Computer Fraud and Security
ISSN (online): 1873-7056

A Novel Garbage Detection Algorithm for Unmanned Surface Vehicles based on Reparameterized Model and Dual Path Feature Fusion

Weina Zhou, Xiang Ren*

College of Information Engineering, Shanghai Maritime University, 1550 Haigang Avenue, Shanghai, 201306, China.

**Corresponding author: Xiang Ren, email: 18317116714@163.com*

Accepted: 30-03-2025 Published: 05-04-2025

Abstract:
With the growing garbage flowing into the ocean through inland rivers, garbage detection and cleanup has become an urgent and necessary task for the safety of the entire ecosystem. Considering the expensive labor cost, Unmanned Surface Vehicle(USV) has been accepted as an important intelligent robot in river management field, and camera is widely adopted as a cost-effective way compared to other sensors like radar and laser for USVs in garbage detection. However, garbage detection based on vision is often affected by factors, such as the small size of distant targets, surface glare, and interference from other floating objects. It is a conflicting issue to achieve a higher detection accuracy with limited resources. To solve these problems, a novel detection algorithm is put forward for USV in garbage cleanup in the paper. Specifically, Dual Block Module(DBM) and QarepC3 modules are proposed based on the reparameterized model and dual-path feature fusion approach. EMIOU and Joint Attention Module(JAM) are constructed according to the characteristics of floating garbage. Subsequent to comprehensive evaluations, the proposed network exhibits not only high detection accuracy and computational efficiency but also robust performance in the complex environments of inland river. Moreover, it outperforms state-of-the-art networks in surface garbage detection for USV in the experiments.

Keywords: Floating garbage detection · Unmanned surface vehicles · Reparameterized model · Dual path feature fusion.

INTRODUCTION

Garbage of water surface has led to increasing severe environmental pollution problem. Garbage on the water surface not only affects the ecological environment of the water area, but also poses a threat to aquatic organisms and human health. Therefore, detection and cleanup of water surface garbage has become an urgent problem to be solved. Traditional detection methods of water surface garbage usually rely on manual inspection and salvage, which are not only inefficient but also costly. Therefore, it is of great practical value to research on an efficient garbage detection system of water surface based on unmanned surface vehicles(USVs), which could free the human and improve the efficiency of garbage cleaning.

As we know, camera detection is a cost-effective solution for vision detection among commonly used sensors for object detection[1]. However, floating garbage such as plastic bottles and cans are small in size and always occupy few pixels in images when they are far from the camera, thus little information about their appearance could be captured in detection. In addition, the environment of inland waters is often complex. Surface glare, reflections from objects on the riverbank, and interference from other floating objects would all bring challenges to the vision-based object detection system. Thus, an accurate and real-time floating garbage detection system is urgently needed for USV to improve cleaning efficiency in inland river.

With the quick development of deep learning, Convolutional Neural Networks (CNNs) have shown great achievements in the field of object detection[2−6]. CNNs-based object detection algorithms are mainly divided into two categories: (1) two-stage algorithms based on region proposal represented by the R-CNN series[7−9]. (2) one-stage algorithms based on regression represented by SSD[10], RetinaNet [11], and YOLO series[12−15]. The former one generates regional proposals before classifying objects. Although it could obtain a high detection accuracy, its real-time performance is poor, and the network is too massive to be deployed. By contrast, the latter one transforms the detection into a regression problem and outputs all predicted bounding boxes directly. Thus, it is more efficient in computation and more suitable to be deployed in mobile and embedded devices.

The YOLO series models of one-stage algorithms are widely used in industry and they also performed well in floating garbage detection. Niu et,al [16] proposed an automated river trash monitoring system called SuperDock. It includes a river trash detection module based on YOLOv3. In addition, a dataset has been generated for training and testing. They improve the loss function and lightweight of the original YOLOv3, and the improved YOLOv3 achieved 81.2% accuracy with an average processing time of only 0.038 seconds. However, the model based on the YOLOv3 has low detection accuracy in the face of small targets and complex environments, due to its limited feature fusion capability. Hasany et,al [17] presented a novel autonomous robotic system equipped with computer vision that helps to detect floating garbage. They first used YOLO and RetinaNet to train on their homemade dataset, but the prediction time for every single image is more than 20 seconds. Then they tried Tiny-YOLO, a smaller network, which achieves 86.9% mAP and takes 3.5 seconds to process a single image on the Raspberry Pi. Kong et,al [18] develop an intelligent water cleaning robot system called IWSCR for collecting floating plastic garbage. The network was trained on the proposed floating garbage dataset with 91% mAP. Aldric SiO et,al [19] developed a system for identifying plastic bottles on the surface of rivers using Raspberry Pi 4B. The detection algorithm is based on YOLOv5, with an overall accuracy of 84.3%. The network was planted with embedded hardware, but its detection accuracy is still low. Nguyen et,al [20] reduced computing costs and improved training and reasoning speed by using MixConv and reduced detector heads based on YOLOv5, which is more suitable to be deployed in embedded device. Recently, transform and attention mechanism began to rise in the field of target detection. Chengwenyuan Huang et,al [21] combined YOLOv5 with CBAM and transform to build a new network to detect surface targets, which exceeded the performance of the original YOLOv5 with its own dataset. All of them improved the detection accuracy of the network based on YOLOv5, which proves its effectiveness in target detection. However, these networks are still not good enough in the trade-offs among model size, inference speed, and detection accuracy. Kaiyuan Dong et,al [22] proposed a cross-layer weighted path aggregation network that incorporates two bottom-up paths for weighted feature fusion, enabling feature interactions across dimensions and spatial domains. Additionally, they introduced and applied a double residual group convolution to both the backbone network and the cross-layer weighted path aggregation between networks. Consequently, the feature layers integrate rich semantic information with fine-grained details. Although these systems obtained some achievements in floating garbage detection, few studies are from the perspective of USVs. Moreover, the balance and trade-offs among model size, inference speed, and detection accuracy needs to improve.

YOLOv5n is a kind of YOLOv5, which not only maintain the efficiency but also has small parameters. It is conducive to be deployed on embedded devices for unmanned surface vehicles. Thus, we adopted it as a baseline in our method for garbage detection which aims to exhibit an optimal trade-off between parameter efficiency, real-time performance and detection accuracy. Its structure is shown in Fig.1,and the contributions of this paper are summarized as follows.

(1) Propose Dual Block Module (DBM) for the backbone based on dual path feature fusion to improve accuracy while keeping real-time performance.

(2) Propose QarepC3 module for the neck of the network based on reparameterized model to improve accuracy with slightly reduction in parameters.

(3) Propose a new attention mechanism named Joint Attention Module(JAM) and EMIOU by making full use of the unique traits of surface garbage.

(4) Propose a much feasible solution by integrating efficient modules for surface garbage detection for unmanned surface vehicle.

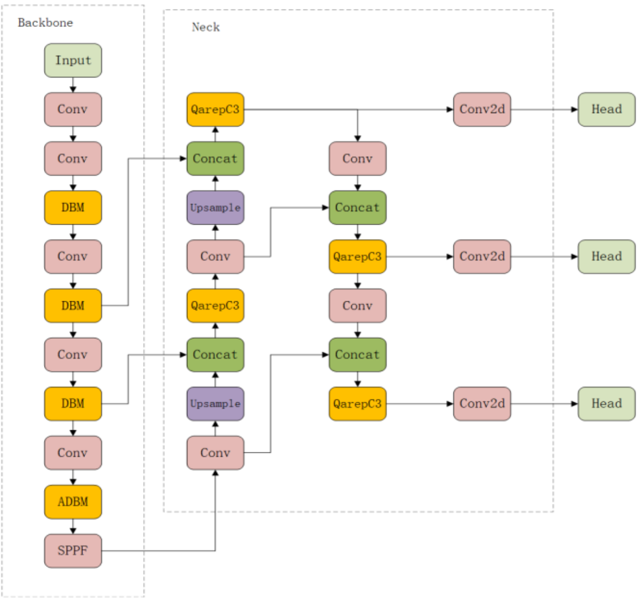

Fig. 1 The proposed network architecture

OUR APPROACH

The DBM structure and a new-style dual backbone

Detecting garbage on the water is a challenging task due to the dynamic surface conditions caused by factors such as weather, sunlight, waves, reflections, and so on. Sometimes, garbage would even share similar colors with the water ripples and be obscured by foam or other objects. These presents a big challenge to the feature extraction of object detection networks. Thus, improving the capability of backbone, which accounts for most of the parameters in the network, is an important mean in feature extraction.

To enhance the network’ s feature extraction ability, DBM, a novel architectural component, was integrated into the backbone network. Inspired by the method of two-path feature fusion, the DBM employs a dual-branch structure. In contrast to the C3 structure, the DBM architecture comprises two branches of equal depth.

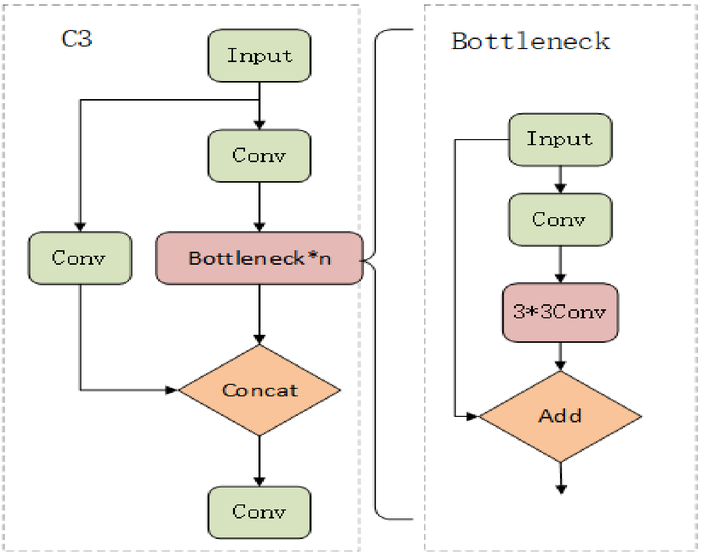

Fig. 2 C3 module

As we know, the C3 structure within the YOLOv5 backbone contains two branches: one branch directly passes through n bottleneck structures, while the other branch undergoes a single 1x1 convolution. Its structure is shown in Fig.2. The green convolution in the figure represents 1*1 convolution. This configuration results in a significant depth disparity between the two branches in the C3 structure. Although this depth disparity facilitates the network ’ s ability to integrate deep and shallow feature information, during actual network inference process, the shallow branch must wait for the deep branch to complete its computations before proceeding with further operations. This results in inefficiencies, such as increased inference time and underutilized hardware resources.

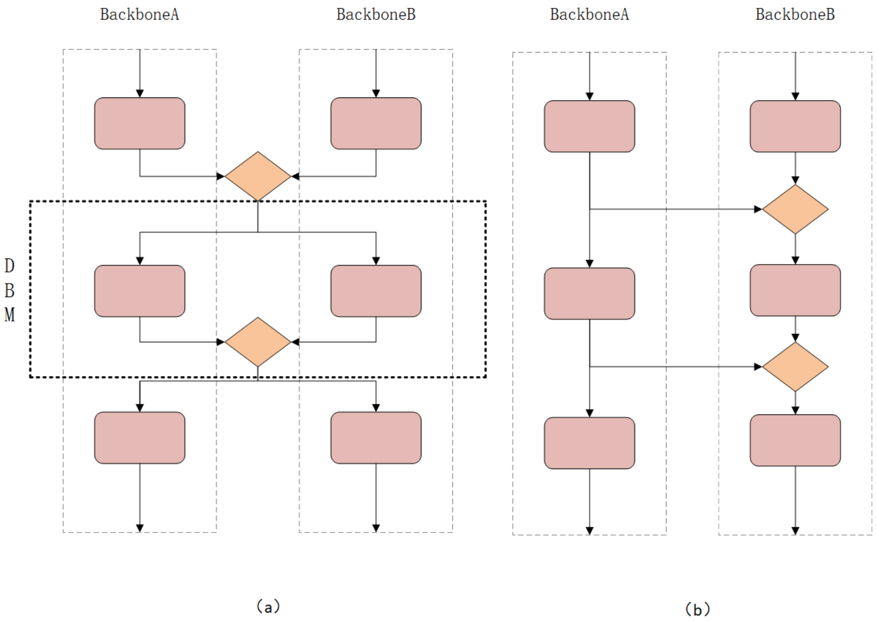

Fig. 3 (a)Proposed dual backbone network (b)Traditional

dual backbone network

Adopting the concept of dual backbone network, the feature extraction module is added in both branches in DBM. Although the C3 structure also has two branches, one of them does not participate in feature extraction, and thus it cannot be considered a dual backbone. This parallel architecture enables the model to learn and extract diverse features from the input data. By combining their outputs, the model leverages these different features to make predictions. This feature fusion enables the model to capture multiple facets of the input data, thereby improving performance of feature extraction.The parallel structure allows the model to flexibly adjust the weights and parameters of different branches according to the task requirements. During the training process, the model can automatically learn how to balance the contributions of different branches to optimize overall performance.

The structures of the proposed and traditional dual backbone network are shown in Fig.3.The newly designed backbone is similar to but different from the traditional dual backbone network, in which one backbone primarily serves the other. The newly network fuses features from both backbones at each stage. These fused features then serve as the common input for both backbones in the next stage, thereby enhancing the network ’ s robustness and feature extraction capabilities.Additionally, our proposed backbone architecture deviates from the combination of two fully independent backbones. Instead, it has two distinct columns of feature extraction modules. These columns share specific intermediate layers, resulting in fewer parameters than the traditional dual-backbone structure.

At the same time, to avoid introducing additional branches that could increase memory usage and slow down the model ’ s running speed, the Qarepconv is employed based on reparameterization. The final DBM built in our network is shown in Fig.4.

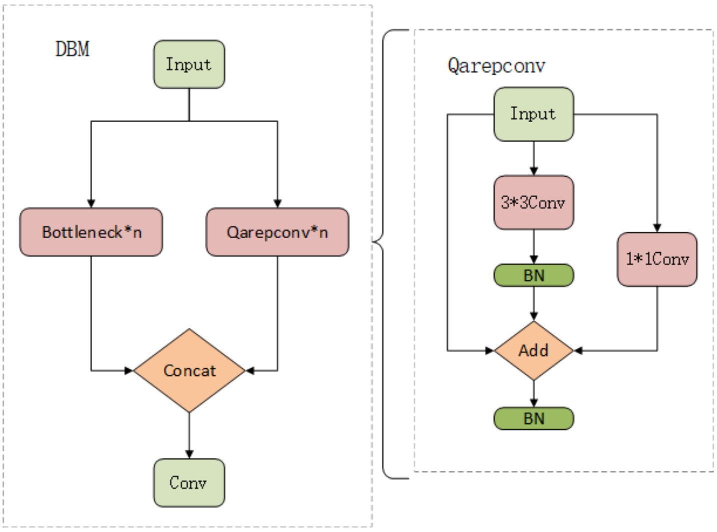

Fig. 4 DBM of backbone

The reparameterized convolution involves using a multi-branch model during training, and merging the branches into a main 3×3 convolution during inference to create a single-path model [23]. This approach combines the higher precision of the multi-branch structure with fewer parameters and higher inference speed of the single-branch structure.However, common reparameterization models suffer a significant loss of accuracy after quantization, which hinders their deployment on hardware platforms. To address this issue, Qarepconv implemented improvements by redesigning the network structure and adjusting parameters through analyzing the conditions influencing quantization error, so that Qarepconv can greatly reduce the quantization error while maintaining the tradeoff between accuracy and speed [24].The transformation process of Qarepconv between inference and training is shown in Fig.5. Its principle is that the parameters of the BN layer can be merged with the parameters of the convolutional layer, resulting in a convolutional layer with bias. At the same time, parallel 3×3 convolutions and 1×1 convolutions can also be equivalent to a single 3×3 convolution [23]. During the training phase, Qarepconv consists of three parallel branches. After training, each of these three branches is individually transformed into a 3×3 convolution. These parallel 3×3 convolutions are then fused and integrated with a BN layer to form a single 3×3 convolution for detection.

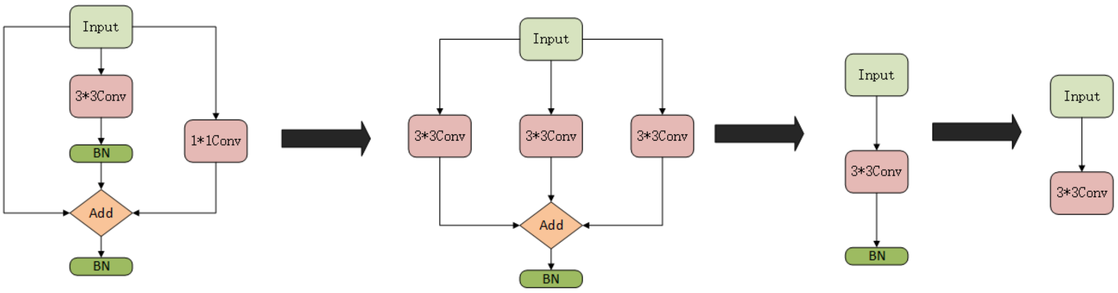

Fig. 5 Qarepconv transformation process

The proposal of the QarepC3 in the neck

In contrast to the backbone, which is mainly responsible for feature extraction, the main function of the neck is feature fusion. However, the two branches of the DBM have similar depths, which prevents the module from blending features in deep layers and shallow layers. To control the overall network parameters, the Cross-Stage Partial (CSP) structure was consistently employed in the neck architecture to combine deep and shallow features.

The proposed QarepC3 structure, based on CSP, is shown in Fig. 6. It consists of two branches: one branch directly passes through n Qarepconv, while the other branch undergoes a single 1x1 convolution. Qarepconv further reduces the risks of gradient vanishing and explosion in the network due to its branch design of merging deep and shallow information. Furthermore, owing to its reparameterization design, it does not increase the number of network parameters.

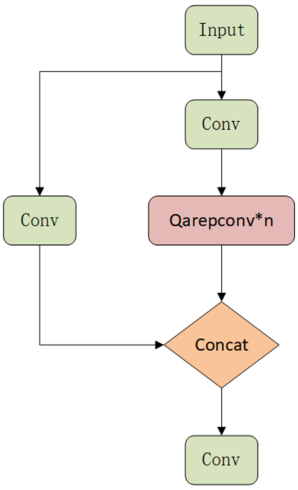

Fig. 6 The architecture of QarepC3

Joint Attention Module

In real-world scenarios, the accumulation of garbage on the water surface is a prevalent issue caused by the water flow. Images captured by USVs often show garbage concentrated in localized areas rather than being evenly distributed. This leads to additional challenges for object detection models, with the spatial attention mechanism emerging as an effective solution.

The spatial attention mechanism introduces attention weights during the feature extraction process so that the model can automatically identify the key regions of the target. This approach is typically based on a convolutional neural network (CNN) architecture that adjusts the weights of different regions by computing the corresponding attention weights. Consequently, the model can pay more attention to the target-related regions to improve detection performance. Channel attention has also been demonstrated effective in enhancing the performance of dual backbone neural networks.The channel attention mechanism adaptively learns the importance of each channel and assigns weights accordingly.

Integrating the spatial and channel attention mechanisms significantly enhances the performance of our network. Furthermore, it has been observed that applying channel attention before spatial attention is more effective than other combinations [25]. Thus, the JAM was proposed, which incorporates channel attention and spatial attention components.

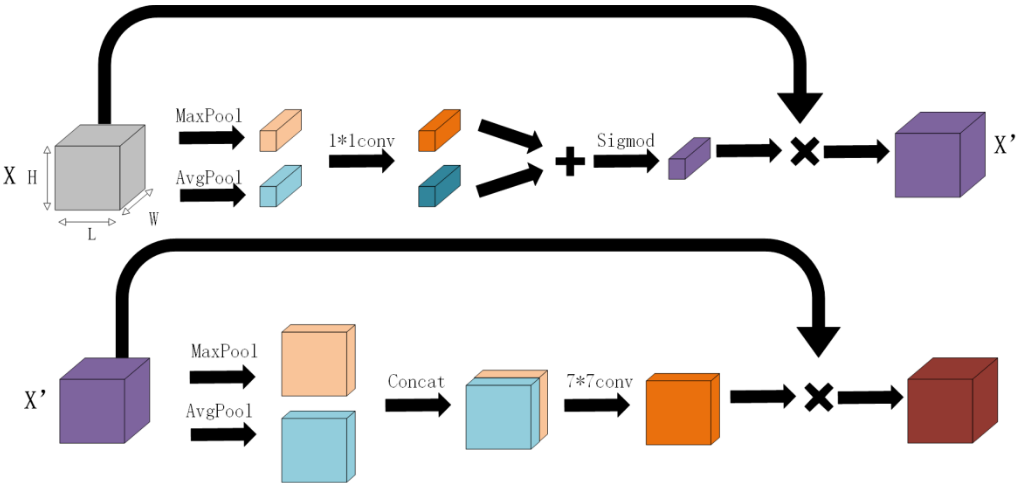

Fig. 7 The architecture of JAM

The structure of JAM is shown in Fig. 7. For an input feature map X of dimensions W×H×L, we initially apply average pooling and max pooling operations to obtain two 1×1×W feature maps, respectively representing the global average and maximum features across each channel. Subsequently, these two feature maps undergo 1×1 convolution, followed by an element-wise addition and a Sigmoid activation function to constrain the weights within the range of 0 - 1, thereby yielding a 1×1×W channel attention feature map. The employment of 1× 1convolutions instead of traditional fully-connected layers aims to reduce the number of parameters while mitigating information loss caused by dimensionality reduction operations [26].This channel attention feature map is then element-wise multiplied with the input feature map X, producing an enhanced channel-attentive feature representation X′ of dimensions W×H×L. Subsequently, average and max pooling operations are performed on X′ to obtain two W×H×L feature maps. These two feature maps are then concatenated, followed by a 7× 7 convolution, producing a single-channel feature map. After applying a Sigmoid activation function, this single-channel feature map is element-wise multiplied with X′ to get the final enhanced feature representation X′ ′ with dimensions W×H×L, incorporating both channel and spatial attention.

However, adding the attention module at the low level of the network is not very effective where the number of channels in the feature map is too small and the resolution is too high. Additionally, adding the attention module in the highest level of the network can easily cause overfitting when the number of channels is excessive. So our JAM, which is connected by the channel attention mechanism and the spatial attention mechanism, is integrated into the final DBM module at the end of the backbone. It is named Attention Dual Block Module (ADBM), as shown in Fig. 8.

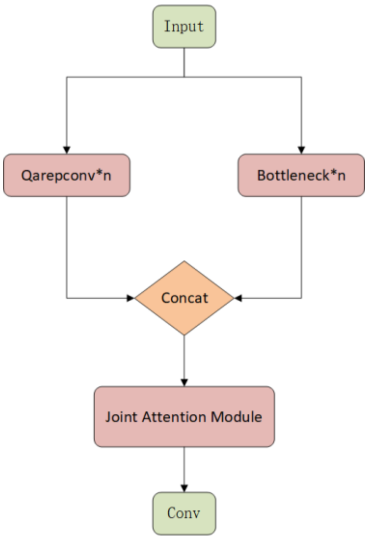

Fig. 8 The flow chart of ADBM

The EMIOU

When performing target detection on USVs, significant differences arise in comparison to fixed target detection equipment or unmanned aerial vehicles. Owing to the motion characteristics of USVs, the distance between the USV and the target is continuously changing, leading to substantial variations in the scale of the object to be detected . These scale variations present a significant challenge for accurate positioning of bounding boxes for networks.

Specifically, when the USV is near the target, the size of the object to be detected in the image becomes relatively large, and when the USV is away from the target, the size of the object to be detected becomes relatively small. These scale variations pose a significant challenge for the network to accurately adjust the size and positioning of the bounding boxes to adapt to objects of varying scales.

The bounding box regression loss of YOLOv5 is calculated by CIOU.The formula for CIOU is as follows:

$$v = \frac{4}{\pi^2}(\arctan\frac{w_A}{h_A} - \arctan\frac{w_B}{h_B}) \tag{1}$$

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{2}$$

$$IOU(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{3}$$

$$CIOU = IOU(A,B) - \frac{\rho^2(A_{ctr}, B_{ctr})}{c} - \alpha v \tag{4}$$

Where A and B represent the ground truth bounding box and the predicted bounding box respectively, A_{ctr} and B_{ctr} represent the center point of A and B respectively, w and h refer to the width and height of the rectangular box respectively,α is a weighting function, and v is used to measure the consistency of the aspect ratio, c is the diagonal distance of the smallest enclosing rectangle.ρ represents the Euclidean distance between two points.

Building upon the IOU, the CIOU additionally accounts for the relationship between the central point positions and the aspect ratios of the predicted and ground truth bounding boxes. However, it may not adequately address the complex scenarios encountered in garbage detection on water surface.

According to the concept of EIOU[27], the aspect ratio component of the CIOU metric is decoupled to independently calculate the length and width ratios between the ground truth and predicted bounding boxes. This directly minimizes the difference between the width and height of the ground truth and predicted bounding boxes, resulting in faster convergence.However, this loss function will lose its optimization ability,when the predicted and ground truth bounding boxes have the same aspect ratio but the height and width are completely different[28].Consequently, we further incorporate the distance loss between the upper left and lower right corners of the ground truth and predicted bounding boxes.We call it EMIOU, and its formula is as follows:

$$L1 = \frac{\rho^2(A_{ctr}, B_{ctr})}{c^2} \tag{5}$$

$$L2 = \frac{\rho^2(A_{UL}, B_{UL})}{c^2} + \frac{\rho^2(A_{BR}, B_{BR})}{c^2} \tag{6}$$

$$L3 = \frac{\rho^2(w_A, w_B)}{w_E^{\ 2}} + \frac{\rho^2(h_A, h_B)}{h_E^{\ 2}} \tag{7}$$

$$EMIOU = IOU(A,B) - L1 - L2 - L3 \tag{8}$$

Where UL and BR stand for Upper Left and Bottom Right respectively, A_{UL} represents the upper left corner of the ground truth bounding box A, while E is the minimum enclosing rectangle of the predicted and ground truth bounding boxes.L1 is the functional component that measures the distance between the center points of two boxes, L2 measures the distance between the vertices of two boxes, and L3 measures the difference in edge lengths between the two boxes.

EXPERIMENT RESULTS AND DISCUSSIONS

Experimental Dataset

The FloW dataset [29] is the world's first floating garbage detection dataset from the viewpoint of USVs. It is published by ORCAUBOAT in collaboration with Turing Award winner Yoshua Bengio and other researchers. FloW includes an image sub-dataset FloW-Img and a multimodal sub-dataset FloW-Radar-Img(FloW-RI). FloW-

Img is adopted as the dataset used in this paper,because our model is based on visual tasks. It was collected by cameras and contains a total of 2000 images and 5217 ground truths, of which small targets (area $<$ 32×32) account for more than half of them. We randomly select 1200 images for training, 400 images for validation, and 400 images for testing.

Evaluation Criteria

To evaluate the pros and cons of the models comprehensively, the evaluation indicators used in this paper include Parameters, Giga Floating Point Operations (GFLOPs), Inference Time and mean Average Precision (mAP).Parameters refer to the total number of parameters in the model. GFLOPs stand for billion Floating-Point operations, which are used to measure model complexity. Inference Time refers to the time taken by the model to infer an image, which is measured in milliseconds. And the mAP is a comprehensive measure of the accuracy of the object detection algorithm. mAP@0.5–0.95 is the average mAP on different IOU from 0.5 to 0.95 with astride of 0.05. Its formula is derived as follows:

$$AP = \int_0^1 P(R)dR \tag{9}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \tag{10}$$

Where i stands for a category, n stands for the total number of categories, P stands for precision rate and R stands for recall. AP stands for the area under the precision and recall curves, which represents the precision value of a single category. Thus, the mAP stands for the average precision value of all category.

Implementation Details

The proposed algorithm is implemented based on Pytorch. The original image sizes are uniformly adjusted to 640 × 640 during training. In the stage of data preprocessing, HSV color space enhancement is used to randomly adjust the hue, saturation, and value of the images to simulate different lighting conditions. And Copy-pasting data enhancement is also used to improve the network's detection effect of small targets. Copy-pasting pastes small targets (< 32 × 32 pixel) to any position in the image and generates new annotations, increasing the number of anchors by enlarging the number of small targets in each image, thus improving the contribution of small targets to the loss calculation. All experiments were performed using SGD optimizer with gradient descent, batch-size is set as 8, initial learning rate as 1e-2, cosine annealing hyperparameter as 1e-1, learning rate momentum as 0.937, and weight decay factor as 5e-4. The network was trained for a total of 300 epochs on a NVIDIA GeForce RTX 3060 GPU in each experiment.

Ablation Experiments

To evaluate the effectiveness of each proposed modules in our algorithm, experiments are conducted on the FloW-Img dataset to quantitatively analyze the detection accuracy, the number of parameters, and the complexity of the model.

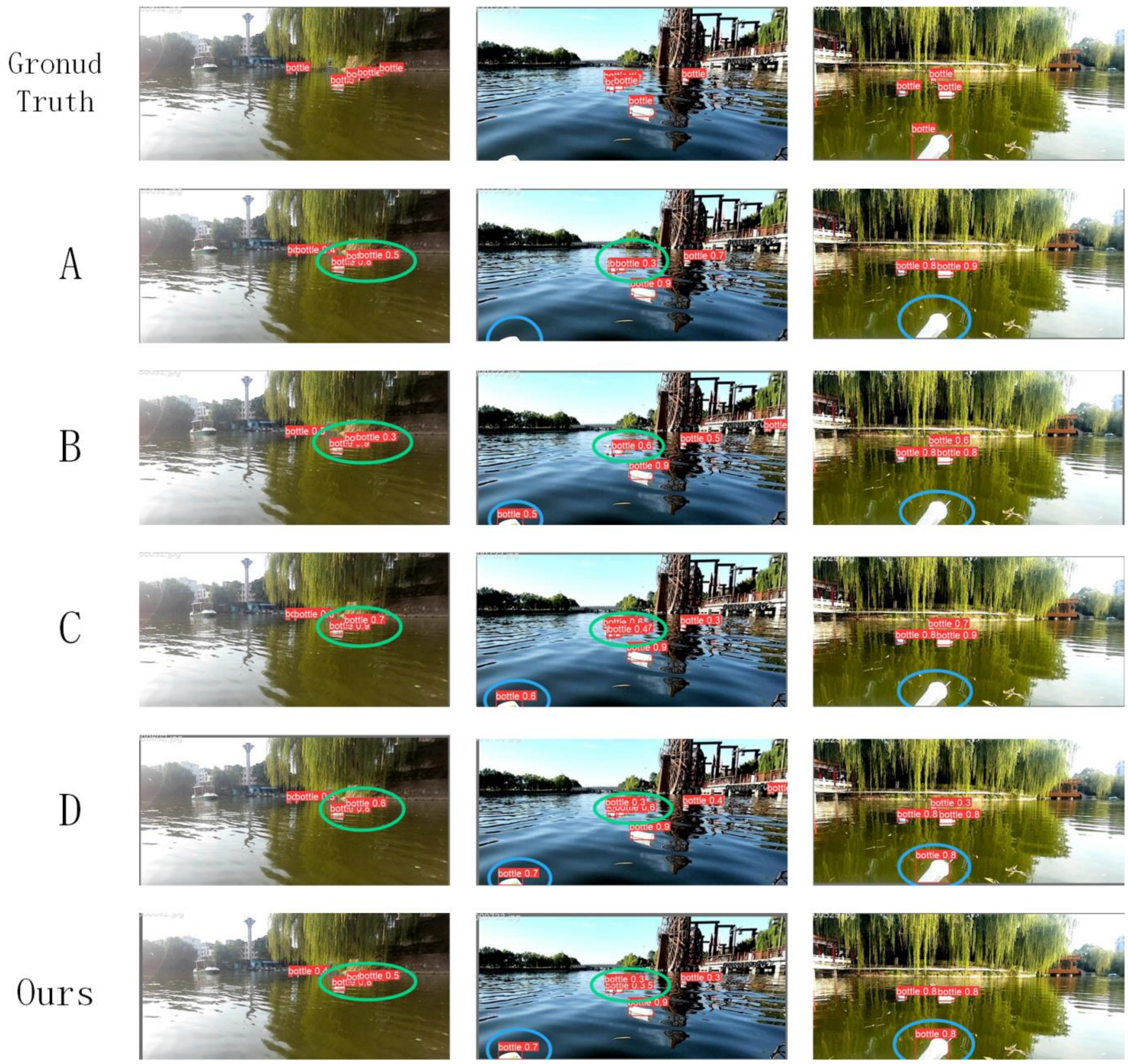

Fig. 9 Visual comparisons of five different methods

Fig.9 shows the visual results of five networks with different modules. Where A represents YOLOv5n,B represents YOLOv5n + DBM,C represents YOLOv5n + DBM + QarepC3,D represents YOLOv5n + DBM + QarepC3 + JAM. It is evident that YOLOv5n struggles with detecting nearby targets that are partially occluded and distant targets that are densely packed. In Fig.9, the blue-circled area highlights the partially occluded targets, while the green-circled area indicates the densely packed distant targets. It can be seen that, with the improvement made by our proposed network, these difficulties have been progressively conquered.

The results of the experiments are also shown in Table1. It can be seen that after upgrading the backbone of YOLOv5n with the DBM module, the network's accuracy increased. Although the number of parameters and calculation amount of the network increased, the inference time of the network did not increase much due to the dual branch with the same depth design. This means that increased computing does not require additional hardware costs to maintain real-time network performance. On this basis, the QarepC3 structure was constructed in the neck. The data shows that the model's accuracy is slightly improved.Due to the reparameterization design, the number of parameters and reasoning time are slightly decreased. Furthermore, adding the JAM and modifying the loss function improved accuracy with a small price. Compared to YOLOv5n, the final network exhibits higher precision. Despite an increase in both parameters and computational complexity, the inference time has not significantly risen.

Table 1 Comparison between networks with different modules

	Parameters (M)	GFLOPs	mAP@0.5–0.95	Inference Time (ms)
A	1.76	4.1	0.342	10.2
B	2.36	5.8	0.356	10.5
C	2.33	6.2	0.357	10.6
D	2.34	6.2	0.363	11.5
Ours	2.34	6.2	0.365	11.5

Comparison with State-of-the-arts

Table 2 Comparison results between state-of-the-arts methods and ours

	Parameters (M)	GFLOPs	mAP@0.5–0.95	Inference Time (ms)
YOLOv8n[27]	3.01	8.2	0.355	14.5
YOLOv7Tiny[28]	6	13	0.337	8.4
YOLOXTiny[29]	5.03	18.93	0.359	27.9
Faster rcnn[30]	41.12	193.78	0.388	120.1
CenterNet[31]	14.43	48.34	0.396	72.5
RetinaNet[32]	36.1	204.36	0.398	121.1
Paper[20]	4.9	13.1	0.360	17.7
Paper[21]	6.6	15.6	0.361	15.2
Ours	2.34	6.2	0.365	11.5

We also compare the proposed model with eight state-of-the-arts object detection methods, including the Faster-RCNN, CenterNet ,RetinaNet, YOLOX-Tiny , YOLOv7-Tiny, YOLOv8n and network proposed by reference[20]and[21]. To ensure the consistency of the comparison, all the models are trained on the FloW-Img dataset. As shown in Table 2, it can be observed that larger networks such as Faster R-CNN, CenterNet, and RetinaNet achieve relatively high accuracy.However, their parameters, GFLOPs, and inference time are considerably much higher, often several times greater than those of lightweight networks. Obviously, they are not suitable to be deployed on USVs with limited hardware performance. At the same time, YOLO series including YOLOv8n, YOLOv7-Tiny, YOLOX-Tiny are relatively lightweight networks, but their Parameters and the GFLOPs still exceed our network. In addition, comparing with other two state-of-the-art models specifically designed for garbage detection on water surface, our model still outperforms in overall performance. Considering the detection accuracy, speed and model size comprehensively, it can conclude that our model is the best choice in garbage detection on water surface for actual deployment.

CONCLUSION

In this paper, we propose a novel lightweight and high-precision network architecture based on YOLOv5n, addressing the challenge of garbage detection on water surface for USVs. In this paper, the DBM structure has constructed and employed in backbone part to effectively enhance the network's performance while preserving its real-time capabilities. Then, QarepC3 constructed by Qarepconv is used to neck part to further improve network performance. Finally, according to the characteristics of floating garbage on the water surface, the JAM and EMIOU are constructed to further improve the network performance.Comprehensive comparative evaluations with state-of-the-art networks suggest that the proposed network offers a potentially optimal solution for garbage detection on water surface in the current field. Our6 further work will be dedicated to model deployment and enabling real-time garbage detection algorithm on embedded devices.

DECLARATION OF CONFLICTING INTERESTS

The author(s) declared no potential conflicts of interest with respect to the research, author-ship, and/or publication of this article.

DATA SHARING AGREEMENT

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

FUNDING

The author(s) received no financial support for the research, authorship, and/or publication of this article.

REFRENCES

[1] Y. Cheng et al., “FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters,” in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada: IEEE, Oct. 2021, pp. 10933–10942. doi: 10.1109/ICCV48922.2021.01077.

[2] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, “TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios,” in 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada: IEEE, Oct. 2021, pp. 2778–2788. doi: 10.1109/ICCVW54120.2021.00312.Velit laoreet id donec ultrices tincidunt arcu non sodales neque. Non curabitur gravida arcu ac tortor dignissim convallis aenean et.

[3] W. Zhou, X. Huang, and X. Zeng, “Obstacle detection for unmanned surface vehicles by fusion refinement network,” IEICE TRANSACTIONS on Information and Systems, vol. 105, no. 8, pp. 1393–1400, 2022.

[4] W. Zhou, Y. Zhou, and X. Zeng, “An Attention Nested U-Structure Suitable for Salient Ship Detection in Complex Maritime Environment,” IEICE Transactions on Information and Systems, vol. 105, no. 6, pp. 1164–1171, 2022.

[5] W. Zhou and K. Chen, “A lightweight hand gesture recognition in complex backgrounds,” Displays, 2022.

[6] W. Zhou and L. Liu, “Multilayer attention receptive fusion network for multiscale ship detection with complex background,” Journal of Electronic Imaging, vol. 31, no. 4, p. 043029, 2022.

[7] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation,” in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580–587. doi: 10.1109/CVPR.2014.81.

[8] R. Girshick, “Fast R-CNN,” in 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1440–1448. doi: 10.1109/ICCV.2015.169.

[9] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.

[10] W. Liu et al., “SSD: Single Shot MultiBox Detector,” in Computer Vision – ECCV 2016, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham: Springer International Publishing, 2016, pp. 21–37.

[11] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll ´ ar, “Focal Loss for Dense Object Detection,” in 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2999 – 3007. doi: 10.1109/ICCV.2017.324.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[13] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6517–6525. doi: 10.1109/CVPR.2017.690.

[14] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv:1804.02767 [cs], Apr. 2018, Accessed: May 06, 2022. [Online]. Available: http://arxiv.org/abs/1804.02767.

[15] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934 [cs, eess], Apr. 2020, Accessed: May 06, 2022. [Online]. Available: http://arxiv.org/abs/2004.10934

[16] G. Niu, J. Li, S. Guo, M.-O. Pun, L. Hou, and L. Yang, “SuperDock: A Deep Learning-Based Automated Floating Trash Monitoring System,” in 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO), Dali, China: IEEE, Dec. 2019, pp. 1035–1040. doi: 10.1109/RO-BIO49542.2019.8961509.

[17] S. N. Hasany, S. S. Zaidi, S. A. Sohail, and M. Farhan, “An autonomous robotic system for collecting garbage over small water bodies,” in 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE), Dalian, China: IEEE, Jul. 2021, pp. 81 – 86. doi: 10.1109/CACRE52464.2021.9501299.

[18] S. Kong, M. Tian, C. Qiu, Z. Wu, and J. Yu, “IWSCR: An Intelligent Water Surface Cleaner Robot for Collecting Floating Garbage,” IEEE Trans. Syst. Man Cybern, Syst., vol. 51, no. 10, pp. 6358–6368, Oct. 2021, doi: 10.1109/TSMC.2019.2961687.

[19] G. Aldric Sio, D. Guantero, and J. Villaverde, “Plastic Waste Detection on Rivers Using YOLOv5 Algorithm,” in 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kharagpur, India: IEEE, Oct. 2022, pp. 1 – 6. doi: 10.1109/ICC-CNT54827.2022.9984439.

[20] T.-T. Nguyen and H.-L. Tran, “An Efficient Model for Floating Trash Detection based on YOLOv5s,” in 2022 9th NAFOSTED Conference on Information and Computer Science (NICS), Ho Chi Minh City, Vietnam: IEEE, Oct. 2022, pp. 230–234. doi: 10.1109/NICS56915.2022.10013413.

[21] C. Huang, Y. Zhu, J. Wang and X. He, ” Water Surface Target Detection Algorithm for Unmanned Cleaning Ship Based on Improved YOLO V5,” 2022 International Conference on Cyber-Physical Social Intelligence (ICCSI), Nanjing, China, 2022, pp. 386-391, doi: 10.1109/ICCSI55536.2022.9970707.

[22] K. Dong, T. Liu, H. Du and Y. Zheng, ” Research on Sea Surface Target Detection Algorithm Based on Deep Learning,” 2023 IEEE 11th International Conference on Computer Science and Network Technology (ICCSNT), Dalian, China, 2023, pp. 258-262, doi: 10.1109/ICC-SNT58790.2023.10334617.

[23] Ding, Xiaohan, et al. “RepVGG: Making VGG-Style ConvNets Great Again.” 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, https://doi.org/10.1109/cvpr46437.2021.01352.

[24] Chu, Xiangxiang, et al. Make RepVGG Greater Again: A Quantization-Aware Approach. Dec. 2022.

[25] Mi, Nanhuan, et al. “CBMA: Coded-Backscatter Multiple Access.” 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS), 2019, https://doi.org/10.1109/icdcs.2019.00084.

[26] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, “ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks,” in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11531–11539. doi: 10.1109/CVPR42600.2020.01155.

[27] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J/OL]. Neurocomputing, 2022: 146-157. http://dx.doi.org/10.1016/j.neucom.2022.07.042. DOI:10.1016/j.neucom.2022.07.042.

[28] SILIANG M, YONG X. MPDIoU: A Loss for Efficient and Accurate Bounding Box Regression[J]. 2023.

[29] Cheng, Yuwei and Zhu, Jiannan and Jiang, Mengxin and Fu, Jie and Pang, Changsong and Wang, Peidong and Sankaran, Kris and Onabola, Olawale and Liu, Yimin and Liu, Dianbo and Bengio, Yoshua, ”FloW: A Dataset and Benchmark for Floating Waste Detection in Inland Waters.” Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021

[30] Dehaerne, Enrique, et al. YOLOv8 for Defect Inspection of Hexagonal Directed Self-Assembly Patterns: A Data-Centric Approach. July 2023.

[31] Wang, Chien-Yao, et al. YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors.

[32] Ge, Zheng, et al. YOLOX: Exceeding YOLO Series in 2021. July 2021.