MARITIME TARGET DETECTION FOR UNMANNED SURFACE VEHICLES BASED ON LIGHTWEIGHT NETWORKS UNDER FOGGY WEATHER

Shuyue Li,* Junjie Wang,* Jinlu Sheng,** Ziyu Liu,*** Shixin Li,**** and Ying Cui****

Abstract

Maritime target detection is a critical component of navigation safety for unmanned surface vehicles (USVs), particularly under foggy weather conditions. We propose an efficient and lightweight method for maritime target detection suitable for foggy weather conditions. This approach aims to address the high cost of data acquisition and to enhance the target detection effectiveness. The proposed method involves several steps to enhance the detection effectiveness and efficiency. Firstly, we improved the accuracy of foggy image synthesis by formulating a more realistic loss function for the GAN model. Secondly, we reduced the model size and number of parameters by introducing depthwise-separable convolution instead of conventional convolution. Finally, we applied a lightweight backbone to improve the high-dimensional maritime target features and accelerate the inference speed. The experimental results demonstrate significant improvement in the accuracy and efficiency of the proposed model. Our model achieved an average accuracy of 89%, which is a significant improvement over the 77% accuracy of the original YOLOv4 model. Additionally, the computational volume of our model was reduced by 83%, and the real-time detection speed reached 45.8 frames per second. This improved accuracy and efficiency make the proposed method more appropriate for complex conditions and enhance the safety of USV navigation.

Key Words

Unmanned surface vehicle, target detection, deep learning, lightweight network

1. Introduction

In the past several years, applications relying on visual systems are becoming more and more popular, such as the automatic detection of objects and the elimination of obstacles in general, both indoors and outdoors, the technique used can also be adapted to the navigation of marine vehicles under dynamic environments. However, the system must be modified so that it can be adjusted to various situations, including changes in the illumination caused by the sun and clouds, the reflections of the sun, the glaring, the partial or complete closure of the system caused by rain or the fog, and other ocean conditions. As an essential intelligent equipment in the emerging marine industry, unmanned surface vehicle (USV) plays a vital role in exploration and research. At present, the vision detection of targets is an important method of perception for the smart navigation of USV, consequently, improving the accuracy and speed of the detection of surface targets has been a major research area for the USV.

Traditional maritime target detection methods, such as mean shift [42], deformable models [1], and support vector machines [45], have been proposed to detect maritime targets, however, they suffer from false detection defects, such as water reflection and wave occlusion. To improve the detection accuracy, deep learning-based methods have been widely used for maritime target detection. Two-stage methods achieve detection and classification by extracting candidate regions and performing with high accuracy, including the regional convolutional neural network (R-CNN), spatial pyramidal ensemble network (SPPNet), and Faster-RCNN [31]. One-stage methods directly calculate the coordinates and category probability of the target and provide results after one detection, which significantly improves the detection speed, including single-time detector (SSD) [21] and YOLOv1-v4 [3], [9], [28]–[30], [39] series methods. Many studies on the detection of maritime targets focus on a single detection stage [4], [5], [11], [13], [19], but the majority of these models are based on static surveillance camera systems, which have difficulties to cope with the speed, the dynamic operation view, and the complex conditions of the USV.

Several researchers are using mathematical methods to simulate fog in clear images, relying on the physical model of this phenomenon and further analysis of the

* Ocean University of China, Qingdao, China; e-mail: shuyue.li@stu.ouc.edu.cn; wjj@ouc.edu.cn
** Chongqing Jiao Tong University, Chongqing, China; e-mail: forwardlulu@163.com
*** Shanghai Marine Electronic Equipment Research Institute, Shanghai, China; e-mail: liuziyu726@outlook.com
**** Qingdao University of Science and Technology, Qingdao, China; e-mail: seawolf_lsx@163.com; yingcui@qust.edu.cn
Corresponding author: Ying Cui

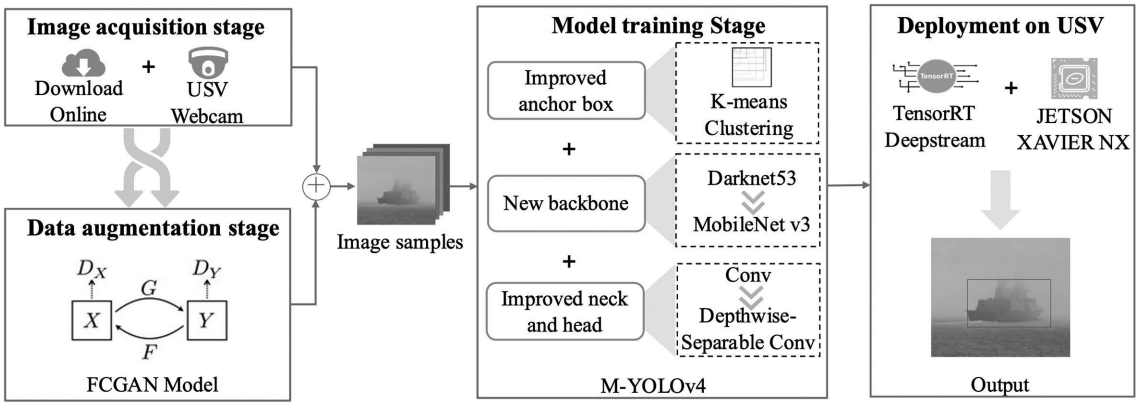


Figure 1. Schematic of the image acquisition system, data augmentation, training stages, and deployment on USV.

image content. However, by employing convolutional neural networks, we can significantly reduce the need for human effort in identifying the appropriate 'fogification' formulas and instead allow our model to determine the optimal approach.

In order to reduce the high rate of false detection in fog, traditional methods of defofogation based upon gammatic correction [18] and *a posteriori* models [17] can be effective in a single frame image, but they increase the computational effort on the edge and result in large delay. Several methods have attempted to simulate the behaviour of fog and superimpose it onto clear images to enhance object detection and recognition systems in adverse weather conditions [33], [34]. These approaches primarily concentrate on the mathematical models that describe the physics behind fog generation and aim to apply it to a fog-free photograph. This process is useful in generating a new set of annotated foggy images from clear ones, which can be used as a training dataset for object detection systems, improving their accuracy by relying on artificially hazy images. The generated model will subsequently be applied to real-time photo frames captured during foggy weather conditions. Therefore, it is crucial for the synthesised images to resemble real foggy ones as closely as possible. This is because better results can be achieved when the trained model is fed with actual foggy images.

Hence, to enhance the target detection and recognition system in adverse weather conditions, some experts resort to simulating and learning image features, such as fog contrast and saturation, using unsupervised learning models to develop synthetic fog datasets for training purposes [33], [34]. However, the synthetic dataset created using this network may not incorporate sea fog features, leaving ample scope for improvement in reducing sea fog characteristics.

Currently, the USV maritime target detection faces two significant challenges. Firstly, due to the limited computational power of the edge computing systems on USVs, large target detection networks cannot be deployed. Secondly, foggy conditions are prevalent at sea, and models trained on high-quality datasets have a high misjudgement rate under cloudy skies, making it challenging to ensure navigation safety. Moreover, obtaining and screening sea fog images suitable for model training poses a significant challenge, hindering the USV target detection process.

To balance the maritime target detection accuracy and speed requirements of USVs and improve the airworthiness in fog, we propose an improved YOLOv4 target detection method. In the data acquisition stage, the sea condition images acquired using a USV and the surface target images acquired online are integrated into a base dataset. Thereafter, the data augmentation method of the foggy cycle consistent generation adversarial network (FCGAN) is used to synthesise a new dataset with different maritime fog intensities. In the model training stage, we adopt the K-means clustering algorithm to extract special features of maritime targets, a depthwise-separable convolution operation to reduce the number of model parameters, and the lighter MobileNetv3 as the backbone network of YOLOv4 to reduce model space complexity. In the model deployment stage, we use TensorRT and Deepstream to increase the speed of inference, reduce latency, and improve the real-time performance of USV detection, as shown in Fig. 1.

The rest of this paper is organised as follows: Section 2 briefly summarises haze removal methods and lightweight surface target detection methods; Section 3 introduces the FCGAN data enhancement and YOLOv4 lightweight improvement methods proposed in this study; Section 4 briefly introduces the USV and data set contents; Section 5 analyses the effect of the sea fog concentration on the detection generalisation performance under foggy weather, compares different lightweight improvement schemes, and verifies the effectiveness of the proposed model; finally, Section 6 provides conclusions and the future scope.

2. Related Work

2.1 Unsupervised Learning-Based Fog Image Synthesis Technique

Unsupervised learning-based methods often use generative adversarial network (GAN) models [10] to learnt image features, such as colour contrast and saturation. Generally, all GAN models comprise a generator and a discriminator.

The generator generates clear images from fogged images learning a set of image features with and without fog (*i.e.*, paired data) [26], [44]. However, the paired images of real sea scenes are difficult to obtain, it is time consuming and labour-intensive to instal fixed cameras and capture videos under foggy conditions at the same location. Therefore, a cycle-consistent adversarial networks (CycleGAN) [50], which does not need to establish a one-to-one mapping relationship between training data, can realise the conversion of two different styles of images. Using unpaired data has contributed to image synthesis, based on CycleGAN, Engin *et al.* [8] proposed the cycle-dehaze method, which uses VGG16 as the backbone network, introduces a perceptual loss function, and compares the original and final images in feature instead of pixel to synthesise fog images. Similarly, Liu *et al.* [22] proposed the Cycle-Defog2Refog method, where the main framework also uses CycleGAN, while the backbone network uses a CNN to remove artefacts and improve defog quality, using the average pixel value of atmospheric light for priori estimation and generating fog images using the atmospheric degradation model. Although the above method using few data samples to a certain extent, it does not address the problem of learning confusion caused by the bi-directional learning framework; furthermore, it does not sufficiently address the importance of global features and leaves the synthesised sea fog dataset with defects, such as poor realism, instability, and low pixels.

2.2 Maritime Surface Target Detection Method

The results obtained from the current high-accuracy target detection models are computed on large servers, many parameters in the network architecture cause severe redundancy. To solve the above problems, more lightweight networks, such as Inceptionv1-v4 [16], [35]–[37], ShuffleNet [25], [46], and Inceptionv 3, apply a batch normalisation layer to increase the speed of gradient descent. ShuffleNet uses group convolution and channel shuffling to reduce computational complexity, which helps to reduce computational requirements. EfficientNet uses a multi-dimensional hybrid model deflation approach to achieve a faster detection algorithm by combining network depth, network width, and image resolution. Zhu and Chen [49] proposed an algorithm to detect moving object in a real-time environment with robustness, such as illumination change, high-speed motion, object occulsion, object distortion, and noisy object. Liu *et al.* improved YOLOv3 which is better than the Faster-RCNN, single shot multibox detector and YOLOv3. In order to solve the camera-based vehicle detection under low illumination, Li *et al.* [24] simulated the information processing and partition functions of its pathway, constructed a vision information processing model, and designed a computing methods for recognition. Chen *et al.* [43] used deep convolution GANs to augment the data and transfer learning method to avoid training from scratch. Experiments show that the improved VGG16 network outperforms the traditional models. Zhou and Yang [48] investigated one-stage and two-stage object detection algorithms on drone-captured images and YOLOv5 shows the best performance on their data. Jiang *et al.* [7] proposed a network that adopts an encoder-decoder structure and introduced the efficient channel attention method. This network leveraged the importance of crack pixels, leading to a lower computation cost. Meanwhile, the superposition of max-pooling and mean-pooling enables the extraction of more features. The lightweight convolution models reduce the computation cost, while the superposition of max-pooling and mean-pooling enables the extraction of more minutiae pixels. Zhang *et al.* [47] used the ENF model to fuse the low-level feature maps several times and then use the neural architecture search technique to automatically search for the most suitable feature extraction network, the improved two-stage model overcomes high missed detection rate.

With the help of lightweight network improvement, Hu *et al.* [14] replaced the backbone network of YOLOv3 with MobileNetv2 to implement a fish morphology detection system. Hu *et al.* [15] used DenseNet to improve the CSPdarknet of YOLOv4 to reduce the loss of the feature layer information in high-dimension features. Li *et al.* [23] performed channel pruning for YOLOv3 to improve the speed of target detection algorithms running in real time on UAVs. Ouyang *et al.* [27] used a lightweight backbone network, including an hourglass convolutional feature extraction module and a parallel expansion convolutional module, for the adaptation of the system to different target sizes. Huang *et al.* [13] replaced YOLOv3's darknet53 with darknet19 to classify ships with a lightweight feature extraction layer to improve the operation speed of maritime surveillance. However, the improved algorithms above still cannot directly apply to the actual sea trials with high integration and energy consumption.

3. Methodology

3.1 Foggy Cycle Generation Adversarial Network for Data Enhancement

Synthesising specific characteristics of sea fog directly using the cycleGAN can result in unrealistic or inaccurate images, as the model may not fully capture the complex and dynamic nature of sea fog. Additionally, the synthesised images may not accurately reflect the diversity of real-world sea fog conditions, which can pose a challenge for the model's generalisation ability. This can lead to reduced performance when detecting maritime targets under different scenarios or when applied to other datasets. To overcome this problem, we propose a more realistic loss function to synthesise the foggy images for the cycleGAN model. This loss function enables the generation of synthetic images that better resemble the specific characteristics of sea fog, such as reduced visibility and scattered light. The proposed loss function is designed to encourage the preservation of low-frequency details and colour distribution, while also reducing artefacts and image blurring caused by the fog. To synthesise more realistic sea fog pictures, we designed two sets of loops, one is fog–clear–fog cycle, another is clear–fog–clear cycle, as shown in Fig. 2. These were drawn on the CycleGAN model to establish a feedback link between

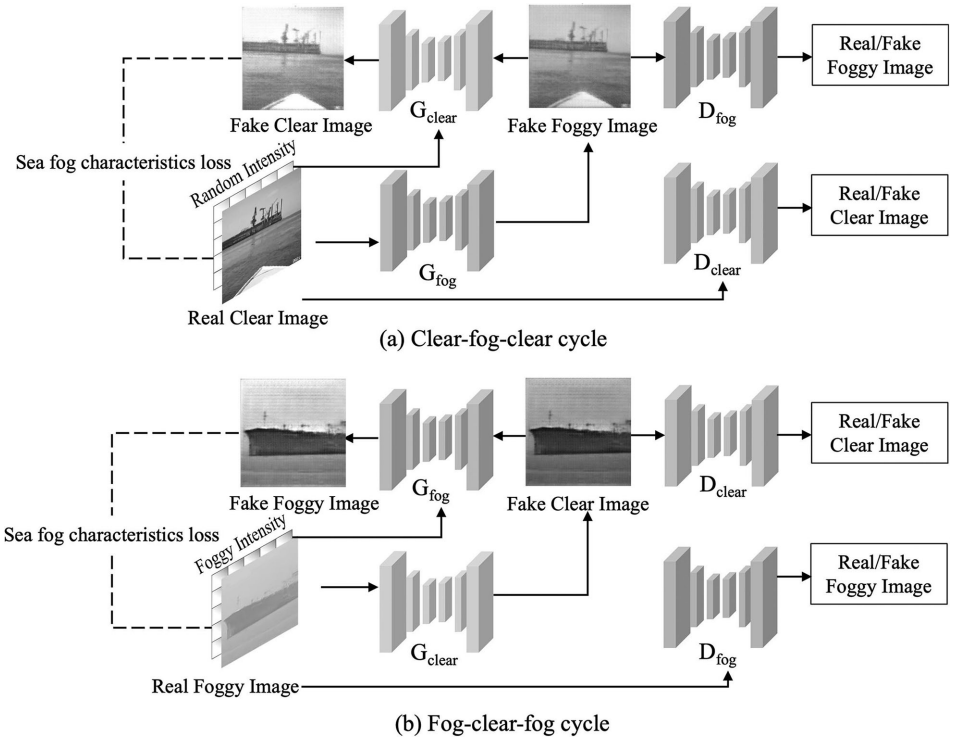


Figure 2. FCGAN structure: (a) clear-fog-clear cycle and .(b) fog-clear-fog cycle.

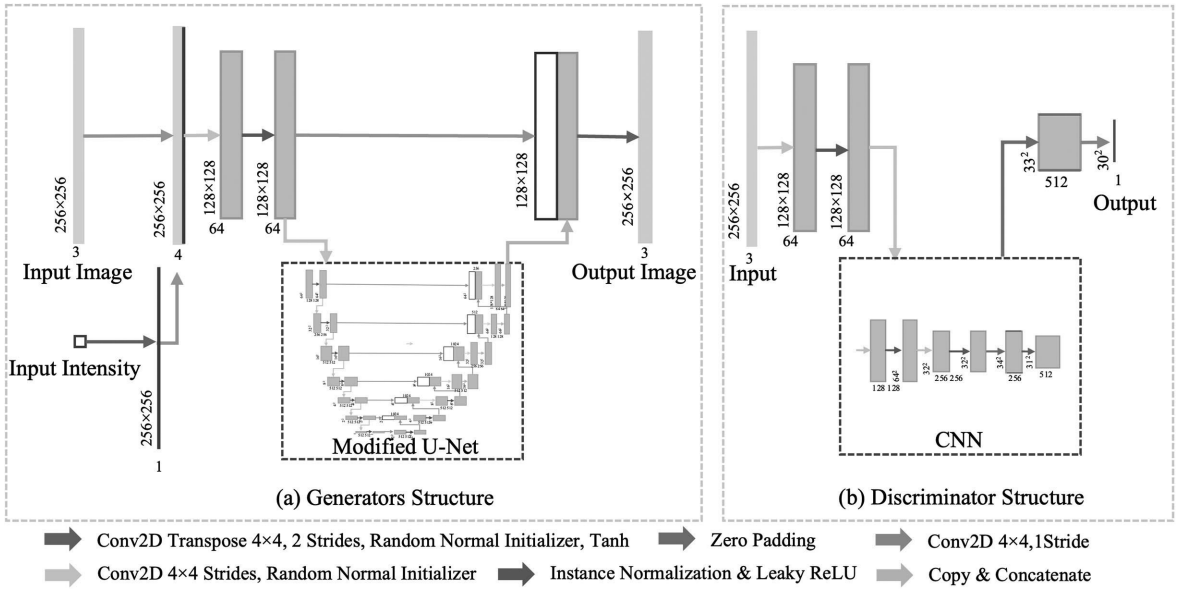


Figure 3. Generator and discriminator structures.

the two loops. Inside the single loop, *via* the adversarial training between the discriminator and the generator, the pixel distribution of the dataset is learnt to generate new images, the learning effect is characterised using the generative adversarial loss function. In addition, to learn global features, we innovatively propose a sea fog characteristics loss function, the fog intensity value is introduced as the fourth channel of training input, it can establish a multi-mapping height constraint to synthesise a more realistic sea fog image. The new model is defined as the FCGAN. Taking the size of the dataset, the complexity of the model architecture, and the foggy task that the model is trained on, the learning rate of the FCGAN model in this article is 0.001.

The initial input values in Fig. 3 include a fog intensity tensor and a 256×256 color image. The real number I representing fog intensity is repeated 256×256 times as a fourth layer channel and added to the input image to

produce a 4-channel image (R, G, B, I). While the fog intensity i is defined as the ratio of the all-white to total pixel values of the image, as (1):

$$i = \frac{\text{white pixel}}{\text{all pixel}} \tag{1}$$

In FCGAN, each loop internally comprises two generators and discriminators, the defog generator G_{clear} and the add-fog generator G_{fog} are used to extract the sea fog characteristics in the image using a modified U-Net [32] structure and the residual connections between the downsampled outputs; furthermore, G_{clear} and G_{fog} are concatenated with the corresponding upsampled outputs, as shown in Fig. 3(a). The clear discriminator D_{clear} and the fog discriminator D_{fog} are responsible for distinguishing between the real and generated images. They comprise a CNN, as shown in Fig. 3(b), where the three downsampling modules produce a convolutional kernel of size 256, and after normalisation, convolution, and zero padding, a 30×30 convolutional kernel is generated. To enhance the constraint of fog intensity on the FCGAN model, a multi-objective optimised sea fog characteristics loss function is proposed, including identity loss, transmission map loss, whitening loss, and RGB ratio losses, as shown in (2):

$$L_{\text{fog}} = L_{\text{identity}} + L_{\text{trans}} + \alpha \cdot (L_{\text{white}} + L_{\text{rgb}}) \tag{2}$$

where α is the penalty factor. Under ideal conditions, the synthetic sea fog image generated by the add-fog generator G_{fog} should be identical to the real fogged image. Similarly, the synthetic clear image generated by the defog generator G_{clear} should be identical to the real clear image, and this loss is defined as identity loss, as shown in (3):

$$\begin{aligned} L_{\text{identity}} = {} & E_f\left[\left\|G_{\text{fog}}\left(f, i_f\right) - f\right\|_1\right. \\ & \left. + E_c\left[\left\|G_{\text{clear}}\left(c, i_c\right) - c\right\|_1\right.\right. \end{aligned} \tag{3}$$

where $c \in C$ refers to the set of clear image samples with concentration $i_c \in I$, $f \in F$ refers to the set of fogged image samples with concentration $i_f \in I$, $I = [0, 1]$, E_f is the expected value of fogged image samples; and E_c is the expected value of clear image samples. To ensure that the clear images can be uniformly distributed in the sea fog concentration i_c when generating the data set with fog images, the transmission map loss is defined in this study using an atmospheric degradation model, as shown in (4):

$$\begin{aligned} L_{\text{trans}} = {} & E_c\left|\text{mean}\left(G_{\text{fog}}\left(c, i_c\right)\right) - \text{mean}\left(c\left(1 - i_c\right) + i_c\right)\right| \\ & + E_c\left[\left\|G_{\text{fog}}(c, 0) - c\right\|_1 - \left\|G_{\text{fog}}(c, 1) - w\right\|_1\right] \end{aligned} \tag{4}$$

The whitening and RGB ratio losses are calculated using the ramp function (ReLU function), which ensures that the generated fogged images generate the correct output under the two extreme conditions of $i_c = 0$ and $i_c = 1$. In (6), r, g, and b represent the three channels of the input image c. Meanwhile, $\widehat{r}$, $\widehat{g}$, cand $\widehat{b}$ correspond to the three-channel values of the fogged images generated from the clear images.

$$L_{\text{white}} = E_c\left[\left\|\text{ReLU}\left(c - G_{\text{fog}}\left(c, i_c\right)\right)\right\|_1\right] \tag{5}$$

$$L_{\text{rgb}} = E_c\left[\left\| r \cdot \widehat{g} - g \cdot \widehat{r} \right\|_1 + \left\| g \cdot \widehat{b} - b \cdot \widehat{g} \right\|_1\right] \tag{6}$$

In summary, the loss function of FCGAN is defined as follows:

$$\begin{aligned} L_{\text{FCGAN}} = {} & L_{\text{GAN}}\left(G_{\text{fog}}, D_{\text{fog}}, C, F, I\right) + \\ & L_{\text{GAN}}\left(G_{\text{clear}}, D_{\text{clear}}, F, C, I\right) + \\ & \lambda L_{\text{cyc}}(C, F, I) + \lambda L_{\text{fog}} \end{aligned} \tag{7}$$

We use more advanced optimisation techniques, such as evolutionary strategies and gradient-based optimisation, to search for the optimal weight combinations for the sea fog loss function. By evaluating the performance of the model with different weight combinations, we identify the weight settings that lead to the best trade-off between different performance metrics, such as realism and artefact reduction. This approach can help to fine-tune the design of the sea fog loss function and improve the performance of the cycleGAN for synthesising realistic sea fog images. In this model, λ is the penalty factor, $\alpha = 10$, and $\lambda = 5$.

3.2 Maritime Target Detection Algorithm Based on M-YOLOv4

The original YOLOv4 network is mainly divided into three major parts: backbone, neck, and head, as shown in Fig. 4. Firstly, the backbone network uses the CSPDarknet53 module to extract preliminary features; secondly, the neck module is used for feature fusion including the path aggregation network (PANet) [20] structure and the spatial pyramid pooling (SPP) module; thirdly, the YOLO head module obtains prediction results with the help of the effective feature layers obtained from the first two parts. Although YOLOv4 adequately performs on small targets with multiscale prediction, the relatively complex CSPDarknet53 backbone network severely delayed the detection speed.

We propose a more lightweight network, as shown in Fig. 4. In the YOLO backbone section, we use the optimal anchor of the maritime target after K-means clustering and replace the CSPDarknet with MobileNetv3. The standard convolution (Conv) in the neck and head are replaced by the depthwise-separable convolution (DW-Conv), to ensure that the high-dimensional features are not lost and to reduce the number of parameters. The improved network is named M-YOLOv4, and the main improvement methods include the following two points.

(1) Anchors after K-means clustering

With the change of the number and proportion of feature layers, the proportion of anchors should accordingly adjusted together, the anchor size close to the maritime target size can reduce traversals number and ensure the accuracy of detection results. Therefore, we use the K-means clustering algorithm [2] to re-cluster the maritime target dataset and calculate the average intersection ratio of anchors when the number of clusters K is taken in the interval [2], [20]. The K-means algorithm is used to

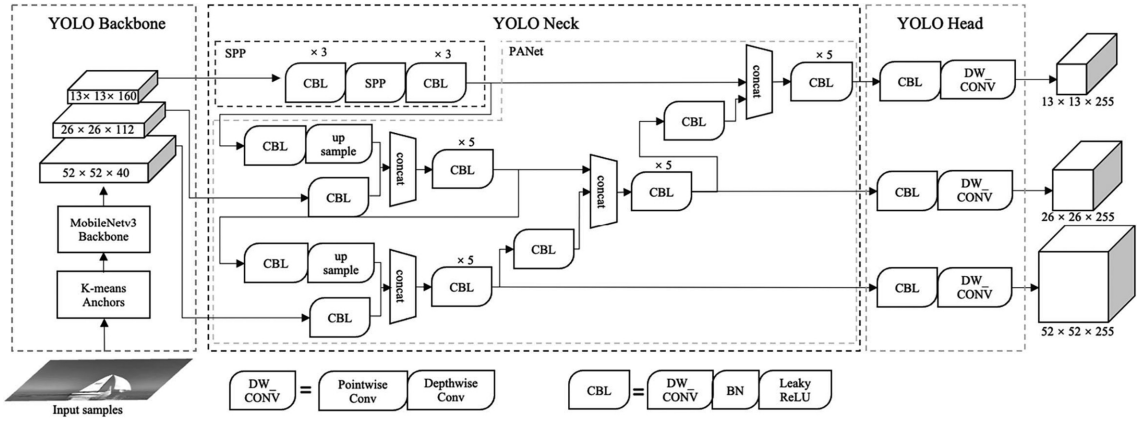


Figure 4. M-YOLOv4 structure.

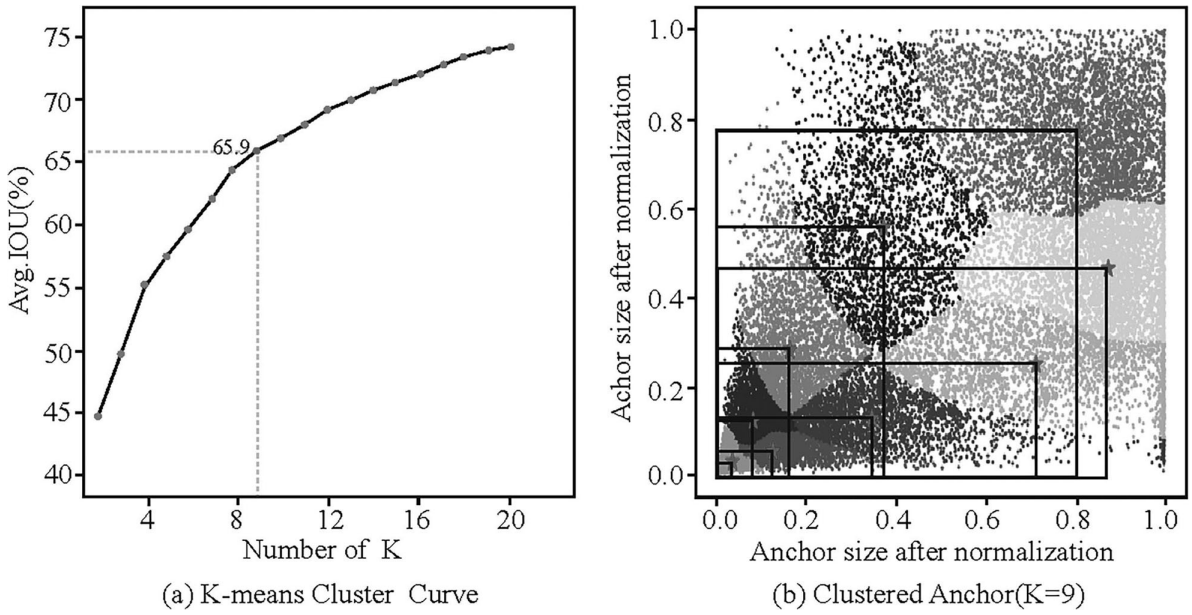


Figure 5. Clustered anchors.

cluster the bounding boxes based on their aspect ratio into K groups, where K is the number of anchor boxes specified in the YOLOv4 model. The multiple k-means sub-detectors first classify the feature space with the diversity of features or multi-subclassed into k clusters [41]. For each cluster, the centroid is calculated as the mean aspect ratio of all bounding boxes in the cluster. The final anchor boxes for YOLOv4 are then determined by sorting the centroids in descending order of their frequency and selecting the top K centroids as the anchor boxes. This ensures that the anchor boxes are representative of the most common object sizes and aspect ratios in the training set.

As shown in Fig. 5(a), the average intersection ratio is 65.9% when $K = 9$ and the curvature is maximum, and the best number of clusters $K^* = 9$, which can be determined using the elbow rule. The red pentagrams in Fig. 5(b) represent the centre point of each cluster, and the corresponding coordinate value is the optimal anchor under the three output feature layers. The optimal anchor value will be used as the initial anchor value for model training, which in turn makes the model easier to converge.

(2) Depthwise-separable convolution and MobileNetv3 backbone networks

The MobileNetv3 backbone network structure with extracted features is shown in Fig. 6(a). MobileNetv3 reduces the model size and the number of parameters mainly through the application of depthwise-separable convolution (DW-Conv) [6], which utilises the idea of group convolution and splits a standard convolution operation (Conv) into two steps, including depthwise convolution and pointwise convolution, as shown in Fig. 6(b) and (c). If the input size of the existing feature layer is D_K and the number of channels is M, an output of size D_F and N channels are required. The ratio of the computation of conventional convolution and depthwise-separable convolution Δ is shown in (8), where $\Delta \approx 1100$

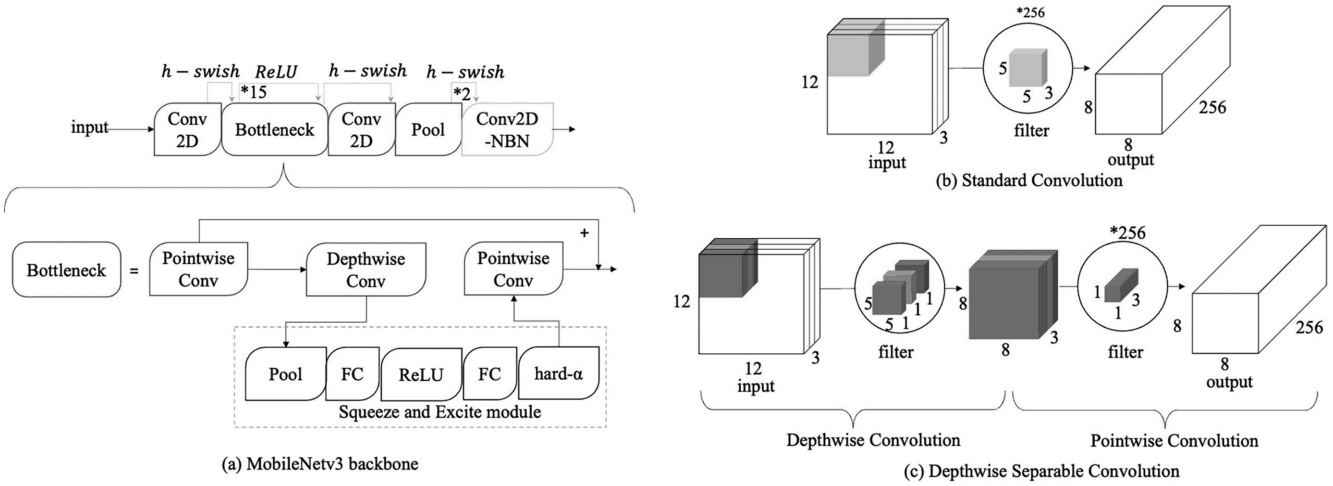


Figure 6. MobileNetv3 backbone and depthwise-separable convolution.

when $N = 256$ and $D_K = 12$. Therefore, using depthwise-separable convolution can significantly reduce computation time.

$$\Delta = \frac{DW-\text{Conv}}{\text{Conv}} \tag{8}$$

This model uses Complete Intersection over Union (CIoU) loss for learning and training the features of the maritime target, as is shown in (9) and (10):

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{9}$$

$$\text{L}_{\text{CIoU}} = 1 - \text{IoU}(\text{A}, \text{B}) + \frac{\rho^2\left(b, b^{\text{gt}}\right)}{r^2} + \beta \cdot \text{v} \tag{10}$$

where b and b^{gt} represent the centroids of the prediction and real frames, ρ represents the Euclidean distance between the two centroids; A and B represent the closure regions of the prediction and real frames, c represents the diagonal distance of the smallest closure region that can contain both the prediction and real frames; β is a positive trade-off parameter; and v is measured by calculating the candidate frame length and height. This loss function enhances the consistency of the aspect ratio, more effectively characterises loss, and avoids gradient explosion.

We used frames per second (FPS) and mean average accuracy (mAP) as the main evaluation metrics. mAP is the average of the AP of the surface target category, and AP refers to the area under the accuracy–recall (PR) curve, as is shown in (11) and (12):

$$\begin{aligned} \text{Precison} &= \frac{\text{TP}}{TP + FP}, \\ \text{Recall} &= \frac{\text{TP}}{TP + FN} \\ AP &= \int_0^1 \text{Precision} \times Recall, \end{aligned} \tag{11}$$

$$AP_{50:95} = \frac{1}{10}\left(AP_{50} + AP_{55} + AP_{95}\right) \tag{12}$$

Consider the category 'yacht' as an example, where true positive (TP) is the number of the predicted and true results of 'yacht' false positive (FP) is the number of the predicted and real results of 'yacht,' false negative (FN) is the number of the predicted and real results of 'yacht.' mAP@50 is the average value of AP when the IoU threshold is 50%, and mAP@50@95 is the average value of AP when the IoU value is taken from 50% to 95% in steps of 5%.

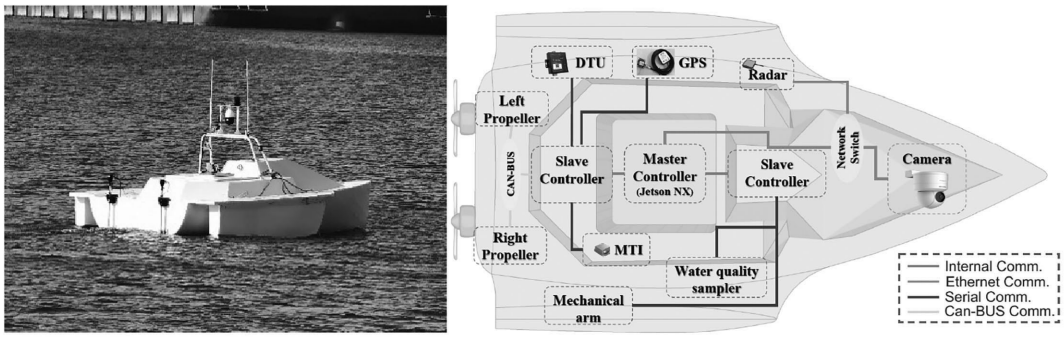
To solve the problem of the difficult deployment of YOLOv4 at the edge computing side of USVs, we replicated six lightweight networks, such as Inceptionv4, under the Pytorch framework, which were trained as the backbone network of YOLOv4 on the target dataset at sea to test the detection speed and accuracy. The computational results are shown in Table 1. At mAP@0.5 and mAP@0.5@0.95 metrics, CSPDarknet53 achieved the optimal accuracy performance of 0.91 and 0.83. In terms of detection speed, MobileNetv3 achieved a detection speed of 44.7 FPS, which is 32.8 FPS higher than that of CSPdarknet. Under the model size metric, MobileNetv3 achieved the optimal result of 123.6 MB, and on balance, MobileNetv3 was used as the backbone network for YOLOv4 in this study.

4. Platform and Basic Dataset

In this study, the acquisition of image data and the real-time detection of lightweight targets were based on the 6-m USV designed and manufactured by Qingdao Intelligent Yacht Manufacturing and Development Technology Engineering Laboratory. The image acquisition uses a 1920×1080 resolution front view camera (HIKVISON DS-2DC4423IW-D), and assuming the camera isn't obscured by water droplets. The communication of the camera, control chip, and other hardware devices is shown in Fig. 7(b). Jetson Xavier NX as the master controller is responsible for receiving the video stream and performing the calculation of the target detection algorithm, while the hull motion control is received by the slave controller with GPS and electronic compass information commands. This master–slave control design ensures the stability and safety of the USV when sailing to the maximum extent.

Table 1
Different Backbone Influence on YOLOv4

Backbone	mAP @0.5@0.95	Model Volume (MB)	FPS
Inceptionv4 [16]	0.69	418.2	12.3
ShuffleNetv2 [25]	0.31	233.2	25.4
EffcientNet [38]	0.64	196.9	19.2
Darknet19 [28]	0.52	257.1	39.7
CSPDarknet53 [40]	**0.83**	362.7	11.9
MobileNetv3 [12]	0.79	**123.6**	**44.7**



(a) USV apperance (b) Internal construction

Figure 7. USV platform.

Table 2
Configuration Information

Configurate	Platform	
	PC	USV
System	Ubuntu20.04	Jetpack4.5.1
Pytorch	1.7.0	1.6.0
Tensorflow	2.6.0	\
CUDA	11.0	10.2
cuDNN	8.0	8.0
TensorRT	\	7.1
Deepstream	\	5.1.0

The algorithms in this study were trained on a personal computer (PC) configured with Intel i5-10400F@2.6GHZ*12 CPU, Nvidia GTX3060@12-GB GPU. The speed and accuracy of the models were tested with a deep learning environment configured on the PC and the USV, respectively, the relevant versions are detailed in Table 2.

The public COCO or VOC datasets is widely used to measure the performance of the network model, however, all kinds of ships are simply categorised as 'ships.' The common targets at sea, such as buoys and islands, are not labelled, besides, pictures are taken under well-lit and distinctive conditions without considering the complex background features in foggy weather. The dataset comprises 14,218 images, wherein 49% are from the website, and 51% are from USV missions. The dataset mainly includes island reefs and ships with different functions, such as the liner, container ship, bulk carrier, sailboat, yacht, fishing vessel, passenger ship, and warship. The maritime target dataset produced in this study is open-source and accessible on GitHub.

5. Experiments and Analysis

5.1 Impact of Synthetic Sea Fog Datasets on Maritime Target Detection Performance

We randomly selected 1000 clear sea images from the USV base dataset as the initial dataset for the FCGAN model, and 10 sets of sea fog datasets with different concentrations were synthesised as shown in Fig. 8. We defined three levels of fog: light, medium, and heavy. Figure 9 shows the loss values of all generators and discriminators after each epoch of training. The decrease in the loss of D_{clear} implies that it is becoming more competent at generating foggy images from clear images, so the loss of D_{fog} increases. Similarly, when D_{fog} is trained, the decrease in the loss of D_{clear} implies that it is becoming more competent at distinguishing between real and generated fog images, increase the loss of the generator. In the 0–30 epoch stage, the loss of the generator and the discriminator

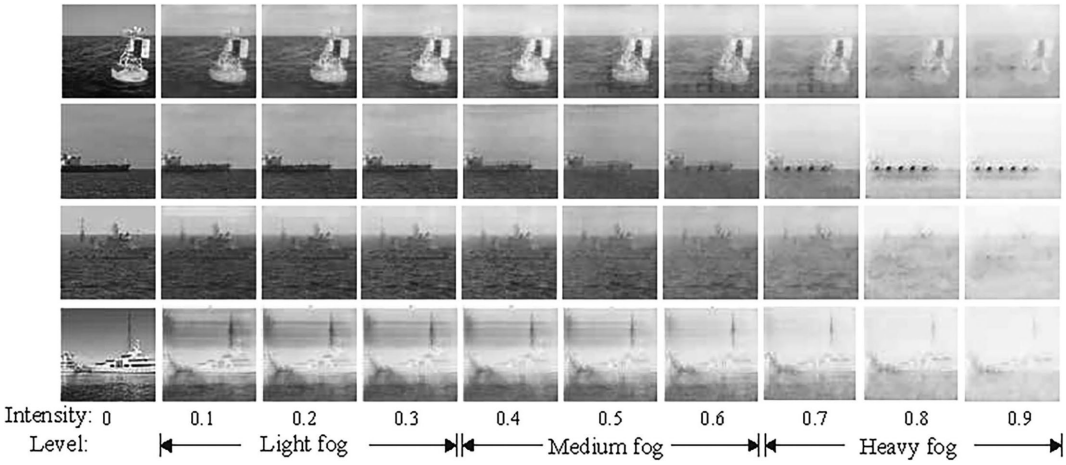


Figure 8. Dataset with foggy intensity.

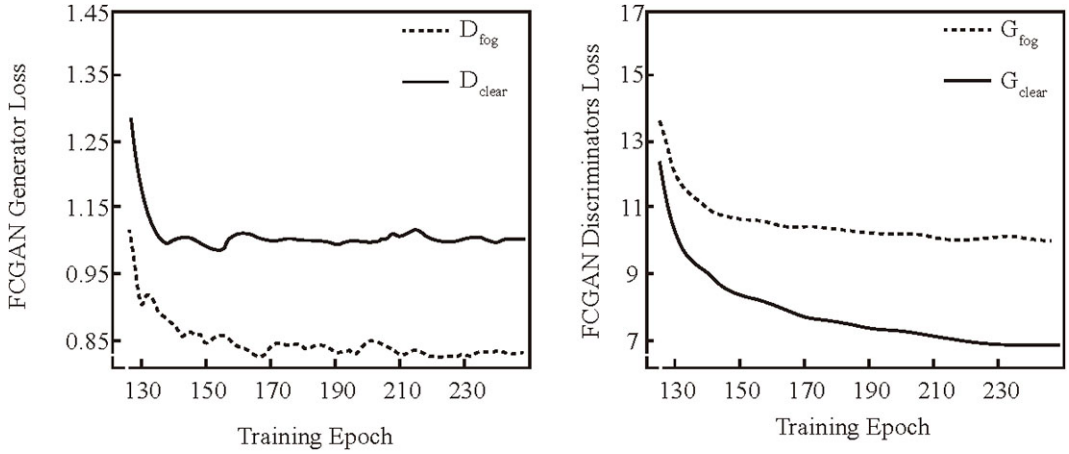


Figure 9. Loss values for the generators and discriminators of FCGAN.

Table 3
Fog Level Effect on Detect Performance

Combo	YOLOv4		M-YOLOv4	
	Recall	Precision	Recall	Precision
1	0.58	0.51	0.76	0.78
2	0.60	0.59	0.81	0.79
3	**0.73**	**0.69**	**0.86**	**0.87**
4	0.32	0.35	0.31	0.43

decrease as the training batch increases, and in the 130–230 epoch stage, the generator loss and the discriminator loss tend to stabilise and can be approximated to converge, the loss function becomes locally optimal. Therefore, we consider the FCGAN training weight at epoch = 230 as the optimal weight.

To evaluate the performance of the target detection algorithm on the sea fog dataset, 3,000 samples generated under the optimal weights were combined with the base dataset. The tag files of the synthesised images were kept consistent with those of the initial dataset, only names were used to distinguish them during training. The recall and precision values in Table 3 were obtained *via* training under two networks, YOLOv4 and M-YOLOv4.

Figure 10 shows results detecting by three different fog weights. The comparison indicates that the results obtained by light fog may have false detection and inaccurate position, while the results obtained by heavy fog weight detection may have missed detection and false detection. Therefore, we show the weight optimal solution obtained by training under medium fog.

Under the M-YOLOv4 model, the maximum recall and precision values appear in combination 2, which improved by 10% and 11%, over the unadded sea fog dataset, while the recall and precision values obtained from combination 1 improved by only 7% and 1%, respectively, over the base dataset. This indicates that the sea fog and base datasets in a suitable combination can improve the M-YOLOv4 detection performance. In combination 3, the recall and precision values of the detection model decreased by 45% and 35%, respectively, so it can be

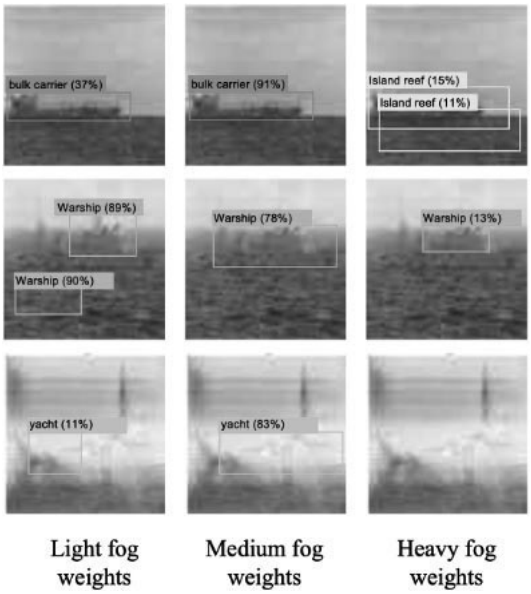


Figure 10. Typical results by three different fog weights.

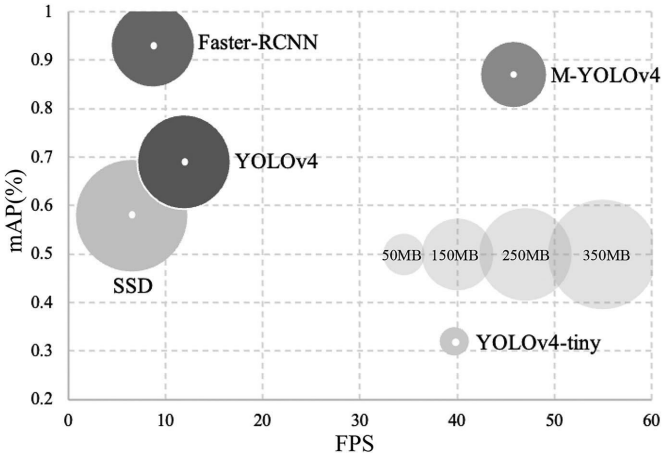


Figure 11. Comparison between different structures.

inferred that the texture features of the image were blurred when the fog concentration was excessively high; thus, the corresponding features of each target were not learnt in the feature extraction stage, which reduced the detection performance. In addition, the recall and precision values obtained by M-YOLOv4 in each combination were larger than those corresponding to YOLOv4, indicating that the improvement of M-YOLOv4 has a certain effect on the precision improvement. Therefore, we adopted the combination scheme of the base dataset plus the medium fog dataset for the final training dataset, then compared and analysed the detection effect of the novel model on USVs.

5.2 Comparison and Analysis of M-YOLOv4 Training Results

In this study, to explore the mechanisms of the effects of improvement points on the improvement of accuracy and speed, we designed ablation experiments according to the single variable principle, as shown in Table 4. After performing the following: (I) clustering optimised initial anchor, (II) new MobileNetv3 backbone network, (III) deeply separable convolutional PANet, and (IV) TensorRT and Deepstream acceleration operations on the original YOLOv4, M-YOLOv4 improved the speed by 33.9 FPS with 2% decrease in mAP, which meets the USV real-time detection requirements. When only improvement I is available, all indicators are not significantly improved, therefore, I is not the improvement factor of this model. Only under the effect of improvement II, the number of parameters decreased by 60% compared to that of the initial model, and the model shrunk by 55%. The model accuracy is reduced by 3% only under the effect of improvement point III, but the number of parameters and the model size are not significantly different from those of II. M-YOLOv4 integrates I–IV, and I–III enables the model to achieve a detection speed of 28.7 FPS with 79% accuracy. The size and number of parameters are not relevant.

We compared the improved M-YOLOv4 with two-stage target detection method Faster-RCNN, SSD300 (another model of single-stage detection method), and YOLOv4-tiny. The accuracy of the M-YOLOv4 model was 4% lower than that of Faster-RCNN, but the speed was 37.1 FPS higher; the model size was 99.3 MB larger than that of YOLOv4-tiny, but accuracy was 57% higher and the performance of detecting types was superior. A typical model is selected for the macro comparison in the speed, accuracy, and model size to obtain Fig. 11, which clearly reflects the superiority of M-YOLOv4 in the model size and the detection speed. When the maritime target detection algorithm is deployed in the USV, to reduce the latency and improve the computational throughput rate, we adopted NVIDIA's TensorRT inference method and the Deepstream module to generate the corresponding engine file to further improve the detection frame rate. Figure 12 shows the effect of the sea trial under clear and foggy weather, respectively. Using the M-YOLOv4 model proposed in this study, the target detection frame rate can reach 45.8 FPS, which is a 30% improvement compared to that of the original YOLOv4 and can meet the real-time effect.

Figures 13 and 14 compare the accuracy and training duration of the latest YOLOv4 model in the same series of training 100 batches, YOLOv4-tiny is a simplified version of YOLOv4, YOLOv4-tiny only retains the 19-layer network in the 53-layer backbone network, YOLOv5 is an improved version of YOLOv4, using a variety of data enhancement methods and parameter iteration methods, it has three different version with different depth and width, named YOLOv5-L, YOLOv5-X, and YOLOv5-S, respectively. The test results show that the accuracy of the M-YOLO model is comparable with that of YOLOv5-X, which is the deepest level of the model. By contrast, the YOLOv5-X has a faster convergence rate, but its training time exceeds 150 h with the weight size of 699.7 MB. However, the M-YOLO only takes 5.6 h to obtain the same detection effect, and its weight size is 123.6 MB. In summary, using the more lightweight MobileNetv3 as the backbone network, the target detection accuracy is not

Table 4
Training Results

Models	Anchor	Backbone	Conv	T&D	mAP	Parameter	Model Volume (MB)	FPS
YOLOv4	✗	✗	✗	✗	0.91	58.3	257.7	11.9
I	✓	✗	✗	✗	0.82	58.4	258.2	20.2
II	✗	✓	✗	✗	0.86	35.1	143.1	27.9
III	✗	✗	✓	✗	0.79	41.9	142.8	28.7
IV	✗	✗	✗	✓	0.89	57.8	257.3	39.7
M-YOLOv4	✓	✓	✓	✓	**0.89**	**19.3**	**123.6**	**45.8**

Table 5
Comparison of Different Models

Models	Recall	Precision	mAP	Parameter (million)	Model Volume (MB)	FPS
Faster-RCNN	0.82	**0.91**	**0.93**	37.1	196.7	8.7
SSD300	0.63	0.58	0.6	39.2	362.7	6.5
YOLOv4	0.73	0.69	0.91	58.3	257.7	11.9
YOLOv4-tiny	0.42	0.31	0.32	31.6	**24.3**	39.7
M-YOLOv4	**0.86**	0.87	0.89	**19.3**	123.6	**45.8**

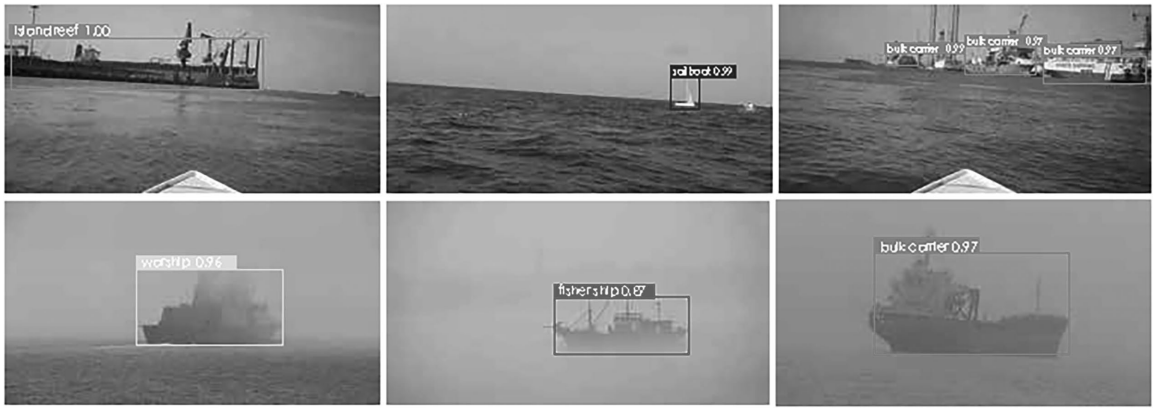


Figure 12. USV detection results.

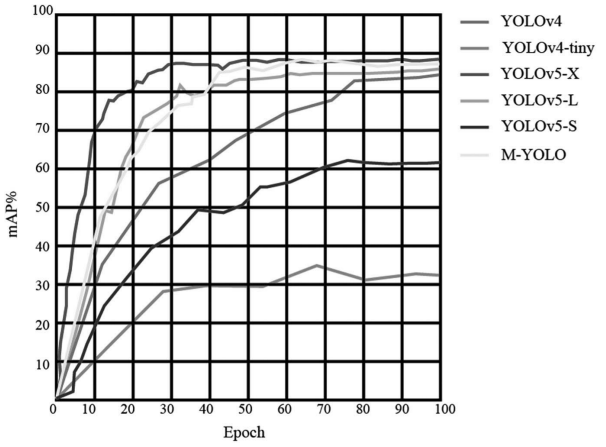


Figure 13. Convergence epoch curves.

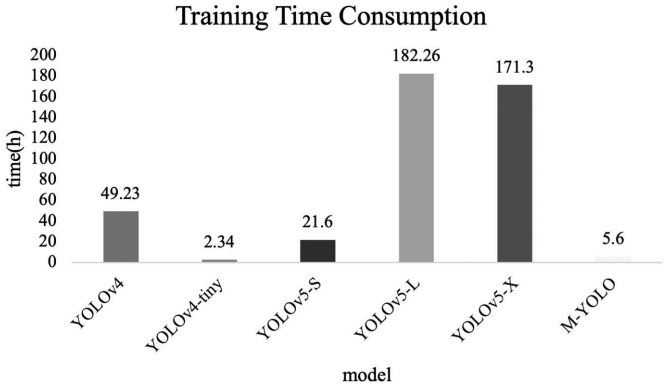


Figure 14. Training time consumption.

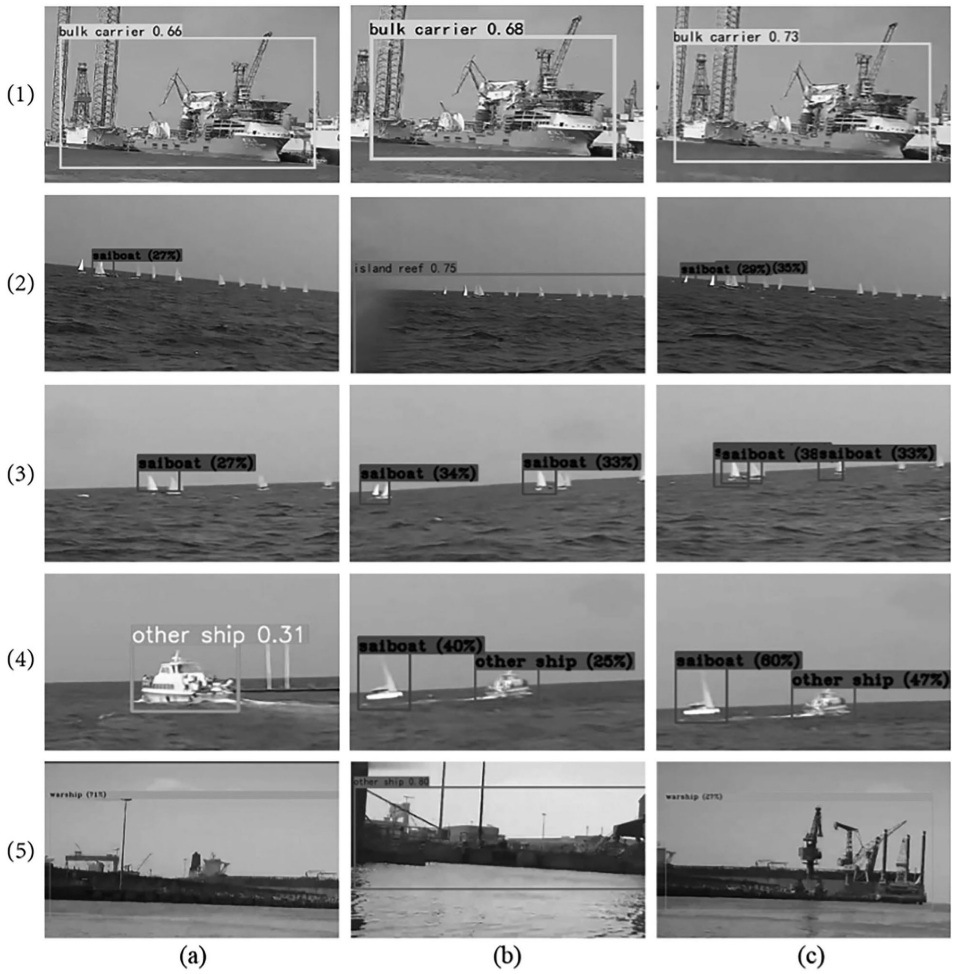


Figure 15. Real-time detection effect: (a) YOLOv4-tiny; (b) YOLOv5-S; and (c) M-YOLO.

greatly reduced compared with YOLOv4, and it can still effectively classify maritime targets and meet the needs of unmanned marine target detection. The reason is that the use of deep separable convolution has the most obvious effect on reducing the number of parameters. Considering the real-time detection effect, only the M-YOLO, YOLOv4-tiny, and YOLOv5-S models are deployed on the edge computing end for testing.

To compare and test the real-time detection effect of the M-YOLO model, three lightweight target detection models are deployed simultaneously on the edge computing end, and the real-time detection screenshots of (a), (b), and (c) in Fig. 15 are YOLOv4-tiny, YOLOv5-S, and M-YOLO, respectively. Due to the small size of some targets, this paper enlarges the panoramic image and captures the local detection results for analysis. In scenario (1), all three models successfully detected the target category, among which M-YOLO had the highest detection accuracy. In scenario (2), the targets are small with a concentrated distribution. By comparison, the YOLOv4-tiny algorithm only detects 1 target. The YOLOv5 detects the sea-sky line as an island and reef and ignores the sailboat, which indicates that it has a poor detection ability for small targets. Similar comparative results have been obtained in scenario (3). In scenarios (4) and (5), all three models fail to identify the target, indicating that the recognition of complex nearshore scenes still needs to be further refined. In summary, under the background of small targets and complex situations, the detection results of M-YOLO are more stable than the others and have higher detection performance.

6. Conclusion

The lightweight target detection method for USVs in foggy weather proposed in this study improves the generalisation of the algorithm and can improve the operational capability of USVs to cope with foggy weather. Under the constraint of sea fog characteristics loss, the FCGAN proposed in this study can synthesise a sea fog dataset more consistent with the actual fog features than other synthetic fog algorithms, and the generated fogged images are more realistic. The re-clustered anchor is more suitable for the detection of targets, such as ships and islands, at sea. Using MobileNetv3 to replace the backbone network of YOLOv4 and applying depthwise-separable convolution, the network model is reduced to 19.3 million parameters with guaranteed accuracy, which has excellent engineering applications. To deploy the studied lightweight target detection network on USVs, we combined the TensorRT acceleration mechanism and Deepstream to achieve adequate performance in model inference and

deployment. Furthermore, this model runs at 45.8 FPS on USVs with effective real-time image information to ensure navigation safety, which provides for the improved real-time fast detection of USV navigation and solutions.

Although the addition of data-enhanced fog pictures to the original data improves the prediction accuracy of USVs in foggy weather, the training batches converge 20% higher than those without the addition of foggy pictures, *i.e.*, the model takes 9 h longer to learnt the features during the training phase because the fog features blur the pixel features of boats and targets. Due to the actual experimental conditions, this solution only considers the effect of foggy weather on the detection performance and fails to consider other extreme conditions, such as insufficient light, blurred rain, and snow. Future work will mainly improve the influence of various factors to obtain the applicable target detection method for USVs under multimodal conditions and match the targets detected by USVs with those detected by radar, combined with path planning algorithms, thus realising the intelligent navigation of USVs.

Acknowledgement

This work is supported by the Key R&D project of Shandong Province (GrantNo. 2018YFJH0704) and International Science and Technology Cooperate Key R & D project of Shandong Province (GrantNo. 2019GHZ007).

References

[1] A. Ali, O.G. Olaleye, and M. Bayoumi, Fast region-based dpm object detection for autonomous vehicles, *Proc. 2016 IEEE 59th International Midwest Symp. on Circuits and Systems (MWSCAS)*, Abu Dhabi, 2016, 1–4.

[2] B. Bahmani, B. Moseley, A. Vattani, R. Kumar, and S. Vassilvitskii, Scalable k-means++, 2012, *arXiv:1203.6402*.

[3] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, Yolov4: Optimal speed and accuracy of object detection, 2020, *arXiv:2004.10934*.

[4] L. Bo, X. Xiaoyang, W. Xingxing, and T. Wenting, Ship detection and classification from optical remote sensing images: A survey, *Chinese J. Aeronaut.*, *34*(3), 2021, 145–163.

[5] Y.-T. Chan, Y.-H. Chu, C.-C. Lee, C.-H. Chen, T.-W. Hou, and C.-H. Huang, Implementation of deep-learning-based edge computing for maritime vehicle classification, *Proc. of the 8th IIAE International Conf. on Industrial Application Engineering 2020*, Matsue, 2020, 247–252.

[6] F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, 1251–1258.

[7] S. Jiang, H. Li, X. Zou, C. Tang, D. Hang, J. Yang, and Die Liu, Lightweight mesh crack detection algorithm based on efficient attention mechanism, *International Journal of Robotics and Automation*, *8*(31), 2023.

[8] D. Engin, A. Genç, and H.K. Ekenel, Cycle-dehaze: Enhanced cycleGAN for single image dehazing, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, UT, 2018, 825–833.

[9] P. Ganesh, Y. Chen, Y. Yang, D. Chen, and M. Winslett, YOLO-ReT: Towards high accuracy real-time object detection on edge GPUs, *Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision*, Waikoloa, HI, 2022, 3267–3277.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial networks, *Communications of the ACM*, *63*(11), 2020, 139–144.

[11] Y. Guo, Y. Lu, Y. Guo, R.W. Liu, and K.T. Chui, Intelligent vision-enabled detection of water-surface targets for video surveillance in maritime transportation, *Journal of Advanced Transportation*, *2021*, 2021.

[12] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le, Searching for mobilenetv3, *Proc. of the IEEE/CVF International Conf. on Computer Vision*, Seoul, 2019, 1314–1324.

[13] Z. Huang, B. Sui, J. Wen, and G. Jiang, An intelligent ship image/video detection and classification method with improved regressive deep convolutional neural network, *Complexity*, *2020*, 2020.

[14] J. Hu, D. Zhao, Y. Zhang, C. Zhou, and W. Chen, Real-time nondestructive fish behavior detecting in mixed polyculture system using deep-learning and low-cost devices, *Expert Systems with Applications*, *178*, 2021, 115051.

[15] X. Hu, Y. Liu, Z. Zhao, J. Liu, X. Yang, C. Sun, S. Chen, B. Li, and C. Zhou, Real-time detection of uneaten feed pellets in underwater images for aquaculture using an improved YOLO-V4 network, *Computers and Electronics in Agriculture*, *185*, 2021, 106135.

[16] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, *Proc. International Conf. on Machine Learning*, (PMLR), Lille, 2015, 448–456.

[17] M. Ju, C. Ding, W. Ren, and Y. Yang, IDBP: Image dehazing using blended priors including non-local, local, and global priors, *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(7), 2022, 4867–4871.

[18] M. Ju, C. Ding, D. Zhang, and Y.J. Guo, Gamma-correction-based visibility restoration for single hazy images, *IEEE Signal Processing Letters*, *25*(7), 2018, 1084–1088.

[19] R.W. Liu, W. Yuan, X. Chen, and Y. Lu, An enhanced cnn-enabled learning method for promoting ship detection in maritime surveillance system, *Ocean Engineering*, *235*, 2021, 109435.

[20] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, Path aggregation network for instance segmentation, *Proc. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, 8759–8768.

[21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A.C. Berg, Ssd: Single shot multibox detector, *Proc. European Conf. on Computer Vision*, Amsterdam, 2016, 21–37.

[22] W. Liu, X. Hou, J. Duan, and G. Qiu, End-to-end single image fog removal using enhanced cycle consistent adversarial networks, *IEEE Transactions on Image Processing*, *29*, 2020, 7819–7833.

[23] Z. Li, X. Liu, Y. Zhao, B. Liu, Z. Huang, and R. Hong, A lightweight multi-scale aggregated model for detecting aerial images captured by UAVs, *Journal of Visual Communication and Image Representation*, *77*, 2021, 103058.

[24] Z. Li, W. Zhou, C. Liukui, and S. Jin, Bio-inspired approach for image vehicle detection under low illumination, *International Journal of Robotics and Automation*, *2*(87), 2020, 332–338.

[25] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, ShuffleNet v2: Practical guidelines for efficient cnn architecture design, *Proc. of the European Conf. on Computer Vision (ECCV)*, Munich, 2018, 116–131.

[26] R.-Q. Ma, X.-R. Shen, and S.-J. Zhang, Single image defogging algorithm based on conditional generative adversarial network, *Mathematical Problems in Engineering*, 2020, 1–8, 2020.

[27] Z. Ouyang, J. Cui, X. Dong, Y. Li, and J. Niu, SaccadeFork: A lightweight multi-sensor fusion-based target detector, *Information Fusion*, *77*, 2022, 172–183.

[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You only look once: Unified, real-time object detection, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, 779–788.

[29] J. Redmon and A. Farhadi, YOLO9000: Better, faster, stronger, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, 2017, 7263–7271.

[30] J. Redmon and A. Farhadi, YOLOV3: An incremental improvement, 2018, *arXiv:1804.02767*.
[31] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *Advances in Neural Information Processing Systems*, *28*, 2015.
[32] O. Ronneberger, P. Fischer, and T. Brox, U-net: Convolutional networks for biomedical image segmentation, *Proc. International Conf. on Medical Image Computing and Computer-Assisted Intervention*, Munich, 2015, 234–241.
[33] C. Sakaridis, D. Dai, and L.Van Gool, Semantic foggy scene understanding with synthetic data, *International Journal of Computer Vision*, *126*(9), 2018, 973–992.
[34] C. Sun, B. Kong, L. He, and Q. Tian, An algorithm of imaging simulation of fog with different visibility, *Proc. 2015 IEEE International Conf. on Information and Automation*, Lijiang, 2015, 1607–1611.
[35] C. Szegedy, S. Ioffe, V. Vanhoucke, and A.A. Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, *Proc. Thirty-First AAAI Conf. on Artificial Intelligence*, San Francisco, CA, 2017, 4278–4284.
[36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, Going deeper with convolutions, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, 2015, 1–9.
[37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the inception architecture for computer vision, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, 2016, 2818–2826.
[38] M. Tan and Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, *Proc. International Conf. on Machine Learning*, (PMLR), Long Beach, 2019, 6105–6114.
[39] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, Scaled-YOLOv4: Scaling cross stage partial network, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, 2021, 13029–13038.
[40] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, CSPNet: A new backbone that can enhance learning capability of CNN, *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, 2020, 1571–1580.
[41] Y. Wang, S. Sun, and J. Zhong, An ensemble anomaly detection with imbalanced data based on robot vision, *International Journal of Robotics and Automation*, *31*(2), 2016.
[42] Z. Xiang, T. Tao, L. Song, Z. Dong, Y. Mao, S. Chu, and H. Wang, Object tracking algorithm for unmanned surface vehicle based on improved mean-shift method, *International Journal of Advanced Robotic Systems*, 17(3), 2020, 1729881420925294.
[43] S. Chen, Y. Feng, T. Tang, and Y. Wu, Automated defect detection based on transfer learning and deep convolution generative adversarial networks, *International Journal of Robotics and Automation*, *7*(35), 2021, 471–478.
[44] Q. Zhang, C. Zhao, X. Zhang, F. Yuan, C. Li, and D. Hao, The generative adversarial network based on attention mechanism for image defogging, *Proc. International Forum on Digital TV and Wireless Multimedia Communications*, Shanghai, 2020, 12–25.
[45] W. Zhang, X.-Z. Gao, C.-F. Yang, F. Jiang, and Z.-Y. Chen, A object detection and tracking method for security in intelligence of unmanned surface vehicles, *Journal of Ambient Intelligence and Humanized Computing*, *13*, 2022, 1279–1291.
[46] X. Zhang, X. Zhou, M. Lin, and J. Sun, ShuffleNet: An extremely efficient convolutional neural network for mobile devices, *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Salt Lake City, UT, 2018, 6848–6856.
[47] Y. Zhang, D. Ren, B. Chen, and J. Gu, Detection of pine wilt disease in autumn based on remote sensing images and enf module, *International Journal of Robotics and Automation*, *37*(6), 2023.
[48] Z. Zhou and X. Yang, Pine wilt disease detection in UAV-captured images, *International Journal of Robotics and Automation*, *37*(1), 2022.
[49] A. Zhu and Y. Chen, A machine-learning-based algorithm for detecting a moving object, *International Journal of Robotics and Automation*, *31*(5), 2016, 402–408.
[50] J.-Y. Zhu, T. Park, P. Isola, and A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proc. of the IEEE International Conf. On Computer Vision*, Venice, 2017, 2223–2232.

Biographies

Shuyue Li received the B.E. degree from the Qingdao University of Science and Technology, China, in 2018, and the M.E.degree from the Ocean University of China, China, in 2022. She is currently pursuing the Doctoral degree with Shanghai Jiao Tong University. Her research interests include computer vision and visual navigation of unmanned surface vehicles.

Junjie Wang received the Ph.D. degree in environmental engineering from the Ocean University of China. He is currently a Lecturer with the College of Engineering, Ocean University of China. His research interests include intelligent nondestructive civil testing and multi-sensor fusion detection technology for internal and external diseases of buildings.

Jinlu Sheng currently works as a Doctor Supervisor with the College of Shipping and Naval Architecture, Chongqing Jiao Tong University, China. He is also the Director of the Institute of Maritime Safety and Environmental Engineering. His research interests include maritime safety and environmental engineering, water traffic safety technology, and management and ship equipment maintenance technology.

Ziyu Liu received the B.E. degree in automation from Qingdao University in 2019, and the M.E. degree in control engineering from the Ocean University of China, in 2022. He is currently working as an Engineer from Shanghai Marine Electronic Equipment Research Institute. His research interest is control technology for underwater vehicle.

Shixin Li received the B.E. degree from Qingdao University of Science and Technology, China, in 2018. He is currently a Junior Engineer with the College of Electromechanical Engineering, Qingdao University of Science and Technology. His current research interests include intelligence algorithm and its application.

Ying Cui received the B.E. degree in Harbor, Channel and Coastal Engineering from the Ocean University of China, China, in 2013, and the Ph.D. degree in Harbor, Coastal and Offshore Engineering from Ocean University of China, China, in 2018. She is currently an Associate Professor with the College of Electromechanical Engineering, Qingdao University of Science and Technology, China. Her research interests include the application of intelligence algorithms, the development of wave energy, fluid mechanics, and computational fluid dynamics.