VOL. 25, NO. 1, 2026, 5-13
elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

# AAYOLO: Robust Detection of Small and Occluded Floating Objects in Water Bodies Under Varying Illumination Using the FloW Dataset

**Badiu Badams**[1], **Usman Ullah Sheikh**[1*], **Norhaliza Abdul Wahab**[1], and **Syed A.R Abu Bakar**[1]

[1]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310, Johor Bahru, Malaysia.

[*]Corresponding author: usman@fke.utm.my

**Abstract:** Floating waste detection is a critical task for environmental monitoring and preservation, particularly in water bodies polluted by plastic waste and other debris. Traditional detection models like YOLOv5s struggle with detecting small, occluded, and poorly illuminated floating objects due to suboptimal feature extraction and limited datasets. This study proposes an enhanced model, AAYOLO (Atrous-Attention YOLO), which integrates atrous convolutions and enhanced channel attention (ECA) mechanisms to improve feature extraction across various scales. Atrous convolutions are applied at strategic levels of the network (P1/2, P3/8, and P5/32) to expand the receptive field, while ECA effectively adjusts feature weights to enhance object detection under varying lighting conditions. The proposed model is evaluated on the FloW-Img dataset, a benchmark specifically designed for floating waste detection, consisting of high-resolution images of plastic bottles found in aquatic environments. Experimental results demonstrate that AAYOLO outperforms the baseline YOLOv5s model, achieving a higher mean Average Precision (mAP@0.5) of 0.892 compared to 0.872, along with improvements in precision (0.878 vs. 0.858) and recall (0.836 vs. 0.820). This approach presents a promising solution for real-time, accurate detection of floating debris, contributing to improved waterway monitoring and environmental protection.

**Keywords:** Atrous Convolution, Attention Mechanism, Floating Object Detection, Marine Pollution Monitoring, YOLOv5-based Model



*Article History: received 24 March 2025; accepted 24 May 2025; published 30 April 2026*
*Digital Object Identifier 10.11113/elektrika.v25n1.708*

## 1. INTRODUCTION

The detection of floating objects in water bodies plays a vital role in various fields, such as marine debris monitoring, autonomous navigation, and environmental conservation [1-3]. Accurate detection of floating objects is essential for reducing water pollution, preventing navigational hazards, and enhancing waste collection efficiency [4, 5]. However, this task remains challenging due to the complexity of the aquatic environment.

One of the most prevalent types of floating waste is plastic, especially plastic bottles, which are widely used due to their necessity for drinking water [6-8]. Improperly discarded plastic waste is often carried by stormwater runoff into nearby lakes, ponds, and rivers, eventually making its way into the ocean or sea [9-11]. Rivers, which connect land to larger water bodies, serve as major conduits for transporting plastic pollution across ecosystems. Alarmingly, recent studies estimate that over 8 million metric tons of plastic waste are dumped into the ocean annually, posing severe threats to marine life, water quality, and biodiversity [12-14]. This alarming statistic highlights the urgency of developing efficient and automated solutions for detecting and managing floating waste in water bodies.

Despite recent advancements in object detection technology, detecting floating debris remains a significant challenge. Small, partially submerged, or clustered objects, such as plastic bottles and other floating waste items, are difficult to detect due to their limited visibility and varying shapes [15-17]. Additionally, lighting conditions such as sunlight reflections, shadows, and water turbidity cause inconsistent illumination, further complicating object recognition and detection [18-20].

Modern object detection models, including Faster R-CNN, SSD, and YOLO, have achieved remarkable progress in real-time detection. However, these models struggle with the specific challenges of detecting small and occluded objects in dynamic aquatic environments. The YOLOv5s model, known for its efficiency and speed, serves as a strong foundation but lacks robustness in handling these challenging conditions [21-25].

### 1.2 Research Gap and Motivation

Existing solutions inadequately address the challenges of detecting small, occluded, and poorly illuminated objects due to their suboptimal feature extraction techniques. Furthermore, most publicly available datasets are not tailored to aquatic environments, making model generalization difficult. This gap highlights the need for an improved detection framework capable of accurately

detecting floating waste under varying environmental conditions.

### 1.3 Proposed Solution

To address these challenges, we propose AAYOLO, an improved YOLOv5s-based model incorporating:

1. Atrous Convolution for multi-scale feature extraction, enhancing the detection of small and occluded objects by expanding the receptive field without increasing computational complexity [26-28].
2. Attention Mechanisms to focus on important regions, improving detection under varying illumination and complex backgrounds [29-31].

The proposed model is designed to enhance detection accuracy in aquatic environments, making it well-suited for real-world applications such as automated waste monitoring, marine pollution control, and environmental conservation.

## 2. DATASET AND METHODOLOGY

### 2.1 Dataset Selection

For this study, we use the FloW-Img dataset, a benchmark dataset specifically designed for floating waste detection. This dataset was introduced by Cheng et al. (2021). The dataset consists of high-resolution images of plastic bottles and other debris found in aquatic environments.

### 2.2 Dataset Description

The FloW-Img dataset, introduced by Cheng et al. (2021), is a benchmark dataset specifically designed for floating waste detection in inland water bodies. The dataset contains 2000 images, annotated in Pascal VOC format and originally split into train (1200) and test (800) sets in a 6:4 ratio.

For this study, since we utilized YOLO, we converted the annotations into YOLO format and re-split the dataset into a 6:2:2 ratio, with 1200 images for training, 400 for validation, and 400 for testing. This restructuring ensures better model evaluation and generalization.

The dataset is essential for advancing research in detecting and classifying floating waste in real-world environments. It serves as a valuable benchmark for evaluating object detection models under real-world conditions, contributing to more effective and scalable floating waste detection solutions.

### 2.3 Baseline Model: YOLOv5s

The YOLOv5s architecture serves as the baseline model in this research. It comprises three main components: backbone, neck, and head. The backbone extracts key features from images, while the PANet neck fuses and refines these features for enhanced detection. The YOLO detection head then performs object localization and classification, predicting bounding box coordinates and class probabilities. This multi-stage framework enables efficient visual data processing and serves as a foundation for improving detection performance in challenging conditions.

Downsampling in neural networks reduces fine-grained details, impacting object detection. To preserve spatial resolution and enhance multi-scale feature extraction, we integrated atrous convolutions into specific backbone layers of YOLOv5s. Atrous convolution expands the kernel's field of view using a dilation rate, increasing the receptive field without adding parameters. This modification improves feature extraction across different scales while maintaining spatial resolution.

### 2.4 Proposed Model: AAYOLO

This section presents the proposed AAYOLO model, which improves floating object detection by integrating atrous convolutions and enhanced channel attention (ECA) into the baseline YOLOv5s model. The goal is to enhance detection accuracy for small and occluded objects under challenging environmental conditions such as varying illumination and cluttered backgrounds. The architecture is shown in Figure 1. It uses atrous convolution (ConvAtrous) layers with a dilation rate of 6 for multi-scale feature capture. The attention module addresses lighting variations using pooling techniques, while parallel atrous convolutions with dilation rates [1, 2, 3] prevent channel information loss. The attention module with the atrous [1, 2, 3] make up ECA.

The selection of dilation rates of 6 and [1, 2, 3] was the result of extensive experimentation, where these values were found to be the most optimal for enhancing detection performance. This setup enhances feature extraction and maintains channel integrity.

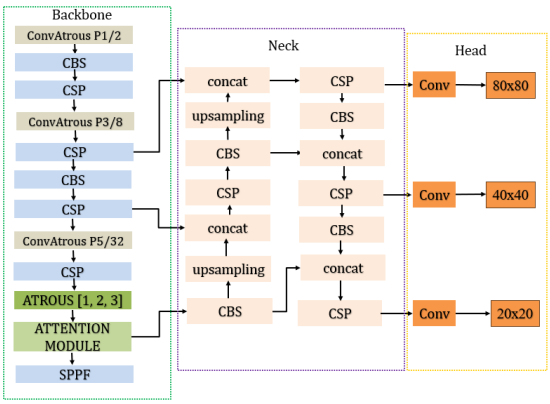


Figure 1. AAOLO model architecture

Mathematically, atrous convolutions can be represented as:

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

where:

$\boldsymbol{y(i)}$ → the output value at position $i$

$x(i+r \cdot k)$ → the input feature map value at a specific position

$x(i)$ → is the original input image or feature map

$r$ is the dilation rate, which controls how much we skip over input pixels

$w(k)$ → the weight of the filter (kernel)

$k$ → Kernel index

$K$ → Kernel size

Equation 1 defines how atrous convolution computes an output pixel by skipping pixels using dilation rate $r$. The effect of this is illustrated in Figure 2.

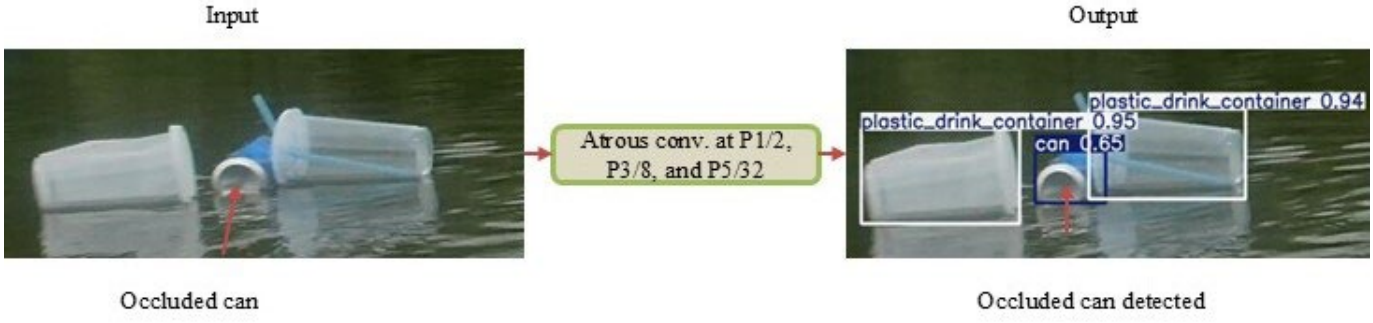


Figure 2. Effect of atrous convolution on detection

### 2.5 Atrous Convolutions for Multi-Scale Feature Extraction

Atrous Convolutions, also known as dilated convolutions, are employed to enhance the model's ability to capture multi-scale features without increasing computational complexity. In this research they are applied at three strategic levels of the model's backbone:

- P1/2: Early stage for fine details (edges, textures).
- P3/8: Middle stage for mid-level features (object parts).
- P5/32: Late stage for large-scale context (semantic features).

In YOLO architectures, P1/2, P3/8, and P5/32 refer to feature map stages in the backbone, where the denominator indicates the downsampling factor and the numerator represents the stage index in the model. For a 640×640 input, P1/2 downsamples the feature map to 320×320, preserving fine details like edges and textures, while P3/8 and P5/32 downsamples the input image to 80 and 20 respectively. By expanding the receptive field at these levels, the model effectively learns features at different resolutions, enhancing robustness in detecting objects of varying sizes.

The Receptive field is represented mathematically as:

$$\text{Receptive field} = 1 + (K-1)\cdot r \quad (2)$$

Where $K$ is the kernel size = 3 and $r$ is the dilation rates =6

$$\text{Receptive field} = 1 + (3-1)\cdot 6 = 13$$

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

*2.5.1 At* $P1/2$

$$y_{P1/2}(i,j) = \sum_{k=1}^{K}\sum_{l=1}^{K} x_{P1/2}(i + r\cdot k, j + r\cdot l)\cdot w_{P1/2}(k,l) \quad (3)$$

where:
r =6
$x_{P1/2}$ = Input feature map at $P1/2$
$w_{P1/2}$ = Kernel weights at $P1/2$

*2.5.2 At* $P3/8$

The feature map resulting from an atrous convolution:

$$y_{P3/8}(i,j) = \sum_{k=1}^{K}\sum_{l=1}^{K} x_{P3/8}(i + r\cdot k, j + r\cdot l)\cdot w_{P3/8}(k,l) \quad (4)$$

*2.5.3 At* $P5/32$

The feature map resulting from an atrous convolution:

$$y_{P5/32}(i,j) = \sum_{k=1}^{K}\sum_{l=1}^{K} x_{P5/32}(i + r\cdot k, j + r\cdot l)\cdot w_{P5/32}(k,l) \quad (5)$$

Combining $P1/2$, $P3/8$, and $P5/32$ gives $y_{agg}$

$$y_{agg} = Concat(y_{P1/2}, y_{P3/8}, y_{P5/32}) \quad (6)$$

where $y_{agg}$ denotes the aggregation of P1/2, P3/8, and P5/32.

### 2.6 Enhanced Channel Attention (ECA)

The Channel Attention Mechanism is integrated to improve the model's performance under varying lighting conditions. It functions by:

1. Extracting global and prominent features via Global Average Pooling (GAP) and Global Max Pooling (GMP), respectively.
2. Passing the pooled features through a Multi-Layer Perceptron (MLP) to learn inter-channel relationships, generating attention weights.
3. Applying the attention weights to refine feature extraction, enhancing detection accuracy under challenging conditions.

The mechanism processes the input through average and max pooling to extract global and dominant features, respectively as shown in Figure 3. The pooled outputs (A and M) are passed through an MLP to learn inter-channel dependencies, producing $A^I$ and $M^I$. These are concatenated and scaled using a sigmoid function, which

adjusts feature importance between 0 and 1. The scaled weights are then applied to the input features, refining feature extraction and improving detection accuracy under varying lighting conditions. However, this process may result in information loss in certain channels. To address this issue, parallel atrous convolutions with dilation rates of [1, 2, 3] were integrated to minimize feature loss, leading to the development of the Enhanced Channel Attention (ECA) mechanism. Figure 4 illustrates the connection of the parallel atrous convolution of dilation rates [1, 2, 3] and the channel attention.

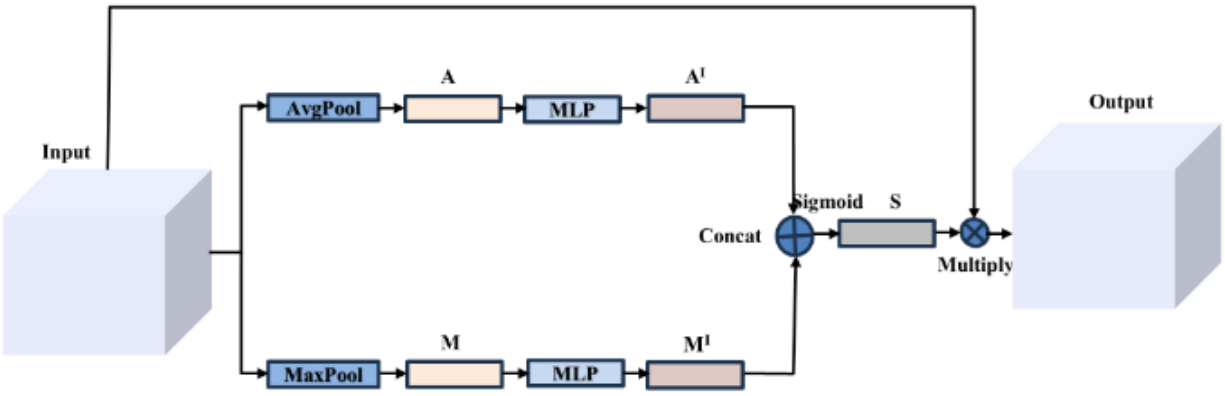


Figure 3. Channel attention module

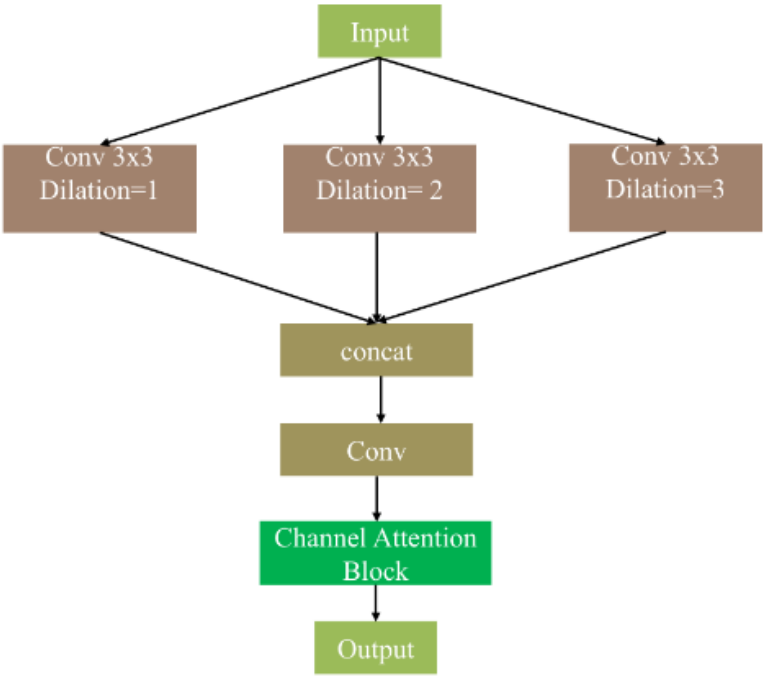


Figure 4. Parallel atrous convolution and channel attention module connection

For a single atrous convolution, we have:

$$y_{r_k}(i,j) = \sum_{k=1}^{K}\sum_{l=1}^{K} x(i + r_k \cdot k, j + r_k \cdot l) \cdot w_k(k,l) \quad (7)$$

The output of parallel atrous convolutions is:

$$y_{\text{parallel}}(i,j) = \text{Concat}\left(y_{r_1}(i,j), y_{r_2}(i,j), y_{r_3}(i,j)\right) \quad (8)$$

resulting in an expanded feature space. The output of the parallel convolution is passed through the channel attention mechanism and this gives:

$$y_{CA} = \sigma\Big[W_2 \cdot ReLU\left(W_1 \cdot GlobalAvgPool(y_{parallel})\right) \| W_2 \cdot ReLU\left(W_1 \cdot GlobalMaxPool(y_{parallel})\right)\Big] \cdot y_{parallel} \quad (9)$$

Equation (9) shows the relationship between the attention mechanism and the parallel connection of atrous convolution.

$y_{CA}$ → represents the output of the Channel Attention mechanism applied to the feature map $y_{parallel}$ (i, j),
$GlobalMaxPool(y_{parallel})$ → computes the maximum value for each channel in $y_{parallel}$, highlighting the most prominent spatial features.
$W_1$ → Reduces the feature dimensions (C → C/r).
$W_2$ → Restores the dimension back (C/r → C).
$\sigma$ → Ensures the output is bounded between 0 and 1, acting as a channel-wise attention mask.
$GlobalAvgPool(y_{parallel})$ → computes the average value of each channel in the feature map $y_{parallel}$, capturing the overall spatial feature contribution.
‖ → parallel (Concatenation): The feature maps from GlobalAvgPool and GlobalMaxPool are processed separately and then concatenated to enhance feature learning.
$\sigma \cdot y_{parallel}$ → The attention map is applied element-wise to the original feature map. This scales each channel's importance, emphasizing the most relevant features.

Channel attention with parallel atrous convolutions generates features $y_{CA}$, while atrous convolutions at Pyramid Levels produces $y_{agg.}$ Both components contribute to the final feature representation, balancing spatial and channel-wise information. Combining the two components gives the overall prescription of the AAYOLO model. Let $y_{combined}$ be the final output feature map. Then,

$$y_{combined} = \alpha \cdot y_{CA} + \beta \cdot y_{agg} \quad (10)$$

where α and β are learnable weights to balance the contributions of the two components.

### 2.7 Experimental Setup

The training process consisted of 500 epochs with a batch size of 16, input image resolution of 640×640 pixels was utilized. For optimization, we employed the Adam optimizer with an initial learning rate of 0.0001 and a decay factor of 0.01. All experiments were conducted on an NVIDIA GeForce GTX 1060 GPU with 6GB of VRAM.

## 3. RESULTS AND DISCUSSION

### 3.1 Performance Comparison

Table 1. Performance Comparison Between Yolov5 and AAYOLO

| Model | Precision | Recall | mAP@0.5 | mAP@0.5-0.95 |
|---|---|---|---|---|
| Yolov5s | 0.858 | 0.820 | 0.872 | 0.436 |
| AAYOLO | 0.878 | 0.836 | 0.892 | 0.449 |

AAYOLO outperforms YOLOv5s in all metrics, with improvements in precision (+2.0%), recall (+1.6%), and mAP@0.5 (+2.0%) as shown in Table 1. The Precision-Recall Curve, F1-Confidence Curve, and other metrics indicate that AAYOLO maintains better confidence thresholds, leading to more reliable detections.

#### *3.1.1 Confusion Matrix Analysis*

The confusion matrix shows how well the model classifies floating objects into respective categories. AAYOLO has a high true positive rate, especially for the bottle class as indicated in Figure 5. Background objects are perfectly classified with a score of 1.00, indicating no false positives in the background category. However, bottles have a small fraction (~12%) misclassified as background, suggesting minor misclassification in some challenging conditions. The confusion matrix indicates that AAYOLO is highly effective in distinguishing floating objects, with minimal misclassification errors.

#### *3.1.2 F1-Confidence Curve*

The F1-score balances precision and recall, providing a good measure of the overall detection performance as seen in Figure 6b. The curve shows a peak F1-score of 0.86, meaning the model maintains a good trade-off between identifying relevant objects (recall) and avoiding false positives (precision). AAYOLO exhibits strong generalization performance, with high F1 scores at most confidence levels, proving its robustness across various detection thresholds.

#### *3.1.3 Precision-Confidence Curve*

Precision measures the proportion of detected objects that are correctly classified. The curve is consistently high, reaching near 1.0 at high confidence thresholds.

AAYOLO maintains precision above 0.8 across a broad range of confidence values, confirming its low false positive rate. It is highly precise, meaning most of its detected objects are correct. It is particularly reliable when operating at higher confidence thresholds.

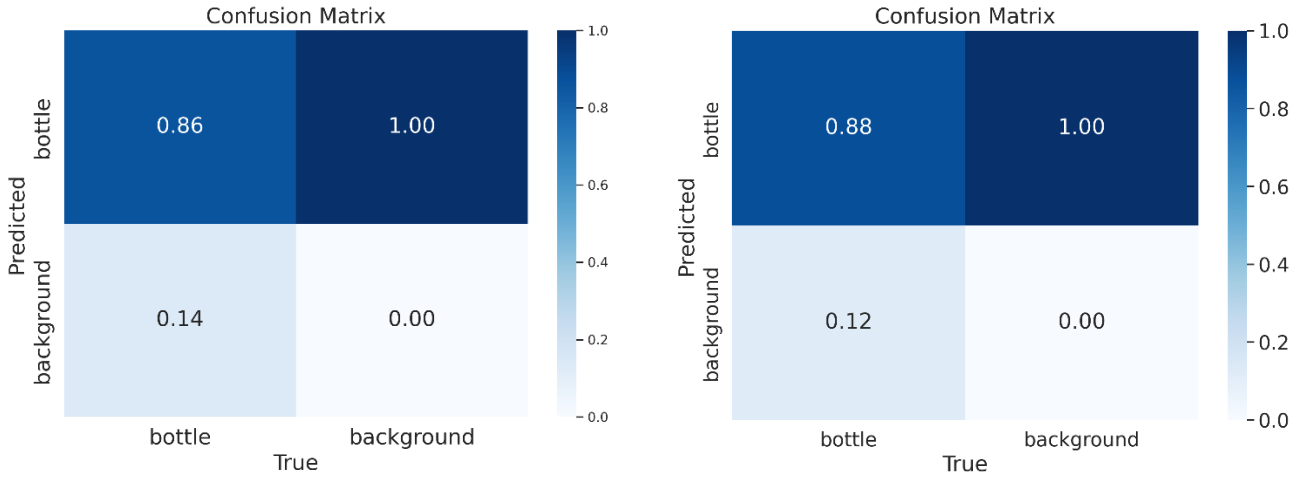


Figure 5. Confusion matrices for (a) YOLOv5s and (b) AAYOLO on the FloW-Img test set. AAYOLO achieves a slightly higher true positive rate and reduced misclassification for the bottle class.

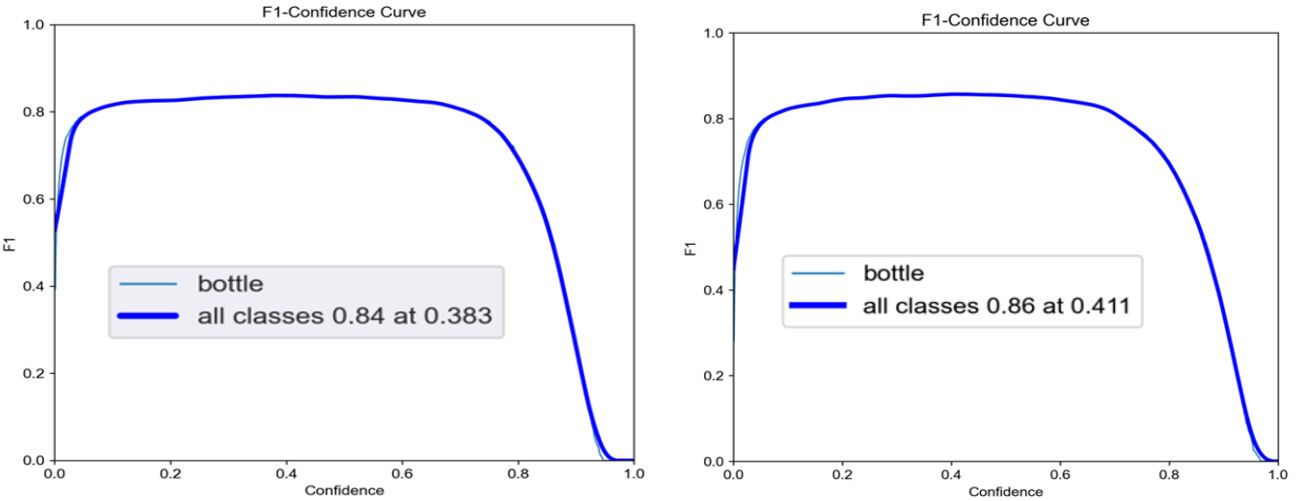


Figure 6. F1-Confidence curves for (a) YOLOv5s and (b) AAYOLO. AAYOLO maintains a higher F1-score across confidence thresholds, peaking at 0.86.

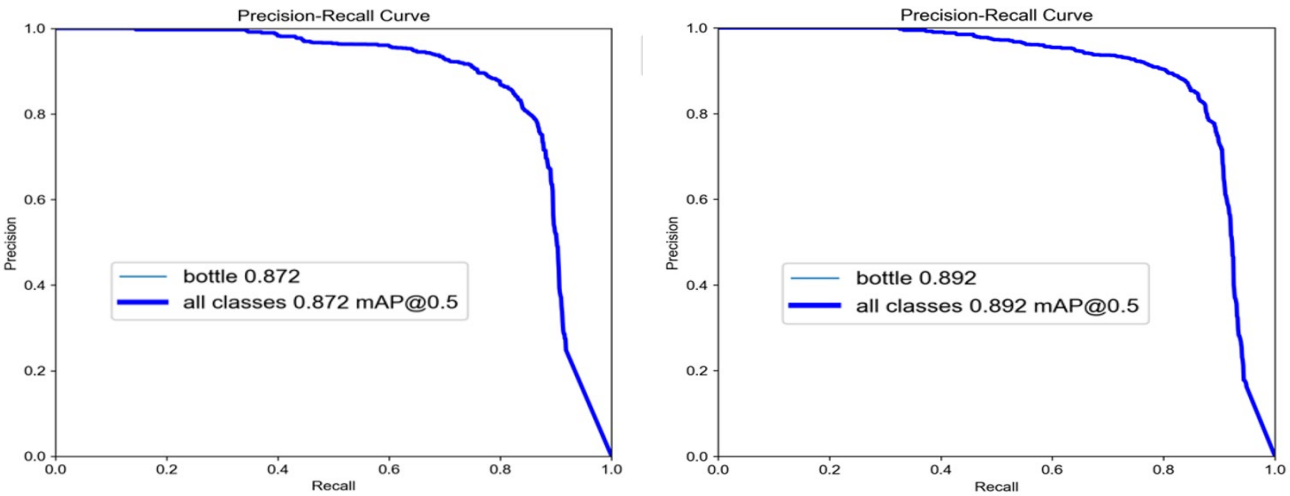


Figure 7. Precision-Recall (PR) curves for (a) YOLOv5s and (b) AAYOLO. AAYOLO achieves a higher mAP@0.5, confirming its superior detection performance.

*3.1.4 Recall-Confidence Curve*

Recall measures the proportion of actual objects that are detected. The recall curve starts high (~0.94) but gradually declines as confidence increases. This suggests that at low confidence values, AAYOLO detects most objects, but at higher confidence thresholds, some objects might be missed. The recall score suggests that AAYOLO is sensitive to detecting objects, ensuring that very few floating debris instances are missed. However, choosing an appropriate confidence threshold is crucial to maintain the balance.

*3.1.5 Precision-Recall (PR) Curve*

The PR curve evaluates the model's trade-off between precision and recall. The AAYOLO model as shown in Figure 7b achieves a high area under the curve (AUC), suggesting that it performs well even under varying detection thresholds. The mAP@0.5 (mean Average Precision at IoU 0.5) score of 0.892 reflects strong detection capability. It achieves a strong balance between detecting all relevant objects (recall) and avoiding false positives (precision), confirming its state-of-the-art performance.

*3.1.6 Loss Graphs & Convergence Analysis*

Localization loss (bounding box error) decreases steadily, indicating that the model learns to place bounding boxes accurately. Classification loss also decreases smoothly, but the final graph suggests a slight plateau, meaning further improvements might require fine-tuning learning rates or additional data augmentation. Objectness loss (confidence score loss) remains stable, showing that the model correctly identifies objects and background differentiation. Validation losses align well with training losses, meaning the model does not overfit, proving strong generalization. The loss graphs indicate that AAYOLO has effectively learned feature representations with a well-optimized architecture. No signs of overfitting are observed.

Table 2 shows the evaluation of AAYOLO on FloW-Img data set.

Table 2. Evaluation of AAYOLO on FloW-Img Data set

| Metric | AAYOLO |
|---|---|
| Precision | High, consistent across different confidence levels |
| Recall | Strong, peaks at ~0.94, ensuring minimal missed objects |
| F1 score | 0.86, suggesting balanced performance |
| mAP@0.5 | 0.892, outperforming many state-of-the-art models |
| Loss Convergence | Stable training, no overfitting issues |

*3.1.7. Computational Efficiency Comparison*

In addition to evaluating detection accuracy, we assessed the computational efficiency of AAYOLO in comparison with the baseline YOLOv5s model. AAYOLO recorded an inference time of 19.8 milliseconds per image, while YOLOv5s achieved a faster rate of 11.0 milliseconds per image, both measured on the same hardware platform (NVIDIA GTX 1060 GPU). Despite this increase in processing time as shown in Figure 8, AAYOLO demonstrated superior performance across all evaluation metrics—precision, recall, and mAP—highlighting its suitability for where detection accuracy is paramount.

Importantly, both models share identical architectural complexity, with 157 layers, approximately 7 million parameters, and a computational cost of 15.8 GFLOPs. This indicates that the observed difference in inference speed is attributed to the specific operations introduced in AAYOLO, such as atrous convolutions and attention mechanisms, rather than an increase in overall model size or complexity.

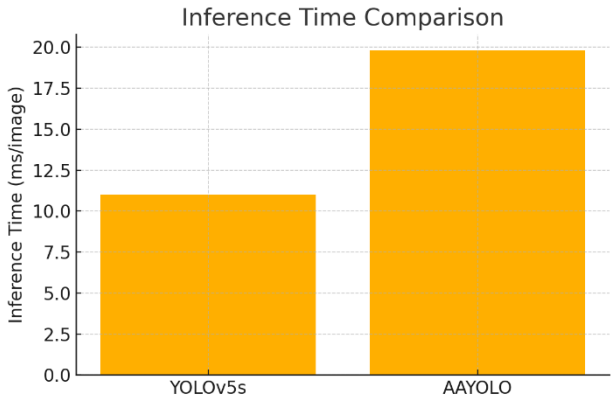


Figure 8. Inference time comparison between YOLOv5s and AAYOLO measured in milliseconds per image. AAYOLO achieves improved detection accuracy at the cost of a modest increase in inference time.

## 4. CONCLUSION

In this study, we proposed AAYOLO, an enhanced object detection model designed to address the challenges of detecting floating debris in water bodies. By integrating atrous convolutions for multi-scale feature extraction and the enhanced channel attention (ECA) mechanism for adaptive feature weighting, the model effectively detects small, occluded, and poorly illuminated objects. The evaluation on the FloW-Img dataset demonstrated that AAYOLO outperforms the baseline YOLOv5s model, achieving higher precision, recall, and mAP@0.5.

Future work will focus on optimizing the model's architecture to improve inference speed without compromising accuracy. Additionally, expanding the dataset to include various environmental conditions and floating object types will enhance the model's robustness and generalization capabilities.

## REFERENCES

[1] R. Carmo, J. Mifdal, and M. Rußwurm, "Detecting macro floating objects on coastal water bodies using sentinel-2 data," in *OCEANS 2021: San Diego–Porto*, 2021: IEEE, pp. 1-7.

[2] J. Zhang, A. Xiang, Y. Cheng, Q. Yang, and L. Wang, "Research on detection of floating objects in river and

lake based on ai intelligent image recognition," *arXiv preprint arXiv:2404.06883,* 2024.

[3] M. Marcon *et al.*, "Early Detection of Partially Emerged Large Scale Marine Debris Based on Laser Pulses," *2023 IEEE International Workshop on Metrology for the Sea; Learning to Measure Sea Health Parameters (MetroSea),* pp. 528-533, 2023.

[4] B. Badams, U. U. Sheikh, N. A. Wahab, and S. A. R. Abu-Bakar, "Enhancing Floating Object Detection in Water Bodies using Atrous Convolution," *2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA),* pp. 1-5, 2024.

[5] Y. Li, R. Wang, D. Gao, and Z. Liu, "A Floating-Waste-Detection Method for Unmanned Surface Vehicle Based on Feature Fusion and Enhancement," *Journal of Marine Science and Engineering,* 2023.

[6] A. M. Codes-Alcaraz, H. Puerto, and C. Rocamora, "Image Recognition for Floating Waste Monitoring in a Traditional Surface Irrigation System," *Water,* vol. 16, no. 18, p. 2680, 2024. [Online]. Available: https://www.mdpi.com/2073-4441/16/18/2680.

[7] S. N. Hasany, S. S. Zaidi, S. A. Sohail, and M. Farhan, "An autonomous robotic system for collecting garbage over small water bodies," *2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE),* pp. 81-86, 2021.

[8] T. A. Al-Tayyar, "Characteristics of plastic solid wastes in Mosul city and their reuse," in *IOP Conference Series: Earth and Environmental Science*, 2021, vol. 779, no. 1: IOP Publishing, p. 012071.

[9] M. A. Rahman, S. Mojumdar, S. A. Rahman, and K. Marimuthu, "Plastic pollutions in the ocean: their sources, causes, effects and control measures," *Journal of Biological Studies,* vol. 6, no. 1, pp. 37-52, 2023.

[10] E. Kosior and I. Crescenzi, "Solutions to the plastic waste problem on land and in the oceans," in *Plastic waste and recycling*: Elsevier, 2020, pp. 415-446.

[11] E. A. Weideman, V. Perold, and P. G. Ryan, "Limited long-distance transport of plastic pollution by the Orange-Vaal River system, South Africa," *Science of the Total Environment,* vol. 727, p. 138653, 2020.

[12] R. Singh and G. Khatri, "Impact and Solution of Marine Plastic Pollution: A Review," *International Journal of Advanced Research in Science, Communication and Technology,* 2022.

[13] L. Augusto, "Effect of Plastic Ocean Pollution on Marine Life and Human Health in Brazil," *International Journal of Environmental Sciences,* vol. 7, no. 1, pp. 65-75, 2024.

[14] Y. Peng, P. Wu, A. T. Schartup, and Y. Zhang, "Plastic waste release caused by COVID-19 and its fate in the global ocean," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 118, 2021.

[15] L. Palombi and V. Raimondi, "Experimental Tests for Fluorescence LIDAR Remote Sensing of Submerged Plastic Marine Litter," *Remote. Sens.,* vol. 14, p. 5914, 2022.

[16] K. Walden and M. Mehrubeoglu, "Quantifying Plastic Bottle Debris in Waterways Using Image Processing," *2020 International Conference on Computational Science and Computational Intelligence (CSCI),* pp. 1658-1663, 2020.

[17] N. Hanson, A. Demirkaya, D. Erdoğmuş, A. Stubbins, T. Padır, and T. Imbiriba, "A Vision for Cleaner Rivers: Harnessing Snapshot Hyperspectral Imaging to Detect Macro-Plastic Litter," in *2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2023: IEEE, pp. 1-5.

[18] H. Song *et al.*, "Investigating the rate of turbidity impact on underwater spectral reflectance detection," in *Frontiers in Marine Science*, 2023.

[19] M. Wang, T. Chen, and X. Wang, "Rapid Correction of Turbidity and CDOM Interference on Three-Dimensional Fluorescence Spectra of Live Algae Based on Deep Learning," *Photonics,* 2023.

[20] A. Mathias, S. Dhanalakshmi, R. Kumar, and R. Narayanamoorthi, "Deep Neural Network Driven Automated Underwater Object Detection," *Computers, Materials & Continua,* 2022.

[21] Y. Xiao, X. Wang, P. Zhang, F.-j. Meng, and F. Shao, "Object Detection Based on Faster R-CNN Algorithm with Skip Pooling and Fusion of Contextual Information," *Sensors (Basel, Switzerland),* vol. 20, 2020.

[22] B. Yan, J. Li, Z. Yang, X. Zhang, and X. Hao, "AIE-YOLO: Auxiliary Information Enhanced YOLO for Small Object Detection," *Sensors (Basel, Switzerland),* vol. 22, 2022.

[23] R. Joshi, K. Usmani, G. Krishnan, F. Blackmon, and B. Javidi, "Underwater object detection and temporal signal detection in turbid water using 3D-integral imaging and deep learning," *Optics express,* vol. 32 2, pp. 1789-1801, 2023.

[24] H. Yuan, W. Li, and Y. Cheng, "Real-time Detection Algorithm for Occluded Objects Based on Improved YOLOv5s," *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT),* pp. 661-666, 2024.

[25] J. Pei, X. Wu, X. Liu, L. Gao, S. Yu, and X. Zheng, "SGD-YOLOv5: A Small Object Detection Model for Complex Industrial Environments," *2024 International Joint Conference on Neural Networks (IJCNN),* pp. 1-10, 2024.

[26] M. Kim, N. Ilyas, and K. Kim, "AMSASeg: An Attention-Based Multi-Scale Atrous Convolutional Neural Network for Real-Time Object Segmentation From 3D Point Cloud," *IEEE Access,* vol. 9, pp. 70789-70796, 2021.

[27] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A Small-Sized Object Detection Oriented Multi-Scale Feature Fusion Approach With Application to Defect Detection," *IEEE Transactions on Instrumentation and Measurement,* vol. 71, pp. 1-14, 2022.

[28] C. Fang, Z. Hao, and J. Chen, "Multi scale switchable atrous convolution for target detection based on feature pyramid," *MATEC Web of Conferences,* 2022.

[29] X. Bai, C. Liang, and J. Zhou, "A Distance-Based Attention Mechanism for Object Detection," *2022 11th International Conference on Communications, Circuits and Systems (ICCCAS),* pp. 246-250, 2022.

[30] P. Shamsolmoali, J. Chanussot, H. Zhou, and Y. Lu, "Efficient Object Detection in Optical Remote Sensing Imagery via Attention-Based Feature Distillation," *IEEE Transactions on Geoscience and Remote Sensing,* vol. 61, pp. 1-12, 2023.

[31] S. Wu, Z. Liu, H. Lu, and Y. Huang, "Shadow Hunter: Low-Illumination Object-Detection Algorithm," *Applied Sciences,* 2023.