A Multiscale Feature Pyramid SAR Ship Detection Network With Robust Background Interference

Shuai Liu, Pengfei Chen, and Yudong Zhang, *Senior Member, IEEE*

***Abstract*—Synthetic aperture radar (SAR) ship detection is widely used in cutting-edge applications such as environmental protection, traffic monitoring, search, and rescue. Lightweight detection algorithms are more important for practical applications. Although there has been extensive research in this field, there are some problems with the existing lightweight algorithms. For example, it is easy to misjudge targets that are mixed with the background, and the detection effect is not ideal for targets with few samples in the dataset. The root cause of these problems lies in the fact that the useless information in the background is relatively close to the target, and existing algorithms are too simplistic in fusing features at different levels, resulting in algorithms not being robust enough when facing these problems. Therefore, this article proposes a multiscale feature pyramid network (FPN)-based detection network (MFPNet), which introduces a spatial information-focusing module in the feature fusion channel to enhance the target's features to suppress interference information in the background and reduce misjudgment. Then, optimize the FPN and extract the importance of different resolution features based on network contribution to identifying multiscale targets. Experiments have shown that the MFPNet has better detection performance compared to existing algorithms on public datasets.**

***Index Terms*—Feature fusion, feature pyramid network (FPN), ship detection, spatial information-focus, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) ship detection is an important research direction in the field of target detection. SAR is a radar imaging technology that can achieve high resolution in, all-weather, and at all times. In the field of ship detection, SAR technology has the following advantages.

1) *All weather:* SAR technology is not affected by weather conditions and lighting and can achieve all-weather imaging, making it suitable for application scenarios such as maritime traffic control and coastline protection.
2) *Wide coverage range:* SAR can scan the sea surface over a wide range, enabling SAR to detect and track ships in larger sea areas.
3) *Multimode imaging:* SAR can perform imaging in different imaging modes, such as squint mode and polarization mode, improving the precision and reliability of ship detection.

The lightweight SAR ship detection algorithm refers to the rapid and accurate analysis and processing of SAR images, achieving efficient ship detection and reducing the algorithm's computational complexity while ensuring detection precision. Lightweight SAR ship detection algorithms have the following advantages.

1) *Improving detection precision:* The lightweight SAR ship detection algorithm can improve detection precision and avoid issues such as missed and false detections by adopting more advanced machine learning, deep learning (DL), and other technologies [1].
2) *Improving algorithm efficiency:* The lightweight SAR ship detection algorithm can achieve fast ship detection in application scenarios with high real-time requirements, greatly improving algorithm efficiency and response speed [2].
3) *Reduce computing resources:* The lightweight SAR ship detection algorithm adopts a lightweight model structure and optimized algorithm design, reducing the consumption of computing resources and improving the practicality and adaptability of the algorithm [3].

However, there are some challenges and difficulties in the existing SAR ship detection algorithms, mainly in the following two aspects.

1) *Background clutter and interference:* There are various background clutter and interference in the imaging environment, such as shore, waves, wind waves, etc. The interference information has similarities with the target in SAR images, leading to algorithm misjudgment of the interference information.
2) The various forms of ship samples are imbalanced, and ships of different sizes exhibit different morphological features in SAR images. The deviation between the various forms of samples makes it difficult for the algorithm to detect targets with large sample deviations. At present, lightweight algorithms are not robust enough to face these two challenges. Therefore, proposing corresponding solutions to these two challenges can help improve the precision and robustness of detection algorithms.

Manuscript received 20 July 2023; revised 21 September 2023; accepted 12 October 2023. Date of publication 17 October 2023; date of current version 31 October 2023. This work was supported in part by the Natural Science Foundation of China under Grant 62207012, in part by the Key Scientific Research Projects of Department of Education of Hunan Province under Grant 22A0049, in part by 2023 Key Supported Project of the 14th Five Year Plan for Education and Science in Hunan Province under Grant ND230795, and in part by the National Social Science Foundation of China under Grant AEA200013. *(Corresponding author: Yudong Zhang.)*

Shuai Liu is with the Key Laboratory of Big Data Research and Application for Basic Education, Hunan Normal University, Changsha 410081, China (e-mail: liushuai@hunnu.edu.cn).

Pengfei Chen is with the College of Educational Science, Hunan Normal University, Changsha 410081, China (e-mail: 202120293780@hunnu.edu.cn).

Yudong Zhang is with the School of Computing and Mathematical Sciences, University of Leicester, LE1 7RH Leicester, U.K. (e-mail: yudongzhang@ieee.org).

Digital Object Identifier 10.1109/JSTARS.2023.3325376

This article proposes a lightweight multiscale SAR ship detection model. We have introduced a spatial information-focusing module to reduce the interference caused by useless background information and improve the extraction effect of target features, thereby solving the problem of misjudgment caused by background information similar to the target. Meanwhile, we optimize the structure of feature pyramid network (FPN) and propose an adaptive weighted fusion mechanism to fully utilize the importance of different resolution features fully, thereby solving the problem of low detection rate for targets with small sample sizes in the dataset. The main contributions of this article are as follows.

1) Introducing a spatial information-focusing module to guide feature extraction and solve the problem of misjudgment caused by background interference. The spatial information-focusing module weakens noise in the background through maximum pooling and global average pooling to suppress useless information in the background. This makes the algorithm pay more attention to the target's spatial location, reducing the negative impact on target feature extraction and enhancing the target features.
2) Optimizing the FPN's structure and adopting a weighted feature fusion mechanism to solve the problem of insufficient robustness in detection caused by sample bias. By learning the feature channels in the FPN network, identifying and eliminating the channels that contribute less to the network, and adding skip channels to connect at the same level to strengthen the features. Then, a weighted feature fusion mechanism is used to fuse features adaptively from different levels, making the extracted target features more sufficient.
3) By comparing it with other advanced lightweight algorithms on the public dataset HRSID, it is demonstrated that the MFPNet performs better than existing advanced lightweight algorithms in terms of four evaluation indicators: precision (P), recall (R), F1score (F1), and mean average precision (mAP). Further promoting the progress of lightweight SAR ship detection algorithms.

II. Related Work

SAR remote sensing has become one of the most important means of ocean monitoring due to its all-weather advantages. Ship detection in SAR images has broad prospects in both military and civilian fields. SAR ship detection algorithms mainly include traditional detection algorithms and DL-based detection algorithms. With the improvement of detection performance, lighter SAR ship detection algorithms have broad application prospects.

The traditional SAR ship detection algorithm mainly relies on the following aspects.

1) *SAR ship detection method based on binary segmentation [4]:* This method divides the SAR image into two regions, the ship and the background, and uses statistical classification methods.
2) *Filter-based SAR ship detection method [5]:* This method uses multiple filters to preprocess SAR images, such as wavelet transform, Gabor filter, and polarization filter, to improve the contrast and edge information of ships.
3) *SAR ship detection method based on template matching [6]:* This method uses predefined ship templates for matching to identify ships in SAR images. This method is suitable for ships with regular shapes and relatively large sizes.
4) *SAR ship detection method based on polarization characteristics [7]:* This method uses the polarization information of SAR images to detect ships, including the polarization scattering matrix and polarization information entropy.

Due to the superior performance of DL-based algorithms compared to traditional detection algorithms, they have been widely applied in the field of SAR ship detection in recent years. Li et al. [8] used an improved faster R-CNN for ship detection in SAR images. Chen et al. [9] proposed a target detection network combined with an attention mechanism and constructed a loss function that combines the generalized cross union (GIoU) loss to reduce the network's scale sensitivity. Dong et al. [10] proposed a multidimensional DL network for SAR ship detection that utilizes complementary spatial and frequency domain features to improve the network's detection performance. Liu et al. [11] proposed a multiscale convolutional network and applied a rotatable bounding box object detection method (DR-Box) to solve the problem of offshore ship detection. Miao et al. [12] modified the state-of-the-art object detection network Faster R-CNN using the standard constant false alarm rate (CFAR) and proposed the target generated by faster R-CNN as the protection window of the CFAR algorithm. Zhang et al. [13] proposed a DL method based on enhanced GPU to detect ships from SAR images. Li et al. [14] used ResNet, an ultradeep network with higher precision and faster training speed, to train the SAR ship detection model. Wang et al. [15] proposed a DL model called RetinaNet object detector for better learning of target features. Chen et al. [16] proposed a multiscale network (MSARN) to detect ships in complex scenes at multiple scales and in any direction. Guo et al. [17] proposed a single-stage detector called CenterNet++, which mainly consists of a feature refinement module, a feature pyramid fusion module, and a head enhancement module.

However, most algorithm models based on DL are relatively large and difficult to apply, so lightweight detection algorithms are crucial for the practical application of algorithms. Therefore, Liu et al. [18] introduced an improved receptive field block (RFB) structure to enhance the network's feature extraction ability. And design a sliding window segmentation method to detect the entire SAR image, solving the problem of image input. Li et al. [19] proposed a lightweight model (LSDM) that improves the feature pyramid network by replacing ordinary convolution with spatial convolution. Zhang et al. [1] designed a lightweight detector (LSSD) and proposed a bidirectional feature fusion module for enhancing low-level and high-level features. Xu et al. [20] designed a lightweight L-CSP module to reduce computational complexity and then used a histogram-based pure background classification (HPBC) module, shape distance clustering (SDC) module, channel and spatial attention

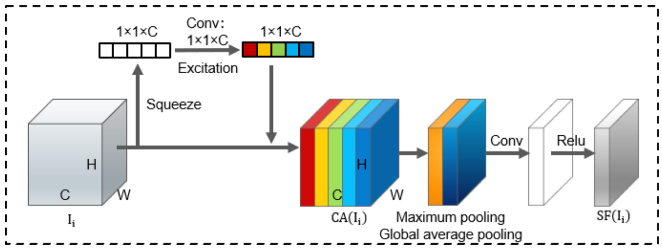

Fig. 1. Spatial information-focusing module based on channel attention. $\mathbf{I_i}$ is the input feature map with the number of channels as C and the height and width as H and W. Squeeze is the dimensional compression operation to obtain a $1 \times 1 \times C$ feature vector. Excitation is the excitation function to obtain a $1 \times 1 \times C$ weight vector. Multiply with $\mathbf{I_i}$ to obtain $CA(\mathbf{I_i})$, which is subjected to maximum pooling and global average pooling, followed by convolution and activation to obtain the spatial information-focused feature map $SF(\mathbf{I_i})$.

(CSA) module, and mixed spatial pyramid pool (H-SPP) module to improve detection performance. Tian et al. [21] proposed a lightweight RetinaNet, where the shallow convolutional layer of the backbone is replaced by ghost modules, and the number of deep convolutional layers is reduced. Embedding spatial and channel attention modules into the model to enhance detectability. The K-means clustering algorithm is used to adjust the initial aspect ratio of the model. Xu et al. [22] proposed a lightweight ship detection algorithm that is mainly based on the standard constant false alarm rate (CFAR). Yun et al. [23] proposed a lightweight algorithm LPEDet to solve the problem of the unclear contour of SAR targets. Li et al. [24] designed a lightweight network with feature amplification and skip connection structure to extract features of targets at various scales in SAR images, thereby improving detection performance. Zhou et al. [25] proposed a lightweight network LiraNet, that combines the ideas of dense connections, residual connections, and group convolution, including stem blocks and extractor modules.

III. Method

A. Spatial Information-Focusing Module

To reduce misjudgment of background interference, we propose a spatial information-focusing module based on channel attention. As shown in Fig. 1, the operation from I_i to $CA(I_i)$ is the original channel attention mechanism. To improve the attention of this mechanism for spatial information, we improve the channel level of the enhanced feature map on top of this to accomplish the focusing of spatial information.

In this article, we consider the DL model as graph $G = \{I, V\}$, where the feature graph nodes $I = \{I_i\}_{i=1\ldots n}$ in the model, and n is the total number of nodes. The node value transfer is represented by $V = \{v_{ij}^k\}_{i,j=1\ldots n}$, where v_{ij} represents the value transfer from I_i to I_j, and k represents different transfer operations.

FPN is the structure of the feature pyramid. FPN consists of bottom-up convolution, top-down upsampling, and lateral feature fusion. This structure is very much in line with the structural characteristics of CNN, which endows each layer of the feature map with a stronger ability to capture semantic information by fusing deep semantic information with shallow texture information. Therefore, we make the following improvements on FPN.

First, in the original FPN, the channel attention mechanism was introduced to enhance the channel containing the target. As shown in Fig. 1, this mechanism first squeezes the dimension of the feature map I_i, and then calculates the incentive factor in obtaining the importance of different channels. Finally, the incentive factor is multiplied by the feature map to assign larger weight coefficients to channels with more target position information. The formulas are as follows:

$$\mathrm{Sq} = \frac{1}{H \times W} \sum_{m=1}^{H} \sum_{n=1}^{W} I_i \tag{1}$$

$$\mathrm{Exc} = \sigma\left(\mathrm{Conv}\left(\mathrm{Sq}\right)\right) \tag{2}$$

$$\mathrm{CA}\left(I_i\right) = \mathrm{Exc} \times I_i. \tag{3}$$

Among them, $\mathrm{Sq}(\cdot)$ is the Squeeze operation, H, W are the length and width of the feature map, and Exc is the excitation factor, σ is the activation function, Conv is a convolutional operation, $\mathrm{CA}(\cdot)$ is the channel attention operation.

The channel attention mechanism is computed using a matrix multiplication operation with a time complexity of $O(C^3)$, where C denotes the number of channels in the input feature map. This is because the attention mechanism involves computing the attention weight matrix, which has a dimension of $C \times C$, so the time complexity of performing the matrix multiplication operation is $O(C^3)$.

However, as shown in Formula (3), after applying the channel attention mechanism to the feature map, the channel containing target information will be enhanced, and the background information similar to the target in the channel will also be enhanced. This leads to the negative impact of useless information in the background on target feature extraction, improving the algorithm's misjudgment rate. To solve this problem, we propose a spatial information-focusing module that directs the detection network to pay more attention to the area where the target is located to suppress the negative impact of useless information in the background and improve the algorithm's detection performance.

As shown in Fig. 1, based on the channel attention of the feature map, we focus the enhanced feature map on spatial information in the channel dimension to suppress useless background information in each channel. The spatial information-focusing operation $\mathrm{SF}(\cdot)$ proposed in this article is shown in the following:

$$\begin{aligned}&\mathrm{SF}\left(I_i\right)\\ &= \sigma(\mathrm{Conv}([\mathrm{AvgPool}(\mathrm{CA}(I_i)); \mathrm{MaxPool}(\mathrm{CA}(I_i))])).\end{aligned} \tag{4}$$

Among them, $\mathrm{SF}(\cdot)$ is the spatial information-focusing operation proposed in this article, $\mathrm{AvgPool}(\cdot)$ is the global average pooling operation, $\mathrm{MaxPool}(\cdot)$ is the maximum pooling operation, σ is the activation function, Conv is a convolutional operation.

In the feature map, the pixel value threshold that satisfies the target feature is α. There will be some disturbing factors in the feature map whose pixel values are less than α, and this part of the disturbing factors will be eliminated by the maximum

pooling operation to eliminate the negative impact. In addition, there will be some interference factors in the feature map whose pixel values are greater than α. The average pooling operation can offset these interference factors because the pixel values around them do not satisfy the threshold value of the target feature. Since most of the interfering information in the feature map will be reduced by the composite pooling operation to reduce its negative impact, more target feature information is retained in the feature map, and the feature information of the background similarities is reduced, which reduces the misjudgment of the background interfering information.

The computational complexity of maximum and average pooling is related to the size of the pooling window. Assuming that the size of the pooling window is $K \times K$, the time complexity of computing the pooling result is $O(H \times W \times K^2)$ for each channel. Since there are C channels, the time complexity of maximum pooling and average pooling is $O(C \times H \times W \times K^2)$.

Then, the two pooling results are spliced in the channel dimension with 1×1 convolution, and the time complexity of this part is $O(C \times H \times W) + O(H \times W \times C^2)$. So, the time complexity of the spatial information focusing module proposed in this article is $O(C^3) + O(C \times H \times W \times K^2) + O(C \times H \times W) + O(H \times W \times C^2)$.

Because noise is usually random, after performing spatial information-focusing operations with average pooling and maximum pooling as the main focus, the impact of noise will be submerged by peak and mean values to maintain the importance of key feature information and make the output feature map more stable and reliable. Moreover, the spatial information-focusing operation adopts the same operation for each position in the feature map, ensuring that the features remain invariant to a certain extent, thereby reducing redundant information in the feature map. Then, the pooled results are concatenated according to channels and then convolved, and then $SF(I_i)$ is obtained through activation function processing to complete the focusing of spatial information.

After being processed by the spatial information-focusing module, the background information similar to the target in the input feature map will be replaced by the peak and mean, thus achieving the goal of suppressing useless background information around the target and achieving the enhancement of the target features. This enables the model to pay more attention to the area where the target is located, improving the model's ability to extract target features. To solve the problem of a high misjudgment rate for interference in the background.

B. FPN Optimization

After focusing on the target's spatial information, to balance the imbalanced samples in the dataset, we also need to optimize the structure and feature fusion mechanism of the FPN. As shown in Fig. 2(a) is the original FPN with only channel attention operation, and (b) is the proposed FPN after adopting the spatial information focusing module and optimizing the network structure.

FPN, as shown in Fig. 2(a), In FPN, when fusing feature maps at different levels, channel attention enhancement is performed

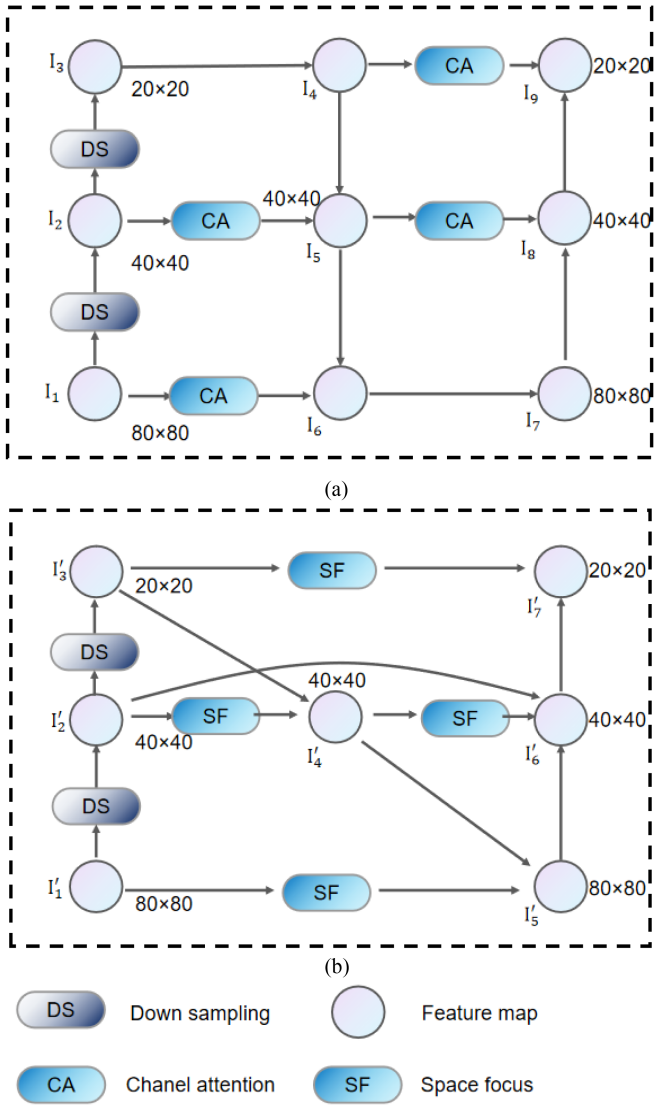

Fig. 2. (a) Original FPN. (b) Optimized FPN.

on the horizontally connected feature maps. The importance of each channel of the feature map is automatically obtained using deep learning, and finally, different weighting coefficients are assigned to each channel to enhance the important features and suppress the unimportant ones. FPN is a network that can extract features at different scales and fuse them. The values of each edge are shown in the following:

$$k = \begin{cases} 1, & v_{ij}^{k} = I_i \\ 2, & v_{ij}^{k} = DS(I_i) \\ 3, & v_{ij}^{k} = CA(I_i) \\ 4, & v_{ij}^{k} = 0 \end{cases}. \tag{5}$$

Among them, $DS(\cdot)$ is a downsampling operation, $CA(\cdot)$ is the channel attention operation, v_{ij} represents the value transfer from I_i to I_j, and k represents different transfer operations.

When $k = 1$, no action is taken. When $k = 2$, perform the downsampling operation. When $k = 3$, perform channel

attention operation. When $k = 4$, the I_i the transfer value is set to zero.

I_1 is the input node, and when $j > 1$, perform the following fusion operations:

$$I_j = \sum_{i=1}^{n} v_{ij}^{k}, \ k \in [1, 4]. \tag{6}$$

I_j represents the node obtained by fusion. However, due to the different resolutions of each input layer in FPN, the impact of each layer on the fused features during fusion is also different. However, FPN only provides an undifferentiated summary of input features, ignoring the importance of features at different resolutions. As shown in Formula (6), it simply adds features together. This simple connection and fusion do not fully utilize the role of different resolutions, making it difficult for FPN to accurately identify targets with small sample sizes when the number of samples at different scales in the dataset is uneven. We have made the following improvements to address this issue.

As shown in Fig. 2(b), we believe that nodes with only one input channel and without feature fusion cannot significantly contribute to the fusion of different features and will increase computational complexity. Therefore, from a lightweight perspective, we removed nodes I_4 and I_6 in Fig. 2(a) to simplify the network.

Then, by adding skip links between input and output nodes at the same scale, the original features are fused with the fused features again to retain more target features at that level, improving the feature extraction effect of the target. The optimized FPN is more concise and efficient for fusing features of different resolutions. At this point, the values of each edge in the graph are shown in the following:

$$k = \begin{cases} 1, & v_{ij}^{k} = I_i^{'} \\ 2, & v_{ij}^{k} = \mathrm{DS}\left(I_i^{'}\right) \\ 3, & v_{ij}^{k} = \mathrm{SF}\left(I_i^{'}\right) \\ 4, & v_{ij}^{k} = 0 \end{cases}. \tag{7}$$

Among them, $\mathrm{SF}(\cdot)$ is the spatial information-focusing operation, and $\mathrm{DS}(\cdot)$ is the downsampling operation.

When $k = 1$, no operation is performed. When $k = 2$, perform the downsampling operation. When $k = 3$, perform the spatial information-focusing operation. When $k = 4$, the $I_i^{'}$ transfer value is set to zero.

At the same time, to balance sample bias in fused features, a weighted feature fusion mechanism is adopted for the fusion of different feature maps. I_1' is the input node, and when $j > 1$, perform the following fusion operations:

$$I_j' = \frac{\sum_{i=1}^{n} w_{ij} v_{ij}^{k}}{\varepsilon + \sum_{i=1}^{n} w_{ij}}, \ k \in [1, 4]. \tag{8}$$

I_j' represents the node obtained by weighted fusion, w_{ij} is the learnable weight corresponding to the edge v_{ij}^{k}, and ε is the learning rate.

We can balance sample bias by learning the importance of the objectives at different levels during the training process and assigning different weights to them. Then, by comparing the weight values obtained from the training, it can be seen which type of scale feature the model is more inclined toward, and it can also indicate the correctness of the experiment for quantitative and qualitative analysis.

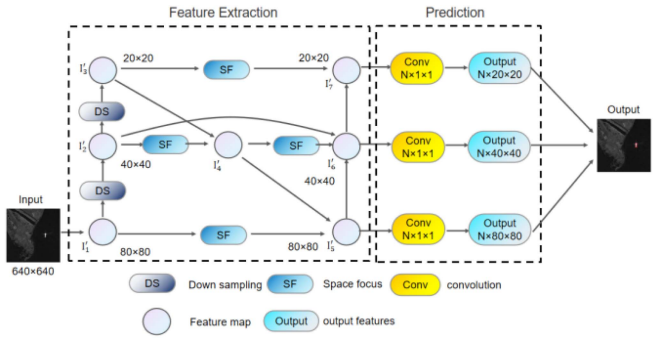

Fig. 3. Overall flowchart. The images are input into the feature extraction module of MFPNet for feature extraction to obtain feature maps at different scales, and the feature maps are input into the prediction module to predict the input images at different scales, where N is the number of target species.

C. Overall Process

In this article, we propose a multiscale feature pyramid SAR ship detection network (MFPNet) that is robust to background interference. Because FPN does not pay enough attention to the target's spatial location, the network is prone to misjudgment of background information similar to the target. Moreover, the FPN network lacks sufficient feature extraction for targets with fewer samples, resulting in unsatisfactory detection performance for targets with fewer samples. Therefore, the introduction of the spatial information-focusing module in the fusion channel makes the network pay more attention to the target area and suppress useless information in the background. Then, we optimized the original FPN network structure. At the same time, a weighted feature fusion mechanism is used to assign weights to features of different scales to fuse features, thereby balancing sample bias adaptively. To improve the network's detection performance when facing the above two problems. The overall flowchart of this method is shown in Fig. 3.

Among them, the feature extraction part is the main improvement part made in this article, and the prediction part acquires the fused features at different scales and then predicts the input image. First, the original image is extracted with multiscale features by a three-layer convolutional neural network on the left side. Then, the spatial information focusing module enhances the three layers' obtained multiscale features. Finally, the weighted fusion of the multiscale features is completed through the horizontal links between the layers and the jump links to obtain the fused feature map.

IV. Experiments and Results

A. Dataset Introduction

In this article, we will compare the MFPNet with other advanced lightweight algorithms on the common datasets HRSID [26], SSDD [8], and SAR-Ship-Dataset [27].

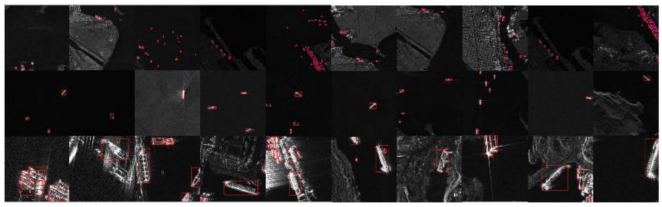

Fig. 4. Ships examples with different scales.

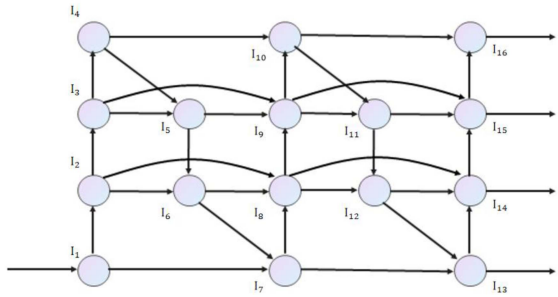

Fig. 5. Network topology diagram. $\mathbf{w_{ij}}$ is the learnable weight of $\mathbf{I_i}$ to $\mathbf{I_j}$.

TABLE I
WEIGHT PARAMETERS

First output	Second output
$w_{15} = 0.7427$	–
$w_{24} = 0.5521$	$w_{26} = 0.1060$
$w_{34} = 0.4479$	$w_{37} = 0.4885$
$w_{45} = 0.2573$	$w_{46} = 0.3720$
$w_{56} = 0.5220$	–
$w_{67} = 0.5115$	–

The HRSID dataset is a high-resolution remote sensing image dataset released by the Institute of Remote Sensing and Digital Earth of the Chinese Academy of Sciences. It contains high-resolution and high-quality remote sensing image data across the Chinese Mainland, covering different areas such as cities, rural areas, and mountains. It is one of the largest high-resolution remote sensing image datasets in China at present. Its imaging sources are Sentinel-1 and Terra-SAR-X, with resolutions of 0.5, 1.0, and 3.0 m, respectively. It includes 5604 SAR images and 16 951 ship instances, including SAR images of different resolutions, polarization, sea conditions, sea areas, and coastal ports. Among them, the target for small ships accounts for 54.5%, the target for medium-sized ships accounts for 43.5%, and the target for large ships accounts for 2%. From the dataset's composition, it can be seen that the scale of a ship instance includes three levels, and the environment in which the ship instance is located is complex.

The SSDD is the first publicly available domestic and international dataset dedicated to detecting ship targets in SAR images that can be used to train and test check algorithms and is used by 30 levels of universities and research institutes. SSDD is obtained by downloading publicly available SAR images on the Internet, cropping the target area to the size of about 500 × 500 pixels, and manually annotating the ship targets' position. The data are mainly from RadarSat-2, Terra SAR-X, and Sentinel-1 sensors, with four polarization modes, HH, HV, VV, and VH, and a resolution of 1–15 m, and there are ship targets in a large area of the sea and near-shore areas.

The SAR-Ship-Dataset takes China's home-made Gaofen-3 SAR data and Sentinel-1 SAR. SAR-Ship-Dataset uses China's domestic Gaofen-3 SAR data and Sentinel-1 SAR data as the main data source, and a total of 102 views of Gaofen-3 and 108 views of Sentinel-1 SAR images are used to construct a high-resolution SAR ship target deep learning sample library. At present, the deep learning sample library contains 43819 ship slices. The imaging models of Gaofen-3 are Strip-Map (UFS), Fine Strip-Map 1 (FSI), Full Polarization 1 (QPSI), Full Polarization 2 (QPSII), and Fine Strip-Map 2 (FSII). The resolutions of these five imaging models are 3 m, 5 m, 8 m, 25 m, and 10 m. The imaging modes of Sentinel-1 are Strip-Map (S3 and S6) and Wide-Map. Therefore, comparative experiments on these datasets can better prove the robustness and precision of the MFPNet.

The preview image of ship instances at different scales in the dataset is shown in Fig. 4. The first row is small scale, the second row is medium scale, and the third row is large scale. The location of the target is marked with a red box.

B. Experimental Parameters

We randomly divided the dataset into training and test sets in the ratio of 8:2 for training MFPNet. We applied left and right, up and down flipping, and image scaling in the training process. MFPNet adopted a three-layer pyramid and trained it with an SGD optimizer, with a momentum of 0.937 and a weight attenuation of 0.0005. We use binary cross-entropy for confidence loss and classification loss, CIOU for localization loss, and set the size of the input image to 640 × 640. We use Adam to update the trainable parameters in the network. Adam hyperparameter is set as: $e = 10^{-8}$, $\beta 2 = 0.999$, $\beta 1 = 0.9$, $\alpha = 0.001$. Initially, set the learning rate to 0.001 and then use the cosine annealing algorithm to gradually attenuate the learning rate to 0.0002.

We train MFPNet for 300 epochs on an RTX3050 graphics card with 8G memory. The network topology diagram is shown in Fig. 5, and MFPNet adopts a three-layer network structure. The weight parameters obtained through training are shown in Table I.

C. Evaluation Indicators

To comprehensively and quantitatively analyze the detection performance, the evaluation criteria of the COCO dataset and PASCAL VOC dataset are used, including precision (P), recall (R), F1 score (F1), mean average precision (mAP), frames per second (FPS), and parameter size (Params). The FPS and parameter size are used to measure the model's lightweight degree.

1) Precision (P) is the ratio of correctly predicted targets to all predicted targets. A higher P represents a higher number of correctly predicted targets. The precision (P) can be expressed as

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{9}$$

where true positive (TP) indicates the number of correctly detected targets, and false positive (FP) indicates the number of false positives.

2) Recall (R) is the proportion of the target correctly predicted by the model among all real targets, which can be expressed as

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{10}$$

where false negative (FN) is the number of undetected targets.

Due to the interaction between precision and recall, the F1 score and mAP are introduced to evaluate the overall performance of lightweight algorithms.

3) F1 score (F1) is the harmonic average of precision and recall. The larger the F1 value, the better the performance of the model. The F1 score can be expressed as

$$\mathrm{F1} = \frac{2{*}\mathrm{P} * \mathrm{R}}{\mathrm{P} + \mathrm{R}}\,. \tag{11}$$

4) Mean average precision (mAP) is a measure of detection precision in object detection, which can be expressed as

$$\mathrm{mAP} = \int_0^1 \mathrm{P}\,(\mathrm{R})\,\mathrm{dR}\,. \tag{12}$$

D. Quantitative Analysis

In this section, we conducted a comparative experiment between the MFPNet and other advanced lightweight algorithms currently available, including Fast R-CNN, SSD, YOLO-V4, YOLO-V7, YOLOX-L, DAPN, HR-SDNet, FASC-NET, and OF-NET. We conducted comparative experiments with these algorithms on the HRSID dataset and analyzed the effects of lightweight algorithms from both quantitative and qualitative aspects. The quantitative analysis mainly compares data from five indicators: precision (P), recall (R), F1 score (F1), mean average precision (mAP), and the number of parameters (Params), and analyzes the reasons. The quantitative results are shown in Table II.

In Tables II–IV, the optimal values are highlighted in bold, and the suboptimal values are underlined. First, this article mainly proposes a lightweight algorithm for SAR ship detection, so the degree of lightweight of the algorithm is particularly important. The main purpose of MFPNet is to achieve better detection results with a higher degree of lightweight.

As described in Section IV-A, large ship targets account for only 2% of the HRSID dataset, so the number of samples at three scales in the dataset is extremely uneven. This leads to the unsatisfactory training effect of FPN on targets with a small number of samples, making the algorithm unable to accurately identify such targets. Therefore, balancing samples during training can help improve the algorithm's detection effect.

TABLE II
COMPARISON WITH OTHER LIGHTWEIGHT ALGORITHMS ON THE HRSID DATASET

Algorithm	P	R	F1	mAP	FPS	Params(10^7)
Faster R-CNN[28]	81.5	73.1	77.1	77.65	10.2	4.13
SSD[29]	80.1	73.2	76.5	81.51	13.8	1.41
YOLO-V4[30]	85.9	74.5	79.8	87.7	14.5	1.12
YOLO-V7[31]	87.5	78.4	82.7	88.1	23.8	3.7
YOLOX-L[32]	84.02	59.91	70	81.8	18.33	5.41
DAPN[33]	83.4	78.54	80.9	81.9	13	2.25
HR-SDNet[34]	86.9	77.44	81.9	88.2	6.8	7.48
FASC-NET[35]	88.09	74.94	80.98	84.66	24.5	**0.064**
OF-NET[36]	88.28	77.29	82.42	88.39	20.1	0.126
MFPNet	**92.88**	**79.50**	**85.67**	**89.85**	**25**	0.1

Where bold is marked as the optimal value and underline is marked as the suboptimal value.

TABLE III
COMPARISON WITH OTHER LIGHTWEIGHT ALGORITHMS ON THE SSDD DATASET

Algorithm	P	R	F1	mAP	FPS	Params(10^7)
Faster R-CNN[28]	87.1	86.0	86.54	89.7	11.9	4.13
SSD[29]	86.2	87.8	86.99	92.3	16.1	1.41
YOLO-V4[30]	96.9	90.4	93.53	95.3	22.7	1.12
YOLO-V7[31]	95.7	89.6	92.54	94.4	30.2	3.7
DAPN[33]	85.6	87.9	86.73	90.6	12.2	2.25
FASC-NET[35]	95.65	90.04	92.76	96.61	42.5	**0.064**
MFPNet	**97.08**	**91.03**	**93.95**	**97.1**	**47.6**	0.1

Where bold is marked as the optimal value and underline is marked as the suboptimal value.

As shown in Table II, in the Params indicator column, we can see that the parameter size of the lightweight algorithm proposed in this article is the second smallest, which is only 0.036×10^7 more parameter counts than the FASC-NET with the smallest parameter counts and achieves the highest FPS, which is sufficient to prove the MFPNet's lightweight degree. Compared to the other eight lightweight algorithms, precision (P), recall (R), F1 score (F1), mean average precision (mAP), and FPS of this article are the optimal values, while the precision (P) is the suboptimal value. It can reflect the comprehensive performance and detection precision of the MFPNet.

As shown in Table II, compared with the other two algorithms, FASC-NET and OF-NET, which have the highest degree of lightweight, the MFPNet has 4.79%, 4.56%, 4.69%, and 5.19%

TABLE IV
COMPARISON WITH OTHER LIGHTWEIGHT ALGORITHMS ON THE SAR-SHIP-DATASET

Algorithm	P	R	F1	mAP	FPS	Params(10^7)
Faster R-CNN[28]	93.0	84.2	88.51	88.0	23.7	4.13
SSD[29]	86.8	**91.9**	89.27	88.5	30.5	1.41
YOLO-V4[30]	93.1	88.8	90.9	89.4	32.6	1.12
YOLO-V7[31]	93.5	90.4	91.92	91.4	40.1	3.7
DAPN[33]	87.3	91.4	89.3	88.2	21.5	2.25
HR-SDNet[34]	93.3	89.4	91.3	88.4	8.9	7.48
FASC-NET[35]	91.1	89.35	90.21	92.4	60.4	**0.064**
MFPNet	**94.0**	90.28	**92.1**	**95.0**	**60.9**	0.1

Where bold is marked as the optimal value and underline is marked as the suboptimal value.

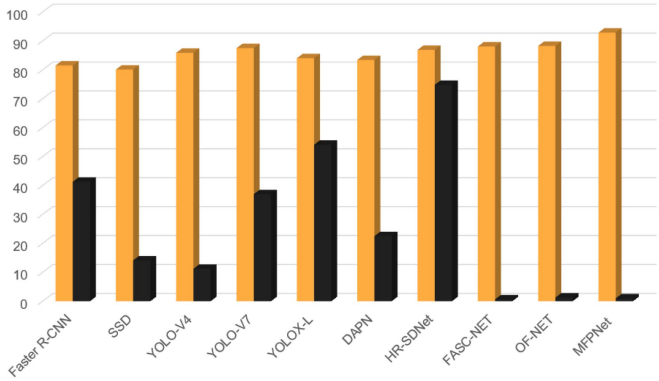

Fig. 6. Comparison of HRSID performance in terms of precision and number of parameters.

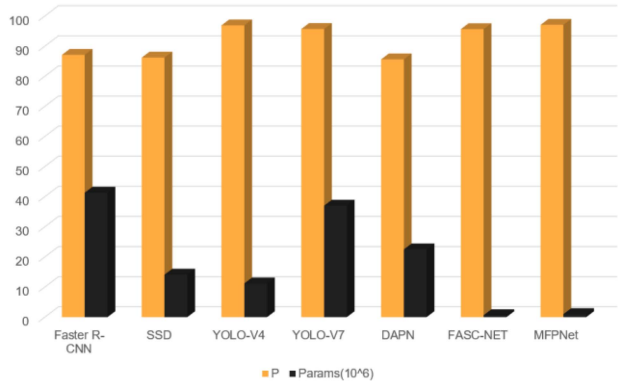

Fig. 7. Comparison of SSDD performance in terms of precision and number of parameters.

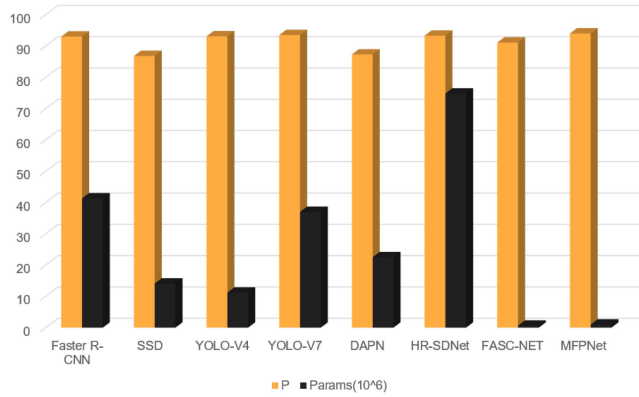

Fig. 8. Comparison of SAR-SHIP-dataset performance in terms of precision and number of parameters.

improvements in precision (P), recall (R), F1 score (F1), and mean average precision (mAP), respectively. Compared to the OF-NET, the MFPNet has 4.6%, 2.21%, 3.25%, and 1.46% improvements in precision (P), recall (R), F1 score (F1), and mean average precision (mAP), respectively. This proves that the MFPNet performs better than lightweight algorithms of the same magnitude.

As shown in Table III, this paper achieves optimal values for each of the metrics of precision (P), recall (R), F1 score (F1), mean detection accuracy (mAP), and FPS compared to the other six lightweight algorithms, and the number of parameters is suboptimal. Compared to the next best-performing algorithm, YOLO-V4, MFPNet has 1.02×10^7 fewer parameters than YOLO-V4 and has 24.9 higher FPS than YOLO-V4. Outside of the first time, MFPNet outperforms YOLO-V4 in all metrics.

As shown in Table IV, compared to the other six lightweight algorithms, this article achieves optimal values for each of the precision (P), F1 score (F1), mean detection accuracy (mAP), and FPS, while the number of parameters is suboptimal. Compared with FASC-NET, the algorithm with the least number of parameters, MFPNet outperforms FASC-NET in the five metrics of P, R, F1, mAP, and FPS, which proves that MFPNet achieves higher detection accuracy and better lightweight effect without adding too many parameters.

This is because after using weighted feature fusion, as shown in Table I, the MFPNet will assign higher weights to targets with fewer samples during the training process, i.e., to large-scale targets, to enhance the training effect on such targets and improve the precision algorithm's precision. At the same time, the spatial information-focusing module can suppress and reduce the negative interference of useless information in the background on feature extraction, effectively reducing the algorithm's misjudgment of background interference. Therefore, the MFPNet outperforms the other two lightweight algorithms in precision, recall, F1 score, and mAP.

Moreover, we can intuitively see from Figs. 6–8 that the number of parameters of MFPNet is sufficiently small compared to other state-of-the-art algorithms, and it makes some performance improvements compared to other algorithms. This is enough to prove that this article implements a lightweight and high-performance detection algorithm.

In conclusion, through experimental validation on three datasets, MFPNet has higher lightweight and better detection performance with good generalization performance compared to other existing lightweight algorithms. This is thanks to the optimized FPN being able to better balance the samples in

TABLE V
ABLATION STUDIES

Method	P	R	F1	mAP	Params (10^7)
FPN	88.09	74.94	80.98	84.66	**0.064**
SF	88.54	72.14	79.50	83.18	**0.064**
WFPN	91.73	75.92	83.07	86.9	0.1
MFPNet	**92.88**	**79.50**	**85.67**	**89.85**	0.1

Where bold is marked as the optimal value.

the dataset and improve the algorithm's precision for targets with fewer samples. Moreover, the introduction of the spatial information-focusing mechanism can make the algorithm pay more attention to the area where the target is located, suppress useless background information around the target, make the extracted target features purer, and thus reduce misjudgment of background interference that has a certain similarity to the target.

E. Ablation Experiments

The two improved schemes in this article can effectively solve the two problems, and the introduction of the spatial information-focusing module can enhance the extraction of multiscale target features by FPN. The joint use of the two options can play a greater role. We have done the following ablation studies on HRSID to validate our idea.

We compare the original FPN with the focusing module that only introduces spatial information, the improved FPN with a weighted fusion mechanism, and the two schemes used together. Further, validate the effectiveness of the MFPNet.

1) Only introducing the spatial information-focusing module (SF): only strengthening the model's attention to the target's spatial position and suppressing useless information in the background.
2) Only improving FPN and using weighted feature fusion mechanism (WFPN): only enhancing the model's feature extraction for targets with small sample sizes.
3) Improving FPN and using a weighted fusion mechanism while introducing a spatial information-focusing module (MFPNet).

As shown in Table V, the first improvement scheme is 4.43% lower in precision, 7.36% lower in recall, 6.17% lower in F1 score, and 6.67% lower in mAP compared to the MFPNet. Compared to FPN, there is only a 0.45% improvement in precision. This is because although using the spatial information-focusing module can suppress useless background information around the target during feature extraction, it needs to be combined with an efficient feature extraction network to produce better results. So, introducing only the spatial information-focusing module will reduce the misjudgment rate of background interference, but the effect is not outstanding.

Furthermore, as shown in Table V, introducing the spatial information focusing module does not significantly increase the number of parameters compared to the FPN. This is because the spatial information focusing module has only one more convolutional layer and one more activation function in the number of parameters than the channel attention mechanism. Moreover, the pooling operation used reduces the dimensionality of the feature map, thus reducing the number of parameters of the module. This results in no significant increase in the number of parameters of the spatial information focusing module.

If WFPN is used, compared to the SF improvement scheme, the precision is 3.19% higher, the recall is 3.19% higher, the F1 score is 3.57% higher, and the mAP is 3.72% higher. Compared to FPN, it has also improved in various indicators. This is because this WFPN can balance targets with small sample sizes in the dataset effectively, improving the algorithm's detection performance for such targets significantly and resulting in significant improvements in various indicators. In addition, as shown in Table V, using only the WFPN, there are 0.036×10^7 more than the FPN in terms of the number of parameters. This is because the WFPN introduces additional learnable weights, which leads to an increase in the number of parameters.

However, compared to the MFPNet, WFPN has a 1.15% lower precision, 3.58% lower recall, 3.6% lower F1 score, and 2.95% lower mAP. This is because although the second improvement scheme can balance samples of different scales, the model has better detection performance for multiscale targets. However, when extracting features from the target, the influence of spatial information is still ignored, resulting in features being interfered with by useless background information around the target. Therefore, compared to the MFPNet, there are slight shortcomings in various indicators.

In summary, using both improvement schemes simultaneously can not only effectively solve the proposed problem but also complement each other to achieve more ideal results. As shown in Table V, as the number of parameters increases, the detection performance of MFPNet is improved. The MFPNet achieved the best results in all indicators. This is because WFPN can effectively solve the problem of uneven sample size and improve the extraction effect of target features with fewer samples. When using MFPNet, it not only balances the samples but also suppresses the negative interference caused by background information in feature extraction, making the extracted features purer. So MFPNet not only improves precision effectively but also reduces misjudgment, enabling it to perform excellently in various indicators.

F. Qualitative Analysis

In this section, we compared the qualitative results with the FPN to verify whether the MFPNet solves two problems.

1) The detection effect of targets with small sample sizes is not ideal. That is, the detection results of large-scale targets are not ideal.
2) It is easy to misjudge background interference, mainly including small objects and inshore objects. Thus fully verifying the correctness of the starting point of this article and the feasibility of the methods in the third part.

First, we conducted a comparison experiment on the HRSID dataset. As shown in Fig. 9, columns (1)–(5) cover small-scale, medium-scale, and large-scale ship targets. From the figure, we can see that MFPNet improves on the false detection of small and near-shore targets and the missed detection of large targets compared to FPN. This is because the spatial information

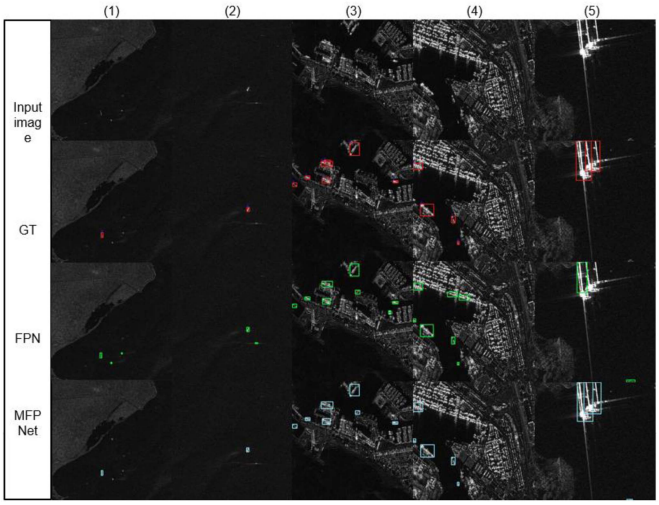

Fig. 9. Synthetic aperture radar ship detection in HRSID dataset.

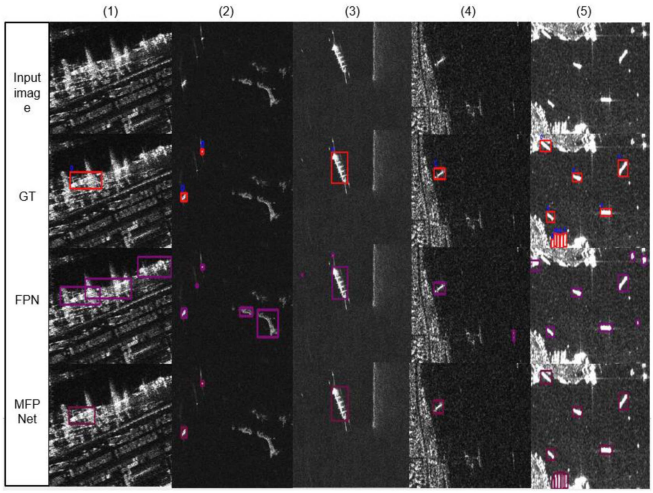

Fig. 10. Synthetic aperture radar ship detection in SSDD dataset.

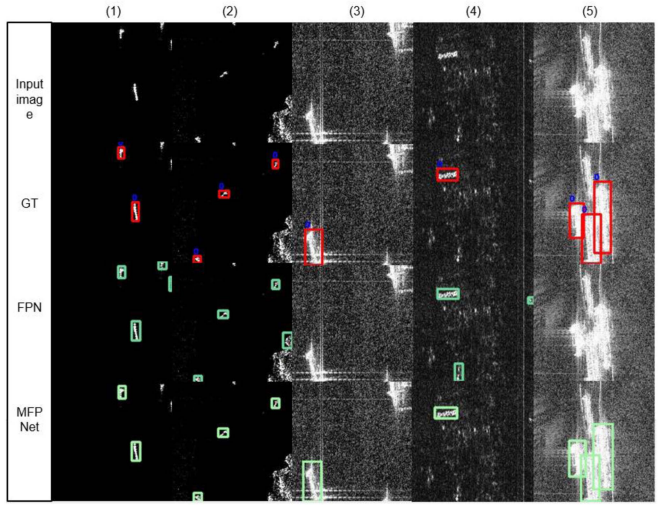

Fig. 11. Synthetic aperture radar ship detection in SAR-SHIP-dataset.

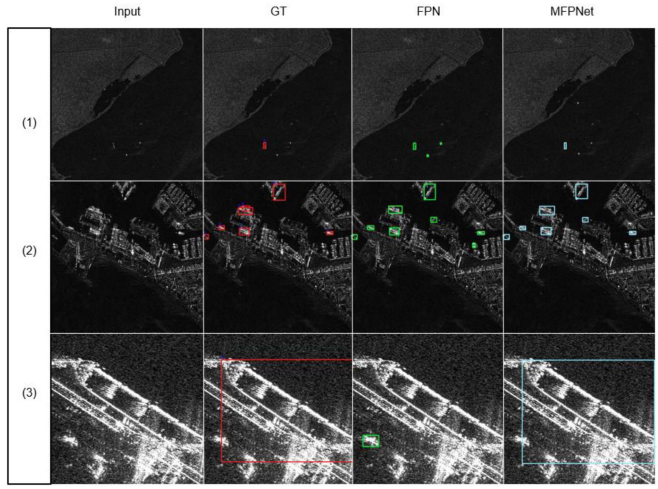

Fig. 12. Comparison of multiscale target detection results.

focusing module introduced by MFPNet makes the model pay more attention to the characteristics of the target and reduces the misdetection. Adjusting the FPN network structure and introducing the weighted feature fusion mechanism enable the model to improve its performance in multiscale target detection. Also, qualitative experiments further corroborate how the performance is improved in quantitative experiments.

Similarly, we have conducted qualitative experiments on multiscale targets on the SSDD dataset. As shown in Fig. 10, the more problems in the SSDD dataset are misjudgment and omission of near-shore targets and misjudgment of tiny targets. MFPNet can effectively deal with the above problems and thus achieve satisfactory detection results.

Fig. 11 shows the qualitative experiment of SAR-SHIP-dataset. As shown in the figure, the same problem exists in this dataset as in the HRSID dataset and the SSDD dataset, which is the main problem affecting the detection performance of the model. MFPNet can effectively deal with the misdetection of tiny targets and the omission of large targets, making MFPNet make some improvements in both qualitative and quantitative experiments. We can also see from the experiments on the three datasets that MFPNet has a certain generalization ability in the face of different datasets.

For problem one, we mainly conducted comparative experiments on the two algorithms at three scales: small, medium, and large. In response to question two, we mainly conducted comparative experiments on inshore targets and small-scale targets.

To verify whether the MFPNet solves problem 1 effectively, we analyzed the comparative detection results of the three scales in Fig. 12. The first behavior is small-scale targets, the second behavior is medium-scale targets, and the third behavior is large-scale targets.

As shown in the third line, FPN not only fails to detect large targets but also misrecognizes objects similar to small and medium-sized ships in the image. This is because large ship targets in the dataset only account for 2%, making the training effect of FPN on large-scale targets with a small sample size not ideal. This validates problem 1 proposed in this article, which is that FPN's detection performance is not ideal when identifying targets with drastic scale changes.

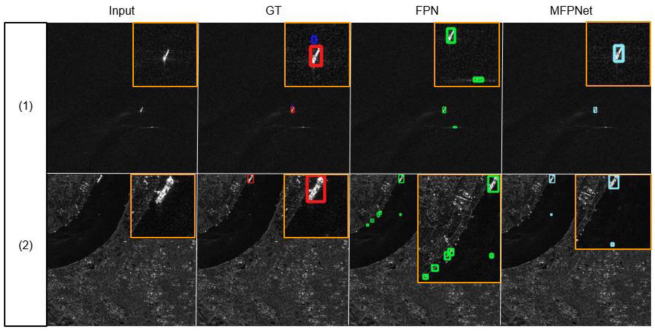

Fig. 13. Comparison of small-scale and inshore target detection results.

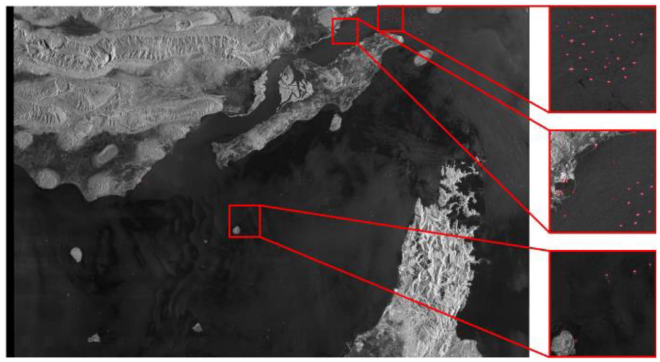

Fig. 14 High-resolution SAR ship detection.

However, after optimizing the FPN structure and using a weighted fusion mechanism, the model can adjust the weight proportion of features with small sample sizes adaptively, enhancing the recognition effect on large-scale targets. As shown in the fourth column of Fig. 12, compared to the third column, the MFPNet outperforms FPN in detecting large-scale targets, which verifies that the proposed improvement scheme can solve problem 1 effectively.

To verify whether the MFPNet solves problem 2 effectively, we analyzed the comparison detection results of two types from Fig. 13. The first behavior is small-scale targets, and the second behavior is inshore targets.

As shown in Fig. 13, it can be seen from one line that FPN has made recognition errors on some small objects that are not targets. From the second line, it can be seen that FPN is also prone to misjudgment of some objects in the inshore area. This is because the imaging characteristics of SAR images enable some small objects and objects on land to have similar features to small-scale targets. Moreover, FPN only focuses on the channel and ignores the negative impact of background information around the target on feature extraction, making the extracted target features relatively complex. Therefore, it is prone to misjudgment when identifying background interference similar to the target.

However, after introducing the spatial information-focusing module, the model can pay more attention to the target's spatial position and suppress useless information in the background, making the extracted target features purer and reducing misjudgment of background interference. Fig. 13 shows that for both small and inshore targets, the MFPNet effectively reduces the misjudgment rate compared to the FPN, fully verifying that the proposed solution can solve problem 2 effectively.

In summary, the optimization of the FPN network structure and the introduction of the weighted fusion mechanism in this article can effectively improve the model's detection performance for targets with small sample sizes. Moreover, the introduction of spatial information-focusing module can reduce misjudgment of background interference effectively. The improvement plan proposed in this article can effectively solve the two problems proposed in this article.

To demonstrate the generalization ability of MFPNet, we also performed SAR ship detection in a larger scene. As shown in Fig. 14, the image has a resolution of $24\,000 \times 16\,000$ and covers most of the ocean and land. We performed the detection on the whole image and selected three parts from it for zoomed-in display. From the figure, we see that MFPNet can still detect ship targets in higher resolution SAR images, proving its generalization ability.

V. Conclusion

In this article, we propose a MFPNet that is robust to background interference. We have made improvements to the feature extraction section of the algorithm. First, adjust the FPN structure to reduce parameters, enhance target features, and introduce a weighted feature fusion mechanism to balance samples adaptively, thereby improving the detection performance of targets with small sample sizes.

At the same time, a spatial information-focusing module is introduced into the optimized FPN channel to make the model pay more attention to the target's spatial information, suppress useless background information around the target, strengthen the model's extraction effect on the target features, and reduce misjudgment of background interference. The algorithm proposed in this article has achieved ideal results in the field of lightweight algorithms.

References

[1] X. Zhang et al., "A lightweight feature optimizing network for ship detection in SAR image," *IEEE Access*, vol. 7, pp. 141662–141678, 2019, doi: 10.1109/ACCESS.2019.2943241.

[2] X. Yang, J. Zhang, C. Chen, and D. Yang, "An efficient and lightweight CNN model with soft quantification for ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5230713, doi: 10.1109/TGRS.2022.3186155.

[3] Y. Li, S. Zhang, and W.-Q. Wang, "A lightweight faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4006105, doi: 10.1109/LGRS.2020.3038901.

[4] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010, doi: 10.1109/TGRS.2010.2046330.

[5] A. Marino, "A notch filter for ship detection with polarimetric SAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 3, pp. 1219–1232, Jun. 2013, doi: 10.1109/JSTARS.2013.2247741.

[6] C. Wang, F. Bi, L. Chen, and J. Chen, "A novel threshold template algorithm for ship detection in high-resolution SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 100–103, doi: 10.1109/IGARSS.2016.7729016.

[7] R. Ringrose and N. Harris, "Ship detection using polarimetric SAR data," in *Proc. SAR Workshop: Committee Earth Observ. Satell.*, 2000, vol. 450, pp. 687–688.

[8] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era: Models, Methods Appl.*, 2017, pp. 1–6, doi: 10.1109/BIGSARDATA.2017.8124934.
[9] C. Chen, C. He, C. Hu, H. Pei, and L. Jiao, "A deep neural network based on an attention mechanism for SAR ship detection in multiscale and complex scenarios," *IEEE Access*, vol. 7, pp. 104848–104863, 2019, doi: 10.1109/ACCESS.2019.2930939.
[10] D. Li, Q. Liang, H. Liu, Q. Liu, H. Liu, and G. Liao, "A novel multidimensional domain deep learning network for SAR ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5203213, doi: 10.1109/TGRS.2021.3062038.
[11] L. Liu, G. Chen, Z. Pan, B. Lei, and Q. An, "Inshore ship detection in SAR images based on deep neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 25–28, doi: 10.1109/IGARSS.2018.8519555.
[12] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, 2017, pp. 1–4, doi: 10.1109/RSIP.2017.7958815.
[13] Y. Chang, A. Anagaw, L. Chang, Y. Wang, C. Hsiao, and W. Lee, "Ship detection based on YOLOv2 for SAR imagery," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 786.
[14] Y. Li, Z. Ding, C. Zhang, Y. Wang, and J. Chen, "SAR ship detection based on resnet and transfer learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 1188–1191, doi: 10.1109/IGARSS.2019.8900290.
[15] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 531.
[16] C. Chen, C. He, C. Hu, H. Pei, and L. Jiao, "MSARN: A deep neural network based on an adaptive recalibration mechanism for multiscale and arbitrary-oriented SAR ship detection," *IEEE Access*, vol. 7, pp. 159262–159283, 2019, doi: 10.1109/ACCESS.2019.2951030.
[17] H. Guo, X. Yang, N. Wang, and X. Gao, "A CenterNet++ model for ship detection in SAR images," *Pattern Recognit.*, vol. 112, 2021, Art. no. 107787.
[18] S. Liu et al., "Multi-scale ship detection algorithm based on a lightweight neural network for spaceborne SAR images," *Remote Sens.*, vol. 14, no. 5, 2022, Art. no. 1149.
[19] Z. Li et al., "Lightweight ship detection methods based on YOLOv3 and DenseNet," *Math. Problems Eng.*, vol. 2020, pp. 1–10, 2020.
[20] X. Xu, X. Zhang, and T. Zhang, "Lite-yolov5: A lightweight deep learning detector for on-board ship detection in large-scene sentinel-1 SAR images," *Remote Sens.*, vol. 14, no. 4, 2022, Art. no. 1018.
[21] T. Miao et al., "An improved lightweight RetinaNet for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4667–4679, 2022, doi: 10.1109/JSTARS.2022.3180159.
[22] P. Xu et al., "On-board real-time ship detection in HISEA-1 SAR images based on CFAR and lightweight deep learning," *Remote Sens.*, vol. 13, no. 10, 2021, Art. no. 1995.
[23] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, no. 8, 2022, Art. no. 1908.
[24] Y. Li, S. Zhang, and W.-Q. Wang, "A lightweight faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4006105, doi: 10.1109/LGRS.2020.3038901.
[25] Z. Long et al., "Lira-YOLO: A lightweight model for ship detection in radar images," *J. Syst. Eng. Electron.*, vol. 31, no. 5, pp. 950–956, Oct. 2020, doi: 10.23919/JSEE.2020.000063.
[26] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020, doi: 10.1109/ACCESS.2020.3005861.
[27] Y. Wang et al., " A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 765.
[28] J. Li, C. Qu, and S. Peng, "A ship detection method based on cascade CNN in SAR images," *Control Decis.*, vol. 34, no. 10, pp. 2191–2197, 2019.
[29] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
[30] A. Bochkovskiy, C. Wang, and H. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[31] C. Wang, A. Bochkovskiy, and H. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 7464–7475.
[32] Z. Ge, S. Liu, and F. Wang, "Yolox: Exceeding yolo series in 2021," 2021, *arXiv:2107.08430*.
[33] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019, doi: 10.1109/TGRS.2019.2923988.
[34] S. Wei et al., "Precise and robust ship detection for high-resolution SAR imagery based on HR-SDNet," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 167.
[35] J. Yu et al., "A fast and lightweight detection network for multi-scale SAR ship detection under complex backgrounds," *Remote Sens.*, vol. 14, no. 1, 2022, Art. no. 31.
[36] W. Yu et al., "A lightweight network based on one-level feature for ship detection in SAR images," *Remote Sens.*, vol. 14, no. 14, 2022, Art. no. 3321.

Shuai Liu received the Ph.D. degree in computer application technology from Jilin University, Changchun, China, in 2011.

He is currently a Full Professor with the School of Educational Sciences, Hunan Normal University, Changsha, China. He has authored more than 80 papers in high-quality journals with more than 7000 citations. His research interests include computer vision, machine learning, and multimodal information processing.

Dr. Liu is currently a Distinguished Member of the China Computer Federation (CCF) and a Senior Member of Chinese Institute of Electronics (CIE). He serves as an editor for many journals in IEEE, Elsevier, and Springer.

Pengfei Chen received the B.A. degree in computer science and technology from the Shandong University of Technology, Zibo, China, in 2021. He is currently working toward the M.S. degree in computer science and technology with the College of Information Science and Technology, Hunan Normal University, Changsha, China.

His research interests include artificial intelligence and computer vision.

Yudong Zhang (Senior Member, IEEE) received the Ph.D. degree in signal and information processing from Southeast University, Nanjing, China, in 2010.

He serves as a Chair Professor with the School of Computing and Mathematical Sciences, University of Leicester, Leicester, U.K. His research interests include deep learning and medical image analysis.

Dr. Zhang's citation reached 28 850 in Google Scholar (h-index 95).