USV-YOLO: An Algorithm for Detecting Floating Objects on the Surface of an Environmentally Friendly Unmanned Vessel

Xuecun Yang, Yijing Song, Lintao He, Hang Xue, Zhonghua Dong, Qingyun Zhang

***Abstract*—Water resources are fundamental to human existence. Precise detection of surface floating objects is the primary prerequisite for environmental protection drones to conduct river cleaning operations. Aiming at the current target detection algorithm's poor adaptability to small targets on the water surface under complex scenes and low feature recognition ability, this paper proposes a water surface floating object detection algorithm USV-YOLO, which realizes the accurate recognition and detection of floating objects under the complex conditions of inland rivers. Initially, a novel C2f-float module is devised. It optimizes the utilization of feature information and boosts the accuracy of detecting floating objects by sequentially fusing and concatenating the feature information emitted from the bottleneck layer; Secondly, the design introduces the GS-EVC module, which improves the utilization of raw feature information of surface floaters by incorporating the GSConv and shuffle operations, strengthens the dependencies between remote feature information, and enhances the feature recognition capability; Ultimately, the standard convolution in the backbone network is substituted with an all - dimensional dynamic ODConv. The weighted attention mechanism within it can accommodate the feature extraction of intricate targets, thereby further enhancing the network's detection precision. Experiments are conducted on open-source datasets, FloatingWaste-I and FloW-IMG, and the experimental results show that the USV-YOLO algorithm in this paper improves the average detection accuracies, mAP_{50} and mAP_{50-95}, by 4.3% and 6.1%, respectively, compared with the original network, which is both better than the other classical target detection algorithms.**

***Index Terms*—Floating Object Detection, YOLOv8, C2f-float, GS-EVC, ODConv.**

I. Introduction

Water resources are the basis for maintaining the natural ecological balance. In recent times, concurrent with the swift progress of industrialization and urban expansion, population growth, and changes in consumption patterns, the amount of floating debris in river water bodies has continued to increase, and this floating garbage not only poses a danger to the ecosystem but also has a significant impact on human activities. At present, river floating debris pollution has become a widespread and increasingly serious environmental problem worldwide.

To protect the ecological environment, it is essential to regularly remove floating debris from the surfaces of rivers and lakes. As scientific and technological progress marches on, artificial intelligence is increasingly reaching maturity in a wide range of domains. Unmanned vehicles, drones, and other autonomous equipment have entered the publics awareness, leading to the emergence of unmanned cleaning boats. These unmanned boats can replace human labor in performing more hazardous and challenging tasks related to floating debris cleanup. Before an unmanned boat can begin cleaning floating objects, target detection technology is first employed to accurately detect and localize the debris. Research into unmanned boat-based floating object detection technology can promote the application of autonomous vessels in watershed management, further advancing the intelligence and automation of such systems. Therefore, efficient and precise floating object detection on water surfaces holds significant theoretical and practical value for both research and engineering applications.

Currently, Algorithms for detecting objects floating on water surfaces can mainly be divided into two types: conventional machine learning techniques and those leveraging deep learning. In traditional machine learning-based detection, pre-processed floating object images are typically feature-extracted based on texture, color, and area. Subsequently, classification algorithms, such as SVM and BP Neural Networks, are employed to classify and identify the floating objects. For example, Kataka et al. [1] used the CIELUV color space to distinguish the sea body from the floating pixels, and this method demonstrated superior performance in the detection of plastic pixels, but it can only be systematically detected for large sea areas, and it cannot be applied to the detection of floating objects in inland waterways. Alid et al. [2] set a threshold value based on the difference between the pixel brightness of the floating wood chips and the water surface as well as the generation time probability map of their motion characteristics, and used this value to segment the water surface from the wood chips, but due to the single color and feature of the wood chips, the method is limited and cannot accurately detect the complex kinds of water surface floaters. In addition to this, Jin et al. [3] introduced an automated IGASM segmentation technique for water sur-

Manuscript received July 14, 2024; revised January 17, 2025. This work is supported by the National Nature Science Foundation of China (51804250).

Xuecun Yang is an Associate Professor of Electrical and Control Engineering, Xian University of Science and Technology, Xian 710054, China (e-mail: 421529497@qq.com).

Yijing Song is a postgraduate student in Electrical and Control Engineering, Xian University of Science and Technology, Xian 710054, China (e-mail: 1318433633@qq.com).

Lintao He is a postgraduate student in Electrical and Control Engineering, Xian University of Science and Technology, Xian 710054, China (e-mail: 1633436409@qq.com).

Hang Xue is a postgraduate student in Electrical and Control Engineering, Xian University of Science and Technology, Xian 710054, China (e-mail: 22206227126@stu.xust.edu.cn).

Zhonghua Dong is a postgraduate student in Electrical and Control Engineering, Xian University of Science and Technology, Xian 710054, China (e-mail: 2286448279@qq.com).

Qingyun Zhang is a postgraduate student in Electrical and Control Engineering, Xian University of Science and Technology, Xian 710054, China (e-mail: 2428331540@qq.com).

face debris, utilizing the Gaussian Mixture Model (GMM). This method projects GMM outcomes into the HSV color space and employs a light-shadow discriminant function to identify highlights and shadows, thereby effectively isolating the water surface debris. However, the method is prone to fail when there are highlights or a lot of shadows on the water surface due to too strong or too dark light. It can be seen that it is difficult to detect floating objects on the water surface of inland waterways, and the traditional target detection methods generally suffer from low accuracy and low efficiency, which cannot meet the requirements of the detection task.

As artificial intelligence continues to evolve, deep learning - based target detection algorithms offer novel technological assistance for detecting floating objects. Following the successive proposal of convolutional neural network CNN and RCNN [4], [5], the three mainstream algorithms of deep learning-based target detection, Faster R-CNN [6], YOLO, and SSD [7], have been widely used by scholars at home and abroad in various visual detection tasks. Take Chen et al. [8] as an example, they put forward an enhanced SSD algorithm to tackle the detection of floating objects on water surfaces, aiming to mitigate the interference from the complex aquatic surroundings, although some parameter decreases have been realized, the method is still not able to meet the real-time detection requirements due to the large parameters of the original network base. Similarly Li et al. [9] used MobileNetV3 [10] instead of the backbone network in SSD in order to reduce the computational cost of the float detection model, which improved the detection speed of the hardware, but it could not satisfy the detection of water surface floats under complex light.Yi et al. [11] proposed a floating object detection and localization algorithm based on Faster R-CNN, which can reduce the localization error without affecting the recognition accuracy, but the network still suffers from a large number of parameters and slow computation speed. Chen et al. [12] utilized the improved YOLOv5 model to detect small water surface floaters in UAV images in real-time, which can well address the missed small target images, but it cannot be adapted to floaters detection in complex water surface environments under unmanned boat view. To better embed the floater network into hardware devices, the YOLOv5 model is improved and pruned in the method in literature [13], [14], which solves the problem of difficult hardware deployment.Along with the gradual maturity of the YOLO series in the past two years, many scholars have adopted the YOLOv8 with better performance to research water surface floating object detection tasks. For example, Zhang et al. [15] proposed a surface small target detection algorithm YOLOv8-WSSOD that improves YOLOv8, which can effectively realize the accurate detection of small targets on the surface of the water, but it is targeted at all targets on the surface of the water such as boats and animals, and cannot be accurately applied to the floating garbage cleaning task. Similarly, literature [16], [17] has also optimized and improved the model parameters of YOLOv8 by considering the problems of small target detection and network lightweight, respectively. In addition to this, to promote the related research on water surface floating debris detection, Cheng et al. [18] from Tsinghua University proposed the first inland water segmentation and water surface floating debris detection dataset and used the classical target detection algorithms to conduct comparative experiments.

Contrasted with conventional machine learning approaches, deep learning's convolutional neural networks (CNNs) boast a more outstanding ability to extract features, typically surpassing most traditional methods. For instance, in detecting floating objects on the water's surface, where the surface is subject to instability due to variations in pixel patterns caused by lighting, weather, and other environmental factors, deep learning models are more adaptable to these changes and demonstrate enhanced robustness. However, most floating objects on the water's surface are complex targets with irregular shapes and sizes. Additionally, the strong mobility of the water's surface causes target overlap, making detection more challenging. As a result, detecting floating objects in the complex environment of an inner river channel demands high detection accuracy. Regrettably, existing deep learning-driven methods for identifying floating objects on water surfaces usually show weak adaptability in detecting small targets and have restricted feature recognition abilities.

Therefore, in this paper, we design a water surface floating object detection algorithm USV-YOLO for the problems of poor adaptation of small target feature information and low detection accuracy in some water surface floating object detection algorithms, and its main contributions are as follows:

1) A novel C2f-float module has been developed to optimize the use of information across the network's feature layers and boost the network's overall feature extraction capacity.
2) A small-target vision center GS-EVC module is designed, which strengthens the dependency between target remote pixels by introducing GSConv [19] as well as the shuffle [20] operation, enabling the network to extract the floating object feature information completely, and significantly improving the detection accuracy of the overall network.
3) Replacing the standard convolution in the original backbone network with the full-dimensional dynamic ODConv [21] enhances the adaptability to the complex feature information of small target floats and greatly reduces the probability of target miss-detection and misdetection by using different attention mechanisms for weighting in each dimension.

II. Related work

In this part, we delve into the architecture and underlying mechanisms of the YOLOv8 object detection network model, as well as the visual display center EVC [22] module, GSConv, and the full-dimensional dynamic ODConv are introduced.

A. YOLOv8

In 2023, Ultralytics proposed a new version of YOLOv8 based on the high efficiency and real-time performance of the previous generations of YOLO series models. Compared with the previous versions, YOLOv8 has been further enhanced and improved in terms of backbone network extraction and feature fusion and possesses stronger multi-scale feature

fusion capability. Therefore, YOLOv8 can significantly improve the performance of target detection and image segmentation tasks while ensuring high efficiency and applying it to more complex water environment scenarios. The YOLOv8 network structure is shown in Fig. 1.

As depicted in the illustration, the YOLOv8 network framework is composed of four key elements: the input layer, the backbone structure, the neck section, and the detection head component. The backbone utilizes the C2f (Cross-Stage Partial Fusion) module, which enhances gradient flow, to improve the model's ability to detect floating objects at various scales. By leveraging the combined capabilities of the convolutional layer, the C2f module, and the SPPF module, the model achieves efficient feature extraction and effective multi-scale feature fusion.

B. EVC

The EVC module is used in the CFPNet proposed by Quan et al. Its structure is shown in Fig. 2.

The visual display center EVC is mainly composed of two modules, LVC and MLP [23], which are used to capture global remote information and local corner information, respectively, and finally, the results of the floater features extracted by the two modules are spliced together along the channel dimensions for output, and the output is shown by the representation of equation 1:

$$X = cat\left(LVC\left(X_{in}\right); MLP\left(X_{in}\right)\right) \tag{1}$$

LVC is a dictionary-containing encoder, in LVC, the features X_{in} output through the Stem module is first encoded through a convolutional layer, after which the input codebook is made to map positional information to each other through the scaling factor S, and then fused using φ and fed into the fully-connected layer and a convolutional layer of 1×1 to predict the key feature information, which is then sequentially multiplied and summed by the input features X_{in} in channels. The output $LVC\left(X_{in}\right)$ is shown in equation 2, where δ is the scale factor and e is the fused codebook output.

$$LVC\left(X_{in}\right) = X_{in} \oplus \left\{X_{in} \otimes \left(\delta\left(Conv_{1\times1}\left(e\right)\right)\right)\right\} \tag{2}$$

After entering the lightweight MLP module, the input feature X_{in} needs to pass through two residual modules, and the two residual structures consist of depth-separable convolution and channel MLP, respectively. When entering each residual structure, it first needs to be normalized, and then after passing through the depth separable convolution or the channel MLP module, it sequentially performs the channel scaling and regularization operation, which is designed to improve the generalization ability of the floater detection model, and finally outputs the result of the serial connection of the two residual structures, whose output $MLP\left(X_{in}\right)$ is shown in equation 4. where X_0 is the output of the first residual structure, as shown in equation 3.

$$X_0 = DConv\left(GN\left(X_{in}\right)\right) + X_{in} \tag{3}$$

$$MLP\left(X_{in}\right) = CMLP\left(GN\left(X_0\right)\right) + X_0 \tag{4}$$

C. GSConv

GSconv, proposed by HU et al, is a hybrid convolution that incorporates SC (the channel-dense convolution operations), DSC [24] (Depth separable convolution), and Shuffle, and its structure is shown in Fig. 3. In GSConv, for the feature layer after channel scaling and dense convolution, depth separable convolution is used again, and the two outputs are shuffled after completing the splicing in the channel dimension, so that the float feature information generated by SC can be completely fused to the output information of DSC, and the two can exchange the local feature information of the small target float uniformly, and the SC and DSC The combination of SC and DSC avoids taking up more computational resources while improving the accuracy of the model.

D. ODConv

The full-dimensional dynamic convolution (ODConv) proposed by Li et al. extends the conventional dynamic convolution by incorporating three additional dynamic dimensions. The enhanced ODConv notably boosts its feature learning capacity through the incorporation of an innovative multi-dimensional attention mechanism and a parallel approach. As a result, it becomes more sensitive to the feature information of small target floaters, particularly in corner regions. Its structure is shown in Fig. 4.

For ODConv, it allocates distinct attention weights to the convolution kernel W_i across the spatial dimension, input channel dimension, output channel dimension, and the dimension of the convolution kernel itself, respectively, and gradually multiplies these four types of attention by the convolution kernel W_i in the corresponding dimensions according to the corresponding order, which can capture the feature information of the small targets in the corner area more efficiently due to the complementary effect of these four attention scalars. The above process is shown in Equation 5, where α_{si}, α_{ci}, α_{fi}, and α_{wi} are the four attention scalars in spatial dimension, input channel dimension, output channel dimension, and convolution kernel dimension, respectively.

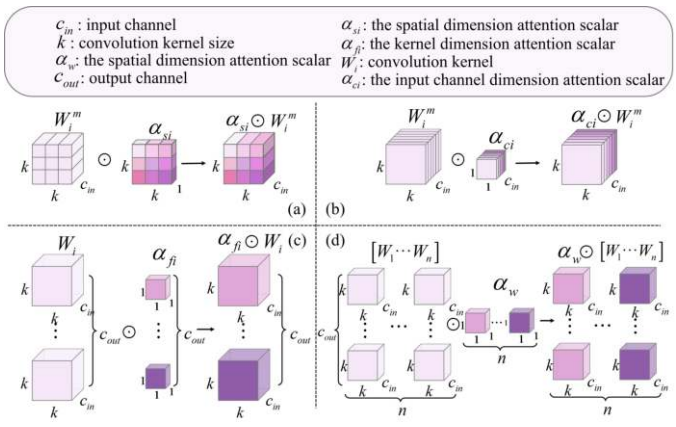

Fig. 4. Schematic diagram of the overall structure of ODConv. (a) multiplication operation of spatial dimension attention with convolution kernel; (b) multiplication operation of input channel dimension attention with convolution kernel; (c) multiplication operation of output channel dimension attention with convolution kernel; and (d) multiplication operation of kernel dimensions in convolution kernel space.

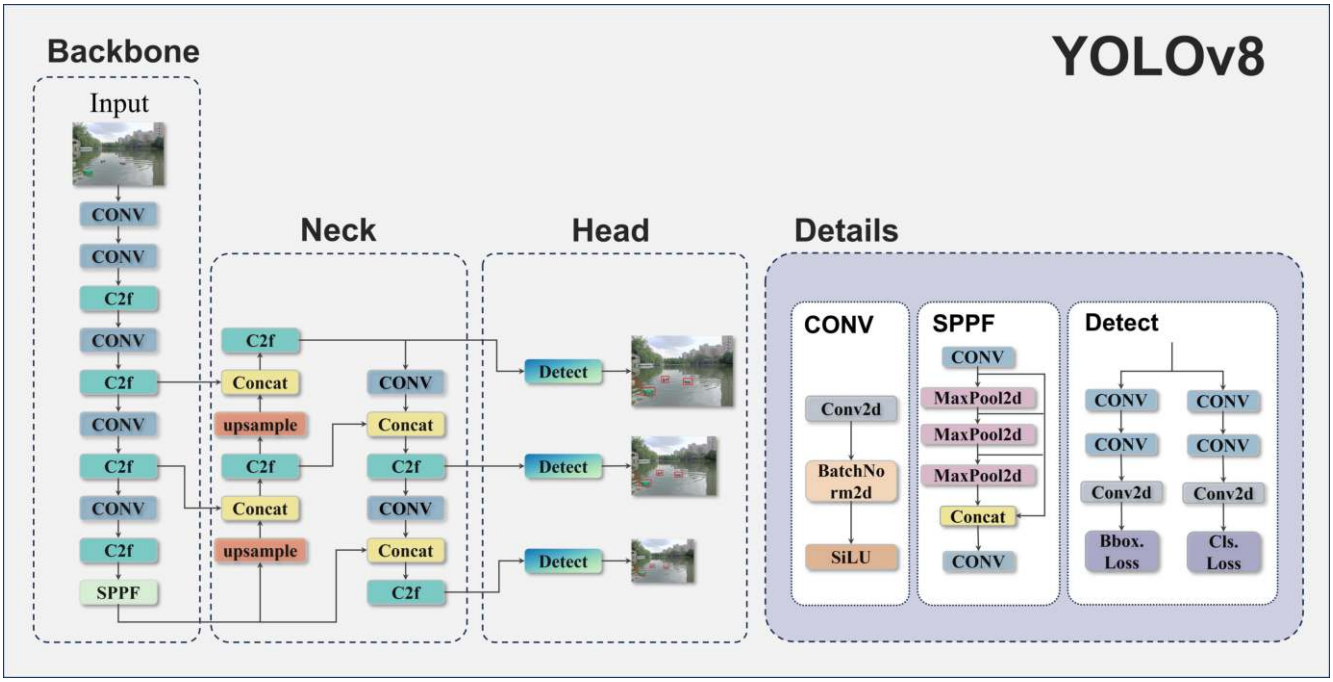

Fig. 1. YOLOv8 network structure.

$$y=(\alpha_{w1}\cdot\alpha_{f1}\cdot\alpha_{c1}\cdot\alpha_{s1}\cdot W_1+\cdots+\alpha_{wn}\cdot\alpha_{fn}\cdot\alpha_{cn}\cdot\alpha_{sn}\cdot W_n)*x \tag{5}$$

III. Algorithm Design

In the task of detecting water surface floating objects for unmanned vessels, the presence of significant noise disturbances in the complex water surface environment can disrupt pixel dependencies, leading to a reduced ability of the model to accurately recognize floating objects. Among various deep learning-based object detection methods, most models struggle to overcome the interference caused by the complex water surface environment and are often ineffective at detecting floating objects with irregular shapes and sizes. To address these challenges and improve the

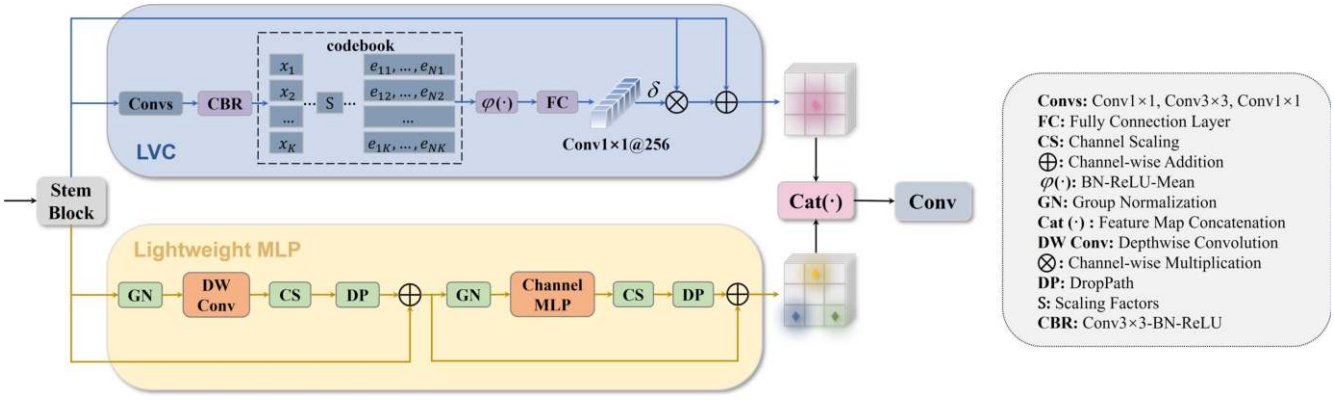

Fig. 2. EVC Module Structure.

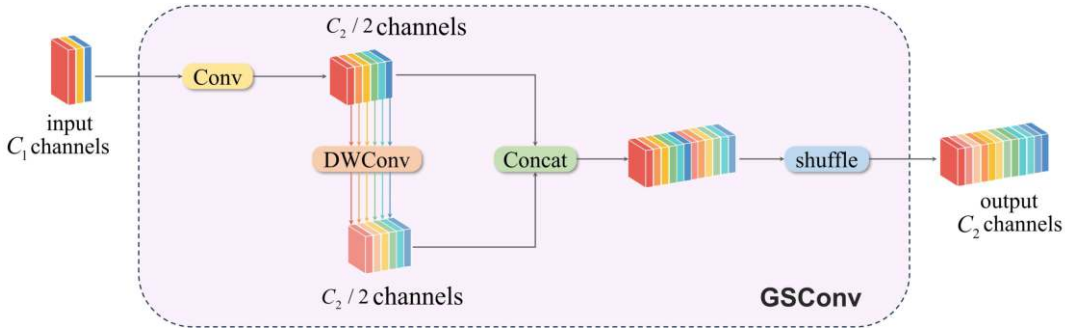

Fig. 3. GSConv structure.

feature extraction capabilities of the network, this paper proposes a novel C2f-float module. This module supplants the initial C2f module within the backbone network, thereby improving the extraction of intricate and edge features, especially for diminutive floating objects. Concurrently, the standard convolution in the backbone network is substituted with an all - dimensional dynamic ODConv. This ODConv is more attuned to the extracted target information and provides more refined feature extraction. Additionally, to better capture long-range dependencies between features and improve the detection of surface floaters with incomplete feature information in corner areas, we propose a novel GS-EVC module. The module is capable of identifying potential information leakage after the initial feature extraction phase.

A. Design of the C2f-Float module

When using the C2f module of the original network to extract features of floating objects, issues such as high light intensity and surface reflections can lead to problems like small target leakage and misdetection. Additionally, because the C2f module in the original network employs a bottleneck structure for information extraction, it tends to overlook the pixel feature information of floating objects in the corners of the water surface. To enhance the precision of detecting small target pixels, this paper's algorithm introduces a novel C2f-float module, with a comparative structural diagram presented in Fig. 5. C2f-float module on the original C2f module to reconstruct the design will be convolved after the input feature layer is divided into four channels, the number of each channel is half of the number of output channels, after that, respectively, using 0, 1, 2, 3 bottleneck, from the first two layers to fusion, and finally the fusion of the three feature results and the output of the first channel for the channel dimensionality of the splicing. Finally the output after completing the convolution, the output result x_{out} can be formulated as 6, where A is the output of all channels as shown in equation 7.

$$x_{out} = Conv_{1\times1}(cat\,(A)) \tag{6}$$

$$A = (x_1; x_1 \oplus x_2; (x_1 \oplus x_2) \oplus x_3; [(x_1 \oplus x_2) \oplus x_3] \oplus x_4) \tag{7}$$

x_n denotes the output of each channel after $n-1$ bottleneck layers, as shown in equation 8. Where x_{in} is the module input and $b\,(x_{in})$ is the output after one bottleneck layer, Each bottleneck contains two 3×3 convolutions. the process is shown in equation 9:

$$x_n = b^{(n-1)}\,(x_{in}) \tag{8}$$

$$b\,(x_{in}) = Conv_{3\times3}\left(Conv_{3\times3}(\frac{Conv_{1\times1}(x_{in})}{2})\right) \tag{9}$$

B. Design of backbone

When ordinary convolution is used for float feature information extraction in the backbone network of the original YOLOV8, it is easy to filter some of the key feature information due to the complex environment in which the float exists and the roughness of the feature area. Upon substituting the original C2f module with the new C2f-float module, to further bolster the backbone network's feature extraction prowess, this paper incorporates dynamic ODConv. It performs weighting from four dimensions respectively, and multiplies with the corresponding convolution kernels one by one. Its structure is shown in Fig. 6. Where (a) is the structure diagram of the original backbone network using standard convolution and the original C2f module, and (b) is the structure diagram of the improved backbone network, in this paper, while replacing the C2f module with the new C2f-float module, the standard convolution is replaced by ODConv.

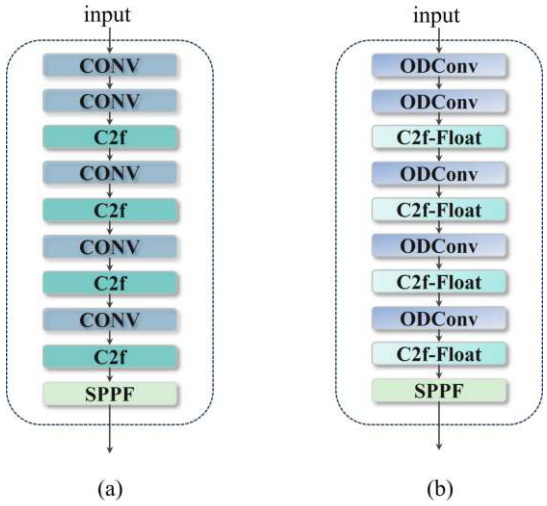

Fig. 6. Comparison diagram of backbone network structure. (a) Diagram of the original backbone network structure; (b) Diagram of the improved backbone network structure, where we made the corresponding block substitutions.

When applied to the backbone network for water surface float detection, ODConv demonstrates superior adaptability in extracting feature information from floats of various sizes and shapes. The full-dimensional dynamic convolution significantly enhances the feature extraction capability for water surface floats with only a minimal increase in parameters. Therefore, in water surface float detection, dynamic convolution offers notable advantages over traditional convolution.

C. Design of the GS-EVC Module

As floating objects are mostly irregular in size and shape, and the image features of some water surface floating objects scattered in the image corner area are not obvious, the problem of missed detection will occur. To detect all the floating objects present in the water surface environment completely, and at the same time, to be able to better adapt to the detection needs of small target floating objects, this paper designs a new GS-EVC (Explicit Visual Center) module and introduces it into the improved network. Its structure before and after improvement is shown in Fig. 7. Where (a) is the structure diagram of the original EVC module and (b) is the structure diagram of the improved GSConv. In this paper, GSConv is used to replace DWConv in the original EVC, while the shuffle operation is added after the original EVC splicing operation.

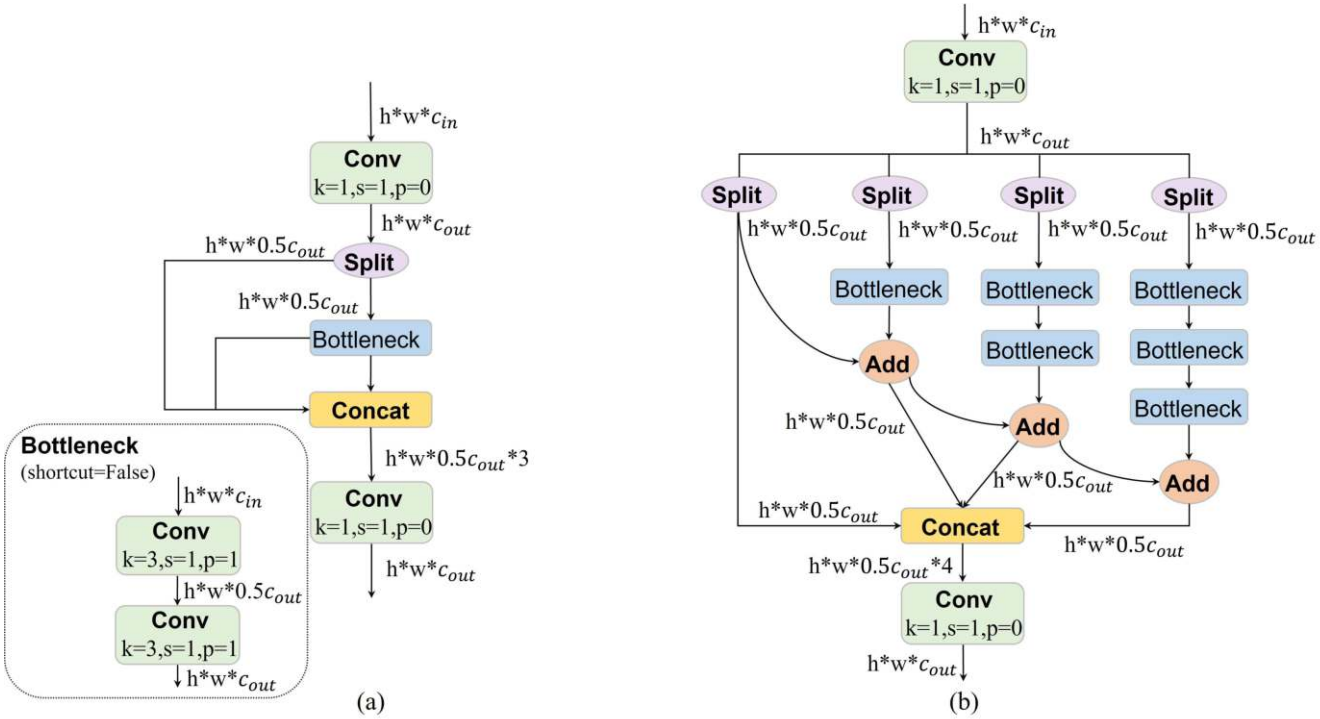

Fig. 5. Comparison diagrams of C2f module structures. (a) Structure of the original C2f module; (b) Structure of the improved C2f-float module.

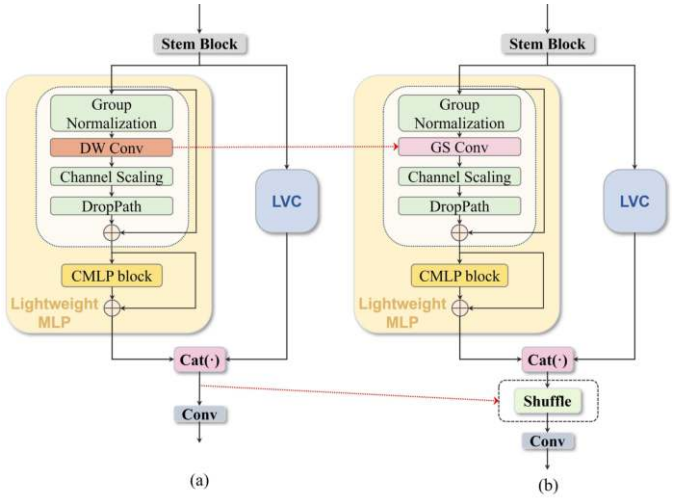

Fig. 7. Comparison diagram of EVC module structure. (a) Structure of the original EVC module; (b) Structure of our GS-EVC.

With almost no increase in computation, DWConv in the original EVC module is replaced by the better-performance GSConv. Compared with DWConv, GSConv can exchange and fuse the feature information of small target floats more uniformly, avoiding the loss of detailed feature information, and thus recovering the accuracy lost in DWConv for the sake of speed improvement. Besides, in GS-EVC, after the lightweight MLP and LVC complete the feature splicing on the channel, this paper adds the Shuffle operation to enhance the mixing and redistribution of feature information, enrich the feature expression, and avoid the problem of insufficient and unbalanced feature information fusion. The improved GS-EVC can be better adapted to the detection of small targets in the complex water surface environment compared with the original EVC module.

D. Overall algorithm structure

When using YOLOv8 for surface floating object detection in complex water environments, the complexity of the surface environment and floating objects often leads to insufficient feature information extraction and low feature recognition ability of the network for small target floating objects. Aiming at such problems, we designed a new surface floating object detection algorithm USV-YOLO based on the YOLOv8 network, and the overall network structure is shown in Fig. 8.

In the overall network design, for the input images of water surface floaters, the primary feature information of the network is firstly extracted using the ODConv and C2f-float modules, the four-dimensional dynamically weighted attention mechanism in ODConv is more adaptive to the target, and the pyramid structure in C2f-float enhances the utilization rate of the output feature layer of each bottleneck. The integration of ODConv and C2f-float in the backbone enhances the overall feature extraction capacity of the backbone network. In the neck section, alongside the C2f-float module, the GS-EVC module is incorporated to further elevate the detection precision for small floating objects. The GSConv and shuffle operations in the GS-EVC module help capture long-range dependencies between floater pixels, while also improving the diversity of feature representations. This alleviates the issue of low feature utilization caused by insufficient fusion of feature layers, which often leads to the loss of original feature information. Finally, the feature layers output from the neck, which contain complete information about the water surface floaters, are passed to the detection head to produce the final floater detection results.

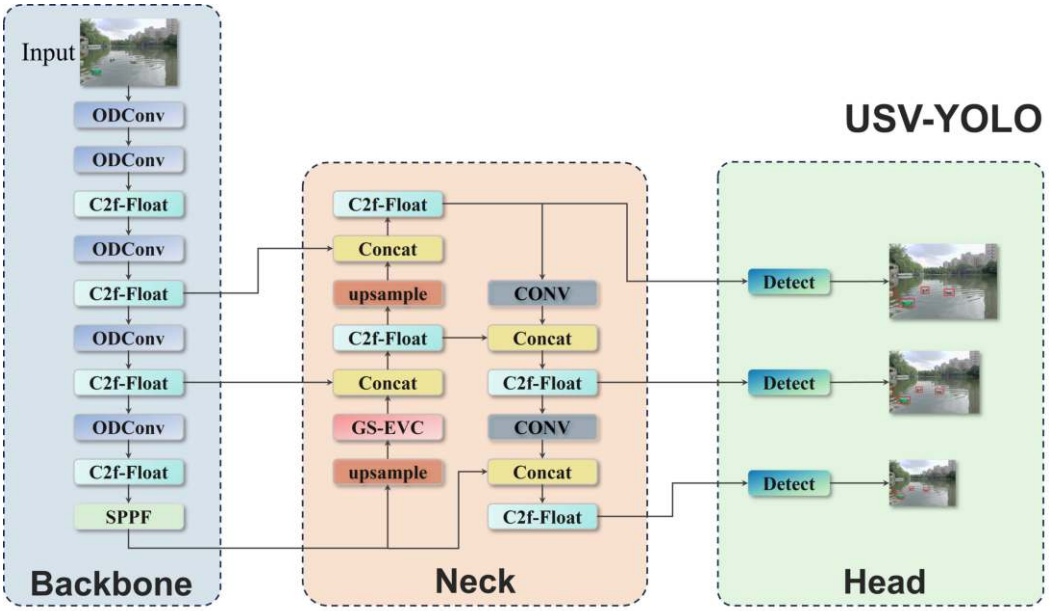

Fig. 8. USV-YOLO network overall structure diagram

IV. Experimental results and analysis

A. Experimental dataset

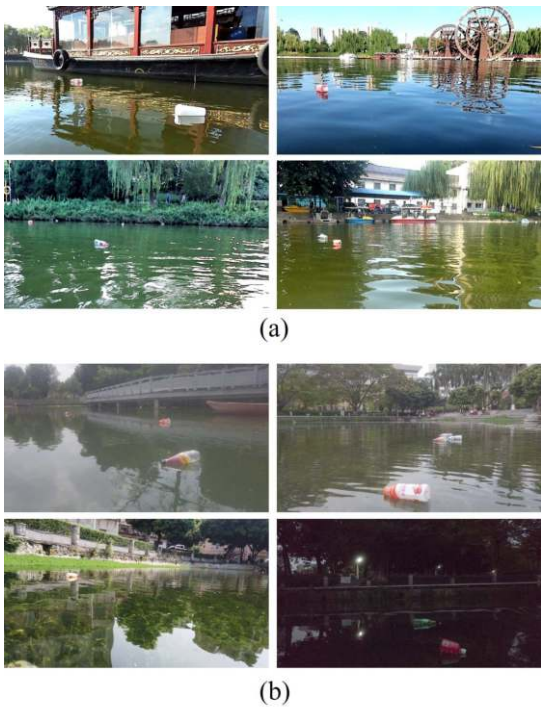

Fig. 9. Example datasets. (a) Example Flow-IMG dataset; (b) Example Floating Waste-I dataset.

In this paper, the methods used are trained and evaluated using the publicly available datasets FloW-IMG and FloatingWaste-I. Both FloW-IMG [25] and FloatingWaste-I [26] are datasets for water surface floating debris detection from the viewpoint of an unmanned boat in an inland waterway scenario, as shown in Fig. 9.

Among them, FloW-IMG is released by Ouka Smart Hublot in 2021, which contains 2,000 images of floating waste, the presence of the scene is mostly a sunny day with sufficient lighting conditions, and the number of targets contained in a frame ranges from 1 to 17, and most of them are small targets (which occupy less than 32×32 pixels); FloatingWaste-I contains 1867 images of floating objects, and this dataset contains more water scenes in complex lighting, including sunny, cloudy, rainy, and nighttime. The final dataset we use contains a total of 3800 float images from FloW-IMG and FloatingWaste-I, which are uniformly labeled into a float category. To fulfill the experimental requirements, in this paper, the 3800 datasets are divided into training, validation, and test sets in the ratio of 6:2:2.

B. Experimental environment and parameter settings

The experimental platform environment as well as the hyperparameter settings in this paper are shown in Table I:

TABLE I
Experimental platform environment and hyperparameter settings

	Designation	Versions/parameters
Experimental environment	Operating System	Windows 11
	GPU	NVIDIA RTX3060
	CPU	i5-12490F
	RAM	16G
	framework	Pytorch
	CUDA version	11.4
	Python	3.9.0
	Batch size	16
	Epoch	1000

C. Evaluation indicators

In this paper, we mainly use the common metrics in target detection models: mean accuracy (mAP), recall, detection precision, and model computation (GFLOPs) to evaluate the models. mAP is used to evaluate the detection accuracy of the model, and its calculation process is shown in Eq. 10,

where N is the number of detection categories, and AP_i denotes the area enclosed under the PR-curve consisting of the precision rate P as the horizontal axis and the recall rate R as the vertical axis. mAP_{50} and mAP_{50-95} represent the average accuracies when the threshold IoU is set to 0.5 and 0.5-0.95, respectively.

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

The recall indicates the proportion of actual positive samples that are correctly identified as positive, and the calculation process is as shown in Equation 11. Precision reflects the ratio of correctly identified positive instances to the total number of predicted positive instances, and the calculation process is as shown in Equation 12.

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$P = \frac{TP}{TP + FP} \tag{12}$$

where TP denotes the number of correctly detected floats, FN denotes the actual number of unpredicted floats, and FP denotes the number of incorrectly detected floats.

D. Analysis of experimental results

1) Feasibility experiment results and analysis: To verify the effectiveness of our water surface floating object detection algorithm, we conducted feasibility validation experiments. Both the original YOLOv8 network and its enhanced version were successively evaluated on this paper's dataset to verify the algorithm's effectiveness and assess its practicality. The evaluation metrics used to analyze the experimental results are mean accuracy (mAP), detection accuracy (precision), and recall, respectively.

As can be seen from Table II, the average precision and detection accuracy are significantly improved after adding the improved C2f-float module and GS-EVC module to the original network, respectively. After adding them to the original network at the same time, their mAP_{50} and mAP_{50-95} are improved by 3.5% and 4.3%, respectively. The detection accuracy is improved by 1.6%. After adding ODConv to the backbone network, it can further improve the detection accuracy by a small margin on the previous basis. As obtained from Table II, the algorithm in this paper significantly improves the average precision, detection accuracy, and recall compared with the original network, with the average precision mAP_{50} and mAP_{50-95} improved by 4.3% and 6.1%, respectively, and the detection accuracy as well as recall improved by 2.1% and 5.9%, respectively, compared with the original network. The experiment demonstrates that the algorithm presented in this paper holds certain merits for the water surface floating object detection network.

The visualization of the results of water surface float detection by the original YOLOv8 network and our USV-YOLO network is shown in Fig. 10. Comparing (a), (b), (c), and (d), our algorithm can detect small-target water surface floaters missed by the original network, and it can be seen that our algorithm extracts feature information of small-target floaters with inconspicuous features in the corner area more abundantly, and detects them better.

To verify the robustness of our algorithm, we use images of floating objects on the water surface taken under different illumination levels for verification, including cloudy days, dusk, evening, sunny days, etc. The verification results are shown in Fig. 11. Our algorithm maintains strong robustness in complex scenarios and is capable of fully detecting floating objects on water surfaces.

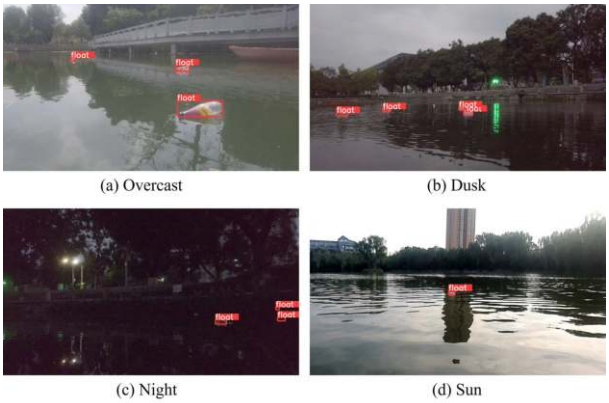

Fig. 11. Plot of experimental results under different light

2) Comparative experimental results and analysis: To further validate the performance of the algorithm, we conduct comparative experiments using the dataset of this paper with other classical target detection networks under the same experimental environment and analyze the experimental results comparatively. The results of the comparison experiments are shown in Table III. The comparison metrics are mean accuracy (mAP), detection accuracy (precision), and model computation (GFLOPs) when the threshold IoU is 0.5 and 0.5-0.95, respectively.

TABLE III
COMPARISON OF EXPERIMENTAL RESULTS

Network	$mAP_{50}(\%)$	precision(%)	GFLOPs
YOLOv5s	81.0	80.0	16.0
YOLOv5m	82.1	83.4	40.5
SSD	70.2	78.7	273.7
Faster R-CNN	73.5	83.9	947
YOLOv7-tiny [27]	77.6	84.5	13.0
YOLOv8n	82.9	88.0	**8.1**
USV-YOLO(our)	**87.2**	**90.1**	34.5

Through Table III, it can be seen that the average accuracy of mAP_{50} and mAP_{50-95} of our algorithm reaches 87.2% and 56.8%, respectively, and the detection accuracy reaches 90.1%, which is improved by 4.3%, 6.1%, and 2.1%, respectively, compared with the original YOLOv8. Contrasted with prevailing target detection networks, our water surface floating object detection algorithm has notable advantages. It meets the detection needs of unmanned boats in inland waterways, largely due to its enhanced network performance.

V. CONCLUSION

This paper introduces a surface-floating object detection network, USV-YOLO, built upon YOLOv8, aiming to resolve the problems of missed and incorrect detections when

TABLE II
COMPARISON OF FEASIBILITY EXPERIMENT RESULTS

Model	$mAP_{50}(\%)$	$mAP_{50-95}(\%)$	precision(%)	recall(%)
YOLOv8	82.9	50.7	88.0	68.9
YOLOv8+C2f-float	84.5	54.5	89.3	68.5
YOLOv8+EVC	83.5	52.8	88.3	69.2
YOLOv8+GS-EVC	84.9	53.7	87.8	70.1
YOLOv8+C2f-float +GS-EVC	86.4	55.0	89.6	73.2
YOLOv8+ ODConv +C2f-float +GS-EVC(our)	87.2	56.8	90.1	74.8

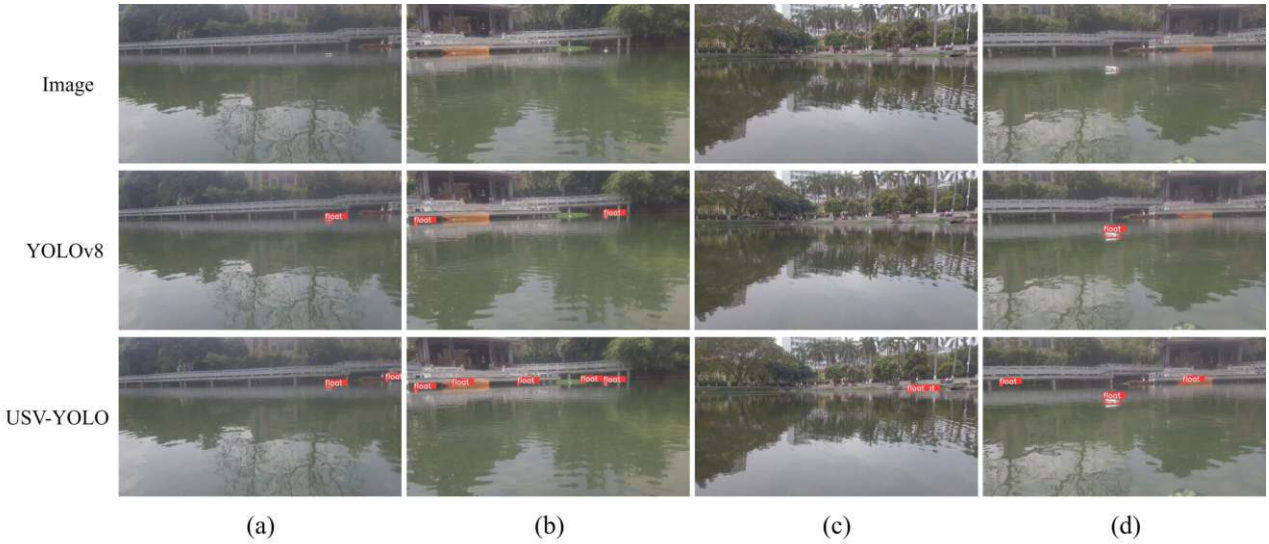

Fig. 10. Comparison chart of network visualization effect

identifying small floating targets on water surfaces in the intricate setting of inland waterways. First, we introduce a new C2f-float module that enhances feature utilization by fusing and concatenating the bottleneck layer across each channel sequentially. Additionally, we design a GS-EVC module that strengthens the dependency relationship among small target features and ensures that corner area information is emphasized during global feature extraction. Finally, we replace the standard convolution in the backbone network with a full-dimensional dynamic ODConv, which offers better adaptability for small target features. This alteration further boosts the feature extraction capacity and elevates the network's overall performance. The experimental outcomes show that our algorithm surpasses the original network and other object detection models in detection accuracy. However, the algorithm still has some limitations, particularly in detecting certain rare and complex types of floating objects in specialized environments, such as lakes or rivers. Therefore, in subsequent research, we intend to enhance the model's capabilities further to tackle these issues.

REFERENCES

[1] T. Kataoka, H. Hinata, and S. Kako, "A new technique for detecting colored macro plastic debris on beaches using webcam images and CIELUV," *Marine Pollution Bulletin*, vol. 64, no. 9, pp. 1829-1836, 2012.

[2] I. Ali, J. Mille, and L. Tougne, "Wood detection and tracking in videos of river," *Image Analysis: 17th Scandinavian Conference, SCIA 2011, Ystad, Sweden, May 2011. Proceedings 17*, pp. 646-655, 2011.

[3] X. Jin, P. Niu, and L. Liu, "A GMM-based Segmentation Method for the Detection of Water Surface Floats," *IEEE Access*, vol. 7, pp. 119018-119025, 2019.

[4] V.T. Hoang, V.D. Hoang, and K.H. Jo, "Realtime Multi-Person Pose Estimation with RCNN and Depthwise Separable Convolution," *2020 RIVF International Conference on Computing and Communication Technologies (RIVF)*, pp. 1-5, 2020.

[5] W. Zhang, S. Wang, S. Thachan, J. Chen, and Y. Qian, "Deconv R-CNN for small object detection on remote sensing images," *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2483-2486, 2018.

[6] R. Girshick, "Fast r-cnn," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.

[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "Ssd: Single shot multibox detector," *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pp. 21-37, 2016.

[8] R. Chen, Y. Peng, J. Wu, W. Ouyang, Y. Li, and T. Yue, "Deep Learning-based Intelligent Detection of Floating Objects on Water Surface," *Engineering Science and Technology*, vol. 55, no. 3, pp. 165-174, 2023.

[9] H. Li, S. Yang, J. Liu, Y. Yang, M. Kadoch, and T. Liu, "A framework and method for surface floating object detection based on 6g networks," *Electronics*, vol. 11, no. 18, pp. 2939, 2022.

[10] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, and P. Pang, "Searching for mobilenetv3," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1314-1324, 2019.

[11] Z. Yi, D. Yao, G. Li, J. Ai, and W. Xie, "Detection and localization for lake floating objects based on CA-faster R-CNN," *Multimedia Tools and Applications*, vol. 81, no. 12, pp. 17263-17281, 2022.

[12] F. Chen, L. Zhang, S. Kang, L. Chen, H. Dong, D. Li, and X. Wu, "Soft-NMS-enabled YOLOv5 with SIOU for small water surface floater detection in UAV-captured images," *Sustainability*, vol. 15, no. 14, pp. 10751, 2023.

[13] G. Qiao, M. Yang, and H. Wang, "A detection approach for floating

debris using ground images based on deep learning," *Remote Sensing*, vol. 14, no. 17, pp. 4161, 2022.
[14] H. Li, S. Yang, R. Zhang, P. Yu, Z. Fu, X. Wang, M. Kadoch, and Y. Yang, "Detection of floating objects on water surface using YOLOv5s in an Edge computing environment," *Water*, vol. 16, no. 1, pp. 86, 2023.
[15] Y. Zhang and Y. Chen, "Improved Small Target Detection Algorithm on Water Surface for YOLOv8," *Computer System Applications*, vol. 33, no. 4, pp. 152-161, 2024.
[16] H. Nie, H. Pang, M. Ma, and R. Zheng, "A Lightweight Remote Sensing Small Target Image Detection Algorithm Based on Improved YOLOv8," *Sensors*, vol. 24, no. 9, pp. 2952, 2024.
[17] Z. Zeng, Y. Xu, and J. Wang, "SOE-YOLO lightweight-based surface target detection algorithm," *Journal of Graphics*, pp. 1-9, 2024.
[18] Y. Cheng, M. Jiang, J. Zhu, and Y. Liu, "Are We Ready for Unmanned Surface Vehicles in Inland Waterways The USVInland Multisensor Dataset and Benchmark," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 3964-3970, 2021.
[19] H. Li, J. Li, H. Wei, Z. Liu, Z. Zhan, and Q. Ren, "Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles," *arXiv preprint arXiv:2206.02424*, 2022.
[20] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6848-6856, 2018.
[21] C. Li, A. Zhou, and A. Yao, "Omni-dimensional dynamic convolution," *arXiv preprint arXiv:2209.07947*, 2022.
[22] Y. Quan, D. Zhang, L. Zhang, and J. Tang, "Centralized feature pyramid for object detection," *IEEE Transactions on Image Processing*, 2023.
[23] I. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit, and others, "Mlp-mixer: An all-mlp architecture for vision," *Advances in Neural Information Processing Systems*, vol. 34, pp. 24261-24272, 2021.
[24] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251-1258, 2017.
[25] Y. Cheng, J. Zhu, M. Jiang, J. Fu, C. Pang, P. Wang, K. Sankaran, O. Onabola, Y. Liu, D. Liu, and others, "Flow: A dataset and benchmark for floating waste detection in inland waters," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10953-10962, 2021.
[26] 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*, vol. 11, no. 12, pp. 2234, 2023.
[27] C. Wang, A. Bochkovskiy, and H. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7464-7475, 2023.