Identifying Ships from Radar Blips like Humans Using a Customized Neural Network

Feng Ma, Zhe Kang, Chen Chen*, Jie Sun, Xiao-bin Xu, Jin Wang

***Abstract*—An experienced helmsman can always distinguish ships from a pile of radar blips in scenarios such as nearshore waters and inland rivers with a single glance. To replicate this intelligence, a novel approach called MRNet based on deep convolutional networks is proposed. It employs a highly customized neural network to extract critical information from successive radar scans, ranging from low-level characteristics to high-level semantics. The feature fusion network of MRNet is also built with a Depthwise Separable Convolution-based network, which reduces parameter size and calculational usage while improving overfitting issues significantly. In the final prediction procedure, a method based on weighted-box fusion and a Scylla-IoU function is used to accelerate convergence. A marine radar image dataset, namely radar3000, was established to validate the proposed approach. In the corresponding experiments, the recall, identification accuracy, and precision of MRNet reached 0.9663, 0.9418, and 0.9267 respectively. On the other hand, the parameter size and calculational consumption were controlled to only 34.41M and 21.55G respectively. Compared with the commonly-used fractal algorithms and the YOLO series, the MRNet can be described as significantly superior in the application of recognizing ships from marine radar blips, especially in crowded scenarios, which is very similar to human eyes, and can be of great use to navigation and coastal surveillance.**

***Index Terms*—Marine radar; Ship identification; Depth convolutional neural network; Feature fusion network; Maritime management**

I. INTRODUCTION

IT is necessary to keep an eye on ships in congested waters, waterways close to harbors, and high-risk areas to ensure their safety. Diverse monitoring approaches are frequently utilized in ship navigation, harbor monitoring, and fishery management [1] to satisfy this demand. Marine radar (MR) systems are typically utilized as the primary sensors in these applications because they are capable of capturing images that are satisfactory in adverse conditions (rain, fog, etc.) or low visibility. Radar systems show any things, including moving and stationary ships, large waves, reefs, rocks, and shorelines, as blips. Thus, distinguishing ships from these blips requires expertise and typically a helmsman. Presently, certain advanced marine radar systems have modules called ARPA, or Auto Radar Plotting Assistance, that allow them to monitor moving blips. Specifically, ARPA can extract abundant ship navigation information by processing the raw radar data [2], but the existence of considerable interference in some complex scenarios results in a significant amount of erroneous information. Consequently, Existing modules are unable to identify blips in crowded waters or inland rivers.

In radar surveillance, ship blips are commonly categorized into high-speed and low-speed items. Presently, the ARPA modules integrated within radar systems demonstrate a notable capability to discern the distinct visual attributes of high-speed objects from the background in the majority of instances, owing to their inherent salience. In addition, these fast-moving objects are likely moving ships when viewed by a marine radar system. Conversely, the berthing and unberthing of low-speed ships are extremely comparable to reefs, noise, and coasts. Hence, a much deeper comprehension of the radar images is necessary for the reliable identification of these items.

Yet, there are further challenges in trying to find ships in radar blips. First of all, radar images are frequently interfered with. Only a small percentage of radar blips in navigable waters correspond to moving ships. Radar photographs have far lower definition than regular images, thus there is not a lot of information there. In addition, it can be challenging to squelch additional contaminants like waves, clouds and rain. The images are greatly influenced by the radar system's angle of observation in relation to the ships. Moreover, ships are typically depicted as small, ineffective objects in radar images, which results in these ship blips containing only a small number of significant features. Identification becomes more challenging when the ships are close to the shore since the shore will effectively alter the ship's blips.

The conventional methods for identifying ships in marine radar images are based on fractal algorithms, filtering, doppler effects, and pattern recognition. They are typically used for target extraction from clutter suppression, motion estimation, and clutter suppression. They are perfectly able to follow

This work was supported in part by the Funds for National Key R&D Program of China under Grant 2021YFB1600400; in part by the National Natural Science Foundation of China under Grant 52171352, and Grant 52201415. *(Corresponding author: Chen Chen.)*

Feng Ma and Zhe Kang are with the Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan 430063, PR China (e-mail: martin7wind@whut.edu.cn; 252128@whut.edu.cn)

Chen Chen is with the School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, PR China (e-mail: chenchen0120@wit.edu.cn)

Jie Sun is with the Nanjing Smart Water Transportation Technology Co., Ltd, Nanjing 210028, PR China (e-mail: sunjie@smartwaterway.com)

Xiao-bin Xu is with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, PR China (e-mail: xuxiaobin1980@hdu.edu.cn)

Jin Wang is with the Liverpool, Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool L3 3AF, UK (e-mail: j.wang@ljmu.ac.uk)

moving ships on the open sea. These techniques, however, do not work well with little radar blips, things moving slowly, or rather significant background disturbances. Instance segmentation, object identification, and tracking technologies have advanced quickly in the past ten years as convolutional neural networks (CNN) and other Deep Learning-based techniques have gained popularity. The extraction of deep semantic information using identification methods based on CNNs, including the YOLO series and the R-CNN-based networks, has clear advantages over traditional methods and is performed admirably in challenging situations like industrial production lines, autopilot, vehicles, and airplanes. These techniques are made for commonplace photos with lots of colors, contours, and spatial information. Radar images are challenging to grasp, even for human eyes, as was previously mentioned. In conclusion, when a dedicated network is created by the characteristics of marine radar, CNNs may exhibit improved proficiency in discerning ships within radar blips.

In this research, a novel MRNet approach based on deep convolutional networks is suggested to extract ships in radar blips for navigation or coastal monitoring. Our approach considerably differs from earlier studies in the following areas, considering the present accomplishments.

1) To extract important ship characteristics and improve the capacity to separate small-scale ships from interference, a distinctive feature network is applied.
2) A unique multi-receptive field fusion structure and a novel receptive field expansion module are created to process the feature output and merge local and global information.
3) To increase the positioning accuracy of the ships, the non-maximum suppression (NMS) and positioning loss computation of the prediction boxes are optimized.
4) To facilitate the algorithms in learning ship features from radar images, a dataset containing multi-scene and multi-type data is constructed.

The remainder of this paper is organized as follows. Section II briefly reviews the related studies on object identification under different scenarios. Section III proposes the enhanced ship identification method based on deep convolutional networks. Experimental results on marine radar images are comprehensively analyzed in Section IV. The main contributions of this work are concluded in Section V.

II. Reference

Finding specific objects in various radar images has been a well-known research topic for decades. Among the most popular techniques are fractal geometry, filtering, motion estimation, and pattern recognition. However, the accuracy of these techniques is typically poor in crowded or complicated situations. For instance, the so-called ARPA function in a marine radar system is created using the aforementioned techniques and is only reliable in open waters, making harbors, rivers, and busy waterways practically useless. It is logical to use CNNs to identify ship blips in marine radar images given that object identification and tracking technologies based on CNNs and other Deep Learning-based algorithms have been shown to outperform conventional methods in a variety of identification scenarios. To serve as a guide for the MRNet algorithm design, this section provides a summary of the state of ship identification techniques in marine radar images, followed by an overview of object identification technology and optimization techniques in other scenarios.

A. Ship identification in marine radar images

Traditional image processing, filtering, pattern recognition, etc. are the main components of traditional identification methods. The main technical details and development of the relevant methods' research will be succinctly summarized in this section.

In many studies, the methods, including threshold binarization and extraction of the image-connected domain, are the beginning of radar image processing. Subsequently, the moving blips will be tracked by morphological characteristics. Chen et al. [3] estimated the ships' position by calculating the features of the ship wakes in the images. However, with the increase of the radar reference distance, the positioning accuracy would decrease observably. Han et al. [4] extracted the coastline outline around the ship from the image, and then matched the coastline features with the pre-constructed map to estimate the ships' position in the specific waterway. These methods perform satisfactorily when the blips of ships are relatively large and distinguishable.

Meanwhile, researchers frequently employ filtering algorithms to determine the location, velocity, and course of ship blips in radar images. Such techniques presuppose that the ship always intends to sail straight ahead and that any deviation in the tracking of a blip is the result of observational noise. When a ship is berthing or unberthing, such an assumption is incorrect. Wen et al. [5] proposed an image clutter suppression and identification method based on time-domain joint filtering. This work converted image signals into a three-dimensional image spectrum through a 3D fast Fourier transform. Then, they designed a time-domain clutter suppression model to filter the image spectrum, achieving high accuracy in the testing image sequence. Wu et al. [6] used preprocessing methods to obtain static feature points in radar images. At the same time, they used synchronous positioning and mapping algorithms to calculate the actual trajectory of ships, which could extract the time and position data of ships from a large number of blips in images. Therefore, accurate ship motion modeling may enable an effective filtering algorithm. However, when ships are moving slowly or there is significant background interference, such filtering algorithms are no longer effective.

Pattern recognition methods are very common in marine radar processing, which can be used to distinguish whether a blip is a ship or not. Such methods have a promising performance on ship identification in radar images. In particular, the extensive use of non-probability has improved recognition accuracy. Ma et al. [7] proposed a method based on evidential reasoning (ER), which obtained the likelihood information of the ships and used ER rules to estimate the

reliability of a specific blip being a ship.

In addition, scholars have built several datasets to accelerate the research of identifying ships in radar systems. Osman et al. [8] established a database that combines synthesizing multi-angle range profile data of ships in a radar system. At the same time, they proposed a method to identify ships in radar images with the help of a novel model and a database. This method analyzed the contour, space, and velocity characteristics of radar blips, then classified multiple types of ships using a so-called object matching method. However, this method needs a huge dataset for learning, which also leads to high computational demands. Moreover, the generalization ability of the proposed method is dissatisfactory.

In recent years, ship identification technologies based on CNNs have been developed from different perspectives in radar systems. Mou et al. [9] adopted hierarchical normalization methods to extract key features of ships in their proposed method. Meanwhile, they used a spatial attention network to suppress clutter and enhance effective feature signals. Chen et al. [10] improved Faster R-CNN from the aspects of the backbone network, the anchor box size, and the feature scale normalization. In the testing, their method achieved high accuracy and promising robustness in the application of identifying ships in radar images. Chen et al. [11] proposed a ship identification method based on a so-called dual-channel convolutional neural network (DCCNN) and a novel false alarm controllable classifier (FACC), which could effectively suppress clutter signals and accurately extract key features of ships. It can be concluded that the use of CNNs is capable of effectively extracting ships from images and achieving high accuracy. However, these methods are all designed and validated in open waters. Their abilities to adapt to complex environments, such as crowded waters and harbors, have not been proven.

The accuracy of identifying objects in radar images is also determined by pre-posed noise suppression, image enhancement, and data fusion after preliminary identification. Mao et al. [12] proposed an efficient marine radar imaging method based on the non-uniform imaging theory, which combined a beam recursive anti-jamming method and a non-uniform sampling model to extract ship distribution with lower computation, providing high-quality images. Zhang et al. [13] used a generated countermeasures network to remove the noise in radar images and utilized the registration methods to remove the imaging differences between radar and chart data. Then, they used the sparse theory and the Fourier transform to process the feature data and to fuse the image.

In summary, the ship identification algorithms utilized deep learning, and CNNs are capable of efficiently suppressing various interference, i.e., complex weather, ocean clutter, islands, and reefs, which have high accuracy for multi-scale or multi-class ships. Such algorithms show more adaptability and resilience for ship identification in marine radar images than classic methods, which could be a fruitful research direction.

B. Ship identification in natural and SAR images

In the last decade, CNN-based methods have been the mainstream of object detection and identification in various fields gradually. These algorithms can be divided into one-stage [14],[15] and two-stage [16],[17] methods. The one-stage algorithms generally have fewer convolutional parameters, which consider inference speed and recognition accuracy at the same time. The two-stage algorithms that can achieve a higher accuracy usually have more parameters. However, these algorithms are slower than one-stage ones. Presently, a considerable proportion of ship identification algorithms are designed for Synthetic Aperture Radar (SAR).

In the research of ship identification in ordinary images, the majority of these methods are designed based on one-stage algorithms. By constructing efficient feature extraction networks and dedicated fusion structures, such methods are capable of extracting features of images at different scales [18]-[20]. In the research of identifying ships in SAR images, a considerable number of researchers chose to optimize the feature extraction structure of the widely-used CNN-based algorithms, aiming to improve their performances [21]-[23]. In addition, other dedicated optimization strategies have been adopted to ensure the generalization ability, e.g., the neural structure search method, the label redistribution strategy, and the feature pyramid network (FPN) [24]-[26]. Moreover, identifying small objects in SAR images seems to be a hot issue of many researchers' study. Their experiments have proved that all the improvements including special structures, dedicated training strategies and robust datasets can increase the accuracy of identifying small objects in radar images [27]-[29].

C. Optimization methods of ship identification

It is a valid and natural way to increase the accuracy of a specific identification algorithm by building a specialized, labeled, and well-balanced dataset [30]. Moreover, methods such as feature enhancement, key point extraction, and interference suppression can also improve the accuracy of object identification [31],[32]. In some ship identification scenarios, the computing resource is limited, since the applications are deployed on embedded devices. Therefore, it is necessary to design lightweight identification algorithms. Yin et al. [33] designed an efficient lightweight feature extraction network using a so-called depthwise separable convolution network to balance accuracy and inference speed. Deng et al. [34] proposed a joint compression method composed of quantitative perception training and structure pruning based on Taylor expansion, reducing the parameters and computation according to hardware constraints.

In conclusion, since identifying small objects is always a bottleneck problem in radar applications, it is essential to build a dedicated network by the imaging characteristics of marine radar systems. An effective feature extraction network and an appropriate predictive structure based on CNNs can significantly increase the accuracy of ship identification in kind of images, obviously superior to traditional methods. In the study, a proposed identification technique that combines a deep convolutional feature network with a fusion structure is capable of incorporating both high-level semantic data and low-level

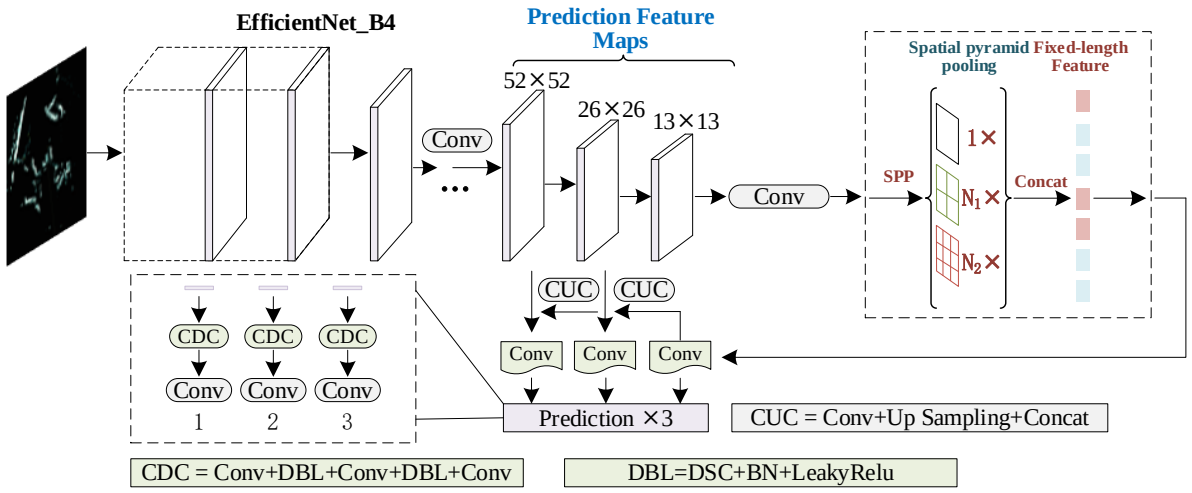


Fig. 1. The flowchart of the ship identification framework

feature information. Such a network can, however, withstand the effects of interference.

III. A Proposed Method

The framework for identifying ships in marine radar images proposed in this research is shown in Fig. 1, which is mainly composed of a feature extraction network and a lightweight feature fusion structure. In particular, the EfficientNet-B4 [35] network is selected as the kernel of the feature extraction network, which has a satisfactory performance in extracting small objects' features at different levels in radar images. Moreover, a lightweight feature fusion network that includes three layers of prediction channels is designed, which respectively cover the identification process of ships at low-to-high scales. Meanwhile, a depthwise separable convolution-based network is integrated into the feature fusion network, which can significantly reduce the parameters and the computational cost. In addition, the spatial pyramid pooling (SPP) module is connected to the EfficientNet-B4 network. Hence, the feature maps of arbitrary sizes can be converted into a vector of a fixed size, which ensures the robustness of the proposed method facing the variable sizes of feature maps. Subsequently, inspired by the YOLO series, the weighted boxes fusion, namely WBF [36], is used to promote the NMS process. Moreover, the Scylla-IoU, namely SIoU [37], is introduced to optimize the positioning loss calculation of the prediction boxes, which accelerates the convergence and increases the accuracy at the backpropagation process.

A. Feature extraction network

Making a trade-off between the depth, width, structure, and resolution of the corresponding convolutional networks is essential to ensure the functionality and speed of the feature extraction network. The EfficientNet-based network scales the aforementioned three dimensions uniformly using a set of fixed coefficients, ensuring the convolutional results of the feature network at various scaling parameters are generally appropriate. A subtype of EfficientNet called EfficientNet-B4

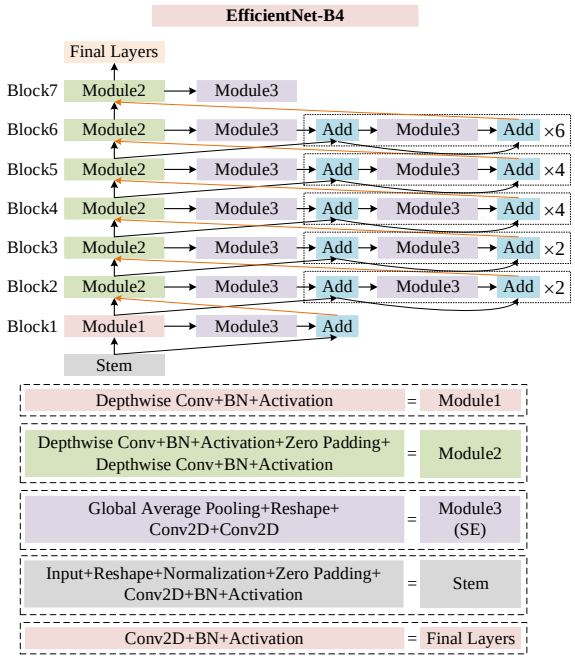


Fig. 2. The structure of the EfficientNet-B4

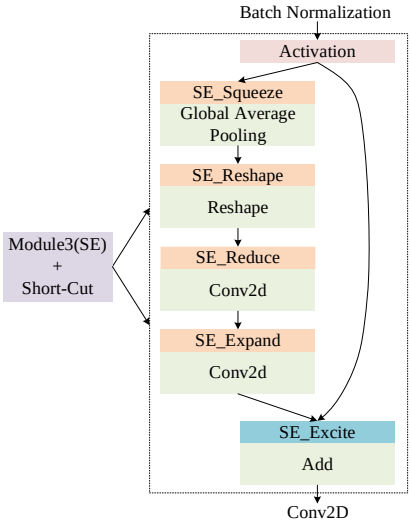


Fig. 3. The structure of the connection between the SE attention mechanism and residual connection (short-cut)

has a relatively high number of convolutional layers, which makes it effective at extracting important characteristics of small ships. Additionally, the EfficientNet-B4 has a lot fewer parameters than the commonly used ResNet 50, ResNet 101, SENet, etc. In other words, the EfficientNet-B4 has managed to strike a balance between inference speed and accuracy.

All objects in a marine radar system, such as ships, shorelines, and rocks, are represented by blips of varying sizes and shapes. To put it another way, unlike ordinary images, radar images contain very little information, thus the identification method must be highly capable of comprehending them, just like human eyes. As a result, the first priority is to extract important ship blip characteristics. To meet this requirement, EfficientNet-B4 adopts a multi-scale feature extraction approach by leveraging appropriate network's depth, width, and resolution parameters to maintain efficiency while accommodating inputs of different scales, as illustrated in Fig. 2. Additionally, the network employs several efficient feature extraction techniques, including depthwise separable convolution (DSC) [38], attention mechanisms, and residual connections. These modules reduce model complexity and parameters, while improving feature extraction efficiency. In particular, the extensive use of attention mechanisms, namely Squeeze and Excitation (SE), and residual connections helps the network to better learn features and reduce the loss of critical information, shown in Fig. 3. These feature extraction methods enable the network to preserve feature information when processing small-scale radar blips.

B. Receptive field expansion module

When performing convolutional calculations on input features of radar images in a feature network, some pixel information may be lost due to stretching, cropping, and other operations that the feature network generally applies to account for differences in input image sizes. Furthermore, small-scale ships often have fewer effective features preserved in deep convolutions, which can lead to lower identification accuracy of such ships by the model. To address these issues, concatenating an SPP module after the feature network can retain more abundant features. This is because SPP, through pyramid-like pooling operations, can increase the network's receptive field without changing the feature map resolution, thereby better capturing object-related features of different scales and enhancing the model's identification capability. The SPP structure is shown in Fig. 4, and it transforms multi-scale feature maps into fixed-size feature vectors by concatenating multiple differently sized max-pooling modules. After numerous tests, the SPP module proved to be more adaptable for ship identification in marine radar images and could produce better experimental results by splicing four layers of max-pooling modules with sizes of 1×1, 5×5, 9×9, and 13×13 [15].

C. Feature fusion network

As previously stated, it is critical to obtain ship features with high sensitivity. Shallow features in convolutional networks are typically represented with a higher resolution that contains

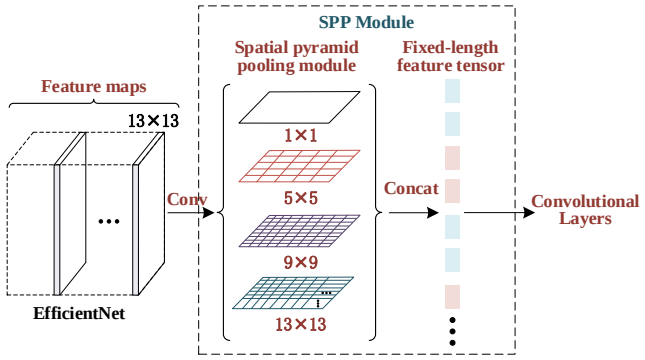


Fig. 4. The structure of SPP

more detailed spatial information. These characteristics are more closely related to target location accuracy in identification. By contrast, deep features have a lower resolution and contain more spatial and semantic information. Due to the absence of color information in radar images, it is advantageous to extract shallow features for ship identification. Considering the characteristics of marine radar images, fully concatenating feature information extracted from different scales of receptive fields into a dedicated feature fusion network is equally critical for ship feature extraction.

The frequently-used feature fusion structures represented by the standard FPN [39] often include redundant calculations and cost much computation resources, which might be acceptable to extract semantic information from ordinary images. As discussed previously, shallow features are more important in the identification of radar blips. On this occasion, a considerable number of invalid convolutional calculations in the standard FPN could take a negative influence on the extraction of ship blips. Particularly in crowded waterways, characterized by a scarcity of distinctive ship features, the excessive convolution computations can potentially lead to the emergence of overfitting issues. In the long-term experiments, it has been found that the 3×3-sized convolutions of FPN are the main reason for the rise of invalid parameters. Moreover, the improved FPN constructed in the paper consists of three individual prediction channels, which predict large, medium, and small ship blips respectively. Thus, the feature tensor dimension of the input is expected to be much larger than that of a single prediction channel. It is acknowledged that the most direct way to reduce the size of model parameters is to reduce the convolutional calculation of feature information. Therefore, this research uses DSCs to replace the traditional 3×3-sized convolutions. This alteration will improve the ability to extract key ship features, which in turn increases the computational efficiency of the proposed MRNet. This newly designed lightweight feature fusion network, namely LightFPN, is shown in Fig. 5. The LightFPN inherits the fundamental structure of the standard FPN, strengthening the ability to extract ship features in low-resolution images.

In a more detailed analysis, due to the fewer pixel features of ships in radar images, the detailed edge information of ships will be diluted after deep convolutional calculation, reducing target positioning accuracy. DSCs are used in this study to

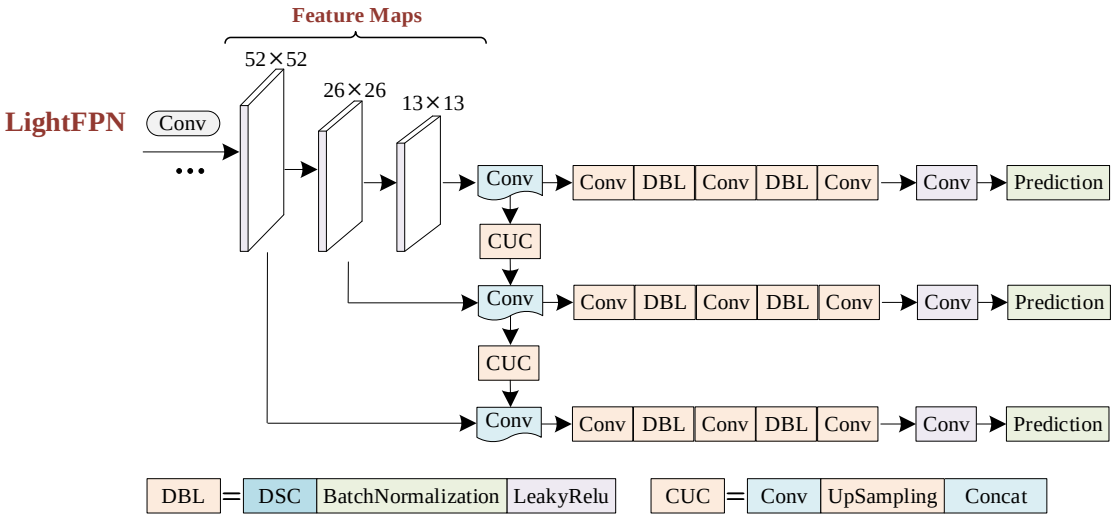


Fig. 5. The structure of LightFPN

simplify the calculational process of deep convolutions, reduce redundant information, and improve the expression of ships' significant features. Meanwhile, the LightFPN improves the expression of ships' edge information by constructing a three-layer feature fusion structure to fuse details of shallow convolutions with semantic information of deep convolutions, improving the accuracy of small-scale ship locations.

It should be noted that the parameter size in DSC is several times smaller than that of standard convolutions [38]. As a result, the DSC greatly simplifies the computation while maintaining the nearly identical performance. In other words, traditional convolution operations can be directly replaced by DSCs without any changes to the hierarchical structure of convolutional networks. In conclusion, the DSC is a technique for convolutional parameter compression. Due to the significant parameter reduction compared to standard convolution, the DSC is better able to determine the relative importance of convolutional features, which contributes to an elevated accuracy in ship identification within radar images.

D. Non-maximum suppression

A so-called LightFPN is used in MRNet to build a three-channel target prediction structure that enables ship blips to possess multiple prediction boxes. The NMS method is used to choose an appropriate prediction box for a ship in order to solve this issue. The WBF has been introduced as the NMS's metric in the research. In contrast to the popular IoU method, the WBF uses weighted adjustment to consider the coordinate position and confidence value of the ship in its entirety, increasing the prediction accuracy of small-scale and crowded blips.

As illustrated in Fig. 6, the MRNet generates multiple prediction boxes for a single ship. Standard NMS or Soft-NMS only filters the prediction boxes, which can result in retained box with suboptimal localization accuracy. In contrast, the WBF method significantly improves the final ship prediction accuracy by fusing the multiple prediction boxes.

E. Position loss function

The research also introduced the SIoU method, an advanced

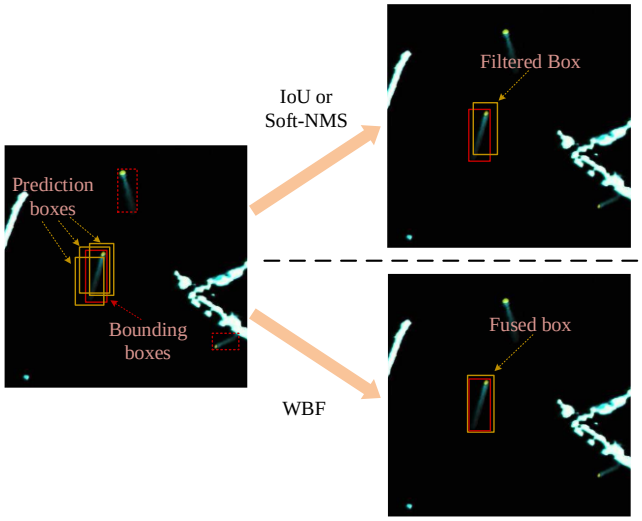


Fig. 6. The actual performance comparison of WBF and other methods in NMS

IoU method, as the calculation standard for the position calculation of objects in order to further increase the positioning accuracy of the prediction boxes. This method fully considers the angle loss, distance loss, shape loss, and IoU loss between the prediction box and the bounding box. Through the consideration of multiple factors, the SIoU loss has a more stable and smoother gradient during algorithm training. Furthermore, this metric provides a more accurate assessment of the overlap between prediction and ground truth boxes, which can improve the effectiveness of the algorithm's training.

In the following section, we will provide an overview of the calculation process for each component, while also referring to Fig. 7 [37]. Firstly, for the angle loss, its calculation process is as follows:

$$\Lambda = \cos(2\times(\arcsin(\frac{C_h}{\sigma})-\frac{\pi}{4}))$$

Where σ represents the distance between the prediction box center and the bounding box center, and C_h represents the distance between the centers of the two boxes in the vertical

direction.

The definition of distance loss is as follows:

$$\Delta = 2 - e^{-\gamma \times \rho_x} - e^{-\gamma \times \rho_y}$$

$$\rho_x = \left(\frac{C_w}{C_w^b}\right)^2, \quad \rho_y = \left(\frac{C_h}{C_h^b}\right)^2, \quad \gamma = 2 - \Lambda$$

Where C_w represents the horizontal distance between the centers of the two boxes, and C_w^b and C_h^b represent the horizontal and vertical distances of the minimum bounding rectangle of the two boxes, respectively.

The definition of shape loss is as follows:

$$\Omega = (1 - e^{-W_w})^\theta + (1 - e^{-W_h})^\theta$$

$$W_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}, \quad W_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}$$

Here, (w, h) and (w^{gt} , h^{gt}) denote the width and height of the prediction and ground truth boxes, respectively, while θ represents a tuning parameter used to balance the importance of shape information in the loss function. In this paper, the value of θ is set to 4. And, the definition of IoU loss is as follows:

$$IoU = \frac{Intersection\ Area}{Union\ Area}$$

The IoU loss is calculated as the ratio of the intersection over the union of the predicted and ground truth boxes. Therefore, the SIoU loss function is finally expressed as:

$$Loss_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2}$$

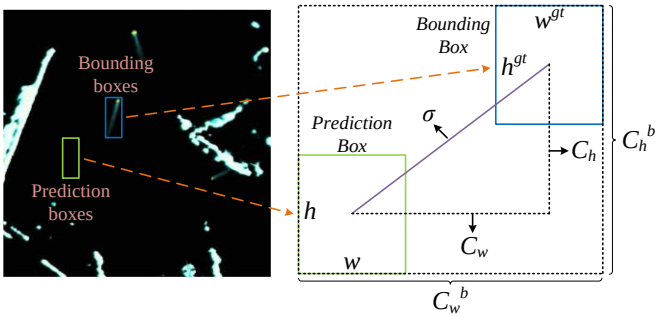


Fig. 7. Explanation of key metrics of the SIoU loss

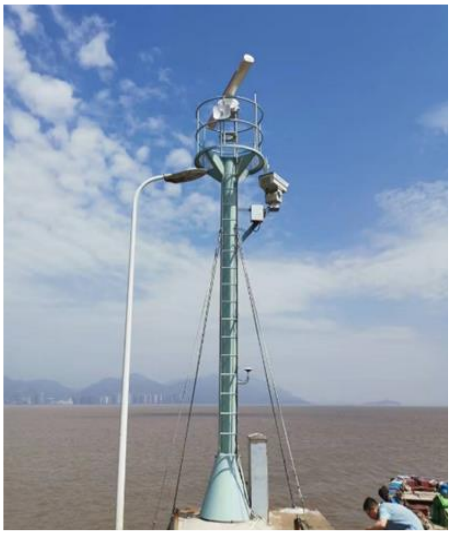
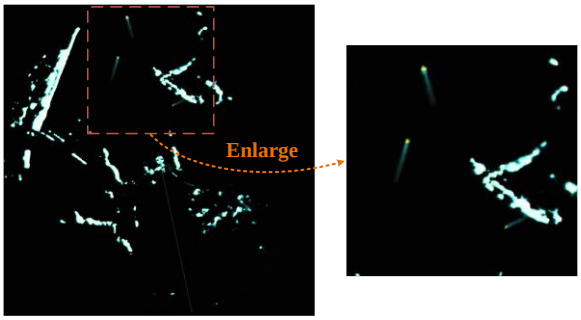
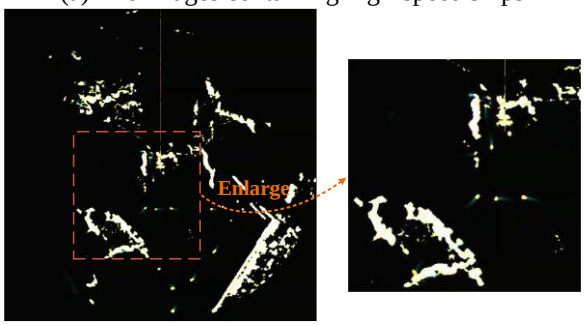


Fig. 8. The shore-based marine radar



(a) The images containing high-speed ships



(b) The images containing low-speed ships

Fig. 9. The marine radar images

IV. A Case Study

A. Dataset

The utilization of a high-quality dataset can greatly augment the experimental results of CNN-based identification algorithms. To establish the groundwork for this research, a radar-image dataset was meticulously curated. The dataset incorporates a JMA5300 marine radar system, depicted in Figure 8, serving as the primary sensor. Precisely, the radar system is installed at Zhujiajian Centipede Temple Pier in Zhoushan City, China. This radar system, in particular, was placed in an ideal location, facing two busy waterways, two harbors, and a 50-kilometer-long shoreline. The water area covered by the radar predominantly includes passenger ships and various auxiliary ships.

We named the dataset as Radar3000, which contains 3000 images resulting from the deep treatment of radar echoes. Ship blips in these marine radar images can be roughly classified as high-speed or low-speed, as shown in Fig. 9. High-speed ones have more distinguishable visual characteristics, such as tadpole-like shapes, yellow heads, and long blue tails, as shown in Fig. 9. By contrast, the visual features of low-speed ones are very difficult to identify at a glance, they are very similar to islands, reefs, noise, and other disturbances. In addition, the Radar3000 dataset collects a considerable num of images under abnormal environments, e.g., bad weather, crowded ships, imaging in multipath interference, etc. Since misidentifications are inevitable when ships are crowded, Radar3000 collected more samples in the corresponding scenarios to ensure performance. In the location of the radar system, a majority of

the captured ships are small ships. Some of them are just 5 meters long. Additionally, a K-means clustering-based pre-processing was carried out, and the results showed that the average ship blip size in these marine radar images was 21×25 pixels or roughly 0.05% of the entire image area. Therefore, the majority of ship blips visible on marine radar images are small-scale targets in a common scenario.

Since the ships are represented as blips in a marine radar system, the ship type, category, and other information cannot be determined. Therefore, the Radar3000 dataset marks all the known ship blips as one category. More precisely, we employed the universal open-source tool, LabelImg, to annotate all the images and draw bounding boxes to indicate the positions of the ships. Each bounding box was annotated with corresponding class information, specifically labeling the ship category as "B". Upon completion of the annotation process for all the images, the annotated results were exported in the standardized Pascal VOC format. Furthermore, all of the images used in the study are split into three groups, referred to as training sets, validation sets, and testing sets, with a ratio of 8:1:1. On the training and validation sets, as well as the testing sets, the proposed MRNet is trained, and its performance is assessed.

B. Training

The training process of MRNet is further improved, which is mainly based on three aspects, i.e., the pre-training, the learning rate attenuation, and the loss calculation. In the training process, the transfer learning method is adopted. Specifically, the network weights of EfficientNet-B4 trained on the ImageNet are used as the initial weights for training [40]. In addition, the algorithm is typically trained with a specific set of initial learning rates. However, the method might cause unstable changes in model convergence [41]. This issue is addressed by the introduction of the learning rate decay method of linear cosine decay, which quickens MRNet convergence. What's more, this study introduced the label smoothing method to enhance the method's loss calculation, ensuring that the training process is maintained in the right direction [42].

C. Experiments

To validate the proposed method, the experiments based on Tensorflow-GPU (2.7.0) in Ubuntu 20.04 have been conducted. In particular, all the experiments were carried out on a server that consists of an Intel(R) Core (TM) i7-12700K CPU and an Nvidia GeForce RTX 3090 GPU. The experiments compare the performances of several state-of-art methods, i.e., the traditional identification method, the YOLO algorithms, and the MRNet. The experiment of each algorithm is based on the same dataset discussed in Section 4.1.

After many experiments and comparisons, the appropriate parameter settings are as follows. During the training processes, the input size is set to 416×416 pixels as customary [15],[16], and the momentum is set to 0.9. As discussed previously, the transfer learning strategy is adopted in this research, hence the proposed MRNet only needs to train the prediction layers first. On this occasion, it is appropriate to freeze the shallow layers to accelerate the training process. Moreover, the batch size is set to 16 after several attempts, which can maximize the use of computing resources. In the comparisons, all the methods had run 15 epochs of iterations, and the initial learning rate was set to 10^{-4}. After the transfer learning, the frozen convolutional layers of the proposed MRNet were activated. The batch size was set to 4 considering the observable increase of convolutional parameters, while the initial learning rate kept the value of 10^{-4}. At this stage, 200 epochs of training iterations had been conducted.

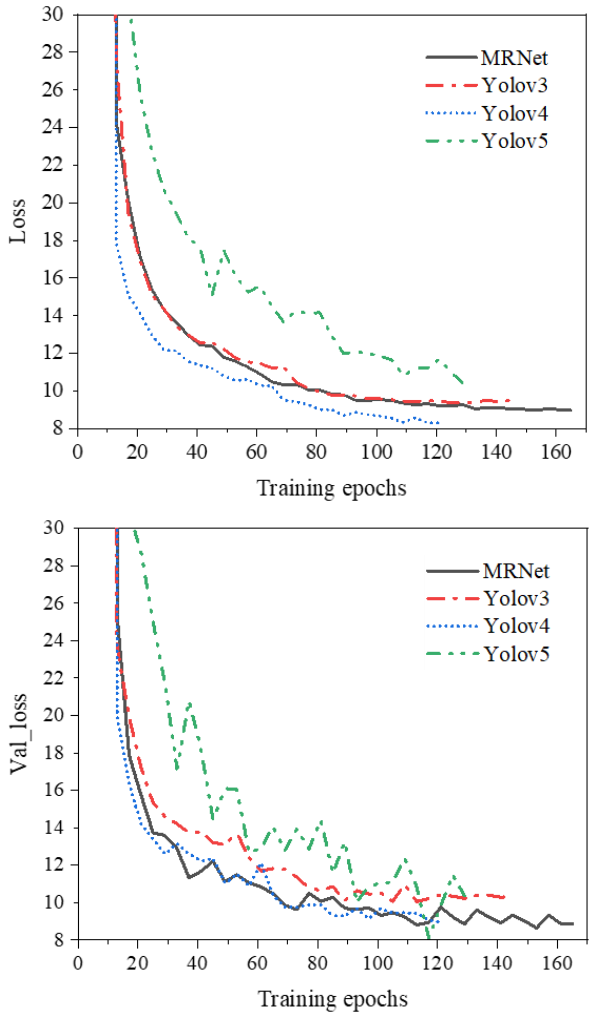


Fig. 10. The comparison of the training process of various algorithms

In the training discussed in the last paragraph, the loss convergences of different algorithms or methods are shown in Fig. 10. The YOLOv3, YOLOv4, YOLOv5(L), and MRNet carried out 145, 121, 129, and 165 epochs of iterations, respectively. Among the training of these algorithms, the losses based on the corresponding training sets converged to 9.52, 8.31, 10.59, and 8.97, respectively. Meanwhile, the losses based on the corresponding validation sets converged to 10.33, 8.98, 10.36, and 8.87, respectively. It can be easily concluded that the loss convergence of MRNet was more stable, and the difference between its two types of losses was minor, which was beneficial to obtain applicable convolutional parameters.

D. Comparisons and Discussions

To make the comparisons more comprehensive, five indicators were selected to quantitatively evaluate the performances of individual algorithms respectively to the data discussed in Section 4.3, i.e., recall, accuracy (Ac), precision (Pr), parameter size (PARAMs), and floating point operations (FLOPs). It should be noted that when calculating recall, Ac, and Pr, the IoU and confidence thresholds are set to 0.25 and 0.3, respectively.

To better explain recall, Ac, and Pr in the context of evaluation metrics for identification algorithms, it is necessary to first define some concepts related to the confusion matrix. The confusion matrix categorizes the classification results of the model into four categories: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). It should be noted that in the experiments focusing on the identification of a single class, the True Negative (TN) value is inherently fixed at 0.

The definitions of the evaluation metrics are as follows:

$$\begin{cases} recall = \dfrac{TP}{TP+FN} \\ Ac = \dfrac{TP+TN}{TP+TN+FP+FN} \\ \Pr = \dfrac{TP}{TP+FP} \end{cases}$$

Moreover, supplementary experiments in four special scenarios were conducted to validate the performance of MRNet in severe environments. Specifically, the performances of typical algorithms were tested on radar images with different resolutions. Then, an experiment was designed to test the actual performance of different algorithms on specific evaluation indicators. In addition, ablation experiments aiming to analyze the effectiveness of each component of MRNet were conducted. In the final step, all the data from these experiments were carefully analyzed.

(1) Different resolutions

In the research, the input radar images of the training process were 416×416 pixels. Therefore, when images with different sizes are input into the algorithm, the accuracy might decrease since re-size operations cause distortions. Considering that several kinds of image sizes are existing in Radar3000, it is necessary to evaluate the adaptability of the MRNet to identify images with different sizes. Specifically, according to the references, YOLOv3, YOLOv4 and YOLOv5(L) used 416×416, 608×608, and 800×800 as their default input sizes, respectively. Therefore, this experiment also used these sizes for comparisons. As shown in Fig. 11, with the increasing size of the input images, these algorithms' inference time also increased while reducing the accuracies of ship blips' identification. Compared with the other algorithms, the MRNet showed many advantages in the validation, which included an observably higher accuracy and relatively shorter inference time. In particular, the proposed MRNet was capable of adapting inputs of different resolutions.

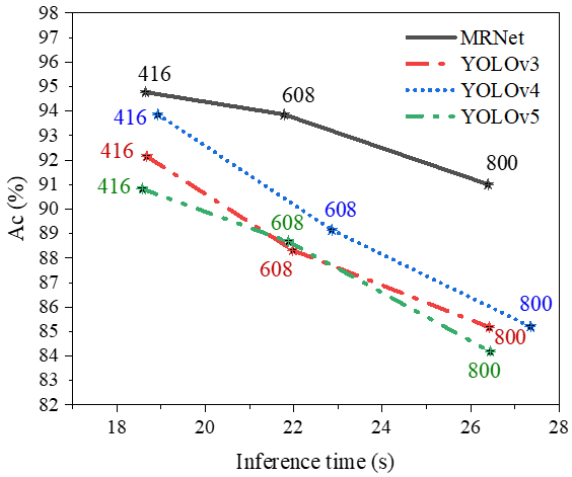


Fig. 11. The changes in identification accuracies of various algorithms under different resolutions

(2) Further discussions

To make the comparisons more comprehensive, more indicators are discussed as follows. The algorithms were all trained and tested using the same initial training coefficients and the same dataset. As shown in Table 1, the recall, identification accuracy, and precision of the proposed MRNet on the testing images reached 0.9663, 0.9418, and 0.9267, respectively. Compared with other algorithms, the recall and accuracy of MRNet were improved at least by 0.33% and 0.28%, respectively, indicating that the MRNet has a more satisfactory performance for crowded waterways and small-scale ships in marine radar images. Meanwhile, the parameter size and the computational consumption of the MRNet are significantly less than the ones of the other state-of-art methods, which are only 34.41M and 21.55G, respectively, less than 45% and 67.29% than the classical YOLOv3.

The traditional method composed of the CV and the GHFilter achieved 0.8910, 0.8815, and 0.8744 in terms of recall [43]-[45], identification accuracy, and precision, respectively for the ship identification, which were 7.53%, 6.03%, and 5.23% lower than MRNet. The traditional method cannot effectively extract the ship edge features due to the small-scale ships with fewer pixels information and the background interference, resulting in the inability to distinguish the target features and background information. Therefore, the traditional method is not credible for ship identification under radar images in many scenarios. This experiment proves that the methods based on depth convolutional neural networks represent a significant advantage for ship identification under marine radar images.

Compared with YOLOv3, YOLOv4, YOLOv5(L), and YOLOv8(L) [46], the recall of MRNet increased by 4.46%, 4.54%, 5.2% and 5.14% respectively on Radar3000, meaning that it performed better than these methods in an overall identification. Given that the dataset only includes labeled and trained samples of a single ship type, the recall values further indicate that MRNet exhibits fewer false negatives in ship identification. Additionally, the proposed method demonstrates a satisfactory capability to discriminate against false targets,

TABLE I
THE EXPERIMENTAL RESULTS OF VARIOUS ALGORITHMS

Algorithms	Recall	Ac	Pr	PARAMs/(M)	FLOPs/(G)
CV+GHFilter	0.8910	0.8815	0.8744	N/A	N/A
YOLOv3	0.9217	0.9195	0.9233	62.57	65.88
YOLOv4	0.9209	0.9286	0.9225	63.94	59.87
YOLOv5(L)	0.9143	0.9082	0.9044	47.39	108.09
YOLOv8(L)	0.9149	0.9107	0.93	43.3	165.7
Faster R-CNN (ResNet50)	0.9277	0.9182	0.9274	48.7	292.9
Mask R-CNN (ResNet50)	0.9118	0.9196	0.9393	61.3	372.2
Swin Transformer(T)	0.88	0.909	0.9361	28.4	4.71
YOLOv3-Eff_B3	0.9213	0.9105	0.9201	60.31	40.57
YOLOv3-Eff_B4	0.9396	0.9390	0.9235	81.96	52.13
YOLOv3-Eff_B5	0.9362	0.9331	0.9247	115.44	75.11
YOLOv3-Eff_B6	0.9206	0.8814	0.8878	151.41	98.25
YOLOv4-Eff_B4	0.9311	0.9104	0.9219	133.58	84.14
TinyYOLO-Eff_B4	0.8908	0.8835	0.8892	56.49	41.37
MRNet_FPN	0.9530	0.9249	0.9293	85.30	56.84
MRNet	**0.9663**	**0.9418**	**0.9267**	**34.41**	**21.55**

effectively mitigating disturbances such as noise, islands, reefs, and clouds. Furthermore, MRNet surpasses the YOLO series in terms of convolutional parameters and real-time computational efficiency, suggesting its potential applicability in embedded systems.

Compared with standard two-stage identification algorithms such as Faster R-CNN [47] and Mask R-CNN, MRNet achieves a recall improvement of nearly 4%, indicating a lower loss rate of ships. This finding further underscores that an excessive number of convolutions or excessively deep convolutional layers may potentially impede the feature extraction and target localization of ships in radar images. Additionally, MRNet has a more significant advantage in terms of model size and computational complexity compared to the two-stage algorithms. Meanwhile, Swin Transformer [48], a novel Transformer model, has been extensively applied in various identification fields and has achieved good experimental results. However, for marine radar images, the algorithm exhibits a higher loss rate in ship detection, resulting in lower recall. The analysis suggests that for small-scale or even tiny ships, the feature resolutions of Swin Transformer may be insufficient to capture subtle target characteristics. The details and edge information of small objects may become blurred or lost, leading to a decrease in identification performance.

Since the YOLO series are the pioneering work of anchor-based identification algorithms, much work had been conducted to improve their performance. The ones which adopt EfficientNet-based structures as the feature extraction network are often used as widely-acknowledged baselines to evaluate a new identification algorithm. In this research, the EfficientNet_B3~B6, namely Eff_B3~B6, is adopted to replace the original networks in various YOLO algorithms, providing 6 individual supplementary baselines. Moreover, such baselines can be used to find which subtype of EfficientNet is more suitable for radar images. Notably, as depicted in Table 1, that algorithms based on Eff_B4 and Eff_B5 demonstrate relatively remarkable ship identification capabilities. The Eff_B4-based network is better in terms of recall and accuracy, which indicates that it can distinguish ships from easily-confused noise more efficiently. However, the Eff_B5-based network performs better in the aspect of identification accuracy. In addition, the convolutional parameters of Eff_B4 are about 40% less than that of Eff_B5, as well as the computation. Consequently, as a result of these findings, it can be concluded that the Eff_B5 and Eff_B6 networks, despite their increased number of convolutional layers and parameters, falls short of that achieved by the Eff_B4 network.

Moreover, Eff_B4 is used as the feature network of YOLOv4 for comparison, meanwhile validating Eff_B4's performance. It can be found that the identification performance of the

TABLE II
THE EXPERIMENTAL RESULTS OF VARIOUS ALGORITHMS

Methods	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
YOLOv3+Eff_B4	★					
+SPP		★				
+WBF			★			
+SIoU				★		
+LightFPN					★	
+Training optimizations						★
Recall	0.9396	0.9448	0.9474	0.9498	0.9621	0.9663
Ac	0.9390	0.9324	0.9369	0.9375	0.9391	0.9418
Pr	0.9235	0.9247	0.9252	0.9258	0.9258	0.9267

modified YOLOv4 associated with Eff_B4 is neck to neck with that of YOLOv3-Eff_B4, which is about 62% higher than that of YOLOv3-Eff_B4 in terms of the parameter size and the calculational consumption. However, experiments showed that the YOLOv4-Eff_B4 was easy to fall into local optimal solutions, which leads to interruption problems in training. In summary, it can be concluded that YOLOv3-Eff_B4 performs perfectly in all aspects of ship identification under marine radar images. Theoretically speaking, Eff_B4 employs multiple convolutional layers with varying depths and widths, each of which can extract target features of different scales. Through cross-layer connections, these layers are combined to enable the network to capture detailed information from images at multiple scales, thereby enhancing its ability to identify small-scale radar blips. Evidently, the experimental results further validate this theoretical proposition. Considering the aforementioned factors, YOLOv3-Eff_B4 can be used as an appropriate benchmark for the MRNet.

To validate the performance of the proposed LightFPN, we slightly modified the FPN network. In this modification, the prediction structure had been simplified, where the prediction channel with the smallest receptive field is removed, testing the corresponding influences of the cutting channel. After this modification, a so-called TinyYOLO-Eff_B4 algorithm is constructed. The results showed that the overall performance of TinyYOLO-Eff_B4 is significantly decreased compared with YOLOv3-Eff_B4 on Radar3000. Surprisingly, the recall and accuracy had reduced by about 5%. The findings suggest that the reduction in prediction channels has a pronounced negative impact on the identification process. Conversely, the reduction in prediction channels does not significantly affect the size of convolutional parameters or computational overhead. Consequently, the retention of all three prediction channels remains crucial in the context of this particular application.

To make the comparisons more persuasive, we designed an algorithm named MRNet_FPN, which still uses standard FPN consisting of three prediction channels. In this experiment, it can be found that LightFPN performed better than FPN. Consequently, standard convolutions may inhibit the feature expressions of the ship targets, which will have a negative impact on the identification. The LghtFPN network, on the other hand, uses the DSCs to streamline the convolutional calculation of image features, preserving more precise ships' edge information and enhancing the identification precision of small-scale ships with fewer features in radar images. Besides, Compared with FPN, the convolutional parameter size, and computation of LightFPN were reduced by 59.67% and 62.09%, respectively.

(3) Ablation Experiments

To analyze the individual performances or influences of each component of MRNet comprehensively, the ablation experiments had conducted based on Radar3000. Meanwhile, all the kernel components of the MRNet were replaced with a traditional module or removed one by one. The results are shown in Table 2, and the discussions are as follows.

This research explores the adaptability of the SPP module in the identification of marine radar images. The results showed that this method increased the recall and identification accuracy of ships by 0.52% and 0.12%, respectively, which indicated that the method enhanced the ability to identify positive samples and decreased the probabilities of mistaking noise for ships. In addition, it should be noted that the SPP can significantly ameliorate the training of the MRNet since this module can improve the search capability of the globally optimal solutions and accelerate convergence. Especially for epochs near the end of the training, the calculated loss can still decrease steadily. Furthermore, the ships' movement trajectories in the images are very different. Particularly for ships moving slowly, image processing seriously dilutes their feature information, making them difficult to distinguish from other interference in the marine environment. The deep convolution output will splice feature maps of various resolutions after the introduction of SPP, which has a good feature enhancement on low-speed ships

in radar images. As a side note, the SPP can integrate global and local features in many experiments, which is essential for reducing overfitting issues during the training process.

In addition, we carried out an analysis of the prediction module. In this structure, this research modified the two modules, i.e., the NMS calculation and the loss calculation of ship positioning. Specifically, the WBF method is used to optimize the candidate results of prediction boxes, meanwhile, the SIoU function is introduced to improve the calculation of the positioning loss of prediction boxes. In the experiments, the WBF method improved the recall, accuracy, and precision indicators by 0.26%, 0.45%, and 0.05%, respectively, which made the prediction boxes more accurate and further improved the overall identification performance in these experiments. In particular, the SIoU function takes a similar influence as the WBF method, which increased the recall by 0.24%. In addition, these experiments also proved that the SIoU method accelerated the training process of MRNet by 11 epochs.

The experimental analysis for the lightweight feature fusion network is as follows. As mentioned previously, the DSCs had been used as the feature fusion structure, which might significantly reduce the convolutional parameters and computational consumption. As shown in Table 2, the application of LightFPN significantly improved the performance of the MRNet, which increased the recall by 1.23%, meanwhile the identification accuracy also increased slightly. Compared with the standard FPN, it can be found that the increase of invalid convolutional parameters had observably reduced the ability to extract ships. The results shown in Table 2 also indicate that the DSC structure is different from the classical methods of compressing identification modules, which not only cuts off redundant convolutional parameters but also improves the ship identification performance in various scenarios.

While the experimental evaluations have demonstrated the relatively superior performance of LightFPN compared to the standard FPN, it remains to be ascertained how these models will fare when applied to real radar images. The results depicted in Fig. 12 indicate that the ships depicted in region 1, characterized by yellow radar blips with a blue tail, possess discernible visual attributes, and both types of feature fusion networks have successfully captured these features. However, when confronted with small ships in regions 2 and 3, the standard FPN exhibits inadequate performance, leading to omission errors. In contrast, the MRNet, leveraging the accurate recognition capability for small-scale detailed features of LightFPN, can precisely locate and identify small targets within radar images.

To improve the performance of the MRNet, a variety of training optimization methods are used. As mentioned previously, several methods or tricks have been adopted as discussed in Section 4.2. As shown in Table 2, these training tricks increase recall by 0.42%. Meanwhile, the identification accuracy also raised a little bit. Perhaps the significant reason is that a step-by-step training process might be partly helpful in avoiding local optima and accelerating the convergences.

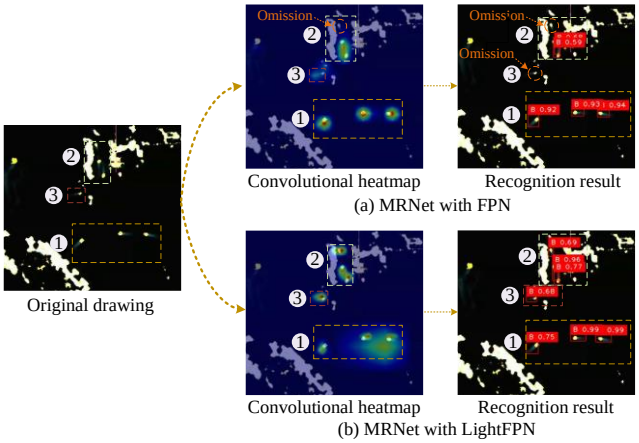


Fig. 12. The comparisons of experimental results of LightFPN and FPN

(4) Comparisons in daily operations

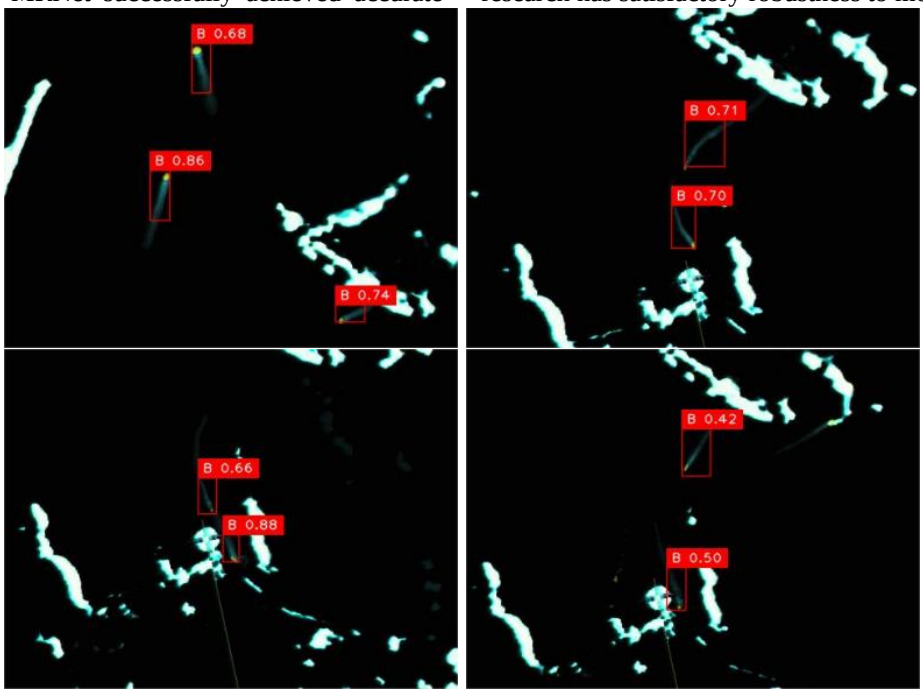
To validate the performances of the MRNet in real applications, more experiments of identifying ships in daily scenarios were carried out. The identification performances of the previously mentioned algorithms when recognizing high-speed and low-speed ships were evaluated respectively. In particular, typical extreme scenarios such as tiny ships, crowded ships, overtaking, and crossing were selected for further comparisons.

Several typical results of the experiments are shown in Fig. 13, which demonstrated the adaptability of the MRNet in different scenarios. As shown in Fig. 13(a), MRNet accurately identified high-speed ships in different waters, which proved that the algorithm had a satisfactory performance when identifying high-speed ships with distinguishable visual features, even if the imaging of the corresponding ship blips was not stable or flickering. Especially, the MRNet achieved promising accuracy when identifying low-speed ships in different environments shown in Fig. 13(b). Notably, in the majority of cases, there was minimal occurrence of omission or misidentification when it came to small-scale ships. This observation suggests that MRNet exhibits a robust capability in extracting features from small targets while effectively suppressing disturbances such as shorelines, reefs, and noise. Furthermore, MRNet demonstrates a somewhat favorable ability to accurately identify and locate individual targets within dense ship scenarios, surpassing conventional fractal-geometry-based methods, as evidenced in Fig. 13(c). Despite the relatively small receptive fields employed by MRNet, it successfully captures the inherent visual features and positional information of ship blips when dealing with small targets.

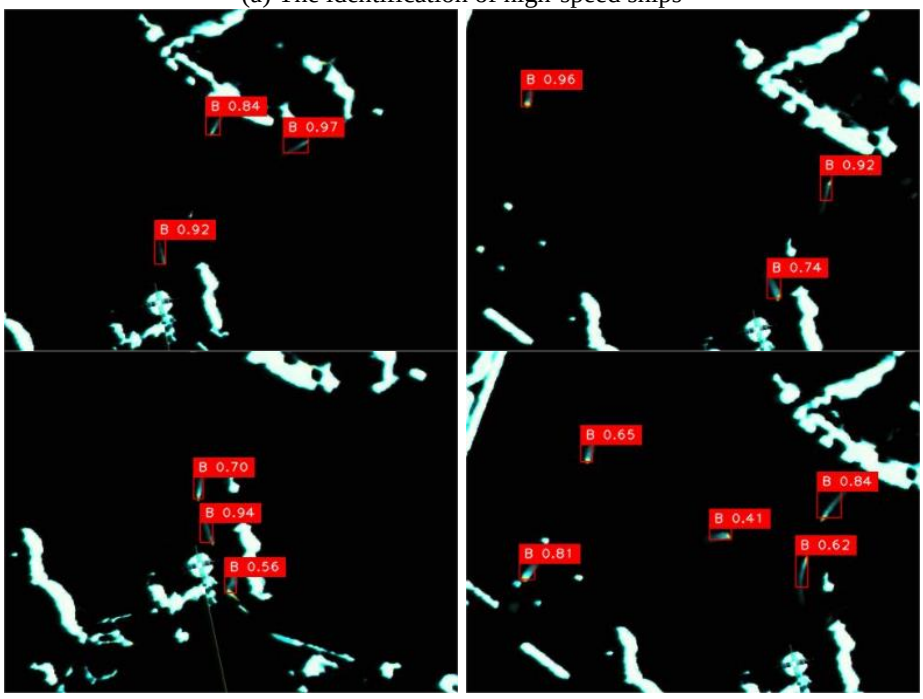
(5) Performances under Interference

As shown in Fig. 14, the original images contain three types of objects, i.e., low-speed ships, high-speed ships, and crowded ships, which are also the most common or challenging scenarios in daily operations. To test the ship identification ability of the MRNet under different interference, raw radar images were processed using three types of interference signals, including Pretzel noises, and Gauss noises, Speckle noises, to simulate the

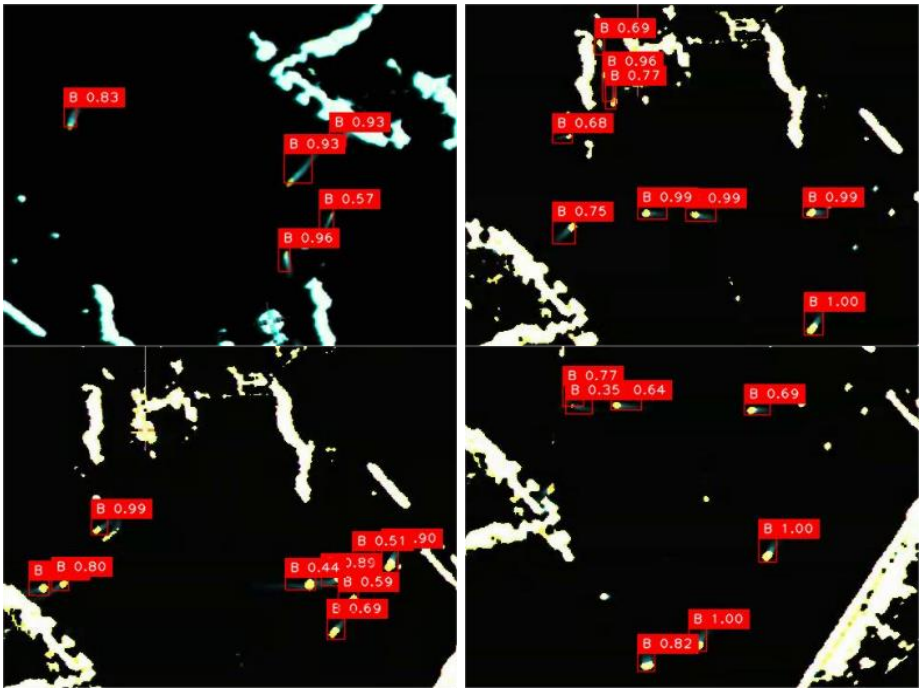
influence of different interference. It can be found from Fig. 14 that Speckle noises had the more significant impact on the radar images, and the pixel characteristics of the ships and other objects are severely disturbed. In contrast, Pretzel noises and Gauss noises had less impact on the original feature information. As anticipated, the experimental results demonstrate that the MRNet successfully achieved accurate ship identification in the presence of both Pretzel noises and Gauss noises, showcasing its enhanced robustness to interference. However, the MRNet exhibited a relatively lower adaptability to Speckle noises, as indicated by a slight decrease in ship identification accuracy under such interference, as illustrated in Fig. 14. In summary, the approach proposed in this research has satisfactory robustness to most interference.



(a) The identification of high-speed ships

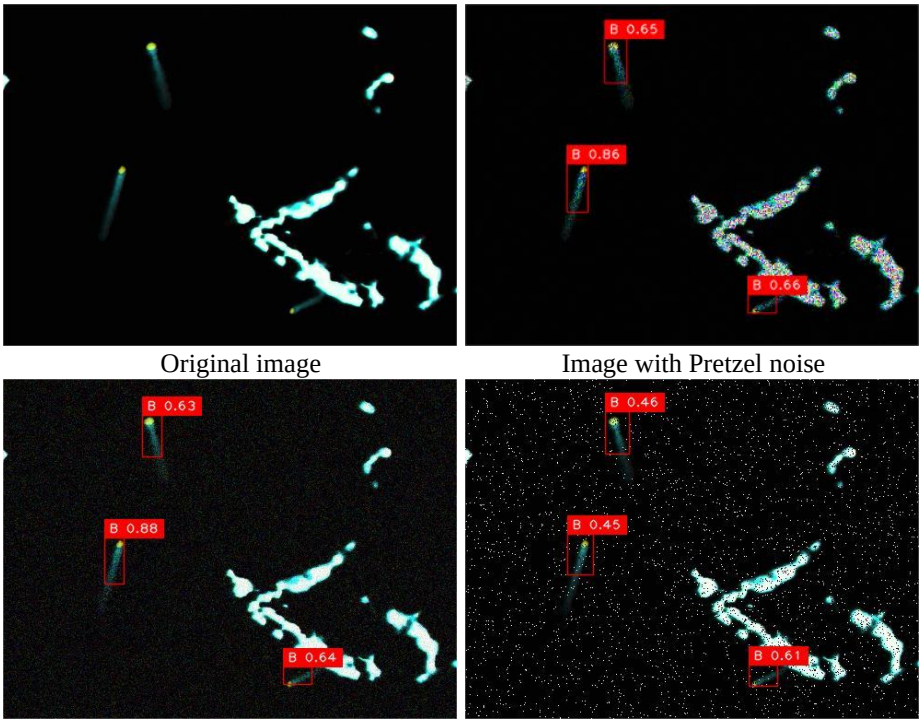


(b) The identification of low-speed ships



(c) The identification of crowded ships

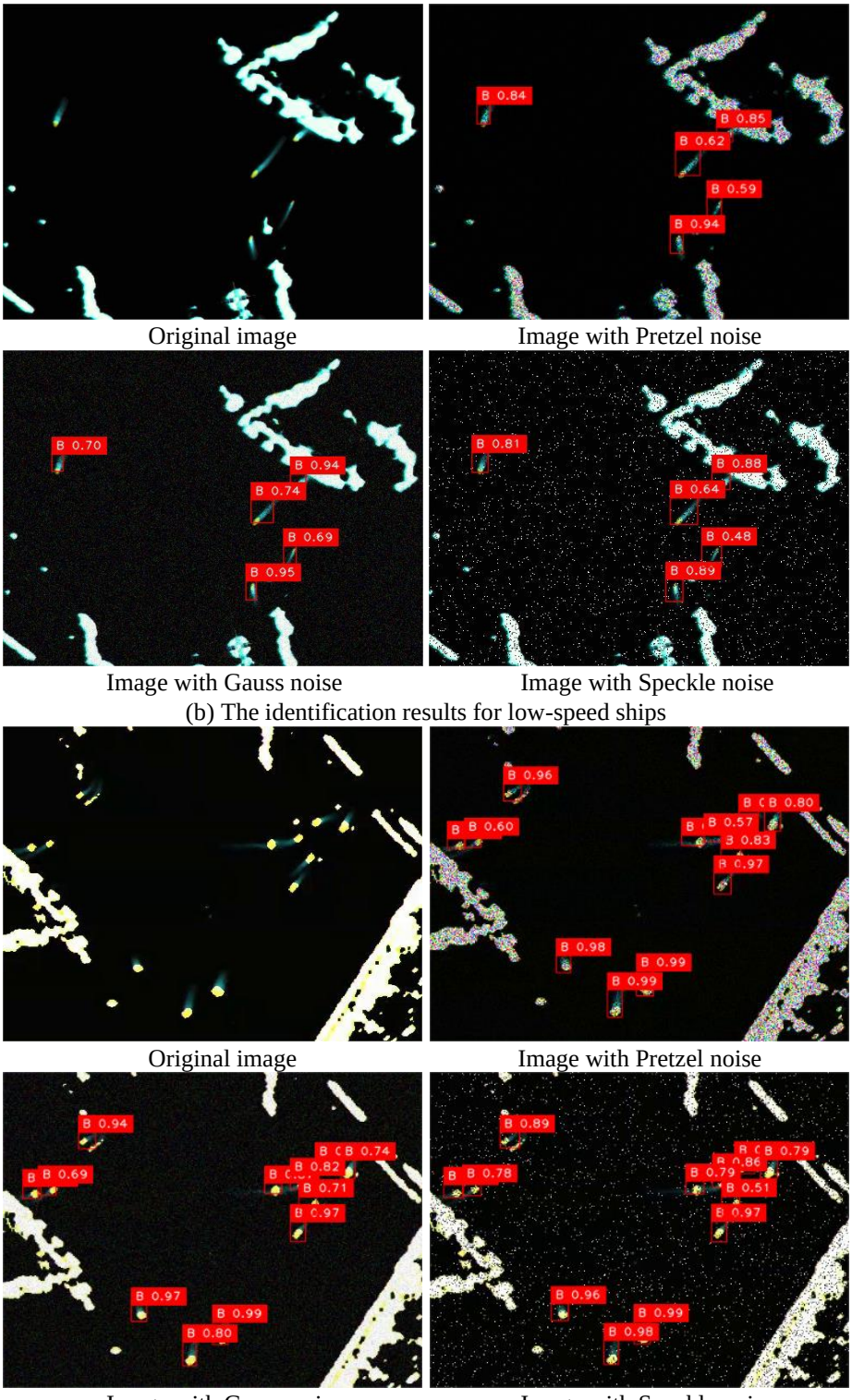
Fig. 13. The ship identification results of MRNet under different scenarios



Original image | Image with Pretzel noise

Image with Gauss noise | Image with Speckle noise

(a) The identification results for high-speed ships



(b) The identification results for low-speed ships

(c) The identification results for dense ships

Fig. 14. The identification comparisons of the MRNet under various noise

(6) Comparisons in extreme scenarios

To validate the performances of the MRNet in extreme scenarios, several supplementary experiments were conducted. To lay the foundation for these experiments, a dedicated sub-dataset of Radar3000, which only collects images from extreme scenarios, is built. In such scenarios, all difficult-to-identify situations, such as crossing and overtaking, are deliberately selected. In these experiments, the commonly-used YOLOv3, YOLOv4, and YOLOv5(L) were chosen as references since they were widely acknowledged as superior tools for identifying small objects [49],[50]. The results are illustrated in Table 3. It can be found that the MRNet performed better than the comparison algorithms in all aspects, proving that MRNet

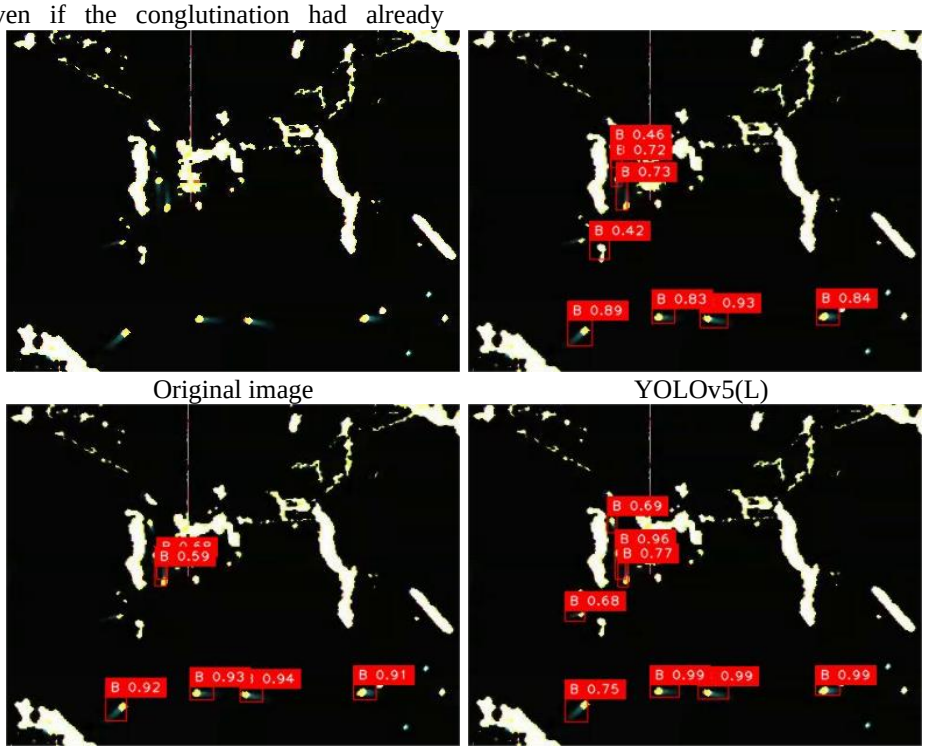
is more effective for ship identification in extreme scenarios.

TABLE III
THE SHIP IDENTIFICATIONS OF DIFFERENT METHODS IN EXTREME SCENARIOS

Algorithms	Recognized Ships	True Ships	False Alarms	Recall	Pr
YOLOv3	1592	1385	207	0.9314	0.8700
YOLOv4	1581	1390	191	0.9348	0.8792
YOLOv5	1633	1371	262	0.9220	0.8396
MRNet	1543	1423	120	0.9570	0.9222

To visualize such an experiment, several typical images of identification are further illustrated in Fig. 15. As shown in Fig. 15(a), the identification ability of YOLOv5(L) was significantly weaker than that of MRNet_FPN and MRNet when recognizing side-by-side blips. Not surprisingly, both the YOLOv5(L) and MRNet_FPN missed some small ship blips occasionally on this occasion. The difference is that MRNet could lock these missing blips in subsequent frames, whereas YOLOv5(L) could not. The scenario of ship crossing is shown in Fig. 15(b). Such a scenario is considered as a difficult problem in processing the radar images, in which ship blips might conglutinate with each other. On this occasion, the tracking of a ship blip is very easy to be disturbed. Surprisingly, the proposed approach, MRNet, identified the ship blips with a large probability even if the conglutination had already happened. This observation indicates that the method, by emphasizing the dissimilarity between the overall features of the ships and their surrounding pixels, significantly enhances the identification accuracy. In addition, the MRNet could recognize every single target from crowded blips most of the time. In comparison, the YOLOv5(L) frequently lost targets on this occasion, MRNet_FPN performed even worse where an overfitting issue might happen, as shown in Fig. 15(c). The performances of YOLOv3 and YOLOv4 will not be discussed here, since they were widely believed to be worse than YOLOv5(L) in identifying micro-objects [29],[51].

In conclusion, when recognizing ship blips in extreme scenarios, the proposed MRNet outperforms other state-of-the-art CNN-based algorithms.

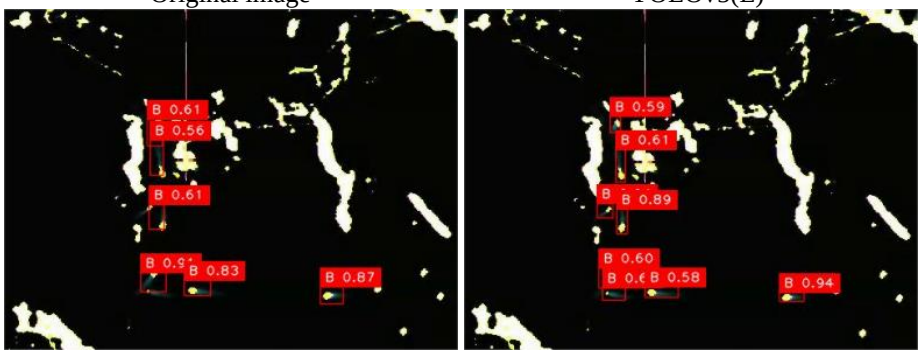


Original image YOLOv5(L)

MRNet_FPN MRNet

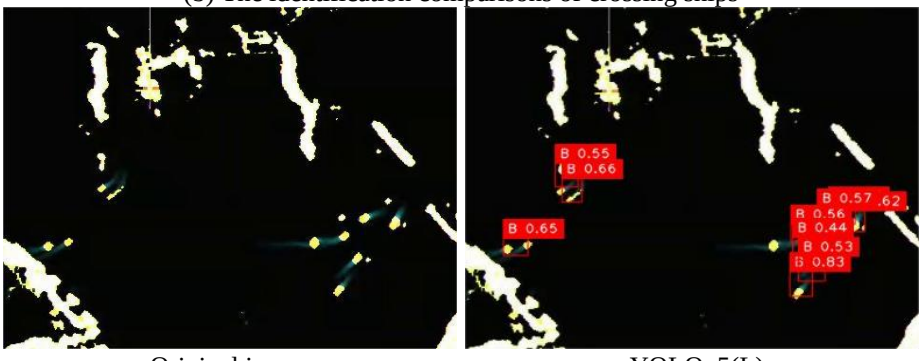
(a) The identification comparisons of micro ships

Original image YOLOv5(L)

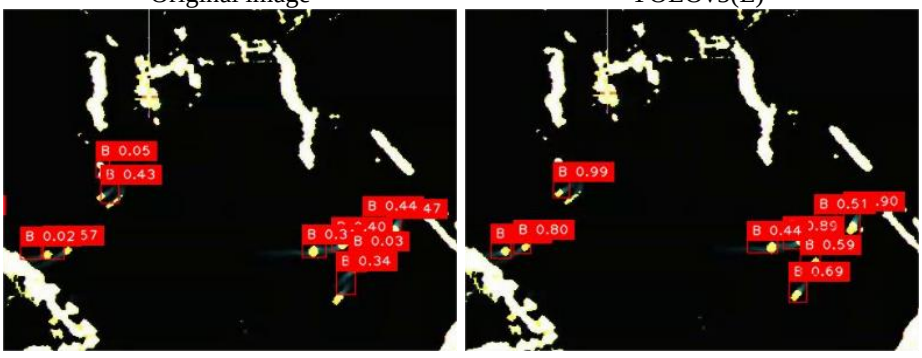


MRNet_FPN MRNet

(b) The identification comparisons of crossing ships



Original image YOLOv5(L)



MRNet_FPN MRNet

(c) The identification comparisons of dense ships

Fig. 15. The identification comparisons of ship images in extreme scenarios

V. Conclusions and Discussions

Considering the challenges associated with ship identification in marine radar images, this paper presents MRNet, a novel neural network designed to accurately identify ships. Additionally, this is the first instance of a customized neural network-based ship identification algorithm created especially for marine radar images. It is commonly believed that ships' blips in radar images have fewer contours and color features, which may lead to excessive convolution calculations and poor identification when using conventional CNN-based methods. To solve this problem, MRNet employs a variety of established network designs and convolution modules that

show superior adaptability in radar images and comparatively high accuracy in ship identification. In field testing, the MRNet achieved recall, identification accuracy, and precision scores of 0.9663, 0.9418, and 0.9267, respectively. Meanwhile, MRNet's parameter size and calculational consumption are only 34.41M and 21.55G, respectively, down 45% and 67.29% from standard YOLOv3.

Specifically, MRNet uses the EfficientNet-B4 network as the feature extraction network due to its deep convolutional structure and ability to extract features from low-resolution images, similar to that of humans. Additionally, MRNet adopts a lightweight feature fusion network that replaces some kernel convolutions with DSCs to reduce convolutional parameters and computation, significantly inhibiting excessive calculations and overfitting problems. Moreover, the SPP module is connected to the EfficientNet-B4 network to improve the adaptability of processing radar images with different input sizes. In the prediction structure, the WBF method and SIoU function are respectively utilized to optimize the calculational process of non-maximum suppression and the localization loss of predicted boxes. Besides, this work constructed a radar image dataset, namely Radar3000, containing rich ship features, which improved the identification accuracy and generalization ability of the algorithm.

In a variety of scenarios, especially in extreme ones like port waters and congested waterways, MRNet consistently demonstrates its capacity to accurately identify both high-speed and low-speed ships. As marine radar systems are often deployed in relatively harsh environments, this necessitates that the equipment itself exhibits high reliability, relatively low power consumption, etc. Therefore, MRNet, which is constructed based on developed modules, may satisfactorily meet these requirements and can be conveniently applied in existing devices. In addition, the MRNet might have a superior capability when identifying small objects with fewer pixel features, such as those in medical images and remote sensing images.

This work can be further improved in the following aspects. The Radar3000 dataset should be further enriched in the future by including more samples from other commonly-used radar systems and images from more circumstances. Moreover, it may be worth combining the proposed approach with multi-object tracking algorithms for further improving performance in identifying ships from radar blips.

APPENDIX

The marine radar image dataset, namely Radar3000, constructed in this paper has been published at the following website:

https://github.com/kz258852/dataset_M_Radar/tree/main

ACKNOWLEDGMENT

This work is financially supported by the Funds for the National Key R&D Program of China (Grant No. 2021YFB1600400), National Natural Science Foundation of China under Grant No. 52171352 and 52201415, Zhejiang Province Key R&D projects (Grant No. 2021C03015).

REFERENCES

[1] 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, vol. 235, p. 109435, Sep. 2021, doi: 10.1016/j.oceaneng.2021.109435.

[2] J. Lisowski, “Artificial Intelligence Methods in Safe Ship Control Based on Marine Environment Remote Sensing,” Remote Sensing, vol. 15, no. 1, p. 203, Dec. 2022, doi: 10.3390/rs15010203.

[3] P. Chen, X. Li, and G. Zheng, “Rapid detection to long ship wake in synthetic aperture radar satellite imagery,” Journal of Oceanology and Limnology, vol. 37, no. 5, pp. 1523–1532, Dec. 2018, doi: 10.1007/s00343-019-8221-y.

[4] J. Han, J. Kim, and N. Son, “Coastal Navigation with Marine Radar for USV Operation in GPS-Restricted Situations,” Journal of Institute of Control, Robotics and Systems, vol. 24, no. 8, pp. 736–741, Aug. 2018, doi: 10.5302/j.icros.2018.0087.

[5] B. Wen, Y. Wei, and Z. Lu, “Sea Clutter Suppression and Target Detection Algorithm of Marine Radar Image Sequence Based on Spatio-Temporal Domain Joint Filtering,” Entropy, vol. 24, no. 2, p. 250, Feb. 2022, doi: 10.3390/e24020250.

[6] C. Wu, Q. Wu, F. Ma, and S. Wang, “A novel positioning approach for an intelligent vessel based on an improved simultaneous localization and mapping algorithm and marine radar,” Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment, vol. 233, no. 3, pp. 779–792, Jul. 2018, doi: 10.1177/1475090218784449.

[7] F. Ma, Y. Chen, X. Yan, X. Chu, and J. Wang, “Target recognition for coastal surveillance based on radar images and generalised Bayesian inference,” IET Intelligent Transport Systems, vol. 12, no. 2, pp. 103–112, Dec. 2017, doi: 10.1049/iet-its.2017.0042.

[8] O. Karabayir, U. Saynak, M. Z. Kartal, A. F. Coskun, T. O. Gulum, and B. Bati, “Synthetic Range Profile-Based Training Library Construction for Ship Target Recognition Purposes of Scanning Radar Systems,” IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 4, pp. 3231–3245, 2020, doi: 10.1109/taes.2020.2972249.

[9] X. MOU, X. Chen, J. Guan, W. Zhou, L. Liu, and Y. Dong, “Clutter Suppression and Marine Target Detection for Radar Images Based on INet,” Journal of Radars, vol. 9, no. 4, Aug. 2020, doi: 10.12000/JR20090.

[10] X. Chen, X. Mu, J. Guan, N. Liu, and W. Zhou, “Marine target detection based on Marine-Faster R-CNN for navigation radar plane position indicator images,” Frontiers of Information Technology & Electronic Engineering, vol. 23, no. 4, pp. 630–643, Apr. 2022, doi: 10.1631/fitee.2000611.

[11] X. Chen, N. Su, Y. Huang, and J. Guan, “False-Alarm-Controllable Radar Detection for Marine Target Based on Multi Features Fusion via CNNs,” IEEE Sensors Journal, vol. 21, no. 7, pp. 9099–9111, Apr. 2021, doi: 10.1109/jsen.2021.3054744.

[12] D. Mao, Y. Zhang, Y. Zhang, J. Pei, Y. Huang, and J. Yang, “An Efficient Anti-Interference Imaging Technology for Marine Radar,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–13, 2022, doi: 10.1109/tgrs.2021.3068787.

[13] C. Zhang, M. Fang, C. Yang, R. Yu, and T. Li, “Perceptual Fusion of Electronic Chart and Marine Radar Image,” Journal of Marine Science and Engineering, vol. 9, no. 11, p. 1245, Nov. 2021, doi: 10.3390/jmse9111245.

[14] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” p. arXiv preprint arXiv:1804.02767, Apr. 2018, doi: 10.48550/arXiv.1804.02767.

[15] A. Bochkovskiy, C. Wang and H. Liao, “Yolov4: Optimal speed and accuracy of object detection,” p. arXiv preprint arXiv:2004.10934, Apr. 2020, doi: 10.48550/arXiv.2004.10934.

[16] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 386-397, 2020, doi: 10.1109/tpami.2018.2844175.

[17] Z. Yu, X. Wang, X. Zhao, L. Wang, L. Wang, and S. Wang, “AT-Cascade R-CNN: a novel attention-based cascade R-CNN model for ovarian cancer lesion identification,” International Journal of Adaptive and Innovative Systems, vol. 3, no. 1, pp. 74-86, Sept. 2021, doi: 10.1504/ijais.2021.117913.

[18] W. Zhou and L. Liu, “Multilayer attention receptive fusion network for multiscale ship detection with complex background,” Journal of Electronic Imaging, vol. 31, no. 04, Jul. 2022, doi: 10.1117/1.jei.31.4.043029.

[19]F. Zhang and X. Hou, “Multi-Site and Multi-Scale Unbalanced Ship Detection Based on CenterNet,” Electronics, vol. 11, no. 11, p. 1713, May 2022, doi: 10.3390/electronics11111713.
[20]W.-J. Lee et al., “Detection and tracking for the awareness of surroundings of a ship based on deep learning,” Journal of Computational Design and Engineering, vol. 8, no. 5, pp. 1407–1430, Sep. 2021, doi: 10.1093/jcde/qwab053.
[21]X. Zhang, X. Yang, D. Yang, F. Wang, and X. Gao, “A Universal Ship Detection Method With Domain-Invariant Representations,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–11, 2022, doi: 10.1109/tgrs.2022.3200957.
[22]S. Chen, R. Zhan, W. Wang, and J. Zhang, “Domain Adaptation for Semi-Supervised Ship Detection in SAR Images,” IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1–5, 2022, doi: 10.1109/lgrs.2022.3171789.
[23]T. Miao et al., “An Improved Lightweight RetinaNet for Ship Detection in SAR Images,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4667–4679, 2022, doi: 10.1109/jstars.2022.3180159.
[24]Y. Wang, H. Shi, and L. Chen, “Ship Detection Algorithm for SAR Images Based on Lightweight Convolutional Network,” Journal of the Indian Society of Remote Sensing, vol. 50, no. 5, pp. 867–876, Jan. 2022, doi: 10.1007/s12524-022-01491-1.
[25]S. Li, X. Fu, and J. Dong, “Improved Ship Detection Algorithm Based on YOLOX for SAR Outline Enhancement Image,” Remote Sensing, vol. 14, no. 16, p. 4070, Aug. 2022, doi: 10.3390/rs14164070.
[26]C. Zhao, X. Fu, J. Dong, R. Qin, J. Chang, and P. Lang, “SAR Ship Detection Based on End-to-End Morphological Feature Pyramid Network,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 4599–4611, 2022, doi: 10.1109/jstars.2022.3150910.
[27]T. Zhang et al., “Region-Based Polarimetric Covariance Difference Matrix for PolSAR Ship Detection,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–16, 2022, doi: 10.1109/tgrs.2022.3146385.
[28]X. Qi, P. Lang, X. Fu, R. Qin, J. Dong, and C. Liu, “A regional attention-based detector for SAR ship detection,” Remote Sensing Letters, vol. 13, no. 1, pp. 55–64, Oct. 2021, doi: 10.1080/2150704x.2021.1987574.
[29]L. Pang, B. Li, F. Zhang, X. Meng, and L. Zhang, “A Lightweight YOLOv5-MNE Algorithm for SAR Ship Detection,” Sensors, vol. 22, no. 18, p. 7088, Sep. 2022, doi: 10.3390/s22187088.
[30]G. Dong and H. Liu, “A New Model-Data Co-Driven Method for Radar Ship Detection,” IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–9, 2022, doi: 10.1109/tim.2022.3169573.
[31]F. Zhang, X. Wang, S. Zhou, Y. Wang, and Y. Hou, “Arbitrary-Oriented Ship Detection Through Center-Head Point Extraction,” IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–14, 2022, doi: 10.1109/tgrs.2021.3120411.
[32]T. Xie, M. Liu, M. Zhang, S. Qi, and J. Yang, “Ship detection based on a superpixel-level CFAR detector for SAR imagery,” International Journal of Remote Sensing, vol. 43, no. 9, pp. 3412–3428, May 2022, doi: 10.1080/01431161.2022.2091966.
[33]Y. Yin, X. Cheng, F. Shi, M. Zhao, G. Li, and S. Chen, “An Enhanced Lightweight Convolutional Neural Network for Ship Detection in Maritime Surveillance System,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 5811–5825, 2022, doi: 10.1109/jstars.2022.3187454.
[34]G. Deng et al., “A Low Coupling and Lightweight Algorithm for Ship Detection in Optical Remote Sensing Images,” IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1–5, 2022, doi: 10.1109/lgrs.2022.3188850.
[35]M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks", arXiv:1905.11946, 2019, [online] Available: http://arxiv.org/abs/1905.11946.
[36]R. Solovyev, W. Wang, and T. Gabruseva, “Weighted boxes fusion: Ensembling boxes from different object detection models,” Image and Vision Computing, vol. 107, p. 104117, Mar. 2021, doi: 10.1016/j.imavis.2021.104117.
[37]Z. Gevorgyan, “SIoU loss: more powerful learning for bounding box regression,” p. arXiv preprint arXiv:2205.12740, May. 2022, doi: 10.48550/arXiv.2205.12740.
[38]A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, et al., "Searching for MobileNetV3", Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), pp. 1314-1324, Oct. 2019.
[39]T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2117-2125, Jul. 2017.
[40]A. Karacı, “VGGCOV19-NET: automatic detection of COVID-19 cases from X-ray images using modified VGG19 CNN architecture and YOLO algorithm,” Neural Computing and Applications, vol. 34, no. 10, Jan. 2022, doi: 10.1007/s00521-022-06918-x.
[41]H. Lin, W. Zeng, Y. Zhuang, X. Ding, Y. Huang, and J. Paisley, “Learning Rate Dropout,” IEEE Transactions on Neural Networks and Learning Systems, no. 99, pp. 1–11, 2022, doi: 10.1109/tnnls.2022.3155181.
[42]C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision", Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 2818-2826, Jun. 2016.
[43]Maritime navigation and radio-communication equipment and systems. Shipborne radar. Performance requirements, methods of testing and required test results, IEC Standard 62388, 2013.
[44]F. Ma, Q. Wu, X. Yan, X. Chu, and D. Zhang, “Classification of Automatic Radar Plotting Aid targets based on improved Fuzzy C-Means,” Transportation Research Part C: Emerging Technologies, vol. 51, pp. 180–195, Feb. 2015, doi: 10.1016/j.trc.2014.12.001.
[45]F. Ma, Y. Chen, X. Yan, X. Chu, and J. Wang, “A novel marine radar targets extraction approach based on sequential images and Bayesian Network,” Ocean Engineering, vol. 120, pp. 64–77, Jul. 2016, doi: 10.1016/j.oceaneng.2016.04.030.
[46]G. Jocher. “YOLO by Ultralytics,” https://github.com/ultralytics/ultralytics, 2023. Accessed: September 21, 2023.
[47]S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/tpami.2016.2577031.
[48]Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 9992-10002, doi: 10.1109/ICCV48922.2021.00986.
[49]Y. Guo, S. Chen, R. Zhan, W. Wang, and J. Zhang, “LMSD-YOLO: A Lightweight YOLO Algorithm for Multi-Scale SAR Ship Detection,” Remote Sensing, vol. 14, no. 19, p. 4801, Sep. 2022, doi: 10.3390/rs14194801.
[50]S. Wang et al., “YOLO-SD: Small Ship Detection in SAR Images by Multi-Scale Convolution and Feature Transformer Module,” Remote Sensing, vol. 14, no. 20, p. 5268, Oct. 2022, doi: 10.3390/rs14205268.
[51]G. Jocher. “YOLOv5 by Ultralytics,” https://github.com/ultralytics/yolov5, 2020. Accessed: September 21, 2023.

Feng Ma received the Ph.D. degree in Vehicle Operation Engineering from Wuhan University of Technology, Wuhan, PR China, in 2013. He is currently a professor with the Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan, China. His research interests include image processing, intelligent ships, reinforcement learning, computer vision and other related researches.

Zhe Kang received the M.Eng degree from the School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan, PR China, in 2022. He is currently pursuing the Ph.D. degree with the Intelligent Transportation Systems Research Center, Wuhan University of Technology, Wuhan, China. His research interests include computer vision, object identification, instance segmentation, and multi-objective tracking.

Chen Chen received the Ph.D. degree in Computer Science and Technology from Wuhan University of Technology, Wuhan, PR China, in 2020. She was a visiting scholar at the University of Manchester in the UK from Nov. 2017 to Oct. 2018. She is currently a lecturer with the School of Computer Science and Technology, Wuhan Institute of Technology, Wuhan, China. Her research interests include artificial intelligence, deep reinforcement learning, biomimetic consciousness modeling, and vehicle autonomy.

Sun Jie received his MSc in Business Administration from Nanjing University, China. He is an engineer in Marine Engineering, also the founder of Nanjing Smart Water Transportation Technology Co., Ltd. His interests include image processing and vehicle autonomy.

Xiao-bin Xu received then Ph.D. degree in power electronics and power transmission from Shanghai Maritime University, China, in 2009.He is a professor of department of automation and China-Austria Belt and Road joint laboratory on artificial Intelligence and advanced manufacturing in Hangzhou Dianzi University. His research interests include evidence theory, fuzzy set theory and applications in the processing of uncertain information, the reliability analysis, safety evaluation and condition monitoring of complex industrial systems.

Jin Wang received his MSc in Marine Engineering and PhD in Maritime Safety from Newcastle University, UK in 1989 and 1994, respectively. He has been Professor of Marine Technology at Liverpool John Moores University (LJMU) since 2002, Associate Dean (Research) in the Faculty of Engineering and Technology of LJMU since 2015 and Director of the Liverpool Logistics, Offshore and Marine (LOOM) Research Institute since 2003. His research interests are in design and operation of large maritime engineering systems.