SHAH, F.M., FOUGH, F., SHAH, Z.M., FOUGH, N. and ZHAO, Y. 2024. Plastic waste over the ocean: an approach to surface recognition. In *Proceedings of the 31st IEEE (Institute of Electrical and Electronics Engineers) International conference on electronics, circuits, and systems (ICECS 2024), 18-20 November 2024, Nancy, France*. Piscataway: IEEE [online], article 10849157. Available from: https://doi.org/10.1109/ICECS61496.2024.10849157

Plastic waste over the ocean: an approach to surface recognition.

SHAH, F.M., FOUGH, F., SHAH, Z.M., FOUGH, N. and ZHAO, Y.

2024

Plastic Waste Over The Ocean: An Approach to Surface Recognition

Faisal Mehmood Shah [1], Faranak Fough [1], Zohaib Mehmood Shah [2], Nazila Fough [1], Yafan Zhao[1]
[1]School of Engineering, Robert Gordon University, Aberdeen, United Kingdom
[2]National University of Sciences and Technology, Islamabad, Pakistan

***Abstract*— To address the challenges of low detection accuracy and slow recognition of floating objects on water surfaces, an enhanced YOLOv5 algorithm has been developed. This improved algorithm incorporates a coordinate attention mechanism to enhance spatial recognition and employs the GIOU loss function with updated anchor matching for quicker decision-making. The new method achieved a 97.39% mAP, marking a 15.28% improvement over the original YOLOv5, demonstrating its effectiveness.**

Keywords—Pattern Recognition, YOLOv5, Sparse Features, Attention Mechanism, CFN

I. Introduction

Plastic pollution threatens health, marine life, and tourism [1]. Various methods have been proposed to monitor ocean plastic effectively [2-4]. This paper presents an enhanced YOLOv5 algorithm to improve detection accuracy and speed for floating objects on water. It features an inception topology for efficient feature extraction, a coordinate attention module for better spatial recognition, and a position attention mechanism to counteract weak signal effects. Additionally, the GIOU loss function with data-matched anchors refines size and location assessments, resulting in a more effective and precise monitoring solution for aquatic plastic pollution.

II. Proposed Methodolgy

Signal strength can affect image capture quality, leading to reduced detection accuracy. To address this, we implemented a position attention mechanism within the CFN (Coordinate Fusion Network) model. This mechanism focuses on key regions of interest, enhancing feature representation and overall model performance by improving how the network processes intermediate information. Figure 1 illustrates how CFN enhances target recognition accuracy through two steps: coordinate data embedding and coordinate attention generation. These steps allow CFN to encode location data, capture channel connections, and track dependencies over time. By refining feature tuning based on location, the model improves its understanding of feature relationships, leading to more accurate target recognition and detection. To embed coordinates, we use a pooling kernel of size (S,1) or (1,V) to encode each channel along the X and Y axes. The output for the C-th channel with height h is calculated using the following formula:

$$P_C^h(h) = \frac{1}{V}\sum_{0\le j\le V} Mc(j,h) \quad (1)$$

The outcome feature of the C-th channel having a height of h is represented by P_C^h(h) in equation 1, while the x-coordinate of extra input data is represented by j. The following equation can be used to get the C-th channel's output having a width of V:

$$P_C^V(V) = \frac{1}{h}\sum_{0\le k\le V} Mc(V,k) \quad (2)$$

Where P_c^v *(V)* denotes the C-th channel's width-wise output characteristic. Two direction-aware feature maps are generated by combining features from two spatial directions using transformations from Equations 1 and 2.

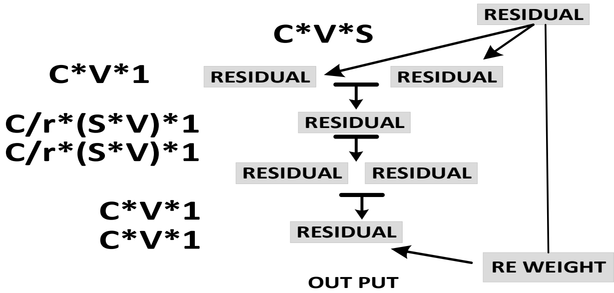

Fig.1 A schematic representing the CA attention mechanism unit

CFN Generation: The 1x1 convolutional transformation function G1 is used to concatenate the embedded data, as shown in the formula.

$$F = \Delta\left(G_1([P^h, P^V])\right) \quad (3)$$

In equation 3, spatial data is processed with an intermediate feature mapping F and a nonlinear activation function Δ, which improves the model's nonlinear fitting. The intermediate feature mapping f is split into two tensors, $G^h \in \left(L^{C/r\times H}\right)$ and $G^v \in \left(L^{C/r\times V}\right)$ and then both are transformed into tensors with the same number of channels using 1×1 convolutions X_h X_Y.

$$T^h = \delta\left(X_h(G^h)\right) \quad (4)$$

$$T^v = \delta\left(X_V(G^V)\right) \quad (5)$$

Attention values for width and height are denoted by th and T^v, respectively. We create a feature map by multiplying and summing with these attention weights.

$$Y_C(i,k) = M_c(i.k) * t_C^h(i) * t_C^V(k) \quad (6)$$

Attention weights Y_C *(i, k),* in Formula 6 enhance YOLOv5s by integrating the CA module with feature fusion. This improves accuracy by focusing on key regions and utilizing multi-scale data. Expanding network width and depth improves accuracy but can cause overfitting. The inception framework uses sparse features to maintain accuracy with fewer parameters (Figure. 2).

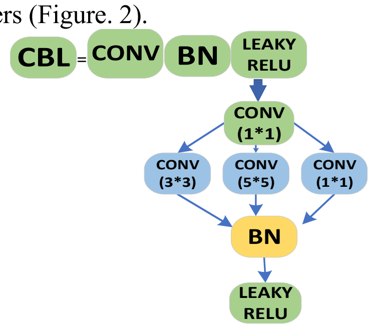

Fig.2 Enhanced convolutional architecture

Multiple scale convolutions (3x3 and 5x5) enhance feature extraction, with layer-by-layer fusion improving feature integration. Figure 3 shows the enhanced bottleneck and CSP trunk structures for effective feature extraction and generalization. The original YOLOv5 location loss function struggles with bounding box prediction. This work proposes adding Generalized Intersection over Union (GIOU) to address this by penalizing discrepancies between actual and predicted bounding boxes.

$$GIOU = IOU - \frac{|Y \setminus (PUB)|}{Y} = IOU - \frac{P^C - u}{P^C} \qquad (7)$$

Figure 4 shows Y as the minimal convex shape enclosing P and B. The area Pc and overlap u are used in the GIOU loss formula.

$$GIOU\ LOSS = 1 - GIOU \qquad (8)$$

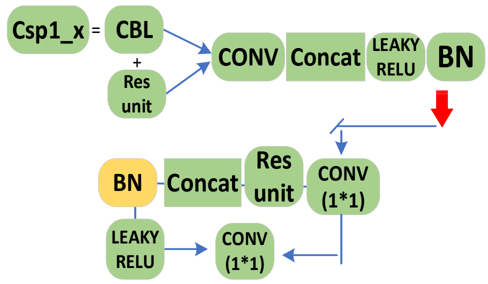

Fig.3 enhanced bottleneck Structure of the CSP network

The GIOU loss function improves anchor box sizes and locations by enhancing bounding box predictions during training. The total loss function used in this study is shown below [5].

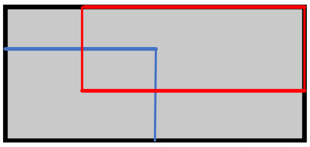

Fig.4 Convex Object Enclosed

$$\begin{aligned} E(Object) = {} & \forall_{coord} \sum_{i=0}^{n\times n} \sum_{j=0}^{f} X_{ij}^{obj} (2 - W_i \times H_i)(1 - GIOU) \\ & - \sum_{i=0}^{n\times n} \sum_{j=0}^{f} X_{ij}^{obj} [T_i log(T_i) \\ & + (1 - T_i) log(1 - T_i)] \\ & - \forall_{NOOBJ} \sum_{i=0}^{n\times n} \sum_{j=0}^{f} X_{ij}^{NOOBJ}\ [T_i log(T_i) \\ & + (1 - T_i) log(1 - T_i)] \\ & - \sum_{i=0}^{n\times n} \sum_{j=0}^{f} X_{ij}^{OBJ} \sum_{z\in classes} \Big[P_i(z) log\Big(P_i(z) \\ & + (1 - P_i(z)) log(1 - P_i(z))\Big)\Big] \end{aligned}$$

III. The Experimental Process and Results Analysis:

The experiment used a 13th Gen Intel® Core™ i9-13900H, Windows 11, NVIDIA® GeForce RTX™ 4080, Python 3.9.1, and Torch 1.8.1. A new dataset of 6000 images, including plastic bags, bottles, cans, and containers, was created, consisting of 2000 test images and 5000 training images (see Table 1).

TABLE I. Quantitative outcomes for each class of target sample

Group/Data Set	Bottles	Cans	Plastic Bags	Containers
Tanning	8502	4854	2980	2900
Testing	2200	1532	1035	982

$$PRECESION = \frac{tp}{fp + tp} \qquad (10)$$

$$RECALL = \frac{tp}{fn + tp} \qquad (11)$$

Algorithm effectiveness is assessed using performance and accuracy metrics. Mean Average Precision (mAP) measures overall probability, with accuracy and recall calculated via formulas (10) and (11) , and it is presented in Table 2.

The enhanced-YOLOv5 algorithm shows a 12.4% accuracy improvement and 4.4-fold speed increase compared to YOLOv5, YOLOv7, YOLOv7-GIoU, and Faster-RCNN. With inception structure and label smoothing, it achieves a 17.2% higher recall rate and an mAP of 97.39%, a 15.28% improvement from the baseline.

TABLE II. A study of floating objects' detection capacities on a water surface

Model Name	Recall%	Precision%	F1-Score
YOLOv7	85.7	86.8	85.9
YOLOv7-GIoU	85.9	87.2	86.5
Faster--RCNN	81.5	83.8	82
YOLOv5	82.4	81.6	81
Our Improved YOLOv5	91.8	93.5	91.6

TABLE III. A comparison of the enhanced model

	AP %			
Model Name	Bottles	Cans	Plastic Bags	Containers
YOLOv7	92.6	91.5	89.2	89.3
YOLOv7-GIoU	91.2	90.6	91.6	88.5
Faster--RCNN	79.8	80.2	79.2	81.2
YOLOv5	85.1	83.2	86.2	89.1
Our Improved YOLOv5	98.8	95.4	95.8	94.6

Table 3 shows that enhanced YOLOv5 offers improved robustness with more consistent AP values across target types. Fig. 5 illustrates effective detection of floating objects, despite limited samples for certain items.

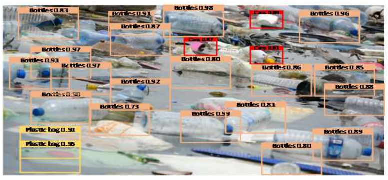

Fig.5 Outcomes of image detection of goods floating on the water's surface

Test results of the enhanced YOLOv5 in low light, obscured targets, and varied angles are shown in Fig. 6

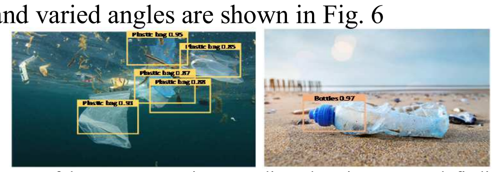

Fig.6 A segment of the test outcome in a complicated environment and findings in a basic environment

IV. Conclusion

The refined YOLOv5 technique improves floating object detection with a CA attention module and optimized anchor boxes. This boosts accuracy, performance, and efficiency. Future work will focus on better recognition and night detection.

V. References

[1] Nadal, M.; Alomar, C.; Deudero, S. High levels of microplastic ingestion by the semipelagic fish bogue Boops boops (L.) around the Balearic Islands. Environ. Pollut. 2016, 214, 517–523.

[2] C. Pang and Y. Cheng, "Detection of River Floating Waste Based on Decoupled Diffusion Model," 2023 8th International Conference on Automation, Control and Robotics Engineering (CACRE), Hong Kong, China, 2023, pp. 57-61

[3] F. Fough, G. Janjua G, Y. Zhao, AW. Don "Predicting and Identifying Antimicrobial Resistance in the Marine Environment Using AI & Machine Learning Algorithms". 2023 IEEE MetroSea 2023 .

[4] X. Wang, F. Wang, H. Dou, J. Ding, S. Miao and Y. He, "Research Progress of Water Surface Object Detection," 2023 IEEE/ACIS 23rd International Conference on Computer and Information Science (ICIS), Wuxi, China, 2023, pp. 20-24

[5] Guo, Zhaojie, et al. "Enhanced Water Leakage Detection in Shield Tunnels Based on Laser Scanning Intensity Images Using RDES-Net." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2024).