOLO outperformed SSD and Faster R-CNN in speed, with YOLOv2 achieving 82.3% mAP for plastic objects (vs. 83.3% of Faster R-CNN), and a fast 74 FPS with Darknet-19 backbone. Tiny YOLO, with fewer layers, achieved only 31.6% mAP but an impressive 205 FPS—ideal for constrained hardware.

In [26], YOLOv3 achieved 61% accuracy and 87% precision, outperforming Xception, ResNet, and VGG16. On

hybrid images with synthetic litter, accuracy dropped to 36%, indicating a need for better synthetic realism in training.

[27] showed YOLOv3 maintained 72% mAP under water turbidity and illumination challenges, but suggested exploring shape-based and attention models instead of just enlarging datasets.

In [28], a pruned YOLOv4 (P-YOLOv4) reduced model size to 7.06% of the original, improving speed from 43.5 to 58.8 FPS. Fine-tuning recovered accuracy close to original (90.6% vs. 91.4% AP), making it suitable for real-time low-resource systems.

[29] tested YOLOv5s and YOLOv5n for medical waste detection. YOLOv5s had higher precision (98.7%) but was more computationally demanding. YOLOv5n, with only 1.9M parameters and 97.5% mAP, was more suitable for AUVs/ROVs.

[16] proposed YOLOX-S with ShuffleNetV2 and ECA attention, improving model size, speed, and precision. Achieving 91.5% mAP and 54.4 FPS, it outperformed MobileNetV3 and MobileViT, offering strong real-time performance.

To provide a clearer comparison of these methods, Table II summarizes the key metrics, including mAP, inference time, and model size when available.

TABLE II
PERFORMANCE COMPARISON OF DEBRIS DETECTION METHODS USING CUSTOM DATASETS

Method	Model	mAP	Inference Time(FPS)	Model Size(MB)
[16]	YOLOX-S	91.5%	54.44	15
[25]	YOLOv2	47.9%	NA	NA
[26]	YOLOv3	80%	NA	NA
[27]	YOLOv3	72%	NA	NA
[28]	Pruned YOLOv4	90.3%	58.82	NA
[29]	YOLOv5s, YOLOv5n	YOLOv5s: 98.7%, YOLOv5n: 97.5%	NA	YOLOv5s: 14, YOLOv5n: 4

Through the analysis of studies that utilized custom datasets for underwater garbage detection, there is a clear growth in the efficiency of YOLO based models given the optimization techniques used and the variety of the datasets. Such custom datasets have been useful in that they contain simulations of the kinds of things one might find in the real world, such as turbid water, variable lighting, and a variety of types of waste. In terms of techniques, model pruning, attention mechanisms and light architectures like ShuffleNetV2 were used to enhance the effectiveness of the models for real time applications. These improvements are not only beneficial for the model accuracy performance but also processing time which is important in resource constrained systems like ROVs/AUVs and addresses the tradeoff with between accuracy and speed/computational complexity.

B. Other Techniques

In addition to YOLO architectures, which dominate most of the reviewed literature due to their real-time performance, a variety of alternative methods for underwater garbage detection have been explored. These include two-stage detectors such as Faster R-CNN, single-stage detectors like SSD and RetinaNet, and deep convolutional networks combined with classical classifiers. This subsection presents the most relevant non-YOLO approaches, highlighting their effectiveness and limitations under realistic underwater conditions.

In [25], Faster R-CNN and SSD algorithms were employed for object detection underwater together with YOLOv2 and Tiny YOLO (see subsection III-A2 for more on these two networks). Faster R-CNN is known to be an accurate two-stage algorithm that first detects Regions of Interest (ROIs) and then objects from those regions. Indeed, Faster R-CNN had the best mAP (81.0%) among all four techniques overall and in each class. However, this high accuracy comes together with high computational complexity, which makes the algorithm unsuitable for real time applications like those required by AUVs. Faster R-CNN had the lowest inference speed of 5.66 FPS on a TX2 hardware. On the other hand, SSD, a single stage algorithm that is relatively fast and almost as accurate as Faster R-CNN. In particular, SSD reached more than double inference speed (11.25 FPS) and the second best mAP of 67.4%.

In both cases, the accuracy of the object detection was found to be significantly reduced by several challenges in the underwater environment e.g., turbid water and low light conditions. The authors also state that it is necessary to reconfigure the algorithms for a particular underwater environment and to use high-quality training datasets to solve the problems specific to such environments.

In [30], the study conducted discusses the problem of recognizing objects in the deep sea using deep convolutional neural networks (CNNs). A new dataset DDI (Deep sea Debris Images) was created using real underwater debris data collected from JAMSTEC to ensure that the data was both realistic and robust. The authors propose a hybrid architecture, Shuffle-Xception, which combines separable convolutions, group convolutions, and channel shuffle to improve the efficiency of the feature learning and classification process. It is claimed that Shuffle-Xception is superior to five benchmark CNN models (ResNet V2 - 2 versions, MobileNet, Xception and LeNet) in terms of accuracy and computational time. It provided very high precision and recall (above 90% for 6 classes) in distinguishing seven classes of debris (metal, plastic, rubber, cloth, fishing net, glass, natural debris) despite the challenges of interclass similarities and intraclass variability.

The results of the experiments show that the choice of the group size for convolution is important for the model's sensitivity and complexity. The study also notes some drawbacks of the current models, such as ResNet and MobileNet, in dealing with the challenging underwater images. Although Xception had comparable performance, Shuffle-Xception had better feature learning ability, as shown by the author's heatmap analysis.

In [31], a number of approaches to underwater marine debris classification using machine learning techniques is evaluated, and specifically, CNNs are used for feature extraction. All the methods were evaluated in the context of real-world application using also the DDI dataset. The authors applied pre-trained CNN architectures such as DenseNet121, ResNet50 and VGG19 along with Support Vector Machines (SVM), Logistic Regression (LR), eXtreme Gradient Boosting (XGB) and Random Forest (RF). Of all the combinations, Neural

Networks (NN) with DenseNet121 produced the best test accuracy of 84%, while other combinations were slightly lower than that, for example, SVM and Logistic Regression. In addition, XGB with ResNet50 also gave a high accuracy of 75% and was also fairly effective in feature selection.

Therefore, the outcomes of the work reveal that it is necessary to combine advanced feature extraction architectures with optimal classifiers in order to address the challenges of underwater debris classification tasks.

In [26], a CNN with four convolutional layers and two fully connected layers was implemented specifically for the task. The network was trained on hybrid data that consisted of real underwater backgrounds on which synthetic images of debris were placed. Although an accuracy of 85% was achieved on test data, overfitting was observed because the dataset was small. The authors also propose increasing the size of the dataset using both real and synthetic images in future work.

On the other hand, [27] assesses the effectiveness of five pre-trained object detection models, including YOLOv3 (see subsection III-A2), SSD, RetinaNet and Faster R-CNN for underwater images. Among the non-YOLO methods, Faster R-CNN had the highest recall of 45.9%, but this came with the lowest precision of 53% due to high probability of false positives. RetinaNet had higher precision of 57.5% but with a drastically decreased recall of 16.6%, thus being less efficient at detecting small or unconventional objects. The authors also introduce the 'plausible-positive' metric to report correctly detected but uncommon objects not present in the ground truth with Faster R-CNN providing the best results in this new metric.

Both studies stress the necessity of optimizing algorithms for particular underwater conditions and the necessity of extensive, realistic datasets for ensuring that models generalize well to real-world settings.

The findings demonstrate that while YOLO-based approaches offer advantages in terms of speed and implementation simplicity, alternative methods provide valuable benefits in specific use cases. Challenges such as turbid water, low visibility, and limited computational resources call for a flexible approach to model selection. Hybrid CNN architectures, classical classifiers, and sensor-specific datasets indicate that there is no universally optimal solution — rather, the choice of method should be guided by the particular demands of the deployment environment.

Table III provides an overview of selected non-YOLO methods, including their task type, evaluation metrics, datasets, and relevant performance observations.

TABLE III
OVERVIEW OF NON-YOLO METHODS FOR UNDERWATER GARBAGE
DETECTION

Model	Task	Metrics	Values	Dataset
Faster R-CNN [25]	Detection	mAP, FPS	81.0%, 5.66	Custom
SSD [25]	Detection	mAP, FPS	67.4%, 11.25	Custom
RetinaNet [27]	Detection	Prec., Rec.	57.5%, 16.6%	Custom
Shuffle-Xception [30]	Classification	Prec., Rec.	>90% (6/7)	DDI
DenseNet121 + NN [31]	Classification	Accuracy	84%	DDI
ResNet50 + XGBoost [31]	Classification	Accuracy	75%	DDI
Custom CNN [26]	Classification	Accuracy	85%	Hybrid

IV. CONCLUSION

This paper highlights the relevance and progress of deep learning techniques for underwater debris detection using visual data. Among the analyzed works, YOLO-based models have proven particularly effective due to their balance between accuracy and speed, with models like YOLOv5s-GTB and YOLOv8n showing strong results. Non-YOLO approaches such as Faster R-CNN, SSD, and hybrid CNN architectures also demonstrate potential, particularly in precision-focused tasks. However, the comparison between methods is often complicated by inconsistent evaluation setups and varying datasets.

In particular, for real-time detection on-board AUV/ROVs, a small, fast YOLO version such as YOLOv5s is a suitable approach (≈ 153.9 FPS, 7 MB, mAP 68.1%). If a slight drop in speed is acceptable in exchange for higher accuracy and smaller model, YOLOv5s-GTB variants reported in the literature offer somewhat higher mAP and a smaller footprint, at the cost of lower FPS compared to v5s. YOLOX-S achieves the best performance on custom datasets by delivering 54.4 FPS while maintaining 91.5% mAP and pruned YOLOv4 reaches 58.8 FPS with 90.6% mAP, both showing a good tradeoff speed-accuracy. YOLO variants with larger sizes and two-stage detectors like Faster R-CNN achieve the highest accuracy but are not fit for real-time operation (mAP 81% at 5.66 FPS). Every situation includes trade-offs between input resolution and model complexity and FPS. High-resolution inputs with complex models produce better accuracy at a cost of lower frame rate but simple models achieve fast performance at the cost of detecting partially hidden objects. It is important to bear in mind that for the papers that use custom datasets, it is impossible to perform a direct comparison and so the numbers serve for qualitative assessment instead of speed-accuracy ranking.

Indeed, most reviewed papers focus on model-level optimizations, often tested under controlled or synthetic conditions. This points to several future research directions: further development of realistic and standardized benchmark datasets, the use of sensor fusion (e.g., combining visual input with sonar or LiDAR), and the design of lightweight, energy-efficient models for embedded platforms such as AUVs and ROVs. Furthermore, detection systems should aim not only to locate debris but also to assess its environmental hazard level, enabling smarter prioritization for removal.

Future work will train and test some of the best techniques identified in this survey in previous data of medical waste and plastic bottles collected both in a swimming pool [32] and at sea [33] in the framework of the MASK project (see Fig.1 for an example). Future work will also explore transfer learning techniques to understand the generalization capabilities of the tested approaches and will test newer CNNs like YOLOv12. Moreover, care will be put into curating diverse and high quality datasets and designing models that achieve a good trade off between speed, accuracy and resource utilization to support the deployment of the models on AUVs and ROVs for real time detection and classification.

REFERENCES

- [1] National Oceanic and Atmospheric Administration, <https://oceanservice.noaa.gov/facts/marinedebris.html>, [mrežno; stranica posjećena: prosinac 2024.].
- [2] U.S. Environmental Protection Agency, <https://www.epa.gov/trash-free-waters/learn-about-aquatic-trash>, [mrežno; stranica posjećena: prosinac 2024.].
- [3] J. Wang, C. Feng, L. Wang, G. Li, and B. He, "Detection of weak and small targets in forward-looking sonar image using multi-branch shuttle neural network," *IEEE Sensors Journal*, vol. 22, no. 7, pp. 6772–6783, 2022. [Online]. Available: <https://doi.org/10.1109/JSEN.2022.3147234>
- [4] M. Valdenegro-Toro, "Submerged marine debris detection with autonomous underwater vehicles," in *2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA)*, 2016, pp. 1–7.
- [5] M. R. Bariq, R. Darmakusuma, and R. Yusuf, "Underwater scanning using lidar," in *2023 International Conference on Electrical Engineering and Informatics (ICEEI)*, 2023, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/ICEEI59426.2023.10346874>
- [6] L. Zotta, S. Matteoli, M. Diani, and G. Corsini, "Afrodite: A fluorescence lidar simulator for underwater object detection applications," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3022–3041, 2015. [Online]. Available: <https://doi.org/10.1109/TGRS.2014.2367549>
- [7] J. Zhang, X. Zhou, Y. Huang, X. Chen, and B. M. Chen, "Uav-enabled smart coastal environment monitoring for marine litter," in *2024 IEEE 18th International Conference on Control & Automation (ICCA)*, 2024, pp. 370–375. [Online]. Available: <https://doi.org/10.1109/ICCA62789.2024.10591895>
- [8] T. Somekawa, S. Kurahashi, S. Matsuda, A. Yogo, and H. Kuze, "Remote detection and identification of plastics with hyperspectral raman imaging lidar," *Opt. Lett.*, vol. 50, no. 1, pp. 57–60, Jan 2025. [Online]. Available: <https://opg.optica.org/ol/abstract.cfm?URI=ol-50-1-57>
- [9] T. Leisheng, X. Hongli, W. Han, T. Dongxu, and G. Lei, "Research on collaborative object detection and recognition of autonomous underwater vehicle based on yolo algorithm," in *2021 33rd Chinese Control and Decision Conference (CCDC)*, 2021, pp. 1664–1669. [Online]. Available: <https://doi.org/10.1109/CCDC52312.2021.9601610>
- [10] U. G. I. Sudha, V. Lakshmanan, F. D. Shadrach, P. M. Krishnammal, and N. T. J., "Crime scene object detection from surveillance video by using tiny yolo algorithm," in *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, 2023, pp. 654–659. [Online]. Available: <https://doi.org/10.1109/ICPCSN58827.2023.00114>
- [11] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37.
- [12] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 06, pp. 1137–1149, Jun. 2017. [Online]. Available: <https://doi.ieeecomputersociety.org/10.1109/TPAMI.2016.2577031>
- [13] Y. Liu and H. Wang, "Yolov5s-gtb: An improved underwater garbage detection algorithm based on yolov5s," in *2024 9th International Conference on Computer and Communication Systems (ICCCS)*, 2024,

- pp. 33–38. [Online]. Available: <https://doi.org/10.1109/ICCCS61882.2024.10602942>
- [14] W. Zhou, F. Zheng, G. Yin, Y. Pang, and J. Yi, “Yolotrashcan: A deep learning marine debris detection network,” *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–12, 2023. [Online]. Available: <https://doi.org/10.1109/TIM.2022.3225044>
 - [15] Y. Li, M. Li, S. Li, and Y. Li, “Improved yolov5 for remote sensing rotating object detection,” in *2021 6th International Conference on Communication, Image and Signal Processing (CCISP)*, 2021, pp. 64–68. [Online]. Available: <https://doi.org/10.1109/CCISP52774.2021.9639292>
 - [16] L. Deng and P. Wang, “Lightweight underwater garbage detection algorithm based on yolox-s,” in *2022 4th International Symposium on Smart and Healthy Cities (ISHC)*, 2022, pp. 85–90. [Online]. Available: <https://doi.org/10.1109/ISHC56805.2022.00026>
 - [17] B. Jalil, L. Maggiani, and L. Valcarengi, “Marine litter classification through multi-camera system: Perception and latency,” in *2024 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, 2024, pp. 154–158.
 - [18] H. Jungseok, F. Michael, and S. Junaed, “Trashcan 1.0: An instance-segmentation labeled dataset of trash observations,” 2020. [Online]. Available: <https://conservancy.umn.edu/handle/11299/214865>
 - [19] C. S. Chin, A. Bo Hui Neo, and S. See, “Visual marine debris detection using yolov5s for autonomous underwater vehicle,” in *2022 IEEE/ACIS 22nd International Conference on Computer and Information Science (ICIS)*, 2022, pp. 20–24. [Online]. Available: <https://doi.org/10.1109/ICIS54925.2022.9882484>
 - [20] D. S. K. R. K. R. DV, S. D. S., and A. J., “A deep learning framework for detecting underwater trash,” in *2024 International Conference on Computing and Data Science (ICCDs)*, 2024, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/ICCDs60734.2024.10560433>
 - [21] S. Saji, M. S. Manikandan, J. Zhou, and L. R. Cenkeramaddi, “Underwater debris detection using visual images and yolov8n for marine pollution monitoring,” in *2024 IEEE 19th Conference on Industrial Electronics and Applications (ICIEA)*, 2024, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/ICIEA61579.2024.10664718>
 - [22] A. Khriiss, A. K. Elmiad, M. Badaoui, A.-E. Barkaoui, and Y. Zarhloule, “Exploring deep learning for underwater plastic debris detection and monitoring,” *Journal of Ecological Engineering*, 2024. [Online]. Available: <https://doi.org/10.12911/22998993/187970>
 - [23] F. Zhao, B. Huang, J. Wang, X. Shao, Q. Wu, D. Xi, Y. Liu, Y. Chen, G. Zhang, Z. Ren, J. Chen, and K. Mizuno, “Seafloor debris detection using underwater images and deep learning-driven image restoration: A case study from koh tao, thailand,” *Marine Pollution Bulletin*, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0025326X25001857>
 - [24] P. Pachaiyappan, G. Chidambaram, A. Jahid, and M. H. Alsharif, “Enhancing underwater object detection and classification using advanced imaging techniques: A novel approach with diffusion models,” *Sustainability*, 2024. [Online]. Available: <https://www.mdpi.com/2071-1050/16/17/7488>
 - [25] M. Fulton, J. Hong, M. J. Islam, and J. Sattar, “Robotic detection of marine litter using deep visual detection models,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 5752–5758. [Online]. Available: <https://doi.org/10.1109/ICRA.2019.8793975>
 - [26] J. Musić, S. Kružić, I. Stančić, and F. Alexandrou, “Detecting underwater sea litter using deep neural networks: An initial study,” in *2020 5th International Conference on Smart and Sustainable Technologies (SpliTech)*, 2020, pp. 1–6. [Online]. Available: <https://doi.org/10.23919/SpliTech49282.2020.9243709>
 - [27] M. Knapik and B. Cyganek, “Evaluation of deep learning strategies for underwater object search,” in *2019 First International Conference on Societal Automation (SA)*, 2019, pp. 1–6. [Online]. Available: <https://doi.org/10.1109/SA47457.2019.8938092>
 - [28] M. Tian, X. Li, S. Kong, L. Wu, and J. Yu, “Pruning-based yolov4 algorithm for underwater garbage detection,” in *2021 40th Chinese Control Conference (CCC)*, 2021, pp. 4008–4013. [Online]. Available: <https://doi.org/10.23919/CCC52363.2021.9550592>
 - [29] A. Lee, B. Jiang, I. Zeng, and M. Aibin, “Ocean medical waste detection for cpu-based underwater remotely operated vehicles (rovs),” in *2022 IEEE 13th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, 2022, pp. 0385–0389. [Online]. Available: <https://doi.org/10.1109/UEMCON54665.2022.9965708>
 - [30] B. Xue, B. Huang, G. Chen, H. Li, and W. Wei, “Deep-sea debris identification using deep convolutional neural networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 8909–8921, 2021. [Online]. Available: <https://doi.org/10.1109/JSTARS.2021.3107853>
 - [31] B. Jalil, L. Maggiani, and L. Valcarengi, “Comparative analysis of machine learning algorithms for the classification of underwater marine debris,” in *2023 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea)*, 2023, pp. 116–120. [Online]. Available: <https://doi.org/10.1109/MetroSea58055.2023.10317314>
 - [32] F. Ferreira, D. Nađ, T. Jaguš, N. Mišković, F. Maurelli, A. Blažina, and L. Maglić, “MASK - increasing sea knowledge awareness among high school students using marine robots,” in *OCEANS 2023 Limerick*, 2023, pp. 1–5.
 - [33] F. Ferreira, D. Nađ, N. Mišković, F. Maurelli, A. Blažina, and L. Maglić, “Teaching AI and marine robotics to high school students - the experience of the MASK project,” in *OCEANS 2025 Brest*, 2025, submitted.