International Journal of Recent Innovations in Academic Research
E-ISSN: 2635-3040; P-ISSN: 2659-1561
Homepage: https://www.ijriar.com/
Volume-9, Issue-4, October-December-2025: 369-374

Research Article

Marine Debris Object Detection Using YOLOv4 Model

Sooyun Bae

Korean Minjok Leadership Academy, HoengSeong, South Korea
Email: yuniebae07@gmail.com

Received: November 21, 2025 **Accepted:** December 12, 2025 **Published:** December 19, 2025

Abstract
Marine debris poses a serious threat to marine ecosystems, necessitating efficient and reliable detection methods to support large-scale monitoring and removal efforts. In this study, an underwater marine debris object detection framework based on the YOLOv4 deep learning architecture is proposed. The AI-Hub marine debris dataset was employed, comprising over 12,000 annotated underwater images. To improve data quality and training consistency, extensive preprocessing was conducted, including label cleaning, class unification, and conversion to YOLO-compatible annotation formats. Experimental results demonstrate that the pretrained YOLOv4 model achieves a mean Average Precision (mAP) of 88.9% and average 2.1ms/img, making it suitable for low-end GPU environments. YOLOv4 can be effectively adapted for real-time underwater marine debris detection while maintaining high detection accuracy, offering a practical solution for deployment in resource-constrained environments.
Keywords: YOLOv4, Deep Learning, Object Detection, Marine Pollution.

Introduction

Environmental pollution has long been recognized as a significant global concern, largely driven by the accumulation of waste, which represents the residual materials generated and discarded through human activities. Unmanaged waste results from factors such as inadequate landfill infrastructure and limited waste processing capacity, leading to its dispersal into terrestrial and aquatic environments. Waste that is discarded or abandoned in marine environments, regardless of intent, is defined as *marine debris* [1]. Marine debris encompasses biologically non-degradable solids, manufactured or processed products, originating from diverse sources such as fishing activities, beach tourism, industrial waste, and improper garbage disposal on land. Marine debris is mainly classified into plastic, metal, glass, rubber, and organic materials. It has been estimated that approximately 14 billion tons of waste enter the ocean annually [2]. Among them, plastic waste dominates as the most prevalent type, constituting around 60-80 percent of total marine debris.

Recently, robotics and artificial intelligence (AI) play critical roles across a wide range of research domains. In underwater research, robotic systems are indispensable for observing environments that are difficult or hazardous for humans to access, including shallow waters, subsea pipeline leak detection, and other constrained conditions. Autonomous Underwater Vehicles (AUVs), equipped with sensors such as side-scan sonar, optical cameras, echosounders, Acoustic Doppler Current Profilers (ADCPs), and Conductivity–Temperature–Depth (CTD) sensors, have proven to be highly effective platforms for underwater observation and data acquisition [3]. As unmanned systems, AUVs operate autonomously under computer-based control, enabling sustained and reliable monitoring without direct human intervention [4].

One of central research areas in AI is computer vision, which focuses on developing algorithms capable of interpreting and extracting meaningful information from images and videos in a manner analogous to human visual perception. Computer vision encompasses a broad range of technologies, including digital image processing, pattern recognition, and data analysis techniques based on machine learning and deep learning frameworks [5]. Among its core tasks, object detection aims to identify both the presence and spatial locations of specific objects within images or video streams. A fundamental task of object detection within computer vision has advanced rapidly since the introduction of Region-based Convolutional Neural Networks (R-CNNs) in 2014 [6].

Literature Review

In robotic detection of marine litter using deep visual detection models, Fulton et al. [7] introduced a marine litter detection dataset and reported that YOLOv2 has a mean average precision (mAP) of 47.9% with a processing speed of 205 frames per second (FPS) on a GTX 1080Ti. Yang et al. [8] evaluated YOLOv3 on the same dataset in Fulton et al. [7] and reported mAP of 76.1%, a recall of 75.6%, and a frame rate of 20 FPS. The study also compared YOLOv3 with Faster R-CNN, analyzing performance differences in accuracy and speed. Tata et al. [9] introduced the open-source *DeepTrash* dataset for marine plastic detection and employed four models: SSD, Faster R-CNN, YOLOv4-tiny, and YOLOv5s. YOLOv5s achieved the highest mAP (85%), while YOLOv4-tiny exhibited the fastest inference time (1.2ms/img). Majchrowska et al. [10] investigated large-scale environmental waste detection using multiple datasets and reported mAP of 7.3% on the Trash-ICRA 2019 dataset with EfficientDet-D2 model using the EfficientNet-B2 backbone. Performance degradation was attributed to poor image quality, while dataset combination with weighted sampling improved final accuracy to 73.02%. Tian et al. [11] introduced a pruning-based YOLOv4 method for underwater garbage detection. While the original YOLOv4 achieved 91.3% mAP at 43.4 FPS, the pruned model attained 90.3% mAP with an increased speed of 58.82 FPS. Recently, Sánchez-Ferrer et al. [12] evaluated Mask R-CNN for underwater debris detection and segmentation using *combined CleanSea* and *JAMSTEC* datasets. While the inclusion of synthetic data yielded an mAP50 of 52.1%, training without synthetic data improved performance to 61.0%.

Methodology

In this work, we perform object detection of marine debris using image processing and computer vision, aiming to identify the presence and position of specific objects in images or videos using You Only Look Once (YOLO) model. The YOLO model adopts a single-stage approach to object detection, offering an efficient and streamlined detection paradigm. Object detection methods are generally classified into two categories: single-stage detectors and two-stage detectors. Unlike two-stage detectors, which rely on a separate region proposal step followed by classification, single-stage detectors perform object localization and classification simultaneously within a single evaluation network. YOLO exemplifies this approach by using a unified neural network, resulting in significantly faster inference speeds compared to conventional object detection methods [13]. However, this architectural simplicity may lead to slightly reduced detection confidence relative to two-stage detectors.

The detection process begins by transforming the input image into a 3D matrix. During training, the image is divided into a grid, and each grid cell predicts multiple bounding boxes along with corresponding confidence scores [14]. Bounding boxes with confidence values below a predefined threshold are discarded, while remaining candidates are ranked according to their confidence scores. These candidates are subsequently refined through the non-maximum suppression (NMS) procedure, which eliminates redundant and overlapping bounding boxes to retain a single representative bounding box for each detected object [15]. The NMS process is governed by the Intersection over Union (IoU) metric, whereby the overlap between bounding boxes is computed, and boxes exceeding a specified IoU threshold are suppressed.

YOLOv4 represents a significant advancement over earlier YOLO architectures by providing flexibility in backbone selection. When configured with CSPDarknet53 as the backbone, YOLOv4 demonstrates superior performance in terms of both mAP and FPS compared to alternative backbones such as CSPResNeXt-50 and EfficientNet-B0 [16]. CSPDarknet53 is an enhanced version of Darknet-53, the original backbone of YOLOv3, in which Cross Stage Partial Network (CSPNet) modules are integrated into residual blocks to improve learning efficiency. Additionally, the activation functions are modified to incorporate both Mish and Leaky ReLU activations [17]. The Mish activation function exhibits self-regularizing and non-monotonic characteristics, producing smoother output distributions than the ReLU activation used in Darknet-53, improving feature representation [18].

In the neck layer, YOLOv4 integrates Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet) modules. SPP integrates a max-pooling layer with three pooling sizes (5×5, 9×9, 13×1, 1×1). Although PANet was originally developed for instance segmentation, its adaptation in YOLOv4 enables effective feature fusion through bidirectional upsampling and downsampling across both low-level and high-level feature maps [19]. The detection head of YOLOv4 closely follows the design of YOLOv3 and consists of three output layers dedicated to multi-scale object detection. These outputs are responsible for detecting small-, medium-, and large-scale objects, respectively [20]. In the backbone, CSPDarknet53 undertakes feature extraction on the image, while the neck incorporates SPP, executing max-pooling with four different pool sizes (5×5, 9×9, 13×13, 1×1) [21]. Post-SPP, another neck component, PANet, enhances the feature map's quality by

performing upsampling and downsampling using both low-level and high-level feature maps, strengthening multi-scale feature representation. The general architecture of YOLOv4 is illustrated in Figure 1.

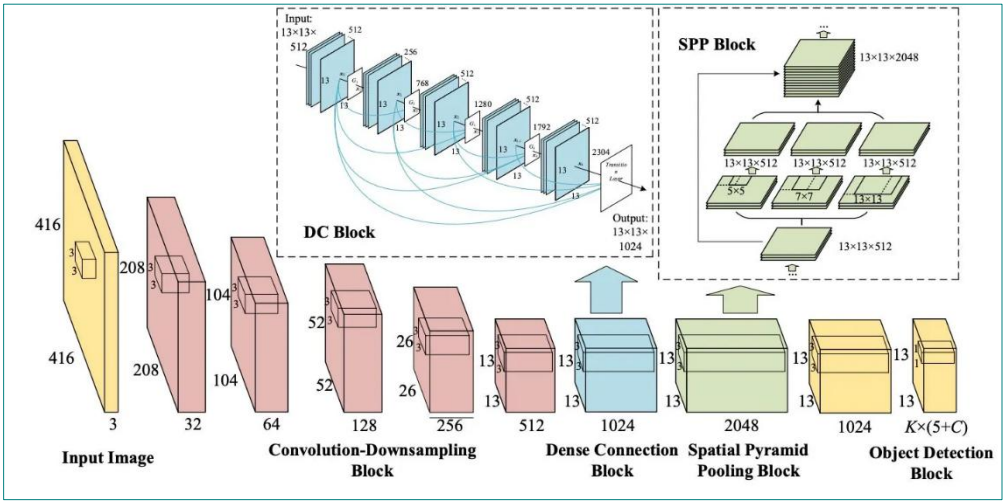

Figure 1. YOLOv4 architecture.

The Complete Intersection over Union (CIoU) loss is a distinctive addition in YOLOv4 and is employed to refine the calculation of IoU scores. The formula of the CIoU loss is given by

$$L_{CIoU} = 1 - IoU + \frac{\rho(b,b^{gt})}{c^2} + \alpha v, \quad (1)$$

Where ρ denotes the Euclidean distance, b and b^{gt} denote the central point of the predicted box and ground truth box, respectively, and c is the length of the shortest enclosing box covering two boxes. Here, α is a positive trade-off parameter and v measures the consistency of aspect ratio. The α is calculated as

$$\alpha = \frac{v}{(1-IoU)+v}, \quad (2)$$

and v is calculated as

$$v = \frac{4}{\pi^2}\left(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h}\right)^2, \quad (3)$$

Where w denotes the width, h denotes the height, and g_t denotes the ground truth. The object confidence loss and no object confidence loss use binary cross entropy to calculate the confidence prediction and confidence ground truth. Confidence ground truth becomes 1 if object is included in the box, and 0, otherwise. The formula of object confidence loss is defined as

$$Obj = \sum_{i=0}^{s^2}\sum_{j=0}^{B} I_{ij}^{obj}\left(\widehat{C}_i \log C_i + (1-\widehat{C}_i)\log(1-C_i)\right), \quad (4)$$

Where s^2 is the total number of grid cells, B is the total number of bounding boxes. i is a specific cell location of the grid, j is the bounding box accessed within the cell. $\widehat{C}_i$ denotes predicted confidence and C_i is actual confidence loss using binary cross entropy to calculate the class prediction that implement one-hot encoding.

I_{ij}^{Noobj} is the function that becomes 1 if the cell i inside the j bounding box contain an object, and 0, otherwise, which is defined as

$$NoObj = \sum_{i=0}^{s^2} I_{ij}^{obj} \sum_{c\in classes}\left(\widehat{C}_i \log C_i + (1-\widehat{C}_i)\log(1-C_i)\right), \quad (5)$$

Finally, classification rule is defined as

$$Class = \sum_{i=0}^{s^2} I_{ij}^{obj} \sum_{c\in classes}(\widehat{p}_i \log p_i + (1-\widehat{p}_i)\log(1-p_i)), \quad (6)$$

Where $\hat{p}_i$ and p_i are defined as predicted classification and actual classification for ith cell location, respectively.

Dataset and Analysis

The dataset was obtained from AI-Hub, the Korea AI Data Platform (https://www.aihub.or.kr/), through its official dataset catalog (https://www.aihub.or.kr/aihubdata/data/list.do), and consists of underwater images annotated in XML format. A total of 9,022 images were used for training and 3,008 images for validation. Ambiguous object classes were removed, and semantically identical labels were unified to ensure annotation consistency. All experiments were conducted using Google Colab with GPU acceleration. Image data and annotations were stored on Google Drive and mounted during runtime. The model was trained using default YOLOv4 hyper-parameters unless otherwise specified.

Table 1. Dataset statistics after preprocessing.

Dataset split	Number of images	Description
Training	9,022	Used for model learning
Validation	3,008	Used for performance evaluation

Figure 2 shows an example of some images containing marine debris including rope, fish traps and tires. Using YOLOv4 model, Figure 3 presents the types of marine debris and their locations. Figure 3 shows that YOLOv4 model accurately classifies the types and the locations of marine debris.

Figure 2. An example of marine debris.

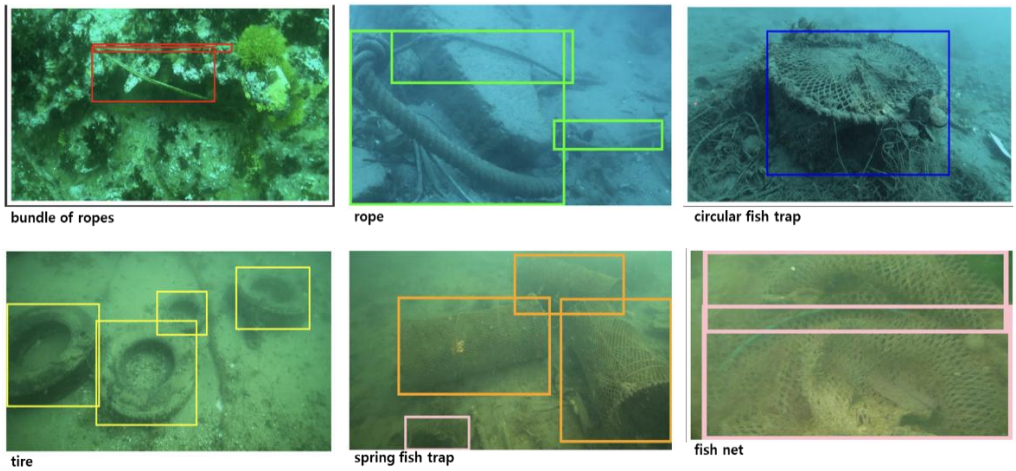

Figure 3. Classification and locating results of marine debris using YOLO model.

Discussion

Based on the experimental results obtained from the six evaluated schemes, it can be concluded that an efficient YOLOv4-based model for marine debris object detection can be achieved. The proposed pruned model attains a mean Average Precision (mAP) of 88.9%, while significantly improving inference speed to average 2.1ms/img. Notably, this optimization increases the computing speed with only a marginal reduction in detection accuracy.

Sometimes, YOLOv4 model fails to accurately classify marine debris owing to poor image quality, and bounding boxes exceeding image size. Bounding boxes that exceed the image boundaries are handled by clipping their normalized parameters such that any value greater than 1 is capped at 1. An alternative normalization strategy must be applied.

Conclusion

This study demonstrates the effectiveness of the YOLOv4-based object detection framework for underwater marine debris detection. Through systematic evaluation of multiple training and optimization strategies, the results confirm that YOLOv4 can achieve high detection accuracy while being adapted for real-time performance. These results highlight the feasibility of deploying YOLOv4-based marine debris detection systems in practical, resource-constrained settings, such as autonomous underwater vehicles and embedded monitoring platforms.

Overall, this work provides a solid foundation for real-time, high-accuracy marine debris detection and contributes to the advancement of AI-assisted marine environmental monitoring. Nevertheless, challenges remain due to underwater image degradation, including poor visibility and inaccurate bounding box annotations exceeding image boundaries. While this study mitigates such issues through bounding box clipping and normalization, further refinement is required to enhance robustness under adverse imaging conditions.

Declarations

Acknowledgements: The author acknowledges that this study was conducted independently and did not receive support from any institution or external organization.

Author Contribution: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Conflict of Interest: The author declares no conflict of interest.

Consent to Publish: The author agrees to publish the paper in International Journal of Recent Innovations in Academic Research.

Data Availability Statement: The dataset used in this study was obtained from AI-Hub, the Korea AI Data Platform, via its official dataset catalog (https://www.aihub.or.kr/aihubdata/data/list.do).

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval was not required for this study.

Informed Consent Statement: Informed consent was not applicable to this research.

Research Content: The research content of this manuscript is original and has not been published elsewhere.

References

1. Iñiguez, M.E., Conesa, J.A. and Fullana, A. 2016. Marine debris occurrence and treatment: A review. Renewable and Sustainable Energy Reviews, 64: 394-402.
2. Hetherington, J., Leous, J., Anziano, J., Brockett, D., Cherson, A., et al. 2005. The marine debris research, prevention and reduction act: A policy analysis (pp. 2–35). Columbia University, New York.
3. Hyakudome, T. 2011. Design of autonomous underwater vehicle. International Journal of Advanced Robotic Systems, 8(1). https://doi.org/10.5772/10536
4. Chang, Y.C., Hsu, S.K. and Tsai, C.H. 2010. Sidescan sonar image processing: Correcting brightness variation and patching gaps. Journal of Marine Science and Technology, 18(6): Article 1.
5. Cosido, O., Iglesias, A., Galvez, A., Catuogno, R. Campi, M., Terán, L. and Sainz, E. 2014. Hybridization of convergent photogrammetry, computer vision, and artificial intelligence for digital documentation of cultural heritage: A case study of the Magdalena Palace. In: Proceedings of the 2014 International Conference on Cyberworlds (pp. 369–376). IEEE.
6. Zou, Z., Chen, K., Shi, Z., Guo, Y. and Ye, J. 2023. Object detection in 20 years: A survey. Proceedings of the IEEE, 111(3): 257-276.
7. Fulton, M., Hong, J., Islam, M.J. and Sattar, J. 2019. Robotic detection of marine litter using deep visual detection models. In: Proceedings of the 2019 International Conference on Robotics and Automation (ICRA) (pp. 5752–5758). IEEE.
8. Yang, H., Liu, P., Hu, Y. and Fu, J. 2021. Research on underwater object recognition based on YOLOv3. Microsystem Technologies, 27(4): 1837-1844.

9. Tata, G., Royer, S.J., Poirion, O. and Lowe, J. 2021. A robotic approach towards quantifying epipelagic bound plastic using deep visual models. arXiv. https://arxiv.org/abs/2105.01882
10. Majchrowska, S., Ferlin, M., Klawikowska, Z., Plantykow, M.A., Kwasigroch, A. and Majek, K. 2021. Waste detection in Pomerania: Non-profit project for detecting waste in the environment. arXiv. https://arxiv.org/abs/2105.06808
11. Tian, M., Li, X., Kong, S. and Yu, J. 2021. Pruning-based YOLOv4 algorithm for underwater garbage detection. In: Proceedings of the 2021 40th Chinese Control Conference (CCC) (pp. 4008–4013). IEEE.
12. Sánchez-Ferrer, A., Valero-Mas, J.J., Gallego, A.J. and Calvo-Zaragoza, J. 2023. An experimental study on marine debris location and recognition using object detection. Pattern Recognition Letters, 168: 154-161.
13. Srivastava, S., Divekar, A.V., Anilkumar, C., Naik, I., Kulkarni, V. and Pattabiraman, V. 2021. Comparative analysis of deep learning image detection algorithms. Journal of Big data, 8(1): 66.
14. Diwan, T., Anirudh, G. and Tembhurne, J.V. 2023. Object detection using YOLO: Challenges, architectural successors, datasets and applications. Multimedia Tools and Applications, 82(6): 9243-9275.
15. Hosang, J., Benenson, R. and Schiele, B. 2017. Learning non-maximum suppression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 4507–4515).
16. Mahasin, M. and Dewi, I.A. 2022. Comparison of CSPDarknet53, CSPResNeXt-50, and EfficientNet-B0 backbones on YOLOv4 as an object detector. International Journal of Engineering, Science and Information Technology, 2(3): 64–72.
17. Bochkovskiy, A., Wang, C.Y. and Liao, H.Y.M. 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv. https://arxiv.org/abs/2004.10934
18. Misra, D. 2019. A self-regularized non-monotonic activation function. arXiv. https://arxiv.org/abs/1908.08681
19. Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. 2018. Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 8759–8768).
20. Redmon, J. and Farhadi, A. 2018. YOLOv3: An incremental improvement. arXiv. https://arxiv.org/abs/1804.02767
21. Qiang, Z., Yuanyu, W., Liang, Z., Jin, Z., Yu, L. and Dandan, L. 2021. Research on real-time reasoning based on Jetson TX2 heterogeneous acceleration YOLOv4. In: Proceedings of the 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA) (pp. 455–459). IEEE.

Citation: Sooyun Bae. 2025. Marine Debris Object Detection Using YOLOv4 Model. International Journal of Recent Innovations in Academic Research, 9(4): 369-374.