(Received 29 November 2024, Accepted 23 January 2025, Scheduled 10 February 2025)

SAR Ship Detection Based on Multi-Scale Feature Cross Fusion

Xiaozhen Ren[1, 2, *], Peiyuan Zhou[1, 2], and Gang Liu[1]

[1]*School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, China*
[2]*School of Information Science and Engineering, Henan University of Technology, Zhengzhou 450001, China*

ABSTRACT: Synthetic aperture radar (SAR) ship detection plays a significant role in ocean monitoring. However, the current SAR ship detection methods face limitations in detecting small and dense ships. To address these issues, a novel SAR ship detection method based on multi-scale feature cross-fusion (MFCNet) is proposed in this paper. In the proposed model, a feature extraction network with a spatial fusion attention mechanism (FESNet) is designed to improve the capability of the backbone network in feature extraction. A multi-intersection spatial pyramid pooling (MISPP) module is proposed to expand the receptive field and enhance the semantic information. Furthermore, a feature cross-fusion network (FCFNet) is designed to comprehensively integrate features of different scales for enhancing SAR ship detection performance. Experimental results demonstrate that the proposed model achieves high detection performance on the SSDD and HRSID datasets, providing more reliable technical support for ship detection in maritime environments.

1. INTRODUCTION

SAR ship detection [1–5] is an important application area in SAR image detection. Traditional methods for SAR ship detection typically utilize image preprocessing and target features. For example, Xu and Liu [6] utilized visual effects and gradient histograms for ship detection. Lee et al. [7] considered that ship detection algorithms are affected by weather and terrain, and combined relevant information from electro-optical satellite images to enhance detection accuracy. Wu et al. [8] improved the representation ability of ship features by fusing multi-level feature information. Yang et al. [9] introduced a unified algorithm that integrates a constant false alarm rate with temporal analysis techniques. Yang et al. [10] studied the geometric optimization method for SAR ship detection and proposed a three-stage detection algorithm based on the relationship among statistical features. However, these techniques only yield accurate detection results under calm sea conditions. When the sea surface is crowded with ships or contains small targets, the detection performance deteriorates significantly. Therefore, traditional ship detection methods are no longer sufficiently effective.

The rapid development of deep learning technology has promoted its widespread application in various fields, such as speech recognition and image processing. Object detection is an important task in image processing. With the development of convolutional neural networks (CNN), object detection technology has also seen significant improvements. Object detection methods mainly include two types: two-stage and single-stage object detection. The two-stage detection methods usually achieve high detection accuracy but are relatively slow. Conversely, the single-stage detection methods improve detection speed and simplify the process, which make them appropriate for real-time applications. The activities of ships on the ocean are complex and variable, and object detection technology enables rapid and accurate identification of ships at sea. This technology not only offers essential data insights for maritime management but also drives advancement in the shipping industry. Therefore, it holds substantial theoretical and practical significance by applying object detection technology to ship detection. Chen et al. [11] effectively utilized the Gaussian mixture Wasserstein generative adversarial network (GAN) and improved it on the generated artificial sample data of small ships. To increase the detection speed, Chen et al. [12] designed a novel network, which detected tilted ships in images through a bidirectional feature network, attention mechanism, and rotation decoupling operation. To address the challenges of ship detection on mobile platforms, Feng et al. [13] constructed a real-time detection network. Shan et al. [14] designed a dense attention detection network to address the impact of inherent speckles in SAR images. In addition, they proposed a new ship detection algorithm aimed at improving both detection speed and accuracy. Considering the potential impact of restricted areas on object detection precision, Li et al. [15] constructed a multi-scale ship detection algorithm using a significance estimation technique. Bai et al. [16] applied the feature enhancement pyramid and shallow feature reconstruction module to design a new SAR ship detection network. Lu et al. [17] established a brightness temperature model for offshore ships and proposed a four-step ship detection tracking algorithm. Niu et al. [18] proposed an efficient encoder-decoder network to extract features. Si et al. [19] considered the diversity of maritime ships and the influence of the sea surface on ship detection. Zhou et al. [20] proposed an edge semantic decoupling module that considers the overlap phenomenon between nearshore marker boxes of ships. Moreover, a semantic segmentation branch was introduced to accurately identify ship targets.

[*] Corresponding author: Xiao-Zhen Ren (rxz235@163.com).

doi:10.2528/PIERB24112910

Published by THE ELECTROMAGNETIC ACADEMY

Xu et al. [21] designed a real-time high-speed ship detection system that effectively utilized algorithms and hardware. Pan et al. [22] effectively utilized scattering mechanisms on ship detection. CNN-based SAR ship detection has been extensively utilized in recent research. However, challenges remain due to high noise in SAR images, which makes it challenging to distinguish ships from their surroundings. The multi-scale nature of ships, particularly small and dense ships, can result in numerous false negatives. To address these issues, a coordinate attention module was introduced by Yang et al. [23] to lessen the interference of backgrounds on ship detection. Li et al. [24] proposed an improved method addressing the blurring of ship contours caused by SAR imaging. This method integrates semantic information through four feature layers and effectively utilizes the channel attention mechanism for feature enhancement. Cui et al. [25] considered that current object detection algorithms have low detection efficiency for small ships and used refined cascaded feature maps to highlight significant features at specific scales. Shan et al. [26] designed a dense multilayer deep network to extract feature information and used an anchor mechanism for classification regression estimation. To tackle the challenges of multi-scale and dense ship detection in complex scenarios, Ma et al. [27] proposed an anchor-free detection method based on keypoint estimation and incorporated a keypoint estimation module to improve detection performance. Li et al. [28] proposed an adaptive superpixel-level detection method to address the challenge in detecting dense ships near the shore. Sun et al. [29] focused on small ships and designed a dual-branch activation network, utilizing feature encoding to obtain more fine-grained features. Ren et al. [30], considering the difficulty of separating targets from backgrounds in complex SAR images, designed an efficient feature extraction network to capture image information and utilized a bidirectional attention module to enhance the extraction of information on small ships. Cui et al. [31] proposed an anchor-free ship detection network that enhances semantic features through a spatial enhancement attention module and utilized keypoint estimation to locate targets. Xie et al. [32] designed an efficient feature pyramid network to model multi-scale dependency and enhance feature representation, aiming to strengthen the dependency in the feature fusion process. Huang et al. [33] designed a feature connection module to distribute semantic features and introduced a self-attention mechanism to enhance feature extraction capability. Xiao et al. [34] combined image pyramid and convolutional neural network to detect surface defects in images and generated image masks within bounding boxes. Gao et al. [35] considered that previous multimodal fusion methods often overlooked inter-layer differences and proposed a perceptual refinement network to preserve the hierarchical information of multimodal data. Although the above methods have improved the detection performance for dense ships, there are still missed and false detection when facing dense and small ships. To tackle these issues, we present a novel SAR ship detection method multi-scale feature cross-fusion (MFCNet) in this study. The primary contributions are as follows:

- To improve the capability of the backbone network in feature extraction, a feature extraction network with a spatial fusion attention mechanism (FESNet) is designed in this paper. Moreover, to eliminate the interference of complex backgrounds on ship features, the spatial fusion attention (SFA) mechanism is proposed and embedded into FESNet by utilizing the cross-dimensional interaction.
- To mitigate the problem of positional ambiguity and loss for small and dense ships in SAR ship detection, a multi-intersection spatial pyramid pooling (MISPP) module is proposed to expand the receptive field and enhance the semantic information. By employing pooling convolutions at different ratios, MISPP expands the receptive field while maintaining high resolution.
- To comprehensively integrate features of different scales for enhancing SAR ship detection performance, a feature cross-fusion network (FCFNet) is proposed in this paper. The proposed FCFNet adopts two fusion paths and introduces cross operations within these paths to fully leverage information from different levels of features.

The remaining sections are organized as follows. In Section 2, we review the relevant work and provide detailed descriptions of the MFCNet network. Section 3 describes the experiments conducted in this study, along with their results. Based on these results, we will analyze the model's performance to further summarize the MFCNet network structure. We will discuss this in Section 4.

2. PROPOSED METHODS

2.1. Overall Network Structure

As one of the mainstream single-stage object detection algorithms, the YOLO series has seen gradual improvements in performance with each version update. YOLOv7l, one of the most advanced versions of the YOLO series, is chosen as the baseline to address the challenge of recognizing small and dense ships in SAR ship detection.

Figure 1 depicts the entire network structure of the MFCNet. First, the input image is passed to the feature extraction network FESNet. In FESNet, the efficient feature extraction block (EFE-Block) is designed for feature extraction. Moreover, to eliminate the interference of complex backgrounds on ship features, the spatial fusion attention (SFA) mechanism is proposed and embedded into FESNet. The refined downsampling (RDS) module is introduced to more accurately locate and identify target ships. Additionally, to more effectively address the challenge of identifying small ships and dense ships, the feature obtained from FESNet is transmitted to MISPP to better capture features of different scales. Next, the feature cross-fusion network (FCFNet) is used to perform up-bottom, bottom-up, and cross-path fusion on the feature layers obtained by the feature extraction network. This operation could integrate features of different scales Furthermore by introducing residual path, the problem of gradient vanishing can be solved while accelerating model convergence. The residual structure also allows the network to directly combine low-level features with high-level

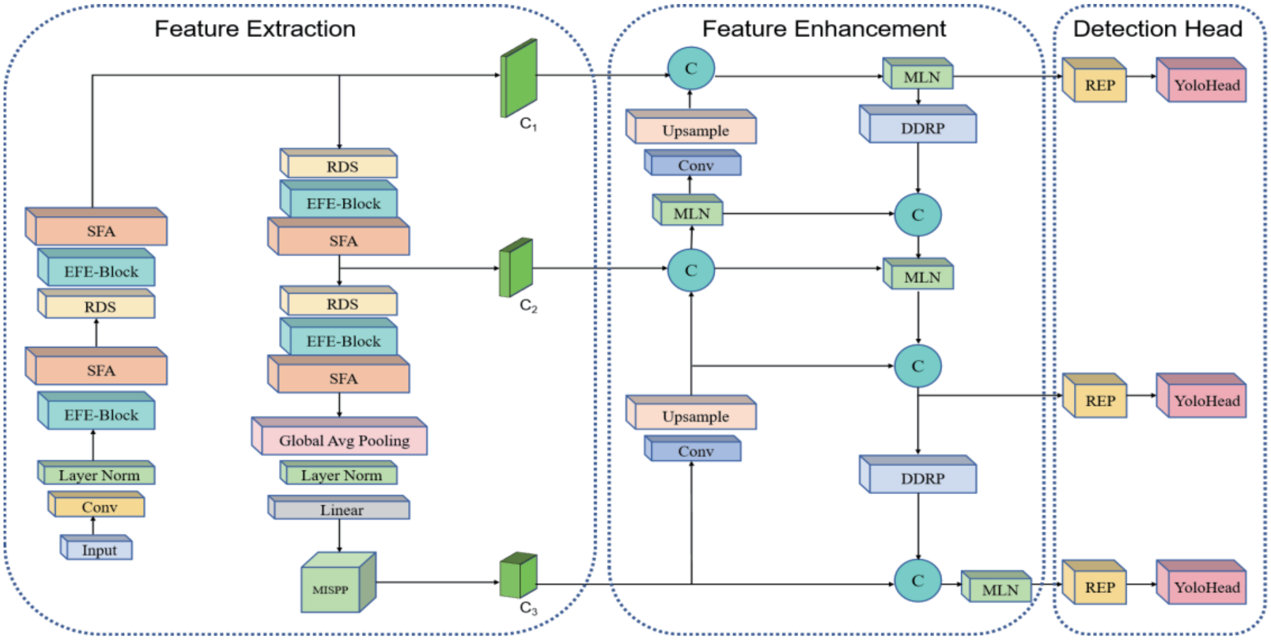

FIGURE 1. Overall structural framework of MFCNet.

features through skip connections, preserving fine-grained information and improving the accuracy of the model in complex scene detection.

2.2. Feature Extraction Network with Spatial Fusion Attention Mechanism FESNet

A feature extraction network with spatial fusion attention mechanism FESNet was designed to improve the capability of the backbone network in feature extraction.

Figure 2 depicts the entire architecture of the FESNet. It is mainly composed of EFE-Block modules, RDS modules, SFA attention modules, convolution layers, and normalization layers. Unlike other images, SAR images have a significant portion of background pixels, which reduces the effectiveness of the model in identifying target features. To reduce computational complexity and improve feature extraction ability, EFE-Block module is designed as an efficient feature block for feature extraction. Furthermore, by stacking multiple EFE-Block modules, the backbone network can learn more feature representations, thereby enhancing the SAR ship detection performance. In addition, the SFA mechanism is incorporated into the FESNet to enhance detection performance.

By utilizing the cross-dimensional interaction, the SFA mechanism strengthens the ability of the backbone to extract critical information and suppress unimportant features, improving model performance in complex environments. The EFE-Block is composed of two paths. The first is the residual path, which accelerates network convergence while promoting effective information transmission and fusion. Furthermore, it improves detection performance by preserving original feature information. For the second path of the EFE-Block module, a 7×7 depthwise separation convolution is first used to greatly reduce the number of parameters. Moreover, the 7×7 depthwise separation convolution captures a large receptive field, helping to extract richer image features.

The layer normalization is then performed to standardize the features, thus ensuring that the inputs of the subsequent layers have a consistent distribution to stabilize and accelerate the model training. Later, the linear transformation is realized through the linear layer, and the input features are mapped to another feature space. Then, GELU combines the nonlinear characteristics of ReLU and the Sigmoid to give the model more expression ability and better capture the complex mode. After that, the linear transformation is realized again to further map and reconstruct the features. Finally, the layer scale is used to scale each characteristic channel, and the drop path layer helps prevent overfitting by randomly discarding certain paths. Moreover, to more accurately locate and detect ships, the RDS module is constructed to enhance the downsampling capability during the feature extraction. This module first applies layer normalization, followed by a convolution operation with both kernel size and stride set to 2, and finally through a maximum pooling layer. This design reduces internal covariate shift while enhancing the network representation capability, thereby better capturing small and dense ships in complex backgrounds.

The structure of SFA is shown in Fig. 3. SFA utilizes four parallel paths to extract attention weights for grouped feature maps. To ensure the uniform distribution of spatial semantic features, the SFA reconstructs some channels into the batch dimension and groups the channel dimension into many subfeature groups. Specifically, SFA enhances the feature extraction capability by introducing cross-dimensional interaction.

In SFA, two parallel paths are located in the 1×1 convolution branch, the third path is in the 3×3 convolution branch, and the fourth path is in the 5×5 convolution branch. To capture dependencies among all channels and reduce computational com-

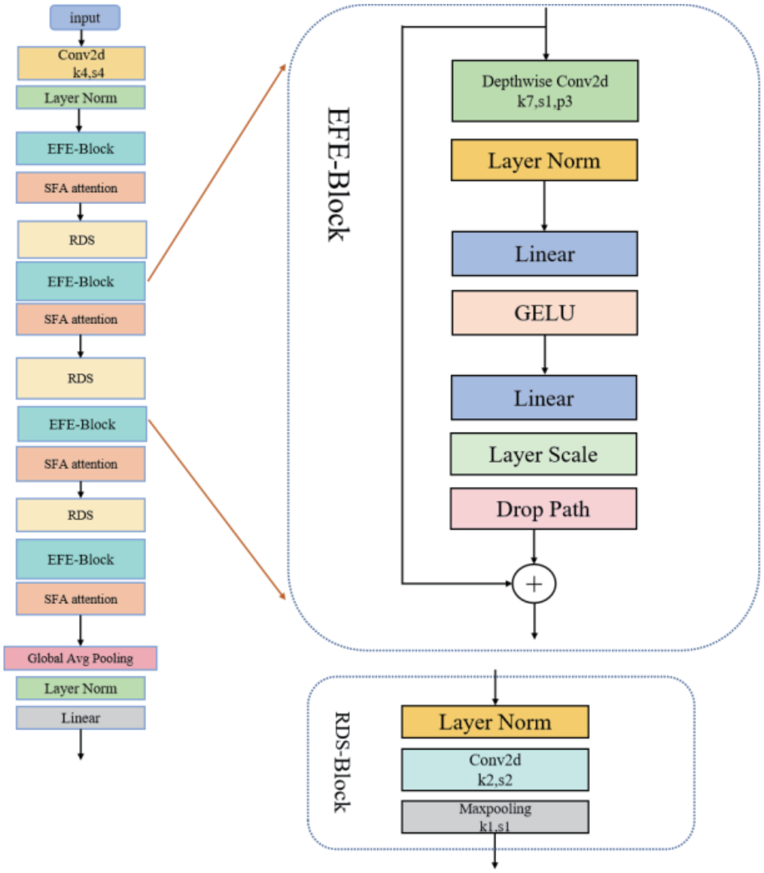

FIGURE 2. The structure of FESNet module.

plexity, we model the interaction of cross-channel data in the direction of the channel. For the input $X \in R^{C \times H \times W}$, SFA partitions X into G sub-features to learn different semantic information. The feature in the 1×1 branch can be represented as $X_{1x} \in R^{C//G \times 1 \times W}$ and $X_{1y} \in R^{C//G \times H \times 1}$, which are encoded along the two spatial directions separately using 1D global average pooling. Subsequently, the features from these two directions are fused using the concatenate function. The fused feature then undergoes 1×1 convolution to extract features and increase the channel dimensions. Then, a nonlinear sigmoid function is applied to compress the input values between 0 and 1, fitting them to approximate a two-dimensional Gaussian distribution over the linear convolution. Afterward, the fitted features are multiplied with the grouped input features using matrix multiplication to obtain a new feature map. Finally, the output feature X^{ω} can be obtained through group normalization operation. This process can be represented by

$$X^{\omega} = GN(\theta(nsf(c1(p)), \quad X_0 \in R^{C//G \times H \times W})) \tag{1}$$

$$p = \text{Concat}(gap(X_{1x}), gap(X_{1y})) \tag{2}$$

where X_0 represents a residual network after channel-wise division; $gap(.)$ represents 1D global average pooling operation; Concat$(.,.)$ represents concatenation operation; $c1(.)$ represents 1×1 convolution operation; $nsf(.)$ represents nonlinear sigmoid function operation; $\theta(.,.)$ represents the matrix multiplication; and GN is the GroupNorm operation.

The features for the 3×3 branch and 5×5 branch can be represented as $X_2 \in R^{C//G \times H \times W}$ and $X_3 \in R^{C//G \times H \times W}$, respectively. The feature X_2 undergoes convolution with a 3×3 kernel to produce X^{∂}, and the feature X_3 undergoes convolution with a 5×5 kernel to produce X^{σ}, which can be represented by

$$X^{\partial} = c3(X_2), \quad X^{\sigma} = c5(X_3) \tag{3}$$

where $c3(.)$ represents the convolution with a 3×3 kernel, and $c5(.)$ represents the convolution with a 5×5 kernel.

Then, the output features of the three branches are aggregated by cross-dimensional interactions to enhance the feature extraction capability. The final output can be represented by

$$X^{*} = \theta(nsf(\Omega(a, b, c))) \tag{4}$$

$$a = \theta(\Theta(mp(X^{\omega})), X^{\partial}) \tag{5}$$

$$b = \theta(\Theta(mp(X^{\partial})), X^{\omega}) \tag{6}$$

$$c = \theta(\Theta(mp(X^{\sigma})), X^{\partial}) \tag{7}$$

where mp(.) represents the 2D global maximum pooling operation; $\Theta(.)$ represents the softmax function operation; $\theta(.,.)$ represents the matrix multiplication; $\Omega(\cdot)$ represents the effective feature fusion operation; and $nsf(.)$ represents the nonlinear sigmoid function operation.

By leveraging the collaborative efforts of these modules, FESNet can effectively handle ship images within complex backgrounds, significantly enhancing the accuracy of SAR ship detection.

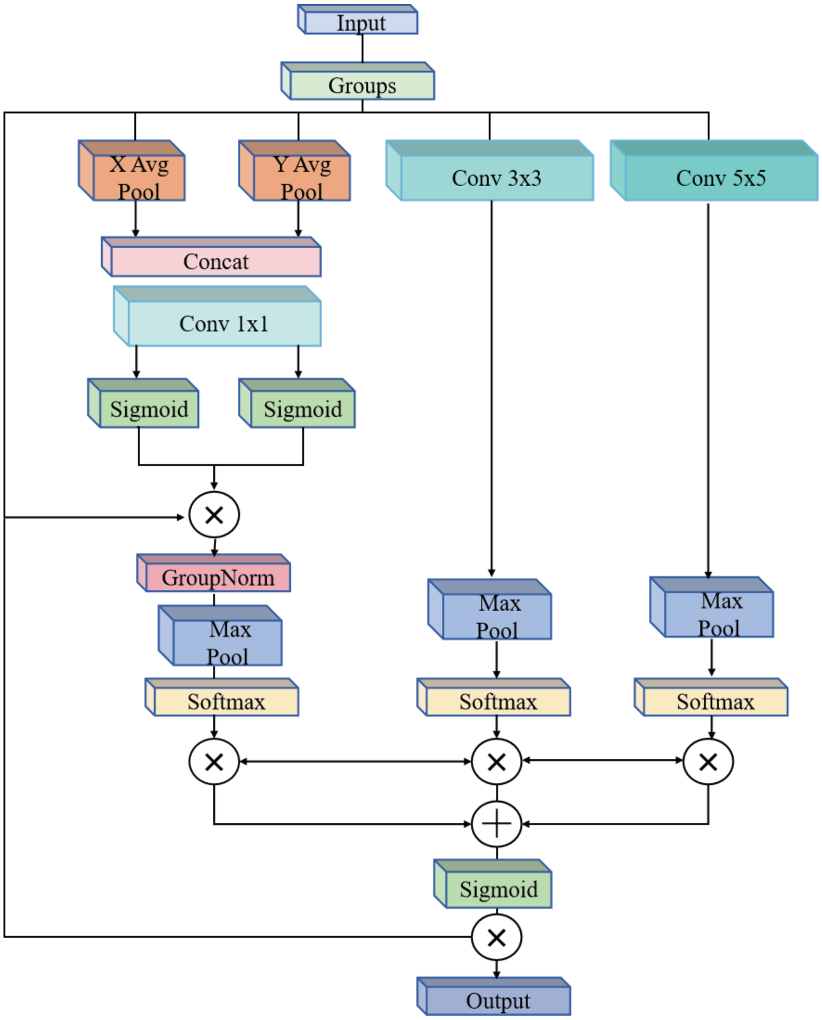

FIGURE 3. The structure of SFA module.

2.3. Multi-Intersection Spatial Pyramid Pooling MISPP

Dense ships and small ships are crucial components in SAR ship images, but their positional information may become increasingly ambiguous or even lost as the depth of the network [27–30]. Conventional spatial pyramid pooling networks divide the image into blocks for pooling, which reduces spatial resolution and leads to loss of fine-grained image details [31–34]. To address the aforementioned issues, the MISPP module is proposed to expand the receptive field and enhance the semantic information. By employing pooling convolutions at different ratios, MISPP expands the receptive field while maintaining high resolution. Moreover, intersection pooling enables better capture of correlations among different features, allowing the network to more accurately understand and differentiate between different classes. By incorporating intersection pooling, the MISPP can maintain detailed information while preserving the receptive field, thereby enhancing the capacity to identify small and dense ships.

Figure 4 depicts the architecture of the MISPP. First, three convolution operations are applied to the input feature map X to obtain the feature map X_1. Each convolution operation comprises three parts. Feature extraction and downsampling are obtained through a 3×3 convolutional kernel. Batch normalization is used to normalize each batch of data, which helps prevent model overfitting. The SiLU activation operation enables the network to better recognize complex data and aids faster convergence while mitigating issues. Next, the feature map X_1 undergoes two-dimensional (2D) maximum pooling with a kernel size of 5 to produce a new feature map, which is then concatenated with X_1 to form the feature map Y_1. Subsequently, Y_1 undergoes 2D maximum pooling with a kernel size of 7 to produce a new feature map, which is then concatenated with X_1 to form Y_2. Then, features Y_3 and Y_4 can be obtained through similar operations. This process can be expressed by

$$X_1 = dbs(X) \tag{8}$$

$$Y_i = \text{Concat}(C_i(Y_{i-1}), X_1), \quad i = 1, 2, 3, 4 \tag{9}$$

where dbs(.) represents a set of convolution operations; $C_i(.)$ represents the corresponding 2D maximum pooling operation; and Concat(., .) represents the concatenation operation.

Each feature map obtained above contains different feature information, and the information in each feature map is interrelated. Therefore, performing cross-fusion operations on the different feature maps can better capture the location informa-

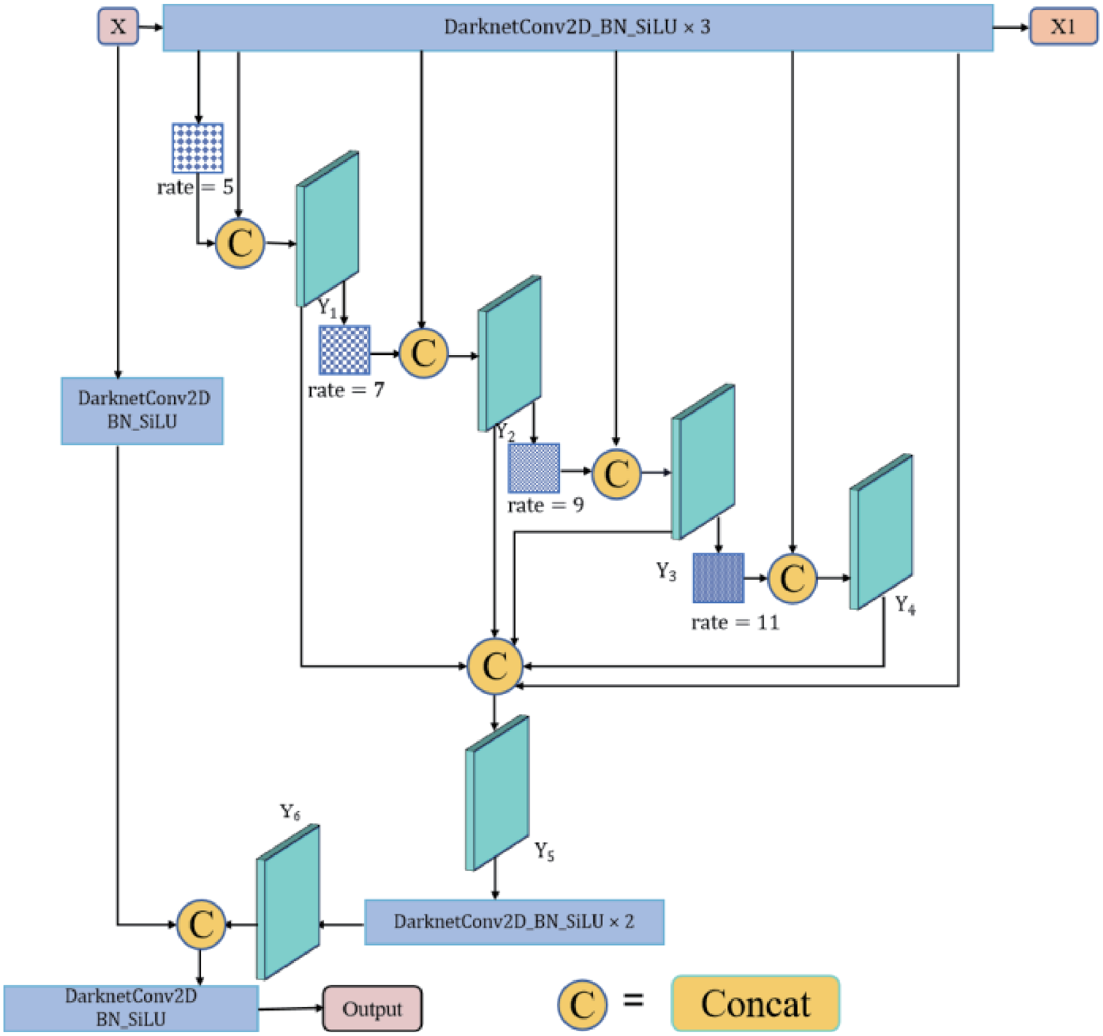

FIGURE 4. The structure of MISPP module.

tion of small and dense target ships. The fused feature Y_5 can be expressed by

$$Y_5 = \text{Concat}(Y_1, Y_2, Y_3, Y_4, X_1) \tag{10}$$

where Concat(., .) denotes the concatenation operation.

After obtaining the feature map Y_5, it undergoes two sets of convolution operations to obtain Y_6, which contains higher-level semantic feature information. Then, by introducing a residual structure, the input feature map X is fused to acquire the final output X_N. The calculation process of the final output X_N is given by

$$X_N = dbs(\text{Concat}(dbs(dbs(Y_5)), dbs(X))) \tag{11}$$

where dbs(.) represents a set of convolution operations, and Concat(.,.) represents the concatenation operation.

2.4. Feature Cross-Fusion Network FCFNet

In object detection, different levels of feature maps are generated after feature extraction. Low-level features might lack semantic information, while high-level features might lose object details. Combining features from different levels helps in better understanding the image content, thereby enhancing object detection performance [35–38]. Therefore, a feature cross-fusion network FCFNet is proposed in this paper. This module adopts two fusion paths and introduces cross operations within them. These paths fully leverage information from different levels of features, improving detection performance. Moreover, the dual downsampling residual path (DDRP) module and the multi-branch learning network (MLN) are designed and embedded into FCFNet during feature fusion to enhance detection performance. The main role of the DDRP module is to perform downsampling by adopting a three-branch structure, which prevents gradient vanishing and enrich feature extraction. The MLN is constructed to learn more features and better capture complex image information by controlling the shortest and longest gradient paths, thus improving the accuracy of SAR ship detection.

Figure 5 illustrates the architecture of FCFNet. First, feature C_3 is passed through the MISPP module to acquire P_1. Next, feature P_1 undergoes convolution and upsampling, and the resulting feature is fused with the convolved feature map C_2 to obtain a new feature. This new feature then passes through the MLN module to capture complex image information, resulting

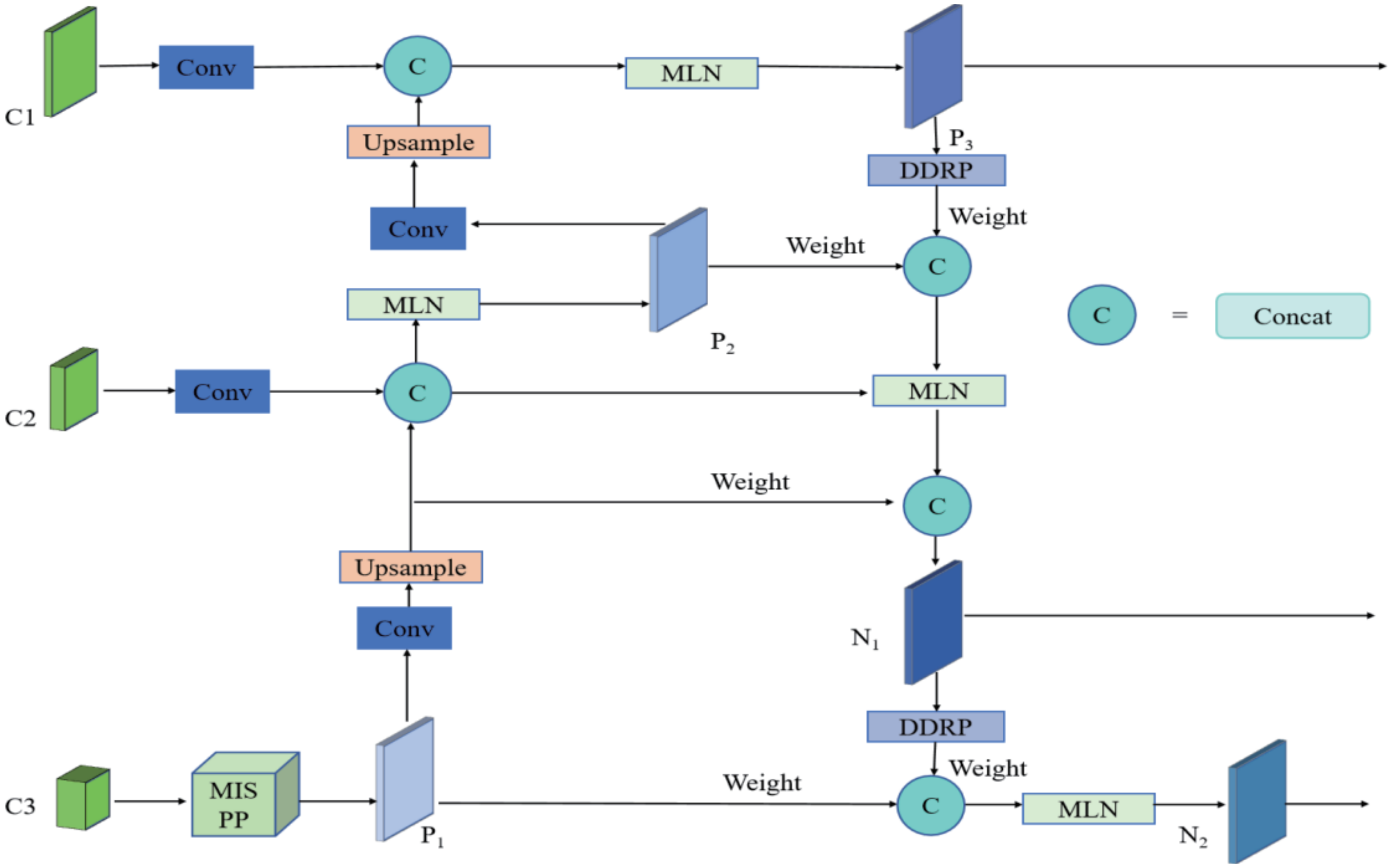

FIGURE 5. The structure of FCFNet module.

in feature P_2. Similarly, the feature P_3 is obtained. Features P_2 and P_3 can be represented by

$$P_2 = mln(\text{Concat}((Up(Conv(mp(C_3)))), Conv(C_2))) \quad (12)$$

$$P_3 = mln(\text{Concat}(Up(Conv(P_2)), Conv(C_1))) \quad (13)$$

where mp(.) represents the operation of the MISPP module; Conv(.) denotes the convolution operation; Up(.) is the upsampling operation; Concat(., .) denotes the concatenation operation; and mln(.) represents the operation of the MLN module.

Then, to generate more accurate results, a bottomup feature fusion path is introduced, incorporating skip connection and weighted operation for feature fusion. Firstly, feature map P_3 undergoes the DDRP module to obtain a new feature map. Subsequently, to fuse more features this new feature is fused with P_2 using skip connection and further passed through the MLN module to capture complex image information. Additionally, weighted operation is performed on different features during this process to discern the influence of various features.

The resulting feature is then fused with the upsampled feature map P_1 to produce feature N_1. Then, feature N_2 is obtained through a similar operation. Features N_1 and N_2 can be represented by

$$N_1 = \text{Concat}(mln(q), \omega_{23} Up(Conv(P_1))) \quad (14)$$

$$q = \text{Concat}(\omega_{21} ddrp(P_3), \omega_{22} P_2) \quad (15)$$

$$N_2 = mln(\text{Concat}(\omega_{31} ddrp(N_1), \omega_{32} P_1)) \quad (16)$$

where ω_{ij} represents the weights of different feature paths, ranging between 0 and 1; ddrp(.) represents the operation of the DDRP module; Concat(., .) represents concatenation operation; Up(.) represents upsampling operation; and mln(.) represents the operation of the MLN module.

In the FCFNet network, the structure of the DDRP module is depicted in Fig. 6. There are three branches in the DDRP module, which are two downsampling branches and one residual branch. The residual path is introduced as the third branch, directly merging from the input to interact with the other two branches. The introduction of the residual path is crucial as it ensures easier gradient propagation during backpropagation, alleviating gradient vanishing issues in deep networks. This three-branch design facilitates downsampling to enhance feature fusion and improve gradient propagation paths, thereby preventing gradient vanishing and enriching feature extraction.

The MLN module is illustrated in Fig. 7. The MLN module controls gradients through different path lengths, enabling the network to learn more features. The final feature is generated by fusing the features from different paths. This multi-branch structure improves the model performance across various scenarios, enhancing the accurate detection of complex targets. Moreover, in SAR ship detection, the MLN module could effectively identify different sizes of ships, reducing detection errors caused by overlapping effects.

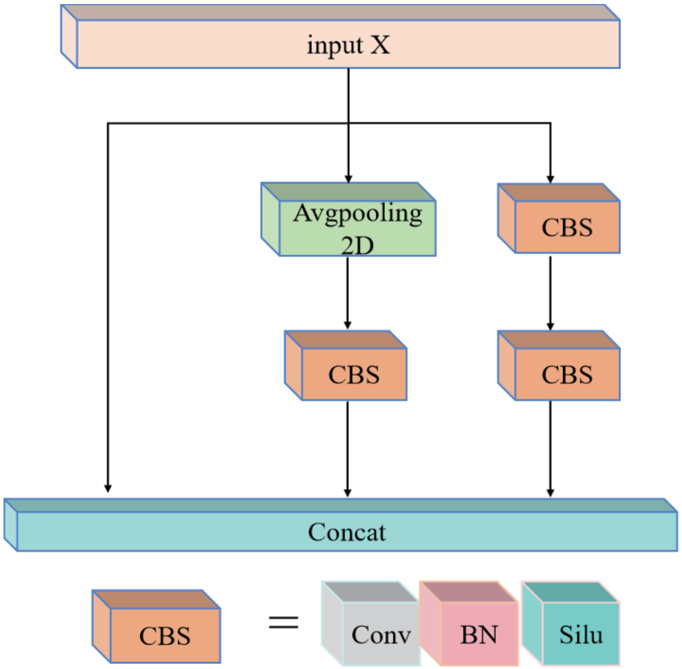

FIGURE 6. The structure of DDRP module.

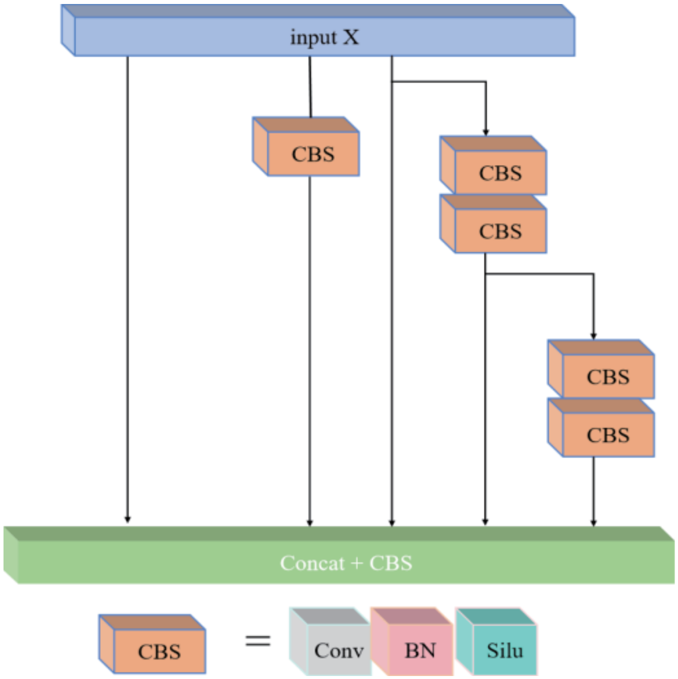

FIGURE 7. The structure of MLN module.

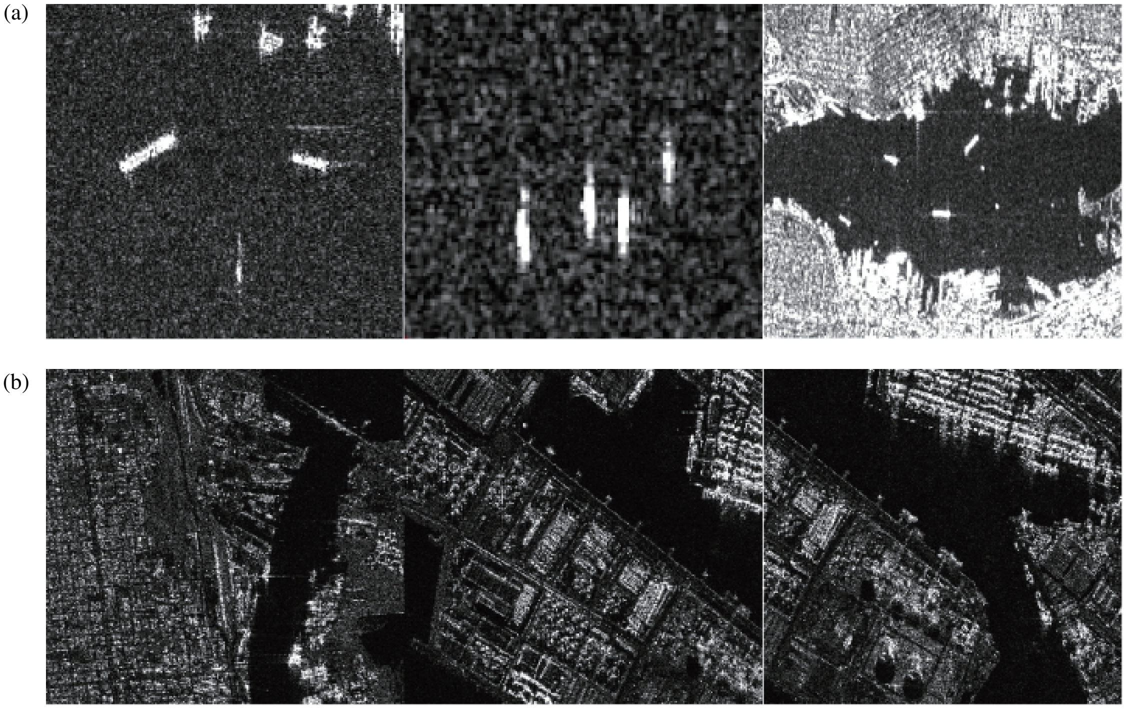

FIGURE 8. Some typical images of SSDD and HRSID datasets. (a) Typical images of SSDD, (b) typical images of HRSID.

3. EXPERIMENTS

3.1. Datasets

Experiments on SSDD [39] and HRSID [40] datasets are carried out to verify the performance of the proposed MFCNet model.

In the area of SAR ship detection, SSDD dataset is frequently utilized. The SSDD dataset comprises a total of 1160 images containing 2456 ships. Additionally, the SSDD dataset annotates only ships with a pixel count greater than 3. In the experiments, we divide the SSDD dataset into a training set, validation set, and testing set with a ratio of 7 : 2 : 1. Figure 8(a) displays typical ship images from the SSDD dataset.

TABLE 1. The performance comparison with raw YOLOv7l on SSDD dataset.

Method	F1 (%)	P (%)	mAP (%)	R (%)	GFLOPs (G)	FPS
YOLOv7l	93	93.24	93.46	92.16	106.5	36.5
MFCNet	96	96.52	96.89	95.87	187.3	18.2

HRSID dataset is a high-resolution dataset used not only for ship detection but also for instance segmentation. Compared to low-resolution SAR ship datasets, HRSID has more detailed ship characteristics and better feature representation. The HRSID dataset is split with a ratio of 6.5 : 3.5 into the training and test sets for the experiments. Figure 8(b) displays typical ship images from the HRSID dataset.

3.2. Experimental Details

Experiments were carried out on a Windows 10 system using an NVIDIA GeForce RTX 3060 12GB graphics card and an Intel Core i5-12490F processor. With Python version 3.8 and CUDA version 11.8, the batch size was set to 4. The learning rate was initially fixed at 0.01.

3.3. Evaluation Metrics

To evaluate the overall performance of MFCNet effectively and intuitively, six metrics are used: precision (P), recall (R), mean average precision (mAP), giga floating-point operations per second (GFLOPs), frames per second (FPS), and F-measure (F1). Their definitions are as follows

$$P = \frac{N_{TP}}{N_{TP} + N_{FP}} \tag{17}$$

$$R = \frac{N_{TP}}{N_{TP} + N_{FN}} \tag{18}$$

$$mAp = \frac{1}{Q} \sum\nolimits_{j=1}^{N} AP_j \tag{19}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{20}$$

where N_{TP}, N_{FP}, and N_{FN} denote the number of true positives, false positives, and false negatives, respectively; Q is the total number of classes; j represents the j-th class; and AP is the average precision.

3.4. Experiments on SSDD Dataset

The experimental outcomes of the YOLOv7l [41] and the MFCNet on the SSDD dataset are displayed in Table 1. As can be seen from the data comparison in Table 1, in contrast to YOLOv7l, although GFLOPs have increased by 80.8 G, and FPS has decreased by 18.3, the MFCNet has shown a detection performance improvement by 3% in F1 score, 3.28% in precision, 3.43% in mAP and 3.71% in recall. Next, the visualization results obtained from the SSDD dataset are shown in Fig. 9. In the visualization results, yellow boxes indicate false positives; blue boxes represent false negatives; and red boxes represent true positives. The visualization experiment conducted a comparative analysis using images from four scenarios. In the scenarios far from the coast, such as the third group of images, there are no other disturbances besides ships in the image, and both the YOLOv7l and MFCNet can work well. However, due to the ships being too small, the YOLOv7l has one missed detection, whereas our proposed MFCNet successfully detected all target ships. Furthermore, from the comparison of the first group images, it can be seen that detection becomes significantly more challenging near the dock due to background interference. The YOLOv7l missed detections twice, whereas our proposed MFCNet performed better in detection. Although there was also one missed detection, the detection performance of MFCNet is better than YOLOv7l. From the comparison of the second group of images, it can be observed that the detection difficulty increased at the image edges due to the lack of ship information. The YOLOv7l missed one detection at the edge of the image, failing to recognize the ship information, while our proposed MFCNet accurately identified all target ships. In dense ships and complex background scenarios such as the fourth group of images, the YOLOv7l had several false positives, whereas our proposed MFCNet model accurately detected all ship targets without missing any or producing false positives. The comparative analysis of the images above demonstrates that the proposed MFCNet exhibits strong detection accuracy in complex coastal backgrounds, as well as scenarios involving small and densely packed ships.

In addition, a more thorough performance comparison of the MFCNet with other excellent models is shown in Table 2. Compared with the SSD, although GFLOPs has increased by 87.8 G, FPS has decreased by 3.5, the MFCNet has shown an improvement by 9% in F1 score, 9.95% in mAP, 5.47% in P. Even compared to the suboptimal YOLOv8l although GFLOPs has increased by 21.9 G and FPS decreased by 8.9, the MFCNet has shown an improvement by 1% in F1 score, 1.07% in mAP, and 0.31% in P. Based on the above analysis, it can be seen that although the computational complexity has increased, our MFCNet has shown more excellent detection accuracy in SAR ship detection than other single-stage ship detection methods.

3.5. Experiments on HRSID Dataset

Table 3 shows the performance comparison results of MFCNet and YOLOv7l on the HRSID dataset. Compared to YOLOv7l, although GFLOPs have increased by 80.8 G, and FPS has decreased by 18.4, the performance metrics $F1$, P, mAP, and R of MFCNet have been significantly improved. Among them, $F1$ score increased by 4%; P increased by 2.2%; mAP increased by 3.96%; and R increased by 4.38%. Then, the visualization operation is performed, and the results are displayed in Fig. 10.

TABLE 2. Comparison of the performance metrics of different models on SSDD dataset.

Method	Backbone	F1 (%)	mAP (%)	P (%)	GFLOPs (G)	FPS
SSD	VGG-16	87	86.94	91.12	99.5	21.7
RetinaNet	ResNet-50	90	91.83	92.95	175.4	19.9
CenterNet	ResNet-50	91	90.45	92.63	45.3	22.3
YOLOv8s	CSPDarkNet	93	93.82	94.65	28.8	90.6
YOLOv8m	CSPDarkNet	94	94.51	95.27	79.1	45.1
YOLOv8l	CSPDarkNet	95	95.82	96.21	165.4	27.1
RT-DETR [42]	ResNet-50	95	93.57	92.68	103.4	25.3
YOLOv9 [43]	CSPDarkNet	94	94.62	95.12	102.1	28.4
MFCNet (Ours)	FESNet	96	96.89	96.52	187.3	18.2

TABLE 3. The performance comparison with raw YOLOv7l on HRSID dataset.

Method	F1 (%)	P (%)	mAP (%)	R (%)	GFLOPs (G)	FPS
YOLOv7l	86	92.84	89.56	80.85	106.5	36.0
MFCNet	90	95.04	93.52	85.23	187.3	17.6

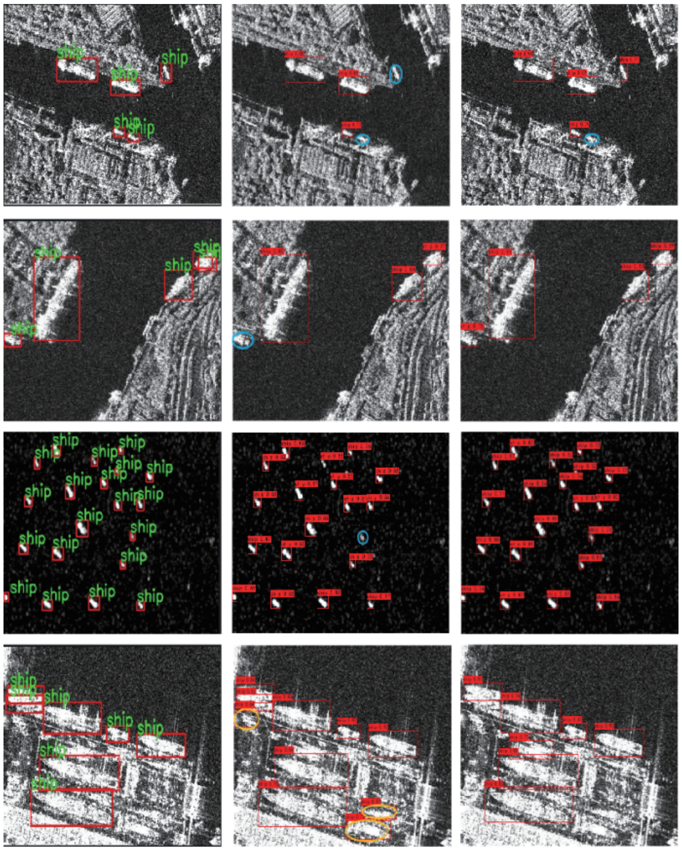

FIGURE 9. SAR ship detection results on SSDD dataset.

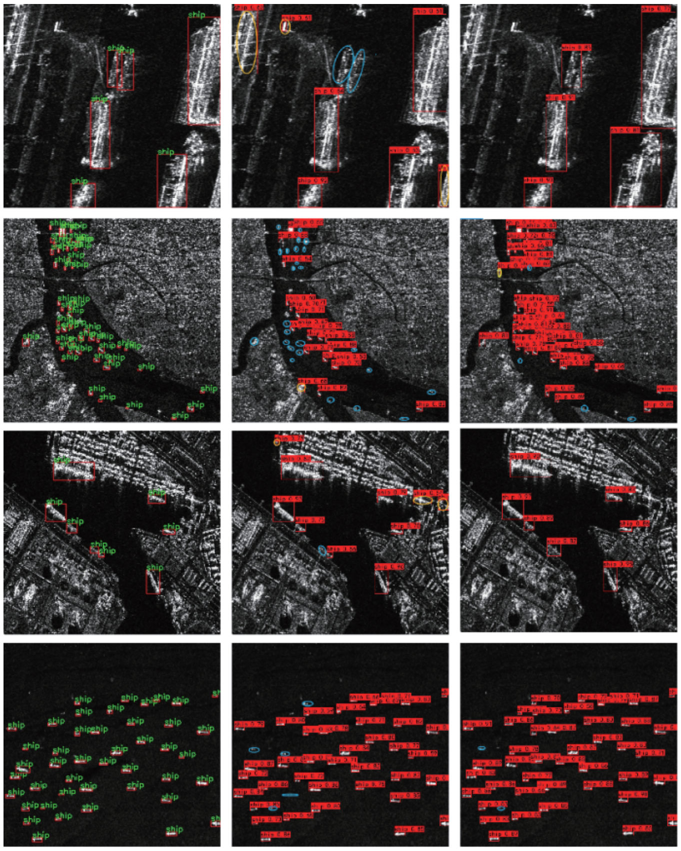

FIGURE 10. SAR ship detection results on HRSID dataset.

The visualization results are compared in four different scenarios. From the results, we can see that YOLOv7l has both missed and false detections for scenes with complex backgrounds of near-coastal ships. In the first group of images, two false negatives and two false positives appear in the detection result of YOLOv7l. However, our proposed MFCNet is able to accurately detect all the target ships without any false detections. In addition, for ships moored in the harbor, the complexity of the image background information and the diverse scales of the ships lead to greater interference in ship detection. Such

TABLE 4. Comparison of the performance metrics of different models on HRSID dataset.

Method	Backbone	F1 (%)	mAP (%)	P (%)	GFLOPs (G)	FPS
SSD	VGG-16	87	85.14	90.07	99.5	21.5
RetinaNet	ResNet-50	88	87.23	91.45	175.4	19.2
CenterNet	ResNet-50	89	88.65	91.82	45.3	20.8
YOLOv8s	CSPDarkNet	87	90.81	92.09	28.8	89.2
YOLOv8m	CSPDarkNet	89	91.16	93.64	79.1	44.9
YOLOv8l	CSPDarkNet	90	92.23	94.17	165.4	27.4
RT-DETR	ResNet-50	88	90.92	92.68	103.4	24.6
YOLOv9	CSPDarkNet	90	92.76	94.35	102.1	27.9
MFCNet (Ours)	FESNet	90	93.52	95.04	187.3	17.6

TABLE 5. The ablation experiment results on SSDD dataset.

FESNet	MISPP	FCFNet	F1 (%)	P (%)	mAP (%)	R (%)	GFLOPs (G)	FPS
—	—	—	93	93.24	93.46	92.16	106.5	36.5
✓	—	—	94	93.87	94.37	93.36	175.4	21.6
—	✓	—	94	94.36	94.95	93.24	108.6	38.9
—	—	✓	94	94.72	94.88	93.68	126.4	30.5
✓	✓	—	95	95.07	95.12	94.84	179.6	23.2
✓	✓	✓	96	96.52	96.89	95.87	187.3	18.2

as in the third group of images, YOLOv7l shows one false negative and three false positives, while MFCNet accurately identifies all the target ships near the harbor. Especially, the detection effect of MFCNet is much better than that of YOLOv7l for the areas with dense small ships, such as the second and fourth groups of images. Due to the large number and concentrated distribution of tiny ships in the images, the detection results of YOLOv7l are very unsatisfactory, with many false negatives and false positives. However, our proposed MFCNet has obvious advantages over YOLOv7l, although it also has several false negatives and one false positive. The results demonstrate that our MFCNet successfully improves the performance of SAR ship detection.

A more detailed performance comparison between our MFCNet and other excellent models is given in Table 4. From Table 4, it can be concluded that MFCNet outperforms other models in F1, mAP, and P metrics. Compared with SSD, although GFLOPs has increased by 87.8G, FPS has decreased by 3.9, the MFCNet has shown an improvement by 3% in F1 score, 8.38% in mAP, and 4.97% in P. Even compared to the suboptimal YOLOv9, mAP is 0.76% higher, and P is 0.69% higher. From the comparison of the above results, it can be concluded that although the computational complexity has increased, MFCNet has achieved the optimal performance in F1, mAP, and P, demonstrating its better detection accuracy for dense and small ships in SAR images. This is mainly because the FESNet and SFA successfully suppress the interference of complex background on ship features, thereby improving the accuracy of feature extraction. Moreover, to tackle the challenge in detecting small and dense ships in SAR images, the proposed MISPP module expands the receptive field while maintaining high resolution, enhancing the semantic information of small ships and significantly reducing positional ambiguity and target loss. In addition, the FCFNet integrates features from different scales, further improving the performance of the model in complex scenarios. Experimental results on the SSDD and HRSID datasets demonstrate that the proposed method effectively reduces the missed and false detections of small and dense ships.

3.6. Ablation Experiment

In this ablation experiment, we evaluated the three key modules of the model, FESNet, MISPP, and FCFNet, one by one, to analyze the independent contribution of each module in performance improvement. Table 5 shows the performance of the model on different indicators after adding these three modules one by one. It can be seen that whether adding modules separately or combining modules, the performance of the model has been improved in F1, mAP, and P metrics. This indicates that the introduction of FESNet significantly enhances the ability of the model to extract basic features and improves its ability to capture small targets. The MISPP module effectively enhances the multi-scale perception ability of the model, making it more robust in dense target scenes. FCFNet module achieves more accurate localization and recognition of targets in complex backgrounds through feature cross-fusion. In addition, when the three modules are combined and used in the model, they exhibit good complementarity among modules, further improving the overall detection performance. This indicates that these modules not only independently demonstrate advantages in their respective design goals, but also achieve collaborative gains when being used together. This ablation experiment design helped us verify the independent value of each module and demonstrate its organic combination effect in improving model performance.

4. CONCLUSION

There are some deficiencies in accurate localization and target recognition for SAR ship detection, and it is easy to miss, misjudge, and lose information in complex nearshore and small ship concentration scenarios. To overcome these deficiencies, a novel SAR ship detection network MFCNet is proposed in this paper. The FESNet module enhances the feature extraction ability, while the SFA and RDS modules are added to better capture and handle small and dense ships in the complex background. Meanwhile, the MISPP module is introduced to expand the range of the receptive field and enhance the positioning accuracy in the high-level features. At last, the FCFNet module is introduced to further enhance the performance of ship detection by making full use of the information of different feature levels. Various experiments were carried out on the SSDD and HRSID datasets to verify the performance of the model. On the SSDD dataset, P reached 96.52%, and mAP reached 96.89%. On the HRSID dataset, P reached 95.04%, and mAP reached 93.52%. In addition, our proposed MFCNet model can effectively detect multi-scale ships in SAR images, especially small ships and ships near the coastline. However, there is still shortcoming in this paper. The computational complexity of MFCNet is a little high. In future research, we plan to explore lightweight SAR ship detection model that improves model accuracy while enhancing model detection efficiency.

ACKNOWLEDGEMENT

This paper was supported by the National Natural Science Foundation of China under Grant U1904120, Henan Science and Technology Department Science and Technology Research Program under Grant 232102210005, and the Project of Henan Key Laboratory of Superhard Abrasives under Grant JDK-FJJ2023010.

REFERENCES

[1] Zhang, X., S. Feng, C. Zhao, Z. Sun, S. Zhang, and K. Ji, "MGSFA-Net: Multi-scale global scattering feature association network for SAR ship target recognition," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 17, 4611–4625, 2024.

[2] Tang, Y., S. Wang, J. Wei, Y. Zhao, J. Lin, J. Yu, and D. Li, "Scene-aware data augmentation for ship detection in SAR images," *International Journal of Remote Sensing*, Vol. 45, No. 10, 3396–3411, 2024.

[3] Zhang, T. and X. Zhang, "ShipDeNet-20: An only 20 convolution layers and < 1-MB lightweight SAR ship detector," *IEEE Geoscience and Remote Sensing Letters*, Vol. 18, No. 7, 1234–1238, 2021.

[4] Tang, G., H. Zhao, C. Claramunt, W. Zhu, S. Wang, Y. Wang, and Y. Ding, "PPA-Net: Pyramid pooling attention network for multi-scale ship detection in SAR images," *Remote Sensing*, Vol. 15, No. 11, 2855, 2023.

[5] Zhang, C., X. Zhang, J. Zhang, G. Gao, Y. Dai, G. Liu, Y. Jia, X. Wang, Y. Zhang, and M. Bao, "Evaluation and improvement of generalization performance of SAR ship recognition algorithms," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 15, 9311–9326, 2022.

[6] Xu, F. and J.-H. Liu, "Ship detection and extraction using visual saliency and histogram of oriented gradient," *Optoelectronics Letters*, Vol. 12, No. 6, 473–477, 2016.

[7] Lee, S.-H., H.-G. Park, K.-H. Kwon, B.-H. Kim, M. Y. Kim, and S.-H. Jeong, "Accurate ship detection using electro-optical image-based satellite on enhanced feature and land awareness," *Sensors*, Vol. 22, No. 23, 9491, 2022.

[8] Wu, Y., W. Ma, M. Gong, Z. Bai, W. Zhao, Q. Guo, X. Chen, and Q. Miao, "A coarse-to-fine network for ship detection in optical remote sensing images," *Remote Sensing*, Vol. 12, No. 2, 246, 2020.

[9] Yang, Z., J. Tang, H. Zhou, X. Xu, Y. Tian, and B. Wen, "Joint ship detection based on time-frequency domain and CFAR methods with HF radar," *Remote Sensing*, Vol. 13, No. 8, 1548, 2021.

[10] Yang, M., D. Pei, N. Ying, and C. Guo, "An information-geometric optimization method for ship detection in SAR images," *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, 1–5, 2020.

[11] Chen, Z., D. Chen, Y. Zhang, X. Cheng, M. Zhang, and C. Wu, "Deep learning for autonomous ship-oriented small ship detection," *Safety Science*, Vol. 130, 104812, 2020.

[12] Chen, X., H. Wu, B. Han, W. Liu, J. Montewka, and R. W. Liu, "Orientation-aware ship detection via a rotation feature decoupling supported deep learning approach," *Engineering Applications of Artificial Intelligence*, Vol. 125, 106686, 2023.

[13] Feng, J., B. Li, L. Tian, and C. Dong, "Rapid ship detection method on movable platform based on discriminative multi-size gradient features and multi-branch support vector machine," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 23, No. 2, 1357–1367, 2020.

[14] Shan, H., X. Fu, Z. Lv, and Y. Zhang, "SAR ship detection algorithm based on deep dense sim attention mechanism network," *IEEE Sensors Journal*, Vol. 23, No. 14, 16 032–16 041, 2023.

[15] Li, Z., Y. You, and F. Liu, "Analysis on saliency estimation methods in high-resolution optical remote sensing imagery for multi-scale ship detection," *IEEE Access*, Vol. 8, 194 485–194 496, 2020.

[16] Bai, L., C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "Feature enhancement pyramid and shallow feature reconstruction network for SAR ship detection," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, 1042–1056, 2023.

[17] Lu, H., H. Li, L. Chen, Y. Cheng, D. Zhu, Y. Li, R. Lv, G. Chen, X. Su, L. Lang, Q. Li, and Y. Zhao, "A ship detection and tracking algorithm for an airborne passive interferometric microwave sensor (PIMS)," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 14, 3519–3532, 2021.

[18] Niu, Y., Y. Li, J. Huang, and Y. Chen, "Efficient encoder-decoder network with estimated direction for SAR ship detection," *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, 1–5, 2022.

[19] Si, J., B. Song, J. Wu, W. Lin, W. Huang, and S. Chen, "Maritime ship detection method for satellite images based on multiscale feature fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, 6642–6655, 2023.

[20] Zhou, Y., F. Zhang, Q. Yin, F. Ma, and F. Zhang, "Inshore dense ship detection in SAR images based on edge semantic decoupling

and transformer," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 16, 4882–4890, 2023.

[21] Xu, M., Z. Zhang, H. Li, Q. Luo, R. Dou, L. Liu, J. Liu, and N. Wu, "Hierarchical parallel vision processor for high-speed ship detection," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 70, No. 3, 1164–1168, 2022.

[22] Pan, X., Z. Wu, L. Yang, and Z. Huang, "Ship detection method based on scattering contribution for PolSAR image," *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, 1–5, 2021.

[23] Yang, X., X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, 1–12, 2021.

[24] Li, S., X. Fu, and J. Dong, "Improved ship detection algorithm based on YOLOX for SAR outline enhancement image," *Remote Sensing*, Vol. 14, No. 16, 4070, 2022.

[25] Cui, Z., Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 57, No. 11, 8983–8997, 2019.

[26] Shan, H., X. Fu, Z. Lv, and Y. Zhang, "SAR ship detection algorithm based on deep dense sim attention mechanism network," *IEEE Sensors Journal*, Vol. 23, No. 14, 16 032–16 041, 2023.

[27] Ma, X., S. Hou, Y. Wang, J. Wang, and H. Wang, "Multiscale and dense ship detection in SAR images based on key-point estimation and attention mechanism," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 60, 1–11, 2022.

[28] Li, M.-D., X.-C. Cui, and S.-W. Chen, "Adaptive superpixel-level CFAR detector for SAR inshore dense ship detection," *IEEE Geoscience and Remote Sensing Letters*, Vol. 19, 1–5, 2021.

[29] Sun, Y., L. Su, S. Yuan, and H. Meng, "DANet: Dual-branch activation network for small object instance segmentation of ship images," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 33, No. 11, 6708–6720, 2023.

[30] Ren, X., Y. Bai, Z. Zhang, W. Xu, and L. Tan, "SEFRNet: An SAR ship target detection network with effective feature representation," *IEEE Sensors Journal*, Vol. 24, No. 6, 8539–8550, 2024.

[31] Cui, Z., X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 59, No. 1, 379–391, 2020.

[32] Xie, J., Y. Pang, J. Nie, J. Cao, and J. Han, "Latent feature pyramid network for object detection," *IEEE Transactions on Multimedia*, Vol. 25, 2153–2163, 2022.

[33] Huang, J., Z. Chen, Q. M. J. Wu, C. Liu, H. Yuan, and W. He, "CATFPN: Adaptive feature pyramid with scale-wise concatenation and self-attention," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 32, No. 12, 8142–8152, 2021.

[34] Xiao, L., B. Wu, and Y. Hu, "Surface defect detection using image pyramid," *IEEE Sensors Journal*, Vol. 20, No. 13, 7181–7188, 2020.

[35] Gao, L., B. Liu, P. Fu, and M. Xu, "Depth-aware inverted refinement network for RGB-D salient object detection," *Neurocomputing*, Vol. 518, 507–522, 2023.

[36] Chen, G., S.-J. Liu, Y.-J. Sun, G.-P. Ji, Y.-F. Wu, and T. Zhou, "Camouflaged object detection via context-aware cross-level fusion," *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 32, No. 10, 6981–6993, 2022.

[37] Ma, M. and B. Sun, "A cross-level interaction network based on scale-aware augmentation for camouflaged object detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, Vol. 8, No. 1, 69–81, 2024.

[38] Liu, Z.-Y. and J.-W. Liu, "Hypergraph attentional convolutional neural network for salient object detection," *The Visual Computer*, Vol. 39, No. 7, 2881–2907, 2023.

[39] Zhang, T., X. Zhang, J. Li, X. Xu, B. Wang, X. Zhan, Y. Xu, X. Ke, T. Zeng, H. Su, *et al.*, "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sensing*, Vol. 13, No. 18, 3690, 2021.

[40] Wei, S., 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, 120 234–120 254, 2020.

[41] Wang, C.-Y., A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7464–7475, Vancouver, BC, Canada, Jun. 2023.

[42] Zhao, Y., W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "Detrs beat yolos on real-time object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16 965–16 974, Seattle, WA, USA, Jun. 2024.

[43] Wang, C.-Y., I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," in *European Conference on Computer Vision*, 1–21, 2024.