A General Multiscale Pyramid Attention Module for Ship Detection in SAR Images

Peng Wang, *Senior Member, IEEE*, Yongkang Chen, Yi Yang, Ping Chen, Gong Zhang, *Member, IEEE*, Daiyin Zhu, Yongshi Jie, Cheng Jiang, and Henry Leung, *Fellow, IEEE*

***Abstract*—Compared with large-scale ships, small-scale ships occupy few pixels and have low contrast, so it poses a great challenge to detect multiscale ships in synthetic aperture radar (SAR) images. In order to improve the accuracy of multiscale ship detection in SAR images, this article designs a general multiscale pyramid attention module (MPAM), which is a plug-and-play lightweight module that can adapt to many ship detection networks. In the MPAM, a deep feature extraction submodule is first designed to use the multiscale pyramid structure to divide the feature map into different levels, extracting rich features with resolution and semantic information for multiscale ship detection. The channel multilayer attention fusion submodule and spatial multilayer attention fusion submodule are then designed to fuse the channel and spatial attention blocks on different level feature maps, which could better learn the dependent features from the channel and spatial dimensions, to enhance the feature representation. Finally, the fused feature map is input into the existing ship detection networks to obtain the detection result. Experiments on SAR datasets containing multiscale ships show that the effectiveness of the MPAM in improving the accuracy of the existing ship detection networks.**

***Index Terms*—Feature map, multiscale pyramid attention, ship detection, synthetic aperture radar (SAR).**

Manuscript received 19 September 2023; revised 16 November 2023 and 16 December 2023; accepted 20 December 2023. Date of publication 29 December 2023; date of current version 12 January 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities in Nanjing University of Aeronautics and Astronautics under Grant NS2023020, Grant NJ2023029, and Grant QZJC20230206, in part by the Open Project Funds for the Key Laboratory of Space Photoelectric Detection and Perception, Nanjing University of Aeronautics and Astronautics, Ministry of Industry and Information Technology under Grant NJ2023029-1, in part by the Key Laboratory of Radar Imaging and Microwave Photonics, Nanjing University of Aeronautics and Astronautics, Ministry of Education under Grant NJ20230005, in part by the Key Laboratory of Ocean Space Resource Management Technology, Ministry of Natural Resources under Grant KF-2023-108, in part by the Open Fund of Hubei Luojia Laboratory under Grant 230100024, in part by the Beijing Key Laboratory of Advanced Optical Remote Sensing Technology under Grant AORS202311, in part by the State Key Laboratory of Geoinformation Engineering and Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology of Ministry of Natural Resources, Chinese Academy of Surveying and Mapping under Grant 2023-03-09, in part by the Shanxi Key Laboratory of Signal Capturing and Processing under Grant 2023-002, in part by the Natural Science Foundation of Jiangsu Province under Grant BK20221478, in part by Hong Kong Scholars Program under Grant XJ2022043, in part by Youth Promotion Talent Project of Jiangsu Association for Science and Technology under Grant TJ-2023-010, in part by the National Natural Science Foundation of China under Grant 61801211, and in part by the Postgraduate Research and Practice Innovation Program of Nanjing University of Aeronautics and Astronautics under Grant xcxjh20230405. *(Corresponding authors: Peng Wang; Daiyin Zhu.)*

Peng Wang is with the Key Laboratory of Space Photoelectric Detection and Perception (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology, Nanjing 210016, China, and with the Key Laboratory of Ocean Space Resource Management Technology, Ministry of Natural Resources, Hangzhou 310012, China, and with the Hubei Luojia Laboratory, Wuhan 430079, China, and also with the Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology of Ministry of Natural Resources, Chinese Academy of Surveying and Mapping, Beijing 100039, China (e-mail: pengwang-b614080003@hotmail.com).

Yongkang Chen, Gong Zhang, and Daiyin Zhu are with the Key Laboratory of Radar Imaging and Microwave Photonics, Ministry of Education, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: 19855530831@163.com; gzhang@nuaa.edu.cn; zhudy@nuaa.edu.cn).

Yi Yang is with the Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology, Ministry of Natural Resources, Chinese Academy of Surveying and Mapping, Beijing 100039, China (e-mail: yangyi@casm.ac.cn).

Ping Chen is with the Shanxi Key Laboratory of Signal Capturing and Processing, North University of China, Taiyuan 030051, China (e-mail: chenping@nuc.edu.cn).

Yongshi Jie and Cheng Jiang are with the Beijing Key Laboratory of Advanced Optical Remote Sensing Technology, Beijing Institute of Space Mechanics and Electricity, Beijing 100094, China (e-mail: jie_yongshi@163.com; cheng3515523@163.com).

Henry Leung is with the Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (e-mail: leungh@ucalgary.ca).

Digital Object Identifier 10.1109/JSTARS.2023.3348269

I. Introduction

SYNTHETIC aperture radar (SAR) is an active microwave sensor that uses pulse compression to improve range resolution and synthetic aperture to improve azimuth resolution, achieving high-resolution imaging. Due to the variability of ocean climate, ordinary optical imaging is disturbed by clouds, rain, and fog, and the imaging effect is not good. SAR is an advanced active microwave Earth observation device that can penetrate clouds, rain, and fog and work throughout the day. SAR has been widely used in military and civilian fields, such as marine monitoring, and Earth observation [1]. In addition, SAR can effectively improve ship transport efficiency and reduce maritime traffic accidents. Due to the uncertainty of sea clutter, the different size of ship targets, and the interference of land clutter, the detection performance of ship target can be deteriorated [2].

Among the traditional SAR ship detection methods, the constant false alarm rate detection (CFAR) is the most widely studied methods [3]. When the radar is affected by noise, clutter, and interference in the detection, the fixed threshold for target detection will produce a certain false alarm. Especially, when the clutter background fluctuation changes, the false alarm rate rises sharply, which can seriously affect the radar detection performance [4]. To solve this issue, CFAR can dynamically adjust the detection threshold according to radar clutter to maximize the target detection probability, while the false alarm

probability remains unchanged. However, the detection methods performance of the CFAR are not satisfactory against a complex background [5], [6].

In recent years, there have been many researches on deep learning in various fields, which allows computational models composed of multiple processing layers to learn representations of data with multiple levels of abstraction [7]. Due to its powerful feature extraction and representation ability, convolutional neural network (CNN) [8] has made brilliant achievements in the field of computer vision, such as image classification [9], [10], [11], image segmentation [12], [13], [14], and object detection [15], [16], [17]. It has become one of the most representative neural networks in the field of deep learning. The object detection methods based on deep learning mainly include one-stage object detection methods and two-stage object detection methods. One-stage object detection methods include single shot detector (SSD) [18], you only look once (YOLO) [19], [20], [21], and RetinaNet [22]. Two-stage object detection methods mainly include region-CNN (R-CNN) [23], fast R-CNN [24], and faster R-CNN [25]. Among them, the two-stage detection method has higher accuracy than the one-stage detector, but the detection speed is slower. However, these methods mainly focus on shallow feature maps and do not consider the deeper ones. To solve this problem, Lin et al. [26] propose a feature pyramid network (FPN), which detects feature maps at different levels. Under the condition of increasing a small amount of calculation, the FPN fuses feature maps with strong low-resolution semantic information and feature maps with weak high-resolution semantic information but rich spatial information, improving the detection accuracy. Dai et al. [27] design a method to fuse bottom-to-top and top-to-bottom feature maps to enhance the semantic features of multiscale ships and improve the detection accuracy of multiscale ships.

The framework proposed by Kang et al. [28] fuses deep semantic and shallow high-resolution features to improve the detection performance of small ships, and the additional contextual features provide supplementary information for classification. Jiao et al. [29] propose a densely connected multiscale neural network based on the faster R-CNN framework to solve the multiscale and multiscene SAR ship detection problem. Zhao et al. [30] propose an end-to-end lightweight network called morphological feature-pyramid YOLOv4-tiny for SAR ship detection. The morphological network is introduced to preprocess the SAR image for speckle noise suppression and edge enhancement, which provides spatial high-frequency information for target detection. The original image and the preprocessed image are combined into multiple channels as the input of the network convolutional layer, and the feature pyramid fusion structure is used to extract high-level semantic features and shallow detail features from the image, Cui et al. [31] propose a multiscale SAR image-based ship detection method based on the dense attention pyramid network (DAPN). The DAPN adopts a pyramid structure, and the convolutional block attention module (CBAM) [32] is inserted into each connection layer of the pyramid network to extract rich features containing resolution and semantic information for multiscale ship detection. At the same time, the connected feature map is refined to highlight the salient features of specific scales. Zhao et al. [33] propose an attention receptive pyramid network (ARPN). The ARPN that is a two-stage detector is designed to improve the performance of multiscale ship detection in SAR images by enhancing the relationship between nonlocal features and refining the semantic information on different feature maps.

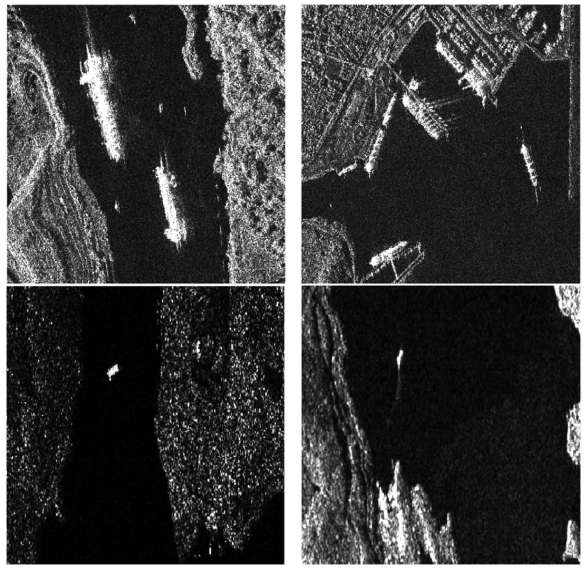


Fig. 1. Multiscale ships in complex background.

Although these methods improve the detection accuracy of ships in SAR images to a certain extent, there are still some problems in multiscale ship detection due to the multiscale and multiangle characteristics of ships, and the scattering characteristics of ships in SAR images will lead to unclear contours of ships. As shown in Fig. 1, it can be seen that large ships and complex ground background are prominent, small ships are not obvious, and common SAR target detection methods tend to focus only on local features of ships, ignoring the relationship between image contexts and extracting less semantic information. Therefore, enriching the extracted semantic information by combining multiscale deep and shallow features is crucial to improve the detection performance of multiscale ships in SAR images.

A general multiscale pyramid attention module (MPAM) for ship detection networks is proposed in this article. At present, there are many methods for ship detection in SAR images, and there are many general attention modules, but there are few attention modules for ships in SAR images. Since the attention module can highlight the features of the target and enhance its feature representation, and the combination of multiscale features can expand the receptive field of the multiscale ship and enhance feature extraction, and can improve the calculation accuracy and model performance with a small number of additional parameters. Therefore, we design a general module MPAM for SAR images. Considering the multiangle and multiscale features of ship targets in SAR images, the low-level feature maps with shallow resolution are suitable for detecting small ships, and the high-level feature maps with more semantic information

are suitable for detecting large ships. Therefore, the deep feature extraction submodule (DFES), channel multilayer attention fusion submodule (CMAFS), and spatial multilayer attention fusion submodule (SMAFS) are designed. In the MPAM, first, the multiscale pyramid structure is used to divide the feature map into different levels through the DFES, and rich features with resolution and semantic information are extracted for multiscale ship detection. Then, CMAFS and SMAFS were used to fuse channel attention blocks and spatial attention blocks on feature maps at different levels, and relevant features were learned from channel and spatial dimensions to enhance feature representation. Finally, the fused feature maps were input into the existing ship detection network to obtain the detection results. DFES, CMAFS, and SMAFS are the key components of the MPAM, and the contribution of this article is reflected as follows.

1) A novel attention module MPAM is proposed for detect multiscale ships in SAR images, and it is a plug-and-play lightweight module that can adapt to many ship detection networks. Whether it is placed in one-stage detection or two-stage detection, the detection performance can be well improved.
2) We designed three submodules in the MPAM. The DFES are used to extract deep features and multiple dimensions to expand the receptive field and improve the detection accuracy of small ships CMAFS and SMAFS are combined to extract and fuse the spatial and channel attention blocks in the multilevel features, which pay more attention to ships of different scales and filter useless information, i.e., land and sea, improving the detection accuracy of multiscale ships.
3) The existing multiscale attention uses multiscale and attention modules separately, while the attention module we designed contains multiscale and dual attention submodules at the same time, and highlights salient features through the combination of multiscale and dual attention submodules. We conduct experiments on three datasets and compare with different attention modules, as well as perform ablation experiments. Experiments show that the proposed module is more suitable for ship detection in SAR images than the existing attention module, and by comparing the detection accuracy with and without MPAM, the detection performance can be improved by using the MPAM, but the impact on computational efficiency is small.

The rest of this article is organized as follows. Section II presents the proposed MPAM. Section III reports the experimental results on the three datasets and the analysis of the results. Finally, Section IV concludes this article.

II. Proposed Method

This section presents the proposed MPAM. Based on the analysis of the existing methods, the main idea of the method is proposed, and its overall framework is given. In the following, we give the detailed description of each submodule of the proposed MPAM to illustrate how it works.

A. Ideas of the MPAM

Because high-level feature maps contain rich semantic information, but lack accurate location information [34]. Low-level feature maps contain rich location information, but lack good semantic information. Lin et al. [26] propose the FPN to solve the multiscale problem in object detection. The FPN can train and test the model by constructing a series of images or feature maps of different scales, improving the robustness of detection. Here, we design a DFES to extract the basic features of the backbone network, and then, extract the deep features hierarchically.

Since the extracted features may be redundant, the important features will be affected by useless information. The CBAM is used to highlight important information at the same scale [32]. It consists of channel and spatial attention mechanism to intelligently refine features, but ignores multiscale feature maps, this approach draws on the idea of CBAM and designs CMAFS and SMAFS; it applies CMAFS and SMAFS to the fusion of multiscale feature maps and pays more attention to the scale features of the target. We leverage multilevel and multiscale channel and spatial attention to impose weights on scale and spatial dimensions, respectively. Channel attention is focused on selecting feature maps of appropriate scale, and spatial attention is focused on salient regions containing objects. Therefore, multiscale ships can be detected more effectively by using CMAFS and SMAFS.

B. Framework of the MPAM

Fig. 2 shows the structure of the MPAM. It is mainly composed of DFES, SMAFS, and CMAFS. After obtaining the basic feature map through the backbone network, the DFES is used to extract the multiscale deep feature map. Then, CMAFS and SMAFS are used to fuse the channels and spatial attention blocks on different level feature maps and multiply them with the basic feature map to obtain the attention weighted feature map.

In the framework of the MPAM in the detection network, the SAR image is fed through the backbone network to obtain a basic feature$\mathbf{F} \in \mathbb{R}^{C\times H\times W}$. In the channel multilayer attention branch, the DFES is used for deep feature extraction and the CMAFS is used for multilayer attention fusion to obtain a channel attention block $\mathbf{F_C} \in \mathbb{R}^{C\times 1\times 1}$. In the spatial multilayer attention branch, the spatial attention block $\mathbf{F_S} \in \mathbb{R}^{1\times H\times W}$ is obtained by DFES and SMAFS, respectively. The refined feature $\mathbf{F_R} \in \mathbb{R}^{C\times H\times W}$ is obtained by multiplying the basic feature $\mathbf{F} \in \mathbb{R}^{C\times H\times W}$ with the channel attention block $\mathbf{F_C} \in \mathbb{R}^{C\times 1\times 1}$ and the spatial attention block $\mathbf{F_S} \in \mathbb{R}^{1\times H\times W}$. $\mathbf{F_R} \in \mathbb{R}^{C\times H\times W}$ is then sent to the detection network to obtain the ship detection results. Here, C, H, and W represent the number of channels, height, and width of the feature map, respectively.

C. Detailed Architecture of the MPAM

As shown in Fig. 2, the MPAM contains the channel multilayer attention branch and the spatial multilayer attention branch. They are used to generate channel attention blocks and spatial attention blocks, respectively, and each branch contains the same DFES for extracting deep features. The difference between

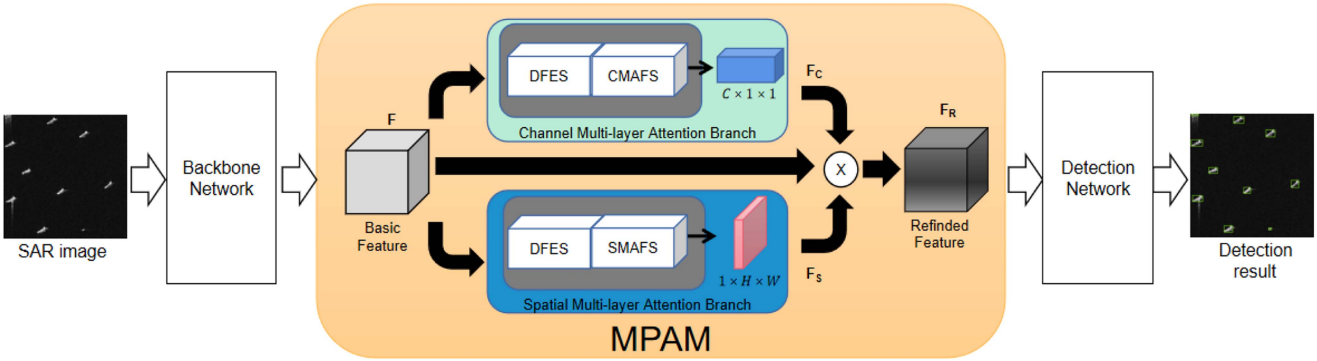


Fig. 2. Framework of the MPAM in the whole detection network.

Fig. 3. Flowchart of the DFES.

two branches is that the attention fusion uses the different sub-modules. Namely, the CMAFS is used in the channel multilayer attention branch, and the SMAFS is used in the spatial multilayer attention branch.

1) Deep Feature Extraction Submodule (DFES): We design the DFES by extracting a large number of features in SAR images, the ship targets of different scales are highlighted and the complex background is suppressed, and the multiscale detection of ship targets is realized. Fig. 3 is the structural diagram of the DFES, where $\mathbf{C}\{\mathbf{x}\}$ represents the basic feature maps from the backbone network and $\mathbf{P}\{\mathbf{x}\}$ is the extracted deep features. To obtain $\mathbf{P}\{\mathbf{x}\}$, we perform a convolution with a kernel size of 3×3 to extract deep features. RELU and batch normalization operations are then performed to normalize the data and optimize the parameters. Max pooling operation is performed for downsampling, which can reduce parameters with retaining features to prevent overfitting, and improve the generalization ability of the model to generate a deep feature map, $\mathbf{P}\{\mathbf{x}\}$with half the size of the shallow feature map. The overall process of DFES is as follows:

$$P\{\mathrm{x}\} = \mathrm{Conv}_{7x7}[\mathrm{ReLU}(\mathrm{BN}(\mathrm{Maxpool}(C\{x\})))] \quad (1)$$

where $\mathbf{P}\{\mathbf{x}\} \in \mathbb{R}^{C\times H\times W}$ and $\mathbf{C}\{\mathbf{x}\} \in \mathbb{R}^{C\times \frac{H}{2}\times \frac{W}{2}}$. Besides, MaxPool and *BN* represent the operations of max pooling and batch normalization, respectively, ReLU is the ReLU function, and $\mathrm{Conv}_{7\times 7}$ is represents the 7×7 convolutional layer.

2) Channel Multilayer Attention Fusion Submodule (CMAFS): The structure of the CMAFS is shown in Fig. 4. Since each channel of the feature map extracts a certain level of feature information, and each feature map has multiple channels, some channel branches contain important feature information, and some contain a small amount of feature information. Therefore, we design the CMAFS, which focuses its attention on which level of feature information is more important in the multilayer branches of the channel. In order to effectively calculate the channel attention, the CMAFS uses

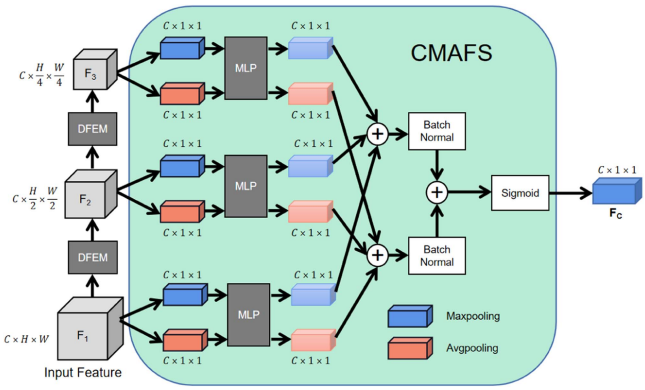


Fig. 4. Overview of the CMAFS.

the spatial dimension method of compressing the input feature map. Compared with other ship detection methods based on single pooling, the CMAFS uses average pooling and max pooling methods to have stronger representation power. This is because the maximum pooling is to take the maximum value of the pixels in the pooling area, and the feature map obtained in this way is more sensitive to the texture feature information, while the average pooling is to take the average value of the image in the pooling area, and the feature information obtained in this way is more sensitive to the background information. Therefore, we choose two pooling methods to extract features and combine them to enhance the features of the image.

We extract the feature maps through the DFES for deep feature extraction and obtain three feature maps $\mathbf{F_1} \in \mathbb{R}^{C\times H\times W}$, $\mathbf{F_2} \in \mathbb{R}^{C\times \frac{H}{2}\times \frac{W}{2}}$,and $\mathbf{F_3} \in \mathbb{R}^{C\times \frac{H}{4}\times \frac{W}{4}}$. These feature maps are then fed into the CMAFS. The features of each layer are compressed in dimension by average pooling and max pooling to obtain two feature maps of size $C \times 1 \times 1$, and the results of average pooling and max pooling are processed by multilayer perceptron (MLP).

Fig. 5 shows the overall structure of MLP. Fig. 5 shows the overall structure of MLP. In the CMAFS, MLP can be used to enable the convolution of feature regions to be sampled to enhance the learning of feature regions, and also to output weights along the channel dimension of the feature map [35]. The number of channels of the feature map is changed to

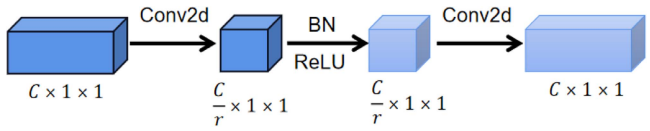


Fig. 5. Overview of MLP.

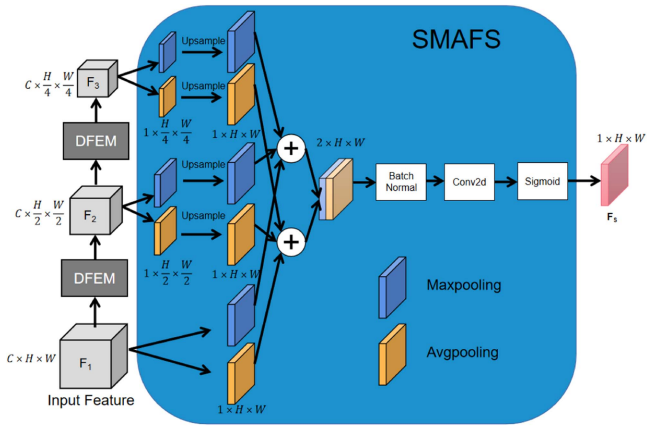


Fig. 6. Overview of SMAFS.

the original *1/r* through convolution operation. The data are then normalized and activated through batch normalization and ReLU, and the original number of channels is changed through the convolution layer. Here, *r* is the reduction ratio, and the overhead of parameters can be reduced by MLP. The results obtained by MLP of average pooling and maximum pooling of each layer are added, respectively, and the data are normalized by batch normalization, it accelerates the convergence speed of the network and prevents gradient explosion, gradient disappearance, and overfitting. The sigmoid activation function is used to obtain the channel attention map $\mathbf{F_C} \in \mathbb{R}^{C\times1\times1}$, that is, the weight (between 0 and 1) of each channel in the input feature layer. The formulas of the CMAFS are expressed as follows:

$$\mathrm{C}_1 = \mathrm{BN}[\sum_{i=1}^{3}\mathrm{MLP}(\mathrm{Maxpool}(\mathrm{F}_i))] \tag{2}$$

$$\mathrm{C}_2 = \mathrm{BN}[\sum_{i=1}^{3}\mathrm{MLP}(\mathrm{Avgpool}(\mathrm{F}_i))] \tag{3}$$

$$\mathrm{F_C} = \mathrm{Sigmoid}\left(\mathrm{C}_1 \oplus \mathrm{C}_2\right) \tag{4}$$

where $i = \{1,2,3\}$, and Maxpool and Avgpool represent max pooling and average pooling operations, respectively. In addition, BN and Sigmoid represent batch normalization operation and activation function, respectively.

3) Spatial Multilayer Attention Fusion Submodule (SMAFS): Its structure is shown in Fig. 6. Since the feature information of ships is very important in ship detection, an image contains a large amount of feature information, but the scattering characteristics of SAR images may cause the location information of ships to be ambiguous. Therefore, we design SMAFS, which can focus on the more important part of the information in the input image, enhance the location information of the ship, and weaken the unimportant feature information, which is a supplement to the multilayer attention branch of the channel.

Like the CMAFS, three different levels of feature maps $\mathbf{F_1} \in \mathbb{R}^{C\times H\times W}$, $\mathbf{F_2} \in \mathbb{R}^{C\times \frac{H}{2}\times \frac{W}{2}}$, and $\mathbf{F_3} \in \mathbb{R}^{C\times \frac{H}{4}\times \frac{W}{4}}$ are obtained through the DFES. The three different levels of feature maps are fed into the SMAFS for the generation of spatial attention blocks. In the SMAFS, two feature maps of size $1 \times H \times W$ are generated by applying max pooling and average pooling (both pooling are applied to the channels of the input feature layer) along the channel directions of the feature maps of each level, the feature maps of each layer through max pooling and average pooling are added and stacked into a feature descriptor of size $2 \times H \times W$. This descriptor is normalized through the BN layer, and then, a standard convolution layer (convolution with channel number 1) is used for connection and convolution (the number of channels is adjusted), the sigmoid activation function is then used to obtain the 2 + D spatial attention map $\mathbf{F_S} \in \mathbb{R}^{1\times H\times W}$, that is, the weight value of each feature point in the input feature map is obtained (between 0 and 1). The formulas of the SMAFS are expressed as follows:

$$S_1 = \sum_{i=1}^{3}\mathrm{Up}_{2^{i-1}}\left(\mathrm{MaxPool}\left(\mathrm{F}_i\right)\right) \tag{5}$$

$$S_2 = \sum_{i=1}^{3}\mathrm{Up}_{2^{i-1}}\left(\mathrm{AvgPool}\left(\mathrm{F}_i\right)\right) \tag{6}$$

$$\mathrm{F_S} = \mathrm{Sigmoid}\left[\mathrm{Con}_{7\times7}\left(\mathrm{BN}\left(S_1 \oplus S_2\right)\right)\right] \tag{7}$$

where $i = \{1,2,3\}$, Maxpool and AvgPool represent max pooling and average pooling operations, respectively, and Sigmoid represents the activation function. In addition, $\mathrm{Up}_{2^{i-1}}$ represents upsampling the feature map by a factor of 2^{i-1}, BN and $\mathrm{Conv}_{7\times7}$ are batch normalization and the 7×7 convolutional layer.

III. Experiments

This section aims to evaluate the performance of different ship detection networks by using the proposed MPAM. Three SAR ship target detection datasets are used here, namely, SAR ship detection dataset (SSDD), high-resolution SAR images dataset (HRSID), and SAR-Ship-dataset. The setup of related experiments and the evaluation criteria are also presented. The detection performance of these methods is compared and discussed.

A. Datasets and Training Settings

In order to evaluate the performance of the MPAM, we conduct experiments using three datasets: SSDD, HRSID, and SAR-Ship-dataset. The SSDD dataset is constructed by Zhang et al. [36] and contains multiscale ships in different environments, including different scenes, sensor types, polarization modes, and image resolutions. The dataset consists of 1160 SAR images of different sizes with 1–15 m resolution. It contains a total of 2456 multiscale ships, and the average number of ships per image is 2.12. And there are four polarization modes,

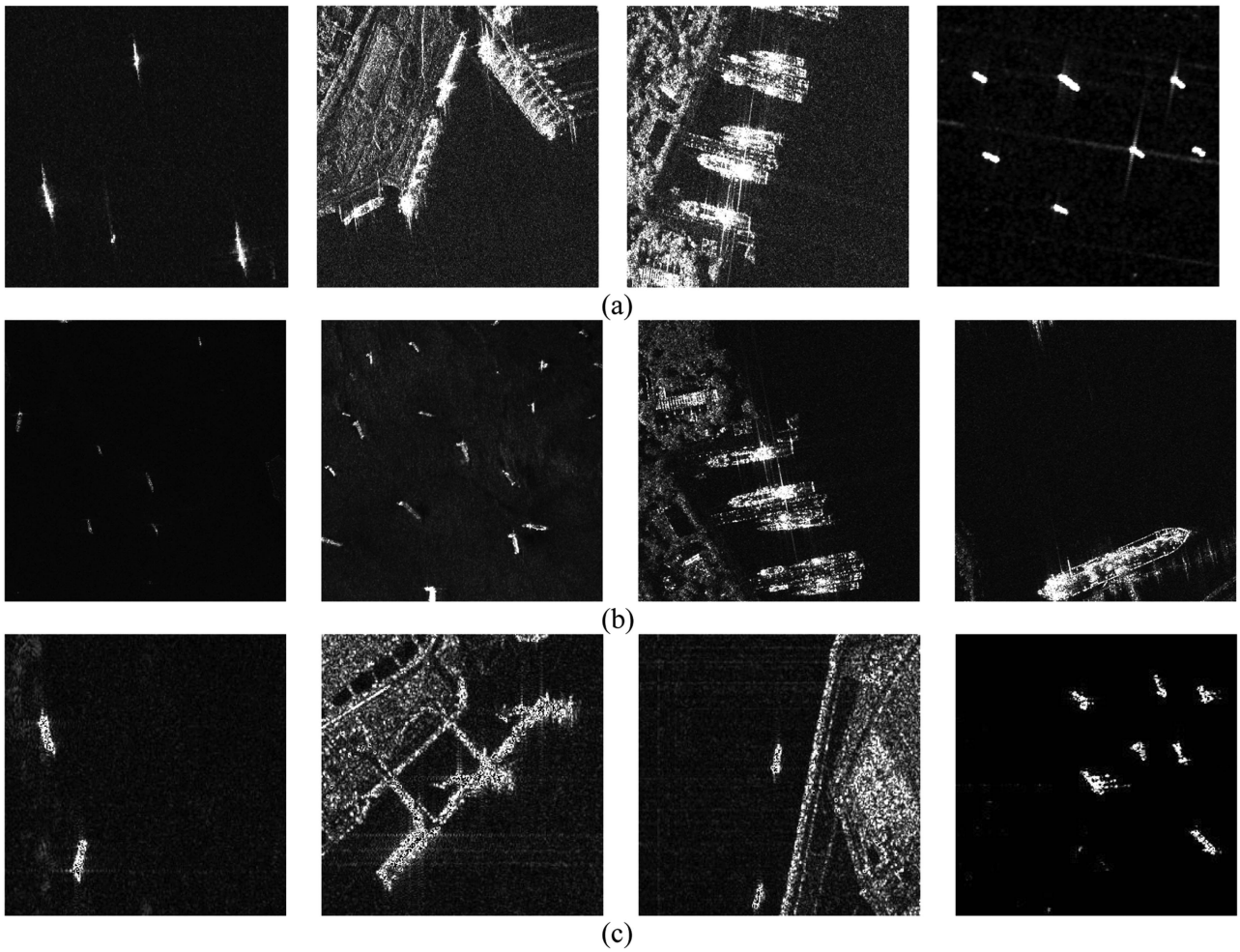


Fig. 7. Some sample images in three datasets. (a) SSDD. (b) HRSID. (c) SAR-Ship-datasets.

including HH, HV, VV, and VH. This dataset uniquely identifies the images with the last digit of the image named as 1 and 9 as the test set, and the rest as the training set. This rule can maintain the distribution consistency of the training set and the test set, which is beneficial to the learning of network features.

The HRSID dataset is constructed by Wei et al. [37] and contains 5604 SAR images with resolutions of 0.5, 1, and 3 m, respectively, all of which are 800 × 800 in size, with a total of 16 951 ships. These images have different environments, polarization rates, and imaging methods. The format of Microsoft Common Objects in Context (MS COCO) is used to label each ship, including area and box (x, y, width, and height). x and y denote the coordinates of the top-left corner of the annotated bounding box, and the width and height denote the width and height of the bounding box. In our experiments, 65% of the dataset is divided into training data and the remaining 35% is divided into testing data.

SAR-Ship-dataset is constructed by Wang et al. [38] in 2019 and extracted from Gaofen 3 and Sentry 1 in the horizontal bounding boxes annotation format. It consists of 39 729 ship chips (remove some repeat clips) of 256 pixels in both range and azimuth. These ships mainly have distinct scales and backgrounds. It can be used to develop object detectors for multiscale and small object detection. We extracted 3198 images, each of size 256 × 256 and divided into training and testing sets by 7:3. Some examples from three datasets are shown in Fig. 7.

The pretrained ResNet101 and YOLOv3 are used as the backbone network of the model, and the stochastic gradient descent method is used to train the network [39]. The weight decay and momentum are 0.0005 and 0.9, respectively. In addition, the learning rate is 0.005. In particular, all the experiments are implemented under the Pytorch framework and performed on a computer using Intel i7-6700HQ.

B. Evaluation Criteria

In this article, we use the COCO dataset evaluation metric and the intersection-over-union (IoU) metric [30]. The concept of IoU is relatively simple, which is to measure the coincidence degree of the detection box and the ground-truth box. The accuracy of the model can be judged by the intersection ratio of the two criteria, the average precision can be divided into AP, AP_{50}, AP_{75}, AP_s, AP_m, and AP_l. AP represents the average

precision when IoU = 0.50:0.05:0.95, that is, IoU starts from 0.5 and increases by 0.05 each time until 0.95. There are ten IoU values in total. AP50 represents the average accuracy at IoU = 0.5. AP_{75} represents the average accuracy at IoU = 0.75. AP_s, AP_m, and AP_l represent the average accuracy for small, medium, and large targets. Generally, an area less than 32×32 belongs to a small object, an area less than 96×96 and greater than 32×32 is a medium object, and an area greater than 96×96 is a large object. When detecting a ship, if the IoU is greater than or equal to 0.5, we consider that the ship detection is correct. AP and IoU are calculated as follows:

$$\text{IoU} = \frac{A_{\text{pred}} \cap A_{\text{true}}}{A_{\text{pred}} \cup A_{\text{true}}} \tag{8}$$

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

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

$$\text{AP} = \int_0^1 P(R)\,\mathrm{d}R \tag{11}$$

where A_{pred} and A_{true} represent the predicted and ground- truth boxes. TP is the sum of correctly detected ships, FP is the sum of incorrectly detected ships, and FN is the sum of missed ships. *P* refers to the proportion of ground-truth ships predicted by the network. *R* refers to the proportion of ground-truth ships predicted by the network among all ground-truth ships.

In addition, frames per second (FPS) shows how fast the detection method is being tested. The larger the FPS, the faster the method will run. The FPS is defined as follows:

$$\text{FPS} = \frac{1}{T} \tag{12}$$

where *T* represents the average inference time to process the image.

C. Results and Analysis

Because the proposed MPAM is a plug-and-play module, we add the MPAM to some popular SAR ship detection methods based on deep learning, such as Faster-R-CNN [25], RetinaNet [22], SSD300 [18], YOLOv3 [21], and improved Faster-R-CNN [39]. These methods are coined as Faster-R-CNN+MPAM, RetinaNet+MPAM, SSD300+MPAM, YOLOv3+MPAM, and improved Faster-R-CNN+MPAM, respectively. We conduct experiments on three datasets: SSDD, HRSID, and SAR-Ship-dataset, and analyze the performance of the MPAM by comparing the accuracy improvement before and after.

1) SAR Ship Detection Dataset (SSDD): Performance evaluation in the SSDD dataset is shown in Table I. It can be seen that when the MPAM is added to the five methods, their detection results are better than those of their original methods. For example, the AP_{50} of Faster-R-CNN+MPAM is 96%, which is 2.2% higher than that of the original Faster-R-CNN. In addition, AP, AP_m, and AP_l are increased by 0.5%, 2%, and 0.7%, respectively. Although AP_{75} and AP_s in Faster-R-CNN+MPAM has some degradation compared to the original Faster-R-CNN, the lower values are few, and overall, the performance of the Faster-R-CNN+MPAM is still better than the original Faster-R-CNN. For the RetinaNet and YOLOv3 methods, using the MPAM improves the accuracy of detection in an all-round way, making each index have different degrees of improvements. According to the most representative index AP, the addition of the MPAM improves the performance of ship detection and improves the accuracy of target localization.

TABLE I
OBJECTIVE EVALUATION OF PERFORMANCE IMPROVEMENT OF DIFFERENT DETECTION METHODS IN SSDD

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Faster-R-CNN	0.654	0.938	**0.788**	**0.638**	0.688	0.659
Faster-R-CNN+MPAM	**0.659**	**0.960**	0.784	0.634	**0.708**	**0.666**
RetinaNet	0.627	0.906	0.738	0.607	0.673	0.328
RetinaNet+MPAM	**0.643**	**0.924**	**0.759**	**0.610**	**0.698**	**0.672**
SSD300	0.541	**0.891**	0.613	0.506	**0.617**	0.302
SSD300+MPAM	**0.553**	**0.891**	**0.657**	**0.526**	0.612	**0.594**
YOLOv3	0.524	0.868	0.579	0.545	0.521	0.218
YOLOv3+MPAM	**0.531**	**0.871**	**0.603**	**0.547**	**0.525**	**0.295**
Improved Faster-R-CNN	0.527	0.864	0.591	**0.526**	0.550	0.428
Improved Faster-R-CNN +MPAM	**0.531**	**0.869**	**0.605**	0.524	**0.553**	**0.498**

The bold entities means best values.

TABLE II
OBJECTIVE EVALUATION OF PERFORMANCE IMPROVEMENT OF DIFFERENT DETECTION METHODS IN HRSID

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Faster-R-CNN	0.652	0.896	0.749	0.511	0.792	**0.851**
Faster-R-CNN+MPAM	**0.655**	**0.905**	**0.751**	**0.514**	**0.796**	**0.851**
RetinaNet	0.615	0.821	0.698	0.385	0.814	0.800
RetinaNet+MPAM	**0.632**	**0.839**	**0.718**	**0.412**	**0.826**	**0.900**
SSD300	0.549	0.789	0.639	0.359	**0.748**	0.484
SSD300+MPAM	**0.555**	**0.793**	**0.652**	**0.366**	0.747	**0.539**
YOLOv3	0.562	0.825	0.644	0.412	**0.728**	0.060
YOLOv3+MPAM	**0.569**	**0.832**	**0.651**	**0.415**	**0.728**	**0.244**
Improved Faster-R-CNN	0.547	0.779	0.630	**0.369**	0.731	0.213
Improved Faster-R-CNN +MPAM	**0.549**	**0.782**	**0.634**	**0.369**	**0.738**	**0.259**

The bold entities means best values.

In order to illustrate the effectiveness of the MPAM, we present the detection effect of an image in each method, as shown in Fig. 8. The most obvious one is that SSD300 only detects four ships in Fig. 8(e), while SSD300+MPAM can detect two more ships in Fig. 8(f). Moreover, Fig. 8(h) also detects three more ships than Fig. 8(g). This shows that the MPAM can capture the characteristics of the ship and accurately locate the smaller target.

2) High-Resolution SAR Images Dataset (HRSID): The comparative results of the detection performance evaluation in the HRSID dataset are shown in Table II. The detection results by adding the MPAM have better accuracy than the original methods. The AP50 of Faster-R-CNN+MPAM reaches 90.5%, which is 0.9% higher than that of the original Faster-R-CNN. In addition, with the help of the MPAM, AP, AP_{75}, AP_s, and AP_m in Faster-R-CNN are increased by 0.3%, 0.2%, 0.3%, and 0.4%, respectively. For the RetinaNet method, the addition of the MPAM

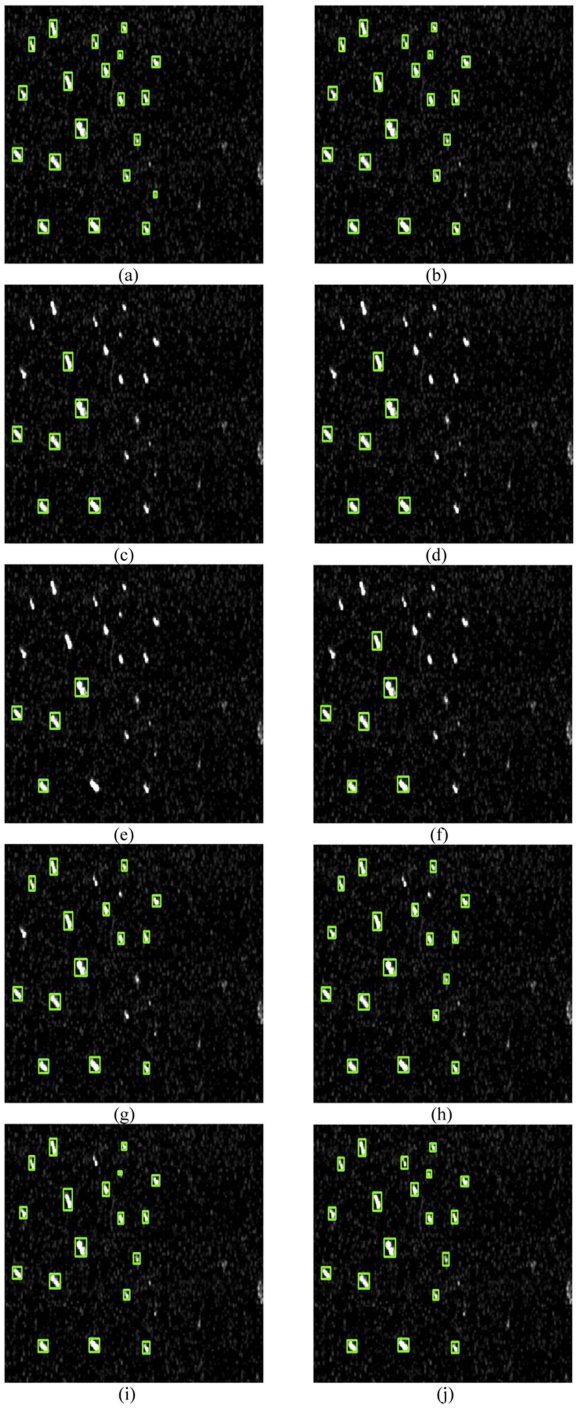


Fig. 8. Comparison of the results of different detection methods and the results after adding the MPAM in SSDD. (a) Faster-R-CNN. (b) Faster-R-CNN+MPAM. (c) RetinaNet. (d) RetinaNet +MPAM. (e) SSD300. (f) SSD300+MPAM. (g) YOLOv3. (h) YOLOv3+MPAM. (i) Improved faster-R-CNN. (j) Improved faster-R-CNN+MPAM.

improves the accuracy of detection, making each index have different levels of improvements, especially the improvement of the detection accuracy of small and large targets, which is increased by 2.7% and 10%, respectively. According to the most representative index AP among the detection results of the five

TABLE III
OBJECTIVE EVALUATION OF PERFORMANCE IMPROVEMENT OF DIFFERENT DETECTION METHODS IN SAR-SHIP-DATASET

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Faster-R-CNN	0.635	0.947	0.779	**0.624**	0.656	0.577
Faster-R-CNN+MPAM	**0.643**	**0.950**	**0.786**	0.618	**0.678**	**0.594**
RetinaNet	0.675	0.961	**0.830**	0.665	0.698	0.543
RetinaNet+MPAM	**0.679**	**0.963**	0.826	**0.667**	**0.703**	**0.562**
SSD300	0.626	0.938	0.769	0.606	**0.660**	0.528
SSD300+MPAM	**0.632**	**0.942**	**0.784**	**0.617**	0.656	**0.556**
YOLOv3	0.392	0.743	0.371	0.437	0.361	0.025
YOLOv3+MPAM	**0.407**	**0.785**	**0.387**	**0.434**	**0.376**	**0.150**
Improved Faster-R-CNN	0.598	**0.938**	0.707	0.583	0.627	0.375
Improved Faster-R-CNN +MPAM	**0.605**	0.937	**0.721**	**0.584**	**0.643**	**0.383**

The bold entities means best values.

methods, the addition of the MPAM improves the performance of ship detection and the accuracy of ship localization.

Fig. 9 shows the ship detection of an image for five methods in HRSID. By adding the proposed MPAM, the visual effects of the five methods are improved. For example, as shown in Fig. 9(g) and (h), the original YOLOv3 detects ten ships in Fig. 9(g), while the YOLOv3+MPAM could detect one more ship in Fig. 9(h). In addition, the ship in the lower right corner of Fig. 9(i) has multiple duplicate detection boxes, and the detection results of Fig. 8(j) remove redundant detection boxes to make the results clearer. This shows that the MPAM can exploit the characteristics of the ship and accurately locate the smaller targets.

3) SAR-Ship-Dataset: Table III shows the detection performance of the SAR-Ship-dataset. Similar to the results of the first two sets of experimental data, when the MPAM is added to the five methods, their ship detection results are improved. For example, AP_{50} in RetinaNet+MPAM reaches 96.3%, which is 0.2% higher than the original RetinaNet, and AP, AP_s, AP_m, and AP_l are increased by 0.4%, 0.2%, 0.5%, and 1.9%, respectively. In addition, in the YOLOv3 method, the MPAM improves the performance of AP, AP_{50}, AP_{75}, AP_m, and AP_l by 1.5%, 4.2%, 1.6%, 1.5%, and 12.5%, respectively. According to the most representative index AP, the addition of the MPAM improves the performance of ship detection in SAR-Ship-dataset, and the detection accuracy of different methods has different degrees of improvement.

As shown in Fig. 10, we give the detection effect of an image for five methods in SAR-Ship-dataset. Similar visual results to previous datasets, the MPAM locates smaller objects and adjusts the size of the detection box more accurately. For example, it can be seen from Fig. 10(g) and (h) that one detection box of YOLOv3 in Fig. 10(g) is redundant and large. However, in Fig. 10(h), after adding the MPAM, the detection box recovers its original size and detects the ship accurately.

In general, the performance of adding MPAM on different datasets may vary due to the differences in the number of ships in SAR images of different datasets, image resolution or ocean conditions, leading to the performance degradation or difference of some methods. But in general, it does improve the

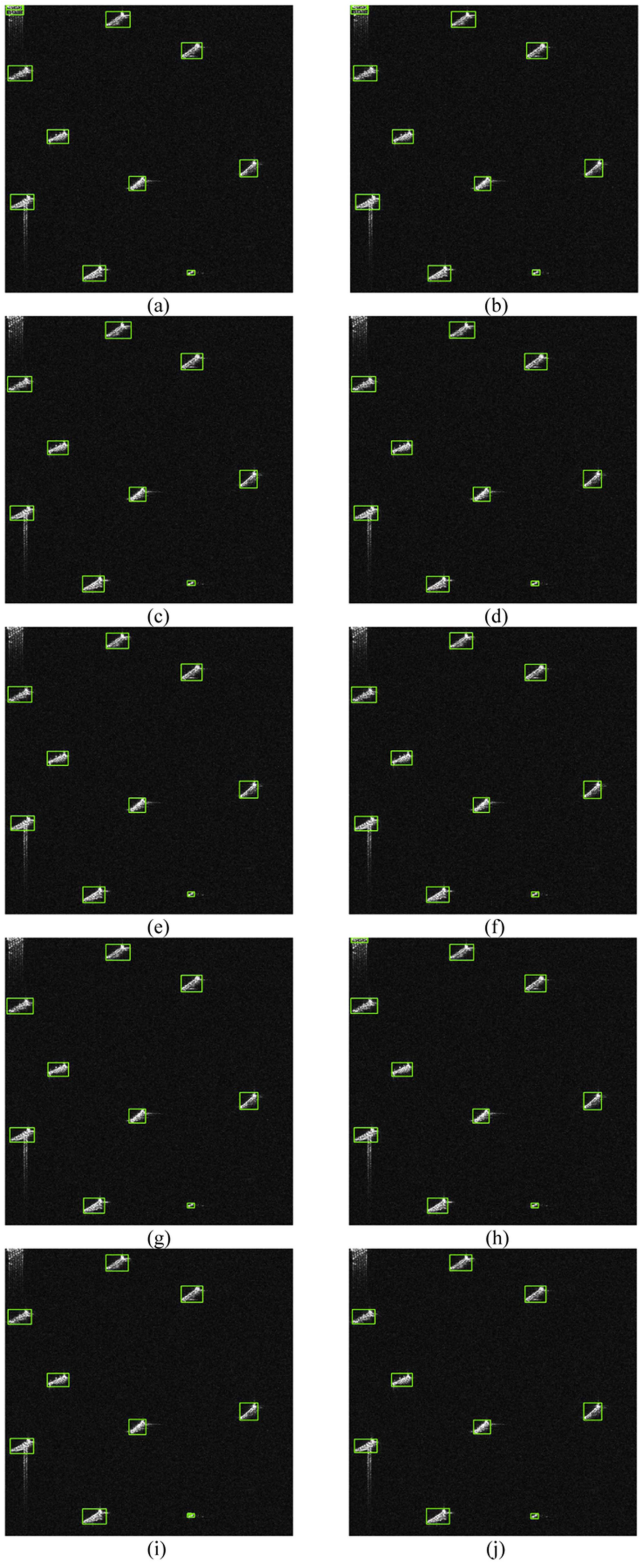


Fig. 9. Comparison of the results of different detection methods and the results after adding the MPAM in HRSID. (a) Faster-R-CNN. (b) Faster-R-CNN+MPAM. (c) RetinaNet. (d) RetinaNet+MPAM. (e) SSD300. (f) SSD300+MPAM. (g) YOLOv3. (h) YOLOv3+MPAM. (i) Improved faster-R-CNN. (j) Improved faster-R-CNN+MPAM.

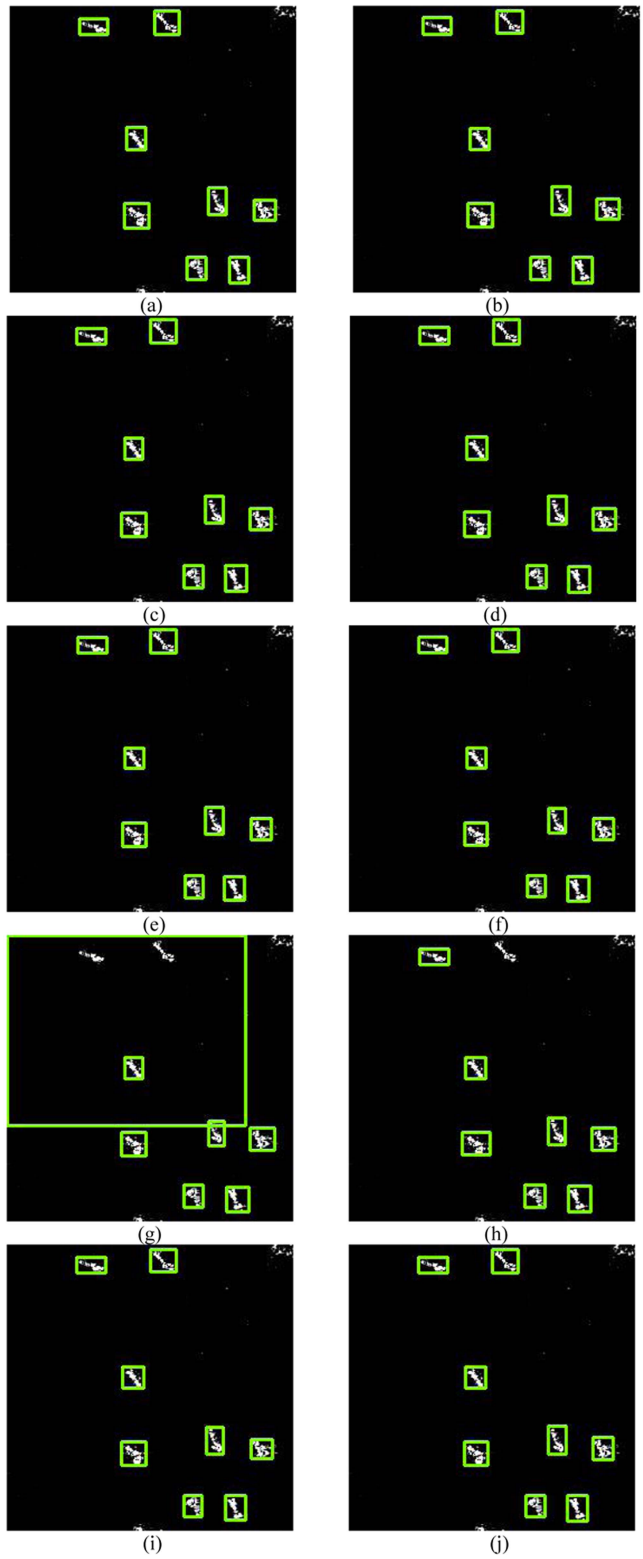


Fig. 10. Comparison of the results of different detection methods and the results after adding the MPAM in SAR-ship-dataset. (a) Faster-R-CNN. (b) Faster-R-CNN+MPAM. (c) RetinaNet. (d) RetinaNet+MPAM. (e) SSD300. (f) SSD300+MPAM. (g) YOLOv3. (h) YOLOv3+MPAM. (i) Improved faster-R-CNN. (j) Improved faster-R-CNN+MPAM.

TABLE IV
COMPARISON OF DETECTION PERFORMANCE OF DIFFERENT LEVELS OF THE MPAM

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
RetinaNet	0.615	0.821	0.698	0.385	0.814	0.800
Two layers	0.583	0.796	0.665	0.343	0.795	0.800
Three layers (ours)	**0.632**	**0.839**	**0.718**	**0.412**	**0.826**	**0.900**
Four layers	0.587	0.795	0.663	0.347	0.801	0.602
SSD300	0.549	0.791	0.632	0.353	0.742	0.525
Two layers	0.552	0.792	0.639	0.362	0.748	0.601
Three layers (ours)	**0.555**	**0.793**	**0.652**	**0.366**	0.747	0.539
Four layers	0.553	0.787	0.637	0.360	**0.753**	**0.752**

The bold entities means best values.

overall performance of each detection method, indicating the applicability and superiority of the MPAM.

D. Discussions

In this part, we discuss the several factors that affect the performance of the proposed MPAM.

1) Number of Feature Extraction Layers: In MPAM, the DFES is used for deep feature extraction. We utilize three layers of features in the DFES. In order to evaluate the detection performance of three layers, we compare it with two layers or four layers in the DFES in HRSID dataset, and the results are shown in Table IV. Taking RetinaNet and SSD300 as examples, the six indicators of the proposed three layers in the DFES is the highest in the six indicators of RetinaNet. In SSD300, the proposed three layers in DFES has the highest four indicators of AP, AP_{50}, AP_{75}, and AP_s compared with two or four layers in the DFES, while AP_m and AP_l reach the highest when four layers in the DFES are used. Although the four layers in the DFES have high detection accuracy for small targets, the comprehensive accuracy of the proposed three layers in the DFES is the highest as a whole. This is because although the features extracted by four-layer feature extraction are deeper, the increase of the number of layers will bring exponential growth of time overhead and memory overhead, and it may be overfitting and difficult to converge, resulting in the decline of detection performance. In addition, the two-layer feature extraction is because the extracted features are shallow and the detailed features are less extracted, which leads to the degradation of detection performance. In summary, we use three layers of feature extraction is the most appropriate.

2) Comparison of Different Attention Modules: To illustrate the advancement of the proposed MPAM, we compare it with the most commonly used and latest attention modules similarity-based attention mechanism (SimAM) [40], squeeze-and-excitation network (SEnet) [41], and CBAM [32] in computer vision methods, and the results for SSD300 and YOLOv3 in HRSID dataset are shown in Table V. Although the highest AP_{50} and AP_l in SSD300 is obtained by adding the CBAM, and the highest AP_{50} and AP_l in YOLOv5 are generated by adding the SEnet and SimAM, respectively. But, whether in the SSD300 method or YOLOv5, after adding the MPAM, the four indicators AP, AP_{75}, AP_s, and AP_m are higher than the other three methods, indicating that our method is more suitable for SAR image ship detection.

TABLE V
COMPARISON OF DETECTION PERFORMANCE IMPROVEMENT BY DIFFERENT MODULES

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
SSD300	0.549	0.791	0.632	0.353	0.742	0.525
SSD300+SimAM	0.553	0.789	0.647	0.363	0.742	0.552
SSD300+SEnet	0.549	0.789	0.633	0.362	0.746	0.550
SSD300+CBAM	0.551	**0.794**	0.638	0.361	0.747	**0.559**
SSD300+ours	**0.555**	0.793	**0.652**	**0.366**	**0.747**	0.539
YOLOv3	0.562	0.825	0.644	0.412	0.728	0.060
YOLOv3+SimAM	0.566	0.827	0.649	0.413	0.725	**0.261**
YOLOv3+SEnet	0.567	**0.835**	0.634	0.412	0.728	0.200
YOLOv3+CBAM	0.550	0.832	0.626	0.406	0.702	0.115
YOLOv3+ours	**0.569**	0.832	**0.651**	**0.415**	**0.728**	0.244

The bold entities means best values.

TABLE VI
IMPACT OF ADDING MPAM ON THE OPERATION TIME IN SSDD

Method	FPS
Faster-R-CNN	**16.13**
Faster-R-CNN+MPAM	15.63
RetinaNet	**16.45**
RetinaNet+MPAM	16.39
SSD300	**12.45**
SSD300+MPAM	10.00
YOLOv3	**11.94**
YOLOv3+MPAM	11.93
Improved Faster-R-CNN	**9.52**
Improved Faster-R-CNN+MPAM	8.26

The bold entities means best values.

TABLE VII
ABLATION EXPERIMENTS IN SSDD

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Baseline	0.627	0.906	0.738	0.607	0.673	0.328
Baseline + DFES	0.638	**0.940**	0.756	0.609	0.695	0.618
Baseline + CMAFS	0.640	0.934	0.755	0.607	0.690	**0.684**
Baseline + SMAFS	0.634	0.926	0.749	0.608	0.694	0.635
Baseline + MPAM	**0.643**	0.924	**0.759**	**0.610**	**0.698**	0.672

The bold entities means best values.

3) Average Inference Time: Because MPAM is a plug-and-play lightweight module, we discuss whether the MPAM affects the average inference time, and the results in SSDD dataset are shown in Table VI. Since larger FPS leads to faster detection, we can see from the table, FPS decreases after adding MPAM, but the decrease is very small. The most influential is the SSD300 method, FPS only decreases by 2.45 after adding MPAM, and the other four methods have little effect, especially the YOLOv3 method, which only decreases by 0.01. This result proves that MPAM is a plug-and-play lightweight module, which has almost no effect on the detection speed.

4) Contributions of Different Modules: In the model, we design three modules DFES, CMAFS, and SMAFS, respectively, and we will conduct ablation experiments on SSDD dataset to discuss the contribution of the three modules, respectively. The results are shown in Table VII. We use RetinaNet as the baseline, and in order to verify the contribution of each module, we add each module to the baseline separately. It can be seen from

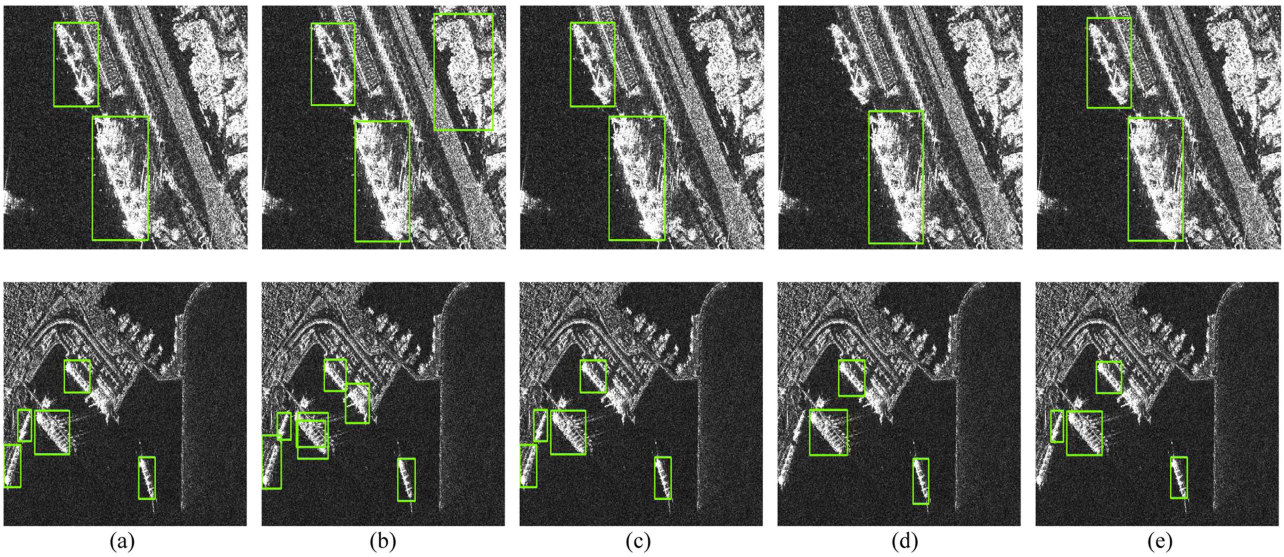


Fig. 11. Detection results under complex background. (a) Ground truth (b) Faster-R-CNN. (c) Faster-R-CNN+MPAM. (d) RetinaNet. (e) RetinaNet+MPAM.

TABLE VIII
EXPERIMENTS WITH COMPLEX BACKGROUNDS

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Faster-R-CNN	0.458	0.778	0.516	0.367	0.556	0.578
Faster-R-CNN+MPAM	**0.490**	**0.812**	**0.533**	**0.436**	**0.576**	**0.643**
RetinaNet	0.401	0.668	0.410	0.351	0.508	0.382
RetinaNet+MPAM	**0.424**	**0.683**	**0.428**	**0.382**	**0.517**	**0.399**

The bold entities means best values.

the table that when we add DFES alone, all its indicators are greater than the baseline, and the AP_{50} value is higher than each module, and 3.4% higher than the baseline, indicating that the deep feature extraction obtains more semantic information and improves the ship detection accuracy. When we add CMAFS alone, each index is greater than the baseline, and the value of AP_l is higher than that of each module, indicating that the attention mechanism of the channel strengthens the acquisition of useful information of the channel, removes redundant background interference, and improves the detection accuracy of large ships. And the most standard index AP is greater than the index after adding DFES and SMAFS. It shows that the CMAFS contributes the most to the whole model. When the SMAFS is added alone, although all indicators are higher than the baseline, there is no prominent indicator, and AP is the lowest among the three modules, indicating that the SMAFS has the lowest contribution to the whole model, and the spatial attention mechanism cannot remove redundant information well, which needs to be improved with the channel attention mechanism. Therefore, we carry out the combination of space and channel, and achieved good results.

5) Effect on the Detection of Complex Background: In order to verify the effectiveness in complex background, we used a dataset of complex background for experiments, and the results are shown in Table VIII. It can be seen from the table that no matter which algorithm is used, all the indicators are improved after adding MPAM, especially the AP indicators, which are improved by 3.2% in Fast-R-CNN and 2.3% in RetinaNet, indicating that our method is also effective in complex backgrounds. Our results are shown in Fig. 11. It can be clearly seen from the figure that the detection effect is greatly improved after adding MPAM. In the first figure, the Faster-R-CNN algorithm detects redundant boxes, while the redundant boxes are filtered after adding MPAM, and the RetinaNet algorithm only detects one ship. However, all of them were detected after adding MPAM. In the second figure, two more redundant boxes are detected by the Fast-R-CNN algorithm, and the redundant boxes disappear after adding MPAM. Only three ships are detected by the RetinaNet algorithm, while one more ship is obviously detected after adding MPAM, although not completely detected. This test result shows that the effect is greatly improved after adding MPAM.

6) Effect in the Anchor-Free Method: The anchor-free object detection method is also a very popular research direction in recent years. It tries to get rid of the dependence on anchor boxes or prior boxes used in traditional object detection methods, which can simplify the model structure and reduce the complexity of training and inference to a certain extent. It achieves good performance on some datasets. Since the previous discussion is based on the object detection with anchor boxes, we choose FCOS [42] and Centernet [43], two methods without anchor boxes, to discuss whether MPAM is also applicable to the object detection method without anchor boxes, and the results are shown in Table IX. As can be seen from the data in the table, for FCOS, the most important index AP is increased by 0.5% after adding MPAM, and the other four indexes AP_{50}, AP_{75}, AP_s, and AP_m are also improved. In Centernet, AP is improved by 0.8% after adding MPAM, and the remaining indicators are also higher than those when using Centernet alone, among which AP_{75} is the most obvious, increasing by 7.6%. In summary, our proposed MPAM is also suitable for anchor-free methods.

TABLE IX
EXPERIMENTS IN ANCHOR-FREE OBJECT DETECTION METHOD

Method	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
FCOS	0.632	0.981	0.777	0.594	0.700	**0.707**
FCOS+MPAM	**0.637**	**0.982**	**0.784**	**0.605**	**0.706**	0.704
Centernet	0.452	0.910	0.346	0.443	0.483	0.541
Centernet+MPAM	**0.460**	**0.911**	**0.422**	**0.452**	**0.495**	**0.576**

The bold entities means best values.

IV. CONCLUSION

In this article, we propose a plug-and-play multiscale ship attention module (MPAM) for detecting ships in SAR images. The MPAM consists of two parts, which are the channel multilayer attention branch and the spatial multilayer attention branch. Both branches contain our designed DFES. Through the DFES, the representative features of multiscale ships can be extracted and the interference of the surrounding environment can be suppressed. In addition, two branches contain CMAFS and SMAFS, respectively, which can fuse channels and spatial attention blocks on different levels of feature maps, so as to better learn relevant features from channel and spatial dimensions and enhance feature representation. The extracted feature maps are fed into different ship detection networks to obtain the detection results. Our experimental results on three datasets (SSDD, HRSID, and SAR-ship-Dateset) show the performance of ship detection methods could be improved by adding the proposed MPAM.

In future works, the computational efficiency of the proposed MPAM will be improved. The specific problems in SAR ship detection after recognition include sidelobes, ghosted ships, and the characteristics of SAR images themselves are also worth further research.

REFERENCES

[1] D. Cerutti-Maori, J. Klare, A. R. Brenner, and J. H. G. Ender, "Wide-area traffic monitoring with the SAR/GMTI system PAMIR," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 3019–3030, Oct. 2008.

[2] P. Wang, L. Wang, H. Leung, and G. Zhang, "Super-resolution mapping based on spatial–spectral correlation for spectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2256–2268, Mar. 2021.

[3] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[4] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.

[5] C. P. Schwegmann, W. Kleynhans, and B. P. Salmon, "Manifold adaptation for constant false alarm rate ship detection in South African oceans," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 7, pp. 3329–3337, Jul. 2015.

[6] X. Qin, S. Zhou, H. Zou, and G. Gao, "A CFAR detection algorithm for generalized gamma distributed background in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 806–810, Jul. 2013.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 6999–7019, Dec. 2022.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.

[10] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Proc. 13th Int. Conf. Control Automat. Robot. Vis.*, 2014, pp. 844–848.

[11] T. Guo, J. Dong, H. Li, and Y. Gao, "Simple convolutional neural network on image classification," in *Proc. IEEE 2nd Int. Conf. Big Data Anal.*, 2017, pp. 721–724.

[12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3431–3440.

[13] K. He et al., "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[14] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.

[15] L. Liu et al., "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, pp. 261–318, 2020.

[16] X. Xie et al., "Oriented R-CNN for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3520–3529.

[17] C. Chen et al., "R-CNN for small object detection," in *Proc. 13th Asian Conf. Comput. Vis.*, 2016, pp. 214–230.

[18] W. WLiu et al., "SSD: Single shot multibox detector," in *Proc. 14th Eur. Conf.*, 2016, pp. 21–37.

[19] J. Redmon et al., "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.

[20] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.

[21] J. Redmon and A. Farhadi, "YOLOV3: An incremental improvement," in *Proc. IEEE Conf. CVPR*, Apr. 2017, pp. 1–6.

[22] T.-Y. Lin et al., "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[23] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.

[24] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.

[25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[26] T.-Y. Lin et al., "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125.

[27] W. Dai et al., "A novel detector based on convolution neural networks for multiscale SAR ship detection in complex background," *Sensors*, vol. 20, no. 9, 2020, Art. no. 2547.

[28] M. Kang et al., "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, Aug. 2017, Art. no. 860.

[29] J. Jiao et al., "A densely connected end-to-end neural network for multiscale and multiscene SAR ship detection," *IEEE Access*, vol. 6, pp. 20881–20892, 2018.

[30] C. Zhao, X. Fu, J. Dong, R. Qin, J. Chang, and P. Lang, "SAR ship detection based on end-to-end morphological feature pyramid network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4599–4611, Feb. 2022.

[31] 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.

[32] S. Woo et al., "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.

[33] Y. Zhao, L. Zhao, B. Xiong, and G. Kuang, "Attention receptive pyramid network for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2738–2756, May 2020.

[34] D. He, Q. Shi, X. Liu, Y. Zhong, and X. Zhang, "Deep subpixel mapping based on semantic information modulated network for urban land use mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10628–10646, Dec. 2021.

[35] D. He et al., "Generating annual high resolution land cover products for 28 metropolises in China based on a deep super-resolution mapping network using Landsat imagery," *GIScience Remote Sens.*, vol. 59, no. 1, pp. 2036–2067, 2022.

[36] T. Zhang et al., "SAR ship detection dataset (SSDD): Official release and comprehensive data analysis," *Remote Sens.*, vol. 13, no. 18, 2021, Art. no. 3690.

[37] 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.

[38] Y. Wang et al., "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, Mar. 2019, Art. no. 765.
[39] B. Chai, L. Chen, H. Shi, and C. He, "Marine ship detection method for SAR image based on improved faster RCNN," in *Proc. SAR Big Data Era*, 2021, pp. 1–4.
[40] L. Yang et al., "SimAM: A simple, parameter-free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11863–11874.
[41] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.
[42] Z. Tian et al., "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 9627–9636.
[43] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv: 1904.07850*.

Peng Wang (Senior Member, IEEE) received the B.E. degree in microelectronics and the Ph.D. degree in information and communications engineering from the College of Information and Communications Engineering, Harbin Engineering University, Harbin, China, in 2012 and 2018, respectively.

In 2016, he was a Visiting Ph.D. Student with the Grenoble Images Parole Signals Automatics Laboratory, Grenoble Institute of Technology, Saint Martin d'Hères, France. He is currently an Associate Professor and Doctoral Supervisor with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. He is also a Hong Kong Scholar with the Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong. He has authored two books and more than 50 articles. His research interests include remote-sensing imagery processing and machine learning.

Dr. Wang is a Reviewer of more than 20 international journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *Remote Sensing of Environment*, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

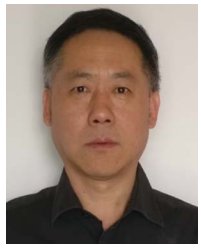
Yongkang Chen received the B.E. degree in electronic information engineering from Anhui University, Hefei, China, in 2022. He is currently working toward the M.S. degree in electronic information with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Yi Yang received the B.E. degree in electronics and information engineering and the M.E. degree in pattern recognition and intellectual system from the Huazhong University of Science and Technology, Hubei, China, in 2009 and 2012, respectively.

He is currently a Research Assistant with Natural Resources Investigation and Monitoring Research Center, Chinese Academy of Surveying and Mapping, Beijing, China. His research interests include remote sensing image processing and deep learning.

Ping Chen received the Ph.D. degree in signal and information processing from the North University of China, Taiyuan, China, in 2011.

He is currently a Professor with the Departments of Information and Communication Engineering, North University of China, and the Director with the Center for Shanxi Key Laboratory of Signal Capturing and Processing. He has undertaken projects funded by a variety of grants and authored/coauthored more than 80 academic articles in the research areas of image processing, image recognition, and X-ray CT imaging



Gong Zhang (Member, IEEE) received the Ph.D. degree in electronic engineering from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2002.

From 1990 to 1998, he was a Member of Technical Staff at the No724 Institute of China Shipbuilding Industry Corporation, Nanjing. Since 1998, he has been working with the College of Electronics and Information Engineering, NUAA, where he is currently a Professor. His research interests include radar signal processing and compressive sensing.

Dr. Zhang is a Member of the Committee of Electromagnetic Information, Chinese Society of Astronautics, and a Senior Member of the Chinese Institute of Electronics.

Daiyin Zhu received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, in 1996, and the M.S. and Ph.D. degrees in electronics from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, in 1998 and 2002, respectively.

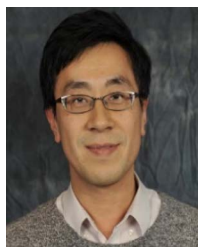
From 1998 to 1999, he was a Guest Scientist with the Institute of Radio Frequency Technology, German Aerospace Center (DLR), Oberpfaffenhofen, Germany, where he was in the field of synthetic aperture radar (SAR) interferometry. In 1998, he joined the Department of Electronic Engineering, NUAA, where he is currently a Professor. He has developed algorithms for several operational airborne SAR systems. His research interests include radar imaging algorithms, SAR ground moving target indication, SAR/ISAR autofocus techniques, SAR interferometry, and multiple-input–multiple-output SAR signal processing.

Yongshi Jie received the Ph.D. degree in signal and information processing from the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2021.

He is currently an Engineer with the Beijing Institute of Space Mechanics and Electricity, China Academy of Space Technology, Beijing. He has authored and coauthored ten articles. His research interests include remote sensing image information extraction and deep learning

Cheng Jiang received the B.S. degree in applied physics and the Ph.D. degree in optical engineering from the College of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, China, in 2007 and 2013, respectively.

He is currently a Professor with the Beijing Institute of Space Mechanics and Electricity, China Academy of Space Technology, Beijing. He has undertaken projects funded by a variety of grants and authored/coauthored more than 30 academic articles in the research areas of remote sensing data processing and satellite-based application.



Henry Leung (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the McMaster University, Hamilton, ON, Canada, in 1991.

Before joining the University of Calgary, Calgary, AB, Canada, he was a Defense Scientist with the Department of National Defence (DND), Canada. His current research interests include information fusion, machine learning, Internet of Things (IoT), nonlinear dynamics, robotics, and signal and image processing.

Dr. Leung is the Associate Editor of *IEEE Circuits and Systems Magazine*. He is the Topic Editor on "Robotic Sensors" of the *International Journal of Advanced Robotic Systems*. He is also an Editor of the Springer book series on "*Information Fusion and Data Science*." In addition, he is a Fellow Member of the Society of Photo-Optical Instrumentation Engineers (SPIE).