EMITTER International Journal of Engineering Technology
Vol. 13, No. 2, December 2025, pp. 214~228
DOI : 10.24003/emitter.v13i2.978

FloYO-Net: Enhancing Small Floating Waste Detection in Natural Waters Using Atrous YOLOv5s

Badiu Badams[1], Usman Ullah Sheikh[1], Norhaliza Abd Wahab[1], Syed Abdul Rahman Bin Syed Abu Bakar[1]

[1]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM, Johor Bahru, Malaysia.
Corresponding Author: usman@utm.my

Received June 20, 2025; Revised July 27, 2025; Accepted September 12, 2025

Abstract

Detecting small and partially hidden objects in rivers and water bodies remains a major challenge for real-time waste detection systems. These objects are often missed due to their small size, low contrast, and cluttered surroundings. Further complicating the task is the lack of dedicated datasets focused on small floating debris, limiting the development of more capable detection models. To bridge this gap, we developed D_six, a custom dataset of 495 high-resolution images capturing six classes of floating waste under real-world conditions. In this study, we improve the YOLOv5s object detection model by integrating atrous convolutions at three key backbone layers: P1/2, P3/8, and P5/32. These layers represent different scales of the feature pyramid, and the strategic placement of atrous convolution at each level plays a crucial role in helping the model recognize small and occluded objects more effectively. Using a dilation rate of 6, the model's receptive field is expanded without increasing its size or slowing it down. When trained and evaluated on the D_six data set, the FloYO-Net (Floating Object YOLO Network) consistently outperformed the standard YOLOv5s, achieving a mean Average Precision (mAP@0.5) of 0.828 and mAP@0.5:0.95 of 0.509, compared to 0.787 and 0.498 respectively. Improvements were especially notable for hard-to-detect items like plastic bottles and plastic drink containers, with average precision gains of 6.6% and 7.1%, respectively. These results demonstrate that atrous convolution — when thoughtfully placed — can significantly improve detection accuracy, making it a powerful enhancement for real-time environmental cleanup systems.

Keywords: Atrous convolution, floating object detection, environmental monitoring, small object detection.

1. INTRODUCTION

The growing presence of floating plastic waste in rivers, lakes, and urban waterways poses a significant threat to aquatic ecosystems, public health, and water management. Items like plastic bottles, take-out containers, and plastic bags accumulate in water bodies due to improper disposal and runoff, creating severe environmental hazards. Microplastic contamination has even been

found in human stools, emphasizing the need for effective waste detection and removal systems [1, 2].

Manual removal is labor-intensive, inefficient, and risky, particularly in large or hazardous bodies of water. Autonomous surface robots offer a scalable solution for automated waste removal, but their success depends on the accuracy and robustness of the detection system [3]. Recent advances in deep learning-based object detection, particularly with YOLOv5, have improved waste detection in aquatic environments [4, 5], balancing speed and accuracy [6-8]. However, standard YOLOv5s models struggle with small, partially submerged, or overlapping objects due to limitations in standard convolutions and downsampling-induced loss of spatial resolution [9].

We propose FloYO-Net, a modified YOLOv5s model that incorporates atrous convolutions at key backbone layers (P1/2, P3/8, P5/32) to expand the receptive field while preserving spatial resolution. This modification improves detection of small and occluded objects in cluttered aquatic environments. Our approach is validated with a custom floating waste dataset, collected using a 5K-resolution camera on a remote-controlled boat. Experimental results show significant improvements in precision, recall, and mAP compared to standard YOLOv5s, demonstrating the effectiveness of atrous convolutions for real-time aquatic waste detection.

2. RELATED WORKS

Object detection in complex environments, such as water bodies, remains a challenging task due to factors such as object occlusion, small object size, cluttered backgrounds, and varying lighting conditions [10]. Traditional two-stage detection frameworks such as R-CNN, Fast R-CNN, and Faster R-CNN offer high detection accuracy by separating region proposal and classification stages, but their slower inference times make them less suitable for real-time applications [11, 12]. One-stage detectors, such as YOLO (You Only Look Once), SSD, and RetinaNet, integrate these stages, offering improved inference speed while maintaining reasonable accuracy, making them ideal for real-time environmental monitoring [13-15].

The YOLO architecture has undergone several improvements since its initial release, evolving from YOLOv3 through to YOLOv5 and most recently YOLOv8. YOLOv5s, a lighter variant, is particularly suitable for edge computing and embedded systems due to its speed and reduced memory footprint[16]. Several studies have explored its use in environmental monitoring.

In [17, 18], an optimized YOLOv5s was applied to riverine solid waste detection, with data augmentation and training optimization techniques to boost performance on occluded objects. However, it struggled with small object detection due to limited receptive field and downsampling losses.

Recent local research has demonstrated substantial progress in the application of YOLO-based models for aquatic and waste detection tasks. YOLOv5 was employed to detect plastic bottles with an accuracy range of 82–93%, while also identifying detection challenges associated with object

occlusion and overlap [19]. Similarly, [20] utilized YOLOv8 for identifying floating debris in Jakarta's Ciliwung River, achieving a precision of 84% and recall of 91%. In the same vein, [21] introduced a tubelet-level bounding box linking mechanism to enhance YOLOv5 performance in aquatic object detection scenarios. Furthermore, [22] achieved a mean Average Precision (mAP) of 96.8% in a multi-class urban waste detection task using YOLOv8 Collectively, these works reflect the increasing contributions of Indonesian researchers to the domain of real-world aquatic debris detection and establish a comparative baseline for evaluating the performance of the proposed FloYO-Net framework.

Additionally, [23] demonstrated the application of stereo vision-based object detection in marine environments, highlighting advancements in aquatic perception technologies. The use of dilated (or atrous) convolution has gained attention for addressing small object detection challenges. Atrous convolution enlarges the receptive field of convolutional kernels without increasing the number of parameters or sacrificing spatial resolution [24-26]. Initially introduced in semantic segmentation tasks such as DeepLab [27], it has since been used in remote sensing [28], lung nodule detection [29], and even underwater object segmentation [30]. These studies demonstrate its effectiveness in extracting contextual information, especially in scenes where objects occupy limited pixels.

Attention-based and multi-scale fusion methods have also been proposed to improve detection of small and occluded objects [31, 32]. While these methods improve accuracy, they often introduce computational complexity, making them less ideal for deployment in real-time or resource-constrained systems such as aquatic drones or floating robots.

Few works have explored combining YOLO with atrous convolution. In [33], atrous convolution was used within SSD-based detection to enhance accuracy in marine debris detection, though inference time was not discussed. In [34], a novel attention-based dilated convolution network was proposed for acoustic scene analysis, achieving high precision. However, both approaches lack real-time validation in actual environmental monitoring setups.

Most datasets used in these studies are either limited to a single class (e.g., plastic bottles) or captured in controlled conditions. The FloW dataset [35] includes plastic waste in canals but lacks diversity in object types and environments, making generalization difficult. Additionally, few studies have explicitly tested on small or overlapping floating objects in natural water bodies.

In this work, we address these gaps by modifying the YOLOv5s backbone with atrous convolutions at three key feature scales (P1/2, P3/8, and P5/32), and evaluate performance using a custom dataset of six classes of floating waste, captured in real-world conditions using a 5K-resolution boat-mounted camera. Our approach preserves YOLOv5s's real-time capability while significantly improving detection accuracy for small and occluded objects, particularly plastic bottles and containers — classes previously shown to be

problematic for standard YOLO models. Table 1 summarizes key published methods—covering mAP@0.5, model footprint, computational cost, inference speed, and dataset used.

Tabel 1. Summary of state-of-the-art object detection models

Model	mAP@0.5 (%)	Model size (MB)	Parameters (M)	GFLOPs	FPS (ms)	Dataset
LFN-Yolo[36]	0.741	5.9	2.7	7.2	58	Trash can, URPC, Brakish
FRL-Yolo[37]	0.793	2.0	0.8	4.6	-	FloW-Img
Dynamic-Yolo[38]	0.686	-	8.21	12.51	60	DUO
Aqua-DETR[39]	0.63	-	50.36	78.33	35	Trash can

3. ORIGINALITY

Current object detection research often focuses on increasing model complexity by deepening networks or adding modules, such as attention layers, to improve accuracy. This study takes a different, more efficient approach. We propose a lightweight enhancement to the YOLOv5s architecture for real-time monitoring in natural environments, incorporating atrous convolutions with a dilation rate of 6 at critical layers (P1/2, P3/8, and P5/32). This modification broadens the receptive field without increasing the number of parameters or compromising resolution, making it especially effective for detecting small or partially occluded objects, such as floating debris in rivers.

What distinguishes this method is its ability to enhance detection performance without sacrificing speed or requiring high-end hardware. By strategically applying atrous convolutions where they are most beneficial, the model improves accuracy while maintaining its efficiency, even in cluttered or occluded scenes. In summary, the novelty of this work lies in striking a balance between model simplicity, computational efficiency, and robust performance, making it a practical, scalable solution for real-world applications like autonomous water-cleaning systems, where computational resources are constrained but high detection accuracy is essential.

4. SYSTEM DESIGN

The proposed system enhances small and occluded object detection in environmental monitoring by modifying the YOLOv5s architecture. It was implemented, trained, and tested using a dataset featuring cluttered riverine environments and partial occlusion scenarios.

The system consists of three core components: Data Preparation, Network Modification, and Evaluation & Inference. The key innovation is the

integration of atrous convolution layers in the YOLOv5s backbone, strategically placed to expand the receptive field without adding computational overhead or reducing spatial resolution. This enables the model to capture broader contextual information, making it particularly effective for detecting small or partially obscured objects in complex environments.

4.1 Dataset Preparation (D_six)

To address the scarcity of realistic aquatic debris benchmarks, we constructed D_six, a six-class dataset of small floating waste designed to emulate real-world waterborne challenges—occlusion, specular glare, dynamic lighting, and partial submersion. All 495 high-resolution frames were captured with a 170° ultra-wide-angle, waterproof 5K camera (24 MP, 30 FPS) equipped with optical stabilization and anti-shake processing. Recordings span three UTM rivers under varied urban–natural contexts, different times of day (morning to late afternoon), and weather states (sunny, overcast, cloudy), ensuring a mix of high-glare and low-contrast scenes. Expert annotators used LabelImg to draw bounding boxes for six debris categories—Plastic Bottles (PB), Plastic Bags (PG), Take-out Containers (TO), Plastic Drink Containers (PD), Styrofoam (SF), and Cans (CN)—with inter-annotator consensus on ambiguous cases. The dataset is split 70/15/15 into Training (341 images, 1 344 objects), Validation (74 images, 294 objects), and Test (73 images, 301 objects) sets. Sample frames by class are shown in Figure 1. D_six is publicly released at: https://doi.org/10.5281/zenodo.15195086

Figure 1. Data set Visualization (D_six)

4.2 YOLOv5s Architecture Modification

Figure 2 shows the YOLOv5s architecture, which forms the baseline for our floating debris detection system. It consists of three main components: the backbone (CSPDarknet53) for feature extraction, the neck (PANet) for multi-scale fusion, and the detection head for bounding box and class predictions. This efficient structure ensures fast, accurate detection, serving as the foundation for our modifications aimed at improving performance in real-world aquatic environments.

In this study, as shown in Figure 3, the backbone was modified by incorporating atrous convolutions (dilation rate = 6) at three key stages of the feature pyramid:

P1/2: A high-resolution map (320×320) that captures fine-grained details for detecting small objects.

P3/8: A mid-level feature map (80×80) that balances fine detail and abstraction.

P5/32: A broader context map (20×20) that captures high-level semantic information.

Atrous convolutions expand the receptive field without increasing parameters or reducing spatial resolution. For instance, a 3×3 convolution with a dilation rate of 2 covers the same area as a 5×5 kernel but with the same parameter count. As illustrated in Figure 4, higher dilation rates (e.g., 3) simulate larger kernels (7×7) while preserving computational efficiency.

This enhancement improves the network's ability to capture contextual information and retain spatial details, particularly in cluttered, low-contrast aquatic environments, enhancing the detection of small or partially submerged debris. Atrous convolution is expressed as:

$$y(i) = \sum_{k=1}^{K} x(i + r \cdot k) \cdot w(k) \tag{1}$$

Where $y(i)$ is the output at position i, x is the input feature map, w is the convolution kernel weights, r is the dilation rate, k is the kernel index and K is the kernel size. For $r = 6$ and $K = 3$ (a 3×3 Kernel), the receptive, R field expands as:

$$R = 1 + (K - 1) \cdot r$$
$$R = 1 + (3 - 1) \cdot 6 = 13 \tag{2}$$

This means that the convolution incorporates information from a 13×13. region, significantly larger than a standard 3×3 kernel. Each modified layer uses:

$$y_{p(i,j)} = \sum_{k=1}^{K} \sum_{l=1}^{K} x_{p(i + r \cdot k, j + r \cdot l)} \cdot w_{p(k,l)} \tag{3}$$

Where y_p is the output from the parallel atrous convolution, k and l are loop indices, $w_p(k, l)$ is the convolutional weight at level p and x_p is the input feature map at level p. We combine outputs of all scales as follows, where y_{agg} is the combined output from all feature levels.

$$y_{\text{agg}} = \text{Concat}\left(y_{P1/2}, y_{P3/8}, y_{P5/32}\right) \tag{4}$$

The multi-scale outputs from each atrous-enhanced stage are concatenated, creating an aggregated feature representation. This aggregation allows the network to retain high-resolution spatial details while incorporating broader

contextual information from deeper layers. This results in enhanced detection capability for floating debris, particularly for small, partially submerged objects or those visually similar to the background.

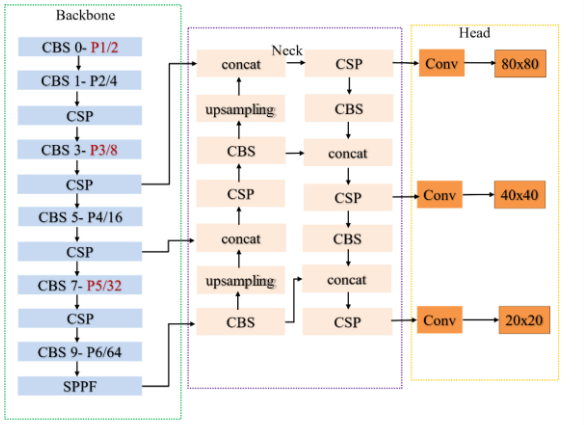


Figure 2. Yolov5s architecture with 640×640 input resolution

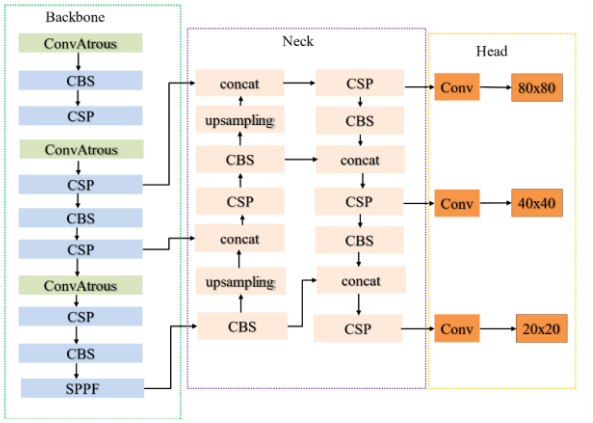


Figure 3. FloYO-Net architecture with 640×640 input resolution

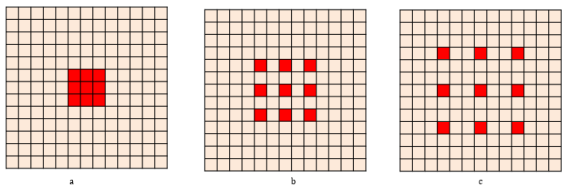


Figure 4. (a)standard convolution with dilation rate 1, (b) atrous convolution with dilation rates 2, and (c) atrous convolution with dilation rates 3

4.3 Training Procedure

Each model was trained for 500 epochs with a batch size of 16 and an input resolution of 640×640 pixels. The Adam optimizer was used with an initial learning rate of 0.0001 and a decay factor of 0.01. Training was conducted on an NVIDIA GeForce GTX 1060 GPU with 6GB of VRAM. Performance was evaluated using Precision, Recall, mAP@0.5, and

mAP@0.5:0.95, and confusion matrices provided further insights into model behavior. The FloYO-Net model was fine-tuned from YOLOv5s pretrained on the COCO dataset.

Precision (P), Recall (R), and mAP were computed using standard definitions, with True Positive (TP), False Positive (FP), and False Negative (FN) values. Average Precision (AP) was calculated as the area under the Precision-Recall curve, while mAP represents the mean of AP across different IoU thresholds. The training setup was built using Python 3.8, PyTorch 1.10, and the Ultralytics YOLOv5 framework, with early stopping and model checkpointing employed to mitigate overfitting and select the best-performing model based on validation mAP@0.5.

$$P = \frac{TP}{TP + FP} \tag{5}$$

$$R = \frac{TP}{TP + FN} \tag{6}$$

$$AP = \sum_{n-1}^{N} (R_n - R_{n-1}) P_n \tag{7}$$

Where P_n and R_n represent the precision and recall at the n^{th} threshold. The mean Average Precision (mAP) is calculated as the average of the Average Precision (AP) values across all classes c ∈C :

$$mAP = \frac{1}{C} \sum_{c=1}^{C} AP_c \tag{8}$$

5. EXPERIMENT AND ANALYSIS

The effectiveness of the FloYO-Net model was evaluated using the custom floating object dataset, with results compared to the baseline YOLOv5s model under identical training and testing conditions for consistency.

5.1 Quantitative Evaluation

Table 1 compares state-of-the-art models for small object detection in aquatic environments, showcasing metrics such as mAP@0.5, model size, parameter count, GFLOPs, inference time (FPS), and datasets used. As shown in Table 2, the FloYO-Net model outperforms baseline Yolov5 with an mAP@0.5 of 0.828, maintaining computational efficiency with a compact size (14.4 MB) and real-time inference (≈39 FPS). Compared to models like FRL-Yolo [27] and Baseline-Yolo, FloYO-Net achieves higher mAP at similar or

lower computational cost, demonstrating its robustness and efficiency for detecting small objects in aquatic environments.

Table 2. Performance Metrics: Baseline YOLOv5s vs. FloYO-Net

Model	mAP@0.5 (%)	Model size (MB)	Parameters (M)	GFLOPs	FPS (ms)	Dataset used
Baseline Yolo	0.787	14.4	7.2	15.8	36	D_six
FloYO-Net	0.828	14.4	7.2	15.8	39	D_six

Table 3 compares the performance across key object detection metrics. The FloYO-Net model outperformed others in most metrics, except precision, with notable improvements in mean Average Precision (mAP) and recall, which are critical for reducing missed detections of small or occluded objects.

Table 3. Overall performance comparison

Metric	YOLOv5s	FloYO-Net	Improvement
Precision	**0.879**	0.867	-0.012
Recall	0.714	**0.734**	+0.02
mAP@0.5	0.787	**0.828**	+0.041
mAP@0.5:0.95	0.498	**0.509**	+0.011

These improvements demonstrate the efficacy of atrous convolution in expanding the receptive field without increasing model size, enhancing object detection accuracy. Figure 5 illustrates the precision, recall, and precision-recall curves for the FloYO-Net model. The model consistently outperforms others in precision, particularly for small or occluded objects. It also maintains higher recall across a broader range of confidence thresholds, underscoring its robustness. These gains, evident in the precision-recall curves, confirm that FloYO-Net offers more reliable and accurate detections in complex aquatic environments.

5.2 Class-wise performance

Table 4 compares average precision per class, highlighting significant improvements in classes with previously lower detection rates, such as plastic bottles (PB) and plastic drink containers (PD), due to object size and visual complexity.

While most classes showed improvement, the "take-out container" class experienced a slight decrease, likely due to class similarity or overlapping annotations. Overall, the FloYO-Net model consistently delivered better or comparable performance.

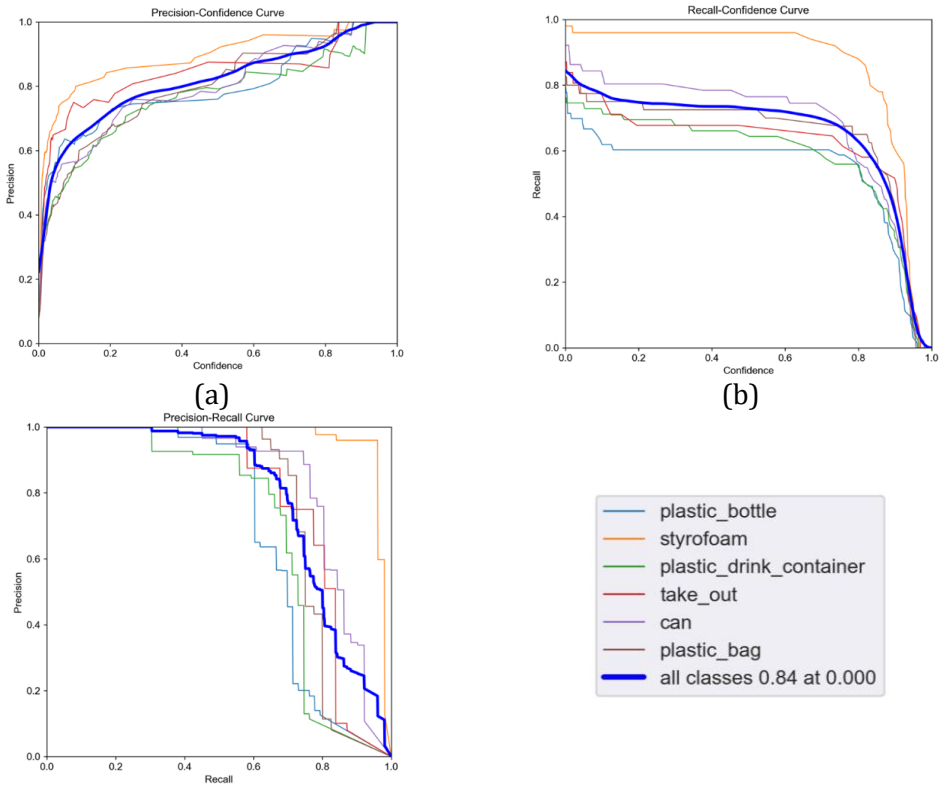


Figure 5. (a) Precision, (b) Recall, (c)Precision-Recall, and (d) legend for FloYO-Net model

Tabel 4. Class-wise mean average precision (mAP@0.5)

Class	YOLOv5s	FloYO-Net	Gain
Plastic Bottle (PB)	0.688	**0.754**	+0.066
Styrofoam (SF)	0.963	**0.984**	+0.021
Plastic Drink Container (PD)	0.693	**0.764**	+0.071
Take-Out (TO)	**0.782**	0.778	-0.004
Can (CN)	0.834	**0.843**	+0.009
Plastic Bag	0.762	**0.848**	+0.086

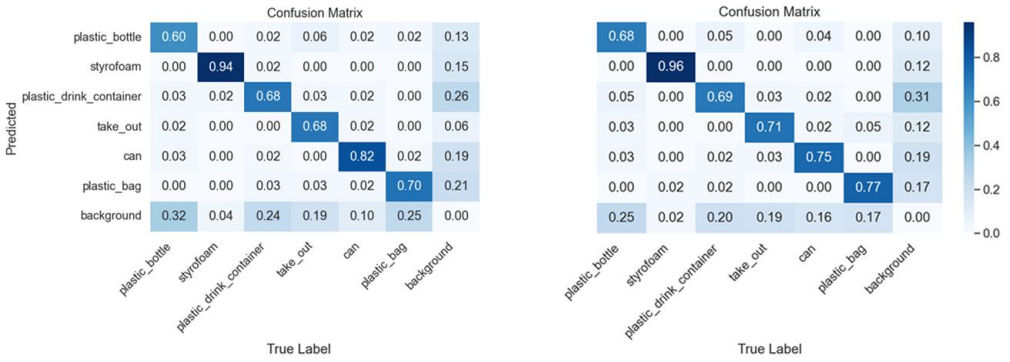
5.3 Confusion matrix analysis

The confusion matrices (Figure 6) reveal that FloYO-Net reduced the false positive rate and misclassifications into the background class. In contrast, the standard YOLOv5s frequently misclassified or missed plastic bottles and drink containers, likely due to their similarity with water surfaces and partial submersion. likely due to their similarity with water surfaces and partial submersion. These errors were significantly reduced in the FloYO-Net.

5.4 Detection Visualization

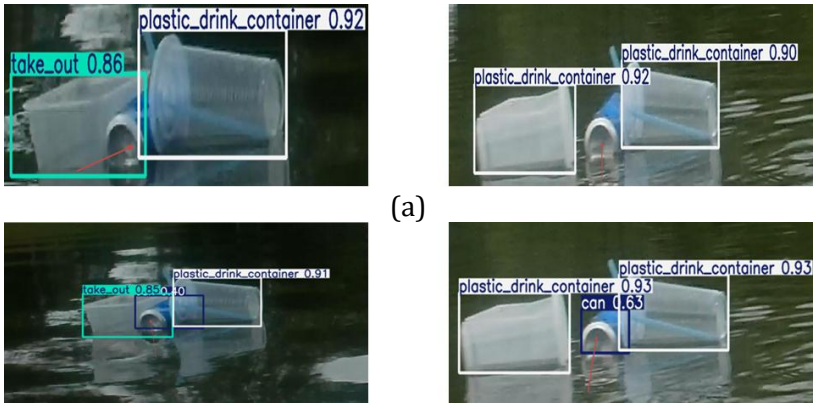
Figure 7 demonstrates qualitative detection improvements. In several test images, the baseline YOLOv5s missed small objects, as shown in Figure 7a.

These objects were accurately detected by FloYO-Net, shown in Figure 7b, highlighting its superior spatial awareness and robustness in real-world scenarios. Despite its strong overall performance, FloYO-Net shows some limitations in specific edge cases. Figure 8 illustrates failure cases, where false positives occur due to background elements resembling target objects, such as plastic drink containers and bottles.



(a) (b)

Figure 6. Confusion matrices for (a) Standard Yolov5s, and (b) FloYO-Net model



(a)

(b)

Figure 7. (a) Overlapping and occluded cans not detected by standard Yolov5s and (b) the detection of overlapping and occluded cans by FloYO-Net model

Figure 8. False positive instances by FloYO-Net

5.5 Performance Interpretation

The observed improvements validate the advantages of atrous convolution, which expands the receptive field without downsampling or

adding parameters, enabling the model to capture multi-scale contextual information essential for detecting small floating debris. Crucially, this enhancement preserves YOLOv5s's real-time inference speed, confirming its feasibility for deployment on embedded systems or autonomous water-cleaning robots. While FloYO-Net performs well on the D_six dataset, its generalizability to broader aquatic environments and diverse debris types remains a challenge. Trained on six predefined categories of floating waste in riverine conditions, the model may face performance degradation when exposed to novel debris types (e.g., fishing nets, organic matter) or environments such as oceans and estuaries, which differ in visual and hydrodynamic properties. Environmental factors like water turbidity, seasonal vegetation changes, and varied urban litter could also impact detection accuracy. To address these issues, future work will focus on cross-domain generalization through domain adaptation, synthetic data augmentation, and semi-supervised learning from unlabeled regional video data. Additionally, we plan to expand the D_six dataset with more debris classes and geographically diverse samples to enhance model robustness. A systematic ablation study will also assess the impact of each atrous convolution stage on detection performance.

6. CONCLUSION

Detecting floating waste in aquatic environments is challenging due to the small size, overlap, and visual similarity of debris with water surfaces. While YOLOv5s performs well in real-time scenarios, it struggles with such cases due to a limited receptive field and loss of fine spatial details during downsampling. This study introduces FloYO-Net, an enhanced YOLOv5s model incorporating atrous (dilated) convolution layers at key feature extraction stages. The model, evaluated on a custom dataset of six floating debris classes from real-world river environments, expands the receptive field with a dilation rate of 6 at scales P1/2, P3/8, and P5/32, improving detection without adding computational cost. Experimental results show FloYO-Net outperforming YOLOv5s across all major metrics, increasing mAP@0.5 from 0.787 to 0.828 and mAP@0.5:0.95 from 0.498 to 0.509. Notable improvements were observed for small, visually ambiguous objects like plastic bottles and drink containers. Confusion matrices and visualizations further confirm enhanced sensitivity to occluded and small targets with minimal false positives. FloYO-Net balances accuracy and speed, making it ideal for real-time environmental monitoring applications, such as autonomous cleanup systems and aquatic drones.

Acknowledgments

The authors thank Universiti Teknologi Malaysia (UTM) for supporting this research work.

REFERENCES

[1] A. Luqman *et al.*, **Microplastic contamination in human stools, foods, and drinking water associated with Indonesian coastal population,** *Environments,* vol. 8, no. 12, p. 138, 2021.

[2] U. R. N. Santoso and F. Gamar, **Deteksi Sampah Botol Plastik di Perairan Menggunakan YOLO v4-Tiny**, *Jurnal Teknologi Dan Sistem Informasi Bisnis,* vol. 7, no. 1, pp. 91-98, 2025.

[3] A. Akib *et al.*, **Unmanned floating waste collecting robot**, in *TENCON 2019-2019 IEEE Region 10 Conference (TENCON), 2019: IEEE*, pp. 2645-2650, 2019.

[4] Q. Li, Z. Wang, G. Li, C. Zhou, P. Chen, and C. Yang, **An accurate and adaptable deep learning-based solution to floating litter cleaning up and its effectiveness on environmental recovery**, *Journal of Cleaner Production,* vol. 388, p. 135816, 2023.

[5] D. Hindarto, **Exploring YOLOv8 Pretrain for Real-Time Detection of Indonesian Native Fish Species**, *Sinkron: jurnal dan penelitian teknik informatika,* vol. 7, no. 4, pp. 2776-2785, 2023.

[6] J. Zhang, J. Jin, Y. Ma, and P. Ren, **Lightweight object detection algorithm based on YOLOv5 for unmanned surface vehicles**, *Frontiers in marine science,* vol. 9, p. 1058401, 2023.

[7] N. D. Ismail, R. Ramli, and M. N. Ab Rahman, **Evaluating YOLOv5s and YOLOv8s for Kitchen Fire Detection: A Comparative Analysis**, *EMITTER International Journal of Engineering Technology,* vol. 12, no. 2, pp. 167-181, 2024.

[8] M. Vijayalakshmi and A. Sasithradevi, **AquaYOLO: Advanced YOLO-based fish detection for optimized aquaculture pond monitoring**, *Scientific Reports,* vol. 15, no. 1, p. 6151, 2025.

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, **Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs**, *IEEE transactions on pattern analysis and machine intelligence,* vol. 40, no. 4, pp. 834-848, 2017.

[10] M. D. Putro, Y. Mose, A. C. Andaria, J. Litouw, V. C. Poekoel, and X. Najoan, **Streamlining Deep Learning Network for Real-time Sea Turtle Detection**, *Jurnal Rekayasa Elektrika,* vol. 20, no. 3, 2024.

[11] D. D. Aboyomi and C. Daniel, **A Comparative Analysis of Modern Object Detection Algorithms: YOLO vs. SSD vs. Faster R-CNN**, *ITEJ (Information Technology Engineering Journals),* vol. 8, no. 2, pp. 96-106, 2023.

[12] W. Dong, **Faster R-CNN and YOLOv3: a general analysis between popular object detection networks**, in *Journal of Physics: Conference Series*, 2023, vol. 2580, no. 1, p. 012016, 2023

[13] T. Palwankar and K. Kothari, **Real time object detection using ssd and mobilenet**, *Int. J. Res. Appl. Sci. Eng. Technol,* vol. 10, pp. 831-834, 2022.

[14] W. Fang, L. Wang, and P. Ren, **Tinier-YOLO: A real-time object detection method for constrained environments**, *IEEE Access,* vol. 8, pp. 1935-1944, 2019.
[15] X. Yan, D. Tian, D. Zhou, C. Wang, and W. Zhang, **IV-YOLO: A Lightweight Dual-Branch Object Detection Network**, *Preprints,* p. 2024082054, 2024.
[16] C. Li, W. Pan, R. Su, and P. Yuen, **Multiple structural defect detection for reinforced concrete buildings using YOLOv5s**, *Transactions Hong Kong Institution of Engineers,* vol. 29, no. 2, 2022.
[17] T. Zhou and J. Yang, **An improved YOLOv5 algorithm for construction solid waste detection**, in *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)*, pp. 473-477, 2023.
[18] H. Li *et al.*, **Detection of floating objects on water surface using YOLOv5s in an edge computing environment**, *Water,* vol. 16, no. 1, p. 86, 2023.
[19] J. R. Yasiri and R. Prathivi, **Detection of Plastic Bottle Waste Using YOLO Version 5 Algorithm**, *Sinkron: jurnal dan penelitian teknik informatika,* vol. 9, no. 1, 2025.
[20] R. T. Hutabarat and R. Kurniawan, **Deteksi Sampah di Permukaan Sungai menggunakan Convolutional Neural Network dengan Algoritma YOLOv8**, in *Seminar Nasional Official Statistics*, 2024, vol. 2024, no. 1, pp. 537-548, 2024.
[21] B. Tjandra, M. S. Negara, and N. S. Handoko, **Deteksi Sampah di Permukaan dan Dalam Perairan pada Objek Video dengan Metode Robust and Efficient Post-Processing dan Tubelet-Level Bounding Box Linking**, *arXiv preprint arXiv:2307.10039,* 2023.
[22] A. Atalarais, K. Saputra, H. Syahputra, S. I. Al Idrus, and I. Taufik, **Automatic Waste Type Detection Using YOLO for Waste Management Efficiency**, *Journal of Artificial Intelligence and Engineering Applications (JAIEA),* vol. 4, no. 2, pp. 883-892, 2025.
[23] H. A. Pratama, B. S. B. Dewantara, and D. Pramadihanto, **Omnidirectional Stereo Vision Study from Vertical and Horizontal Stereo Configuration**, *EMITTER International Journal of Engineering Technology,* pp. 294-310, 2022.
[24] H. Chen and H. Lin, **An effective hybrid atrous convolutional network for pixel-level crack detection**, *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-12, 2021.
[25] K. R. Ahmed, **Dsteelnet: a real-time parallel dilated cnn with atrous spatial pyramid pooling for detecting and classifying defects in surface steel strips**, *Sensors,* vol. 23, no. 1, p. 544, 2023.
[26] Y. Jiang, M. Ye, D. Huang, and X. Lu, **AIU-Net: An Efficient Deep Convolutional Neural Network for Brain Tumor Segmentation**, *Mathematical Problems in Engineering,* vol. 2021, no. 1, p. 7915706, 2021.

[27] Y. Huang, Q. Wang, W. Jia, Y. Lu, Y. Li, and X. He, **See more than once: Kernel-sharing atrous convolution for semantic segmentation**, *Neurocomputing,* vol. 443, pp. 26-34, 2021.
[28] T. Panboonyuen, K. Jitkajornwanich, S. Lawawirojwong, P. Srestasathiern, and P. Vateekul, **Semantic labeling in remote sensing corpora using feature fusion-based enhanced global convolutional network with high-resolution representations and depthwise atrous convolution**, *Remote Sensing,* vol. 12, no. 8, p. 1233, 2020.
[29] A. Halder and D. Dey, **Atrous convolution aided integrated framework for lung nodule segmentation and classification**, *Biomedical Signal Processing and Control,* vol. 82, p. 104527, 2023.
[30] V. V. Y. Le Thanh Viet, V.-T. Pham, and T.-T. Tran, **A Fully Convolutional Network with Waterfall Atrous Spatial Pooling and Localized Active Contour Loss for Fish Segmentation**, 2023.
[31] X. Chen, Y. Li, and Y. Nakatoh, **Pyramid attention object detection network with multi-scale feature fusion**, *Computers and electrical engineering,* vol. 104, p. 108436, 2022.
[32] Y. Zhang *et al.*, **Small object detection based on hierarchical attention mechanism and multi-scale separable detection**, *IET Image Processing,* vol. 17, no. 14, pp. 3986-3999, 2023.
[33] X. Xu, J. Zhao, Y. Li, H. Gao, and X. Wang, **BANet: A balanced atrous net improved from SSD for autonomous driving in smart transportation**, *IEEE Sensors Journal,* vol. 21, no. 22, pp. 25018-25026, 2020.
[34] Z. Ren, Q. Kong, J. Han, M. D. Plumbley, and B. W. Schuller, **CAA-Net: Conditional atrous CNNs with attention for explainable device-robust acoustic scene classification**, *IEEE Transactions on Multimedia,* vol. 23, pp. 4131-4142, 2020.
[35] Y. Cheng *et al.*, **Flow: A dataset and benchmark for floating waste detection in inland waters**, in *Proceedings of the IEEE/CVF international conference on computer vision,* 2021, pp. 10953-10962, 2021.
[36] M. Liu, Y. Wu, R. Li, and C. Lin, **LFN-YOLO: precision underwater small object detection via a lightweight reparameterized approach**, *Frontiers in Marine Science,* 2025.
[37] R. Xian, L. Tang, and S. Liu, **Development of a Lightweight Floating Object Detection Algorithm**, *Water,* vol. 16, no. 11, p. 1633, 2024.
[38] J. Chen and M. J. Er, **Dynamic YOLO for small underwater object detection**, *Artificial Intelligence Review,* vol. 57, no. 7, p. 165, 2024.
[39] Z. Xiao, Z. Li, H. Li, M. Li, X. Liu, and Y. Kong, **Multi-Scale Feature Fusion Enhancement for Underwater Object Detection**, *Sensors,* vol. 24, no. 22, p. 7201, 2024.