An Improved Lightweight RetinaNet for Ship Detection in SAR Images

Tian Miao, HongCheng Zeng, *Member, IEEE*, Wei Yang, Boce Chu, Fei Zou, Weijia Ren, and Jie Chen, *Senior Member, IEEE*

***Abstract*—The rapid development of remote sensing technology has led to a sharp increase in the amount of synthetic aperture radar (SAR) measurements, which put forward higher requirements for remote sensing image processing. As an important application of SAR, fast and accurate ship detection has always been a research hotspot. In this article, an improved lightweight RetinaNet for ship detection in SAR images is proposed. Compared with the standard RetinaNet, the shallow convolutional layers of the backbone are replaced by ghost modules and the number of the deep convolutional layers is reduced. The spatial and channel attention modules are embedded into the model to enhance detectability. *K*-means clustering algorithm is applied to adjust the initial aspect ratios of the model. The effectiveness and robustness of the proposed method is demonstrated by numerical experiments with SSDD dataset, Gaofen-3 mini dataset, and a large-scale SAR image of Hisea-1 satellite, it is shown that the proposed method can significantly reduce the floating-point operations and the number of parameters without decreasing the detection accuracy and recall ratio. Moreover, the experimental results also show the proposed model's robustness and the ability to detect ship targets in small datasets.**

***Index Terms*—Lightweight design, Retinanet, synthetic aperture radar (SAR), ship detection.**

I. Introduction

GIVEN its advantages of working all-day and all-weather, synthetic aperture radar (SAR) has become an important tool for earth observation [1]. At present, SAR is widely used in environmental monitoring, marine monitoring, resource exploration, ocean monitoring, military, surveying, and mapping [2].

Ship detection is an important application for SAR images due to its high economic and military value. SAR receives the information of the earth's surface through microwave that can penetrate water vapor and clouds, which makes SAR a competitive tool for reliable ship detection, in contrast to limited optical images [1], [2]. At present, ship detection is mainly based on single-channel SAR images and polarimetric SAR images [3], [4]. Polarimetric SAR image contains more information than single channel SAR image, but the amount of data and application of single channel SAR are more extensive [3], [4].

The traditional method for ship detection counting on the statistical modeling of the distribution of background clutter requires high sea surface conditions, unaffected ship surroundings, and high-class satellite imaging quality. Constant false-alarm rate (CFAR) is a traditional SAR ship detection method, based on which many scholars carry out SAR ship detection research [7]. An *et al.* [8] proposed an improved method which reduces the influence of target returns on the estimation of local sea clutter distributions. To avoid detecting noise as ships without using spatial domain information, a novel bilateral CFAR algorithm proposed by Leng *et al.* [9] has combined the intensity distribution and the spatial distribution together, as a result, the detection rate is effectively improved and the false alarm rate is reduced. To improve model's ability to detect targets of different sizes, Dai *et al.* [10] presented to use the target proposal generator to generate region proposals of different sizes first, and apply CFAR on detecting ships. Some researchers also focus on ship detection in the complex backgrounds like harbors and busy shipping lines [11]–[13]. Ai *et al.* [11] presented a new two-parameter CFAR detector to improve the traditional CFAR to enhance its detection ability in complex environment. Zhai *et al.* [12] generated superpixel regions and detect inshore ships based on salient region detection. In addition, some researchers consider using different distributions to deal with sea clutter. Liao *et al.* [14] used an alpha stable distribution to model the distribution of sea clutter, and image segmentation is applied to the algorithm to improve the homogeneity of each region. To avoid the interference of land targets on detection, some researchers consider suppressing false targets on land before detection. In addition to the general image segmentation algorithms, researchers have also carried out relevant research works based on SAR images [15]–[17]. In [15], the researchers improved the standard generalized mean shift algorithm according to the characteristics of polarimetric SAR and verified the effectiveness of the algorithm through real radar data. Jin *et al.* [16] developed a level set segmentation algorithm for polarimetric SAR images based on a heterogeneous cluster model, and experiments show that it has more

Manuscript received January 22, 2022; revised March 25, 2022 and April 30, 2022; accepted May 26, 2022. Date of publication June 3, 2022; date of current version June 17, 2022. This work was supported by the National Nature Science Foundation of China under Grant 62101014. *(Corresponding author: Wei Yang.)*

Tian Miao is with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China, and also with the CETC Key Laboratory of Aerospace Information Applications, Shijiazhuang 050081, China (e-mail: miaot@buaa.edu.cn).

HongCheng Zeng, Wei Yang, and Jie Chen are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: zenghongcheng@buaa.edu.cn; yangweigigi@sina.com; chenjie@buaa.edu.cn).

Boce Chu is with the CETC Key Laboratory of Aerospace Information Applications, Shijiazhuang 050081, China (e-mail: bocc012628077@163.com).

Fei Zou is with the Beijing Institute of Remote Sensing Information, Beijing 100192, China (e-mail: z_feifei@126.com).

Weijia Ren is with the Spacety Company, Ltd. (Changsha), Changsha 410221, China (e-mail: rwj@spacety.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3180159

accurate segmentation results than conventional segmentation methods.

In recent years, target detection technology has made great progress with the development of deep learning. In 2013, the region-convolutional neural network (R-CNN) successfully applied deep learning to target detection [18]. Since then, many advanced target detection networks have been proposed and widely used, and target detection methods in deep learning started to be applied to SAR image processing. Some researchers attempted to combine the traditional method with the deep learning model to optimize its performance. For instance, in [19], the faster R-CNN was modified through reevaluating bounding boxes with low scores by CFAR to gain better performance. In [20], the researchers used CFAR to detect inshore ships on the basis of sea-land segmentation, and the CNN network is used to identify false targets and suppress false alarms. Some researchers have explored the improvement of target recognition network in SAR ship detection. Li *et al.* [21] used feature fusion, transfer learning, hard negative mining, and other implementation details to improve the standard faster-RCNN, Wang *et al.* [22] used RetinaNet to detect targets in a large SAR dataset and explore the influence of different parameters. In [23], Zhang *et al.* proposed a quad feature pyramid network (FPN) for ship detection in SAR images, which suppresses the background interference and overcomes the challenge brought by multiscale ships.

While using CFAR to detect ships in SAR images, the amplitude of radar signal is the main decisive factor, and the shape of ships is difficult to be used by the model. Other extra modules are needed to be added to further improve the model's accuracy. Compared with those CFAR based methods, the deep learning-based algorithms can achieve a higher detection rate with a lower false alarm in SAR ship detection. However, its excellent performance is at the cost of a great amount of calculation, which brings difficulties to the practical application of ship detection. Image processing based on deep learning relies on a large number of operations to reduce loss to approximate the optimal solution. Therefore, in addition to pursuing a higher detection rate and recall rate, the lightweight design of the model is also a research hotspot of researchers. Many explorations have been made in target detection of optical images [24], [25]. However, there are fewer related studies for SAR images. This article focuses on ship detection in SAR images and explores enhancing the detection ability of the model under the condition of reducing the amount of model parameters. We chose RetinaNet as the basic framework of target detection, and made a series of modifications to reduce the parameters of the model and improve the detection ability of the model. The effectiveness of the model is evidence-based and tested on multiple datasets. The contributions of this article are as follows.

1) An improved lightweight backbone based on ResNet-50 for SAR images is proposed. In the proposed backbone, part of the shallow convolution layers is replaced by ghost modules, and the number of deep convolution layers is reduced. As a result, the number of parameters of the model is reduced by 25% compared with ResNet-50, and the accuracy and recall rate are improved on multiple datasets.
2) Based on the proposed lightweight backbone, a backbone with better detection effect is proposed. Several modules and methods are applied in the model. The spatial attention modules and channel attention modules are embedded in the detection model. As optional blocks, attention modules can improve the detection effect without increasing the computational cost. The initial aspect ratios of anchors are reset by the *K*-means clustering algorithm, through which the model adapts to the size of the targets better.
3) Compared with previous models, the proposed model has been evaluated on several datasets. SAR ship slices in the datasets are from Radarsat-2, TerraSAR-X, Sentinel-1, and GF-3 of different polarization modes and resolutions. A large scene SAR image of Hisea-1 has also been used to test the robustness of the model.

The rest of this article is organized as follows. Section II introduces the related work. Section III presents the specific parts of the improvement and explains the structure of the model. Section IV demonstrates and analyzes the experimental results. Section V discusses the selection of model parameters and the lightweight design of the model. Finally, Section VI concludes this article.

II. Related Work

A. Focal Loss and RetinaNet

To thoroughly learn the information of targets and attain a good detection effect, models often need to train a large number of samples. In order to enhance the robustness of the models, targets of the same type in a dataset tend to have a variety of characteristics. However, different samples have the same impact on most detection models in the training process, meaning that the loss weights of easy and difficult samples are the same. Though the loss of easy samples is lower in target detection, the huge number of easy samples will still dominate of the total loss, which reduces the model's ability to detect hard samples. To reduce the impact of samples' imbalance on the model, Lin *et al.* [26] proposed focal loss and applied it to their model: RetinaNet.

Focal loss makes the model focus on hard samples by changing the loss weight of different samples. Compared with cross-entropy loss, an adjustment factor is added to the focal loss. The expressions of standard cross-entropy loss and focal loss are as follows:

$$L_{\mathrm{ce}} = -\log(p) \tag{1}$$

$$L_{\mathrm{fl}} = -(1-p)^{\beta}\log(p) \tag{2}$$

where L_{ce} is the standard cross-entropy loss, L_{fl} is the focal loss, *p* represents the probability that the sample belongs to the true class, and β is the adjustment factor of focal loss. When β takes 0, focal loss becomes standard cross-entropy loss, and as β increases, the suppression of simple samples by focal loss also increases, which means hard samples play a more important role in the training of the model.

RetinaNet is a detection network that combines backbone, FPN, and task-specific subnets together. Focal loss has been

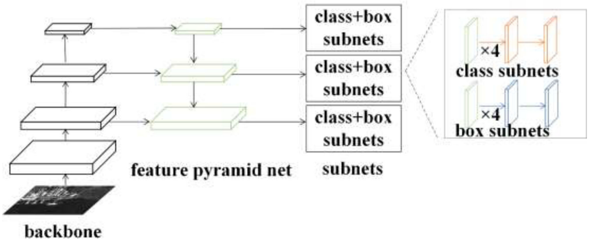


Fig. 1. Structure of RetinaNet.

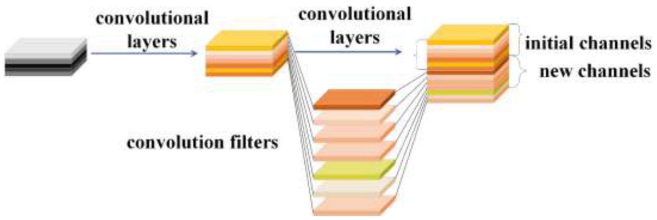


Fig. 2. Structure of ghost module.

applied in subnets to improve the model's detection ability. The structure of RetinaNet is shown in Fig. 1. RetinaNet can be divided into three parts. In its first part, input images are processed by CNN to obtain the corresponding feature maps. In the second part, the feature maps are entered into FPN, in which features of different scales are extracted and recombined. At last, the feature maps of three different scales output by FPN are sent into "box+class" subnets to get targets' classes and locations. RetinaNet calculates loss of location through smooth L1 loss, while calculating loss of classification, to improve the detecting ability of the model, the focal loss is used to replace standard cross-entropy loss.

Compared with the previous target detection networks, RetinaNet has made significant progress. However, it is mainly used in optical images [27]. Some researchers have tried to apply it to SAR images for ship detection, but most of these studies are aimed at adjusting the parameters of the network, or testing RetinaNet in different scenarios, and there are few studies on improving the structure of RetinaNet model based on the characteristics of SAR image, further research lacks [22], [28].

B. Ghost Module

In recent years, with the wide application of artificial intelligence in the industry, many researchers have explored in lightweight model design. GhostNet is a novel and effective, lightweight network proposed in 2020 [24].

In CNN, images' information is often transmitted in feature maps in convolutional layers and pool layers. Different channels of feature maps contain different information of the input image. According to the research, Han *et al.* [24] found a large redundancy among different channels of the feature map, which means it is unnecessary to use a convolutional layer to generate all channels' information of the feature map. The structure of the Ghost module is shown in Fig. 2 [26]. To reduce resource consumption, the researchers only use standard convolutional layers to generate part of channels of the feature map, which are called initial channels. For the rest channels, the researchers use ordinary convolution filters with fewer floating-point operations (FLOPs) and parameters to generate, which is called new channels. Since new channels are generated based on the initial channels, they are also regarded as the "ghost" of the initial channels. In GhostNet, ghost modules replace the traditional convolutional layers to make the network more lightweight.

Ghost module has been proved to be an effective, lightweight method, but most of the research on the Ghost module is based on classification tasks. There are few researchers on the improvement of target detection modules based on ghost module, especially for SAR images.

III. Methodology

As an efficient one-stage target detection network, RetinaNet has a strong detection ability for hard targets due to the focal loss, which is suitable for ship detection in SAR images. We choose RetinaNet as the basic framework of the model and improved it in several aspects. Section III-A introduces the backbone of the proposed model, then the attention modules of the model are presented in Section III-B. Section III-C introduces the setting rules of the initial aspect ratios of the bounding boxes, and Section III-D shows the whole structure of the model.

A. Lightweight Backbone: Ghost-ResNet-41

In RetinaNet, the input images are processed by the backbone to get their feature maps first. As an efficient neural network for feature extraction, ResNet [29] is often used as the backbone of target detection models. A large number of residual structures are used in ResNet to overcome the degradation of the network. Experiments on many traditional optical datasets show that taking ResNet as the backbone of the detection model can achieve excellent detection results [30], [31]. For SAR images, some researchers also use ResNet as the backbone, and compared with other backbones, ResNet can still get good result [32]. However, compared with optical images, the single-channel amplitude images of SAR contain less information, the standard ResNet designed for optical images is not the best choice for the backbone. The different imaging methods of optical sensors and SAR lead to the difference between optical images and SAR images. Specifically, ship targets have a clear outline in the optical image, and the optical images can directly reflect ships' color information and shapes. However, SAR images mainly reflect the backscattering coefficients in different regions, as a result, it is difficult to distinguish ships from other targets in SAR images visually, what's more, the SAR images lack color information, which makes the neural network designed for the optical image cannot achieve the best effect.

To make ResNet extract features from SAR images more effectively, the structure of standard ResNet-50 is improved in the proposed model for SAR images.

For SAR images, the basic features extracted by the shallow convolutional layers are not as complex as those of optical images, therefore, it is not necessary for the shallow network to use convolutional layers to generate all channels of feature maps. The shallow convolutional layers of ResNet-50 can be replaced

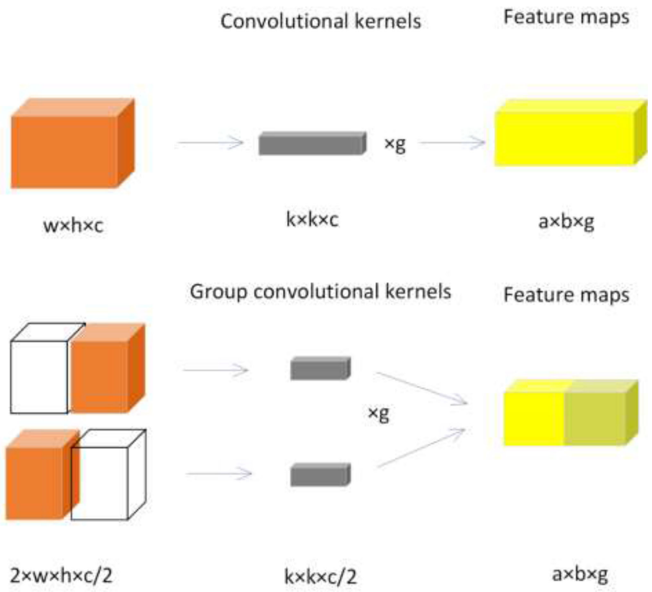


Fig. 3. Comparison of standard convolutional layers and group convolutional layers. In which w, h, and c represent the length, width, and depth of the input feature map, respectively, k and c represent convolutional kernels' length and depth, g represents the number of convolutional kernels/output feature maps' depth, and a, b represent output feature maps' length and width, respectively. The figure shows that group convolutional kernels require fewer parameters than standard convolutional kernels to generate a feature map of the same size.

by ghost modules, through which convolutional layers generate only part of channels of feature maps, the rest of the channels are generated through filters, which requires less computing resource. In the proposed backbone, group convolutional layers are used as the filters to generate half of the feature channels in the ghost module. Compared with standard convolutional layers, using group convolution to generate the same amount of feature maps requires fewer parameters.

As shown in Fig. 3, the size of the input feature map is $w \times h \times c$, in convolutional layers, the input feature map is processed by g convolutional kernels, assuming that the size of the convolutional kernel is $k \times k \times c$, and the size of the output feature map is $a \times b \times g$. The parameter amount of the standard convolutional layer is shown in the following formula:

$$p_c = k^2 \times c \times g \tag{3}$$

where p_c represents the parameter of the standard convolutional kernel

$$p_g = k^2 \times \frac{c}{2} \times \frac{g}{2} \times 2 = \frac{k^2 cg}{2} \tag{4}$$

where p_g represents the parameter of the group convolutional layer. Compared with p_c, p_g is reduced by half. Since half of the feature maps in ghost module are generated by group convolutional layers, the parameter can be reduced compared with standard convolutional layers.

The feature map generated by deep convolutional layers has sizeable receptive fields and high-level semantic information. Since single-channel SAR images contain less information than optical images, the depth of the model for SAR images is supposed to be less than that of optical images. In the proposed model, the number of deep convolutional layers is reduced. The structure of standard ResNet-50 and the proposed backbone is shown in Table I.

TABLE I
COMPARISON OF RESNET-50 AND GHOST-RESNET-41

Layer name	Output size	ResNet-50	Ghost-ResNet-41
Conv1	112×112	7×7, 64, stride 2	
Conv2_x	56×56	3×3 max pool, stride 2 $\begin{bmatrix}1\times 1,64\\ 3\times 3,64\\ 1\times 1,256\end{bmatrix}\times 3$	$\begin{bmatrix}1\times 1,64\\ 3\times 3,64\ (Ghost\ Module)\\ 1\times 1,256\end{bmatrix}\times 3$
Conv3_x	28×28	$\begin{bmatrix}1\times 1,128\\ 3\times 3,128\\ 1\times 1,512\end{bmatrix}\times 4$	$\begin{bmatrix}1\times 1,128\\ 3\times 3,128\\ 1\times 1,512\end{bmatrix}\times 4$
Conv4_x	14×14	$\begin{bmatrix}1\times 1,256\\ 3\times 3,256\\ 1\times 1,1024\end{bmatrix}\times 6$	$\begin{bmatrix}1\times 1,256\\ 3\times 3,256\\ 1\times 1,1024\end{bmatrix}\times 4$
Conv5_x	7×7	$\begin{bmatrix}1\times 1,64\\ 3\times 3,64\\ 1\times 1,256\end{bmatrix}\times 3$	$\begin{bmatrix}1\times 1,512\\ 3\times 3,512\\ 1\times 1,2048\end{bmatrix}\times 2$

As Table I shows, compared with ResNet-50, the Conv1, Conv3_x, and the fully connected layers are retained; the proposed backbone: Ghost-ResNet-41 replaces the 3 × 3 convolutional layers in the Conv2_x with ghost modules, and reduces the number of the residual blocks in Conv4_x and Conv5_x from 6 and 3 to 4 and 2.

B. Attention Module

After being processed by convolutional layers, the input images become feature maps with multiple channels. Take some common feature extraction networks as examples: VGG16, ResNet-18, and ResNet-34's output feature maps have 512 channels; ResNet-50, ResNet-101, and ResNet-152's output feature maps have 2048 channels [33]. These channels contain different information that contributes to feature extraction. To make full use of each channel's information, greater weight is supposed to be given to the more important channels for target detection. Therefore, embedding channel attention modules into the original model can improve the ability to detect.

The channel attention module usually compresses the feature map into one-dimensional vector and processes it to obtain the weight of each channel. In this article, we use squeeze-and-excitation (SE) blocks to improve the model. [34] SE block is a lightweight module that can be directly embedded into the existing convolutional neural network. The structure of the SE block is shown in Fig. 4.

SE block can be divided into two parts: squeeze and excitation. In the squeeze part, the feature map will be processed by global pooling layers to get a one-dimensional vector, in which each channel's information will be squeezed as a channel descriptor. The equation of the squeeze part is as follows:

$$z_{i,j}(c) = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} u_c(i,j) \tag{5}$$

where u_c represents the pixel value of the original feature map, W and H are the width and height of the feature map, and $z_{i,j}$ is the channel descriptor.

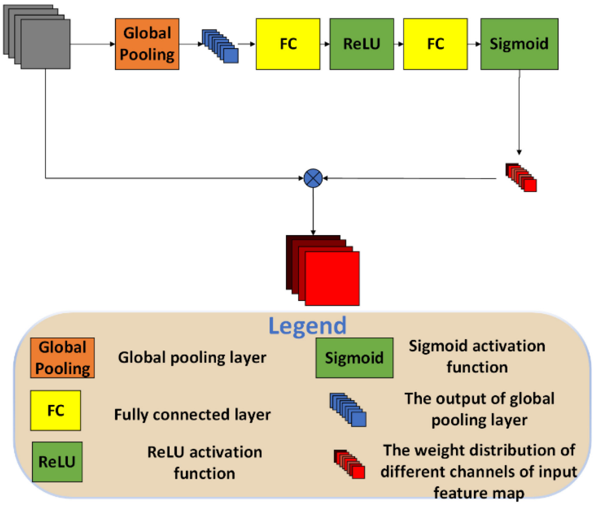

Global Pooling
FC
ReLU
FC
Sigmoid
Legend
Global Pooling
Global pooling layer
Sigmoid
Sigmoid activation function
FC
Fully connected layer
The output of global pooling layer
ReLU
ReLU activation function
The weight distribution of different channels of input feature map

Fig. 4. Structure of SE block.

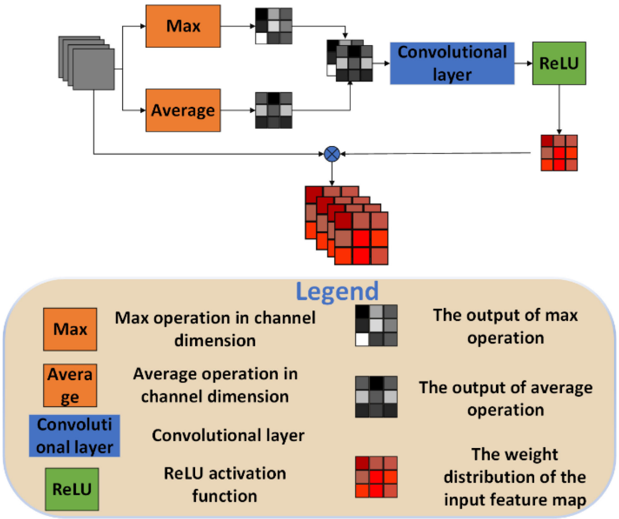

Max
Average
Convolutional layer
ReLU
Legend
Max
Max operation in channel dimension
The output of max operation
Avera ge
Average operation in channel dimension
The output of average operation
Convoluti onal layer
Convolutional layer
ReLU
ReLU activation function
The weight distribution of the input feature map

Fig. 5. Structure of spatial attention block.

In the excitation part, the one-dimensional vector will be processed by two fully connected layers. To enhance the model's ability to fit nonlinear relationships, activation functions are added after each fully connected layer. Through the excitation part, each channel's weight will be obtained as an element in the output one-dimensional vector. The output vector will be applied as each channel's weight to modify feature map's channel relationship.

Compared with ordinary optical images, targets in remote sensing images usually occupy fewer pixels, making it more difficult for models to detect the target in remote sensing images. In SAR images, ships are generally small targets. After being processed by multiple convolutional layers and pooling layers, the information of ship targets can be easily lost, leading to a decline in accuracy and recall. To reduce the influence of irrelevant areas in the images on target detection, a spatial attention module is inserted into the backbone of the model [35].

The spatial attention module can also be divided into the squeeze part and the excitation part. Unlike the SE module, the feature map is processed in the spatial dimension. First, in the squeeze part, the feature map is compressed into a feature map with two channels. As (6) and (7) show, each pixel's information will be squeezed as two spatial descriptors. The first spatial descriptor is obtained by maximization, which represents the salient feature of the pixel and the second spatial descriptor is obtained through averaging, which means the overall feature of the pixel

$$z_{\max}(i,j) = \frac{1}{C}\sum_{i=1}^{C} u_{i,j}(c) \tag{6}$$

$$z_{\text{avg}}(i,j) = \max\left[u_{i,j}(c)\right] \tag{7}$$

where $u_{i,j}$ represents the pixel value of the original feature map, C is the depth of the feature map, and $z_{\max}$ and z_{avg} are two spatial descriptors of the feature map.

Second, in the excitation part, the two-channel feature map is processed by a convolution layer, and the output is the weight matrix of the input feature map. Last, the input feature map is multiplied by weight matrix. Through the spatial attention module, the weight of the area near the target becomes larger, which makes the model pay more attention to the target area in spatial dimension. The structure of spatial attention module is shown in Fig. 5.

C. *Setting Priori Anchors by* K*-Means Clustering Algorithm*

Target detection models usually set prior anchors on the last few feature maps and get the targets' position based on the prior anchors. For RetinaNet, the output feature map of each scale corresponds to 3 scales and 3 ratios, as a result, each anchor size will generate 9 priori boxes. RetinaNet sets the initial anchor's aspect ratios as 0.5, 1.0, 2.0, respectively, to adapt targets of different ratios. To make the model adapt to the aspect ratios of the targets better, some models like YOLO-v3 and YOLO-v4 adjust the aspect ratio of the initial anchors by clustering to achieve better detect results [31], [36].

To get the exact aspect ratios of the targets, the proposed method follows YOLO-v4's setting rules to reset the initial aspect ratios of anchors by the *K*-means clustering algorithm. *K*-means clustering algorithm is an unsupervised learning algorithm that can divide datasets into several clusters according to their similarity. The steps of *K*-means clustering algorithm are as follows: First, *K* objects are selected from datasets randomly as the initial centers; second, the distance between each object and each center is calculated, and each object is reassigned to the nearest clustering center, once all objects are assigned, the clustering center of each cluster will be recalculated. The algorithm will end until no clustering centers change again.

To accurately define the similarity between objects and initial anchors' aspect ratios, the proposed model uses Intersection over Union (IOU) as the distance. As shown in Fig. 6, the red box represents the ground truth of the ship target in SAR image, and the blue box represents the bounding box of the target detect model. To calculate their IOU, the two boxes need to be extracted, and their upper left corners are placed at the same position, which is set as the origin.

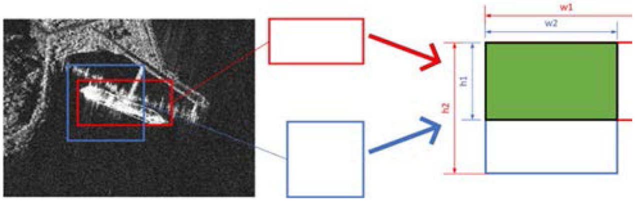


Fig. 6. Schematic diagram of IOU.

To calculate the IOU, the space of intersection area and union area need to be calculated first, as shown in the following equations:

$$S_i = \min(w_1, w_2) \times \min(h_1, h_2) \tag{8}$$

$$S_u = w_1 \times w_2 + h_1 \times h_2 - S_i \tag{9}$$

Where w_1 and h_1 represent the width and the height of the ground truth, w_2 and h_2 represent the width and the height of the initial anchor, S_i is the space of the intersection area, and S_u is the area of union area. The calculation formula of IOU is as follows:

$$\text{IOU} = \frac{S_i}{S_u}. \tag{10}$$

Before training the model, the targets' aspect ratios of the training set are clustered by *K*-means clustering algorithm to obtain the cluster centers, when the deep learning model is trained and tested, the computing resources occupied by the clustering algorithm will be released, which means that the clustering algorithm will not reduce the deep learning model's computing resources when it is processed; then the cluster centers are set as the initial anchor boxes' aspect ratios.

D. Structure of the Proposed Model

The structure of the proposed model is shown in Fig. 7. The proposed model replaces the second convolutional layer with a ghost module, and reduces the number of convolutional layers from 6 and 3 to 4 and 2. the SE modules are embedded at the end of Conv_2, Conv_3, Conv_4, and Conv_5 to correct the weight distribution of the channels; as for the spatial attention module, to improve the detection structure without significantly increasing the amount of calculation, it is embedded at the end of the backbone. The feature map output by backbone is input into FPNs, in which the fusion of low-dimensional features and high-dimensional features helps the model to detect targets of different sizes. The feature maps output by FPN are input into three different modules for classification and detection. Each module composes a subnet for classification and a subnet for box positioning. To meet different detection requirements and adapt to various application scenarios, we also provide a lightweight version of the model. In the lightweight version, all attention modules of the model have been deleted. As shown in the red box on the left in Fig. 7, the feature map output by Ghost-ResNet-41 is sent directly into FPN.

TABLE II
HARDWARE CONFIGURATION FOR EXPERIMENT

Hardware	Configuration
CPU	Intel Core i7-10870H
CPU frequency	2.20GHz
GPU	NVIDIA RTX 3070
Memory	32G
Video Memory	8G

IV. EXPERIMENTS

A. Dataset and Implementation Details

To evaluate the model comprehensively and accurately, two datasets of ship targets in SAR images are used for detection. The first dataset is SSDD, a common SAR ship dataset which contains 1160 slices and 2456 targets [19]. The slices of SSDD come from images of Radarsat-2, TerraSAR-X, and Sentinel-1, the polarization mode of images include HH, HV, VV, and VH, and the resolution of images is between 1 and 15 m. Some slices of SSDD are shown in Fig. 8.

A small SAR ship dataset based on Gaofen-3 is also constructed for evaluation. The dataset is composed of 226 slices with 478 targets. The resolution of the images is between 1 and 10 m, and all the slices are cut through rectangular windows with a size of 500 × 500. Some slices of the Gaofen-3 SAR ship dataset are shown in Fig. 9.

Pytorch 1.7.1 was used as the framework of all experiments. Pycharm 2021.1 and JupyterLab 3.0.11 was used as the integrated development environment. A work station with Ubuntu 16.04 and a laptop with Windows 10 are used as the platform for all experiments. The hardware configuration is given in Table II.

The initial learning rates of the models are set as 0.0005 and batch size are set as 8. Stochastic gradient descent is selected as the optimizer and momentum is set as 0.9. We adjust the learning rate at equal intervals. The adjustment multiple is 0.33 and the adjustment interval is 3.

B. Selection of Adjustment Factors for Focal Loss

There are two parameters to be determined in focal loss applied in the model. As shown in the following formula, α is the weighting factor to balance positive and negative examples

$$a_t = a \times p_t + (1 - a) \times (1 - p_t) \tag{11}$$

$$L_{fl} = -a_t \times (1 - p)^b \log(p). \tag{12}$$

To confirm the best parameters, we tested on the validation set of the SSDD dataset and selected mean average precision (mAP) as the evaluation index. The results are shown as shown in Table III.

As shown in Table III, mAP reaches the maximum validation set in the fifth experiment. Therefore, α takes 0.25, and β takes 2 will be better in the model.

C. Parameters of the Model's Structure

The specific parameters of the model's structure need to be determined by experiments. In the proposed network, the number

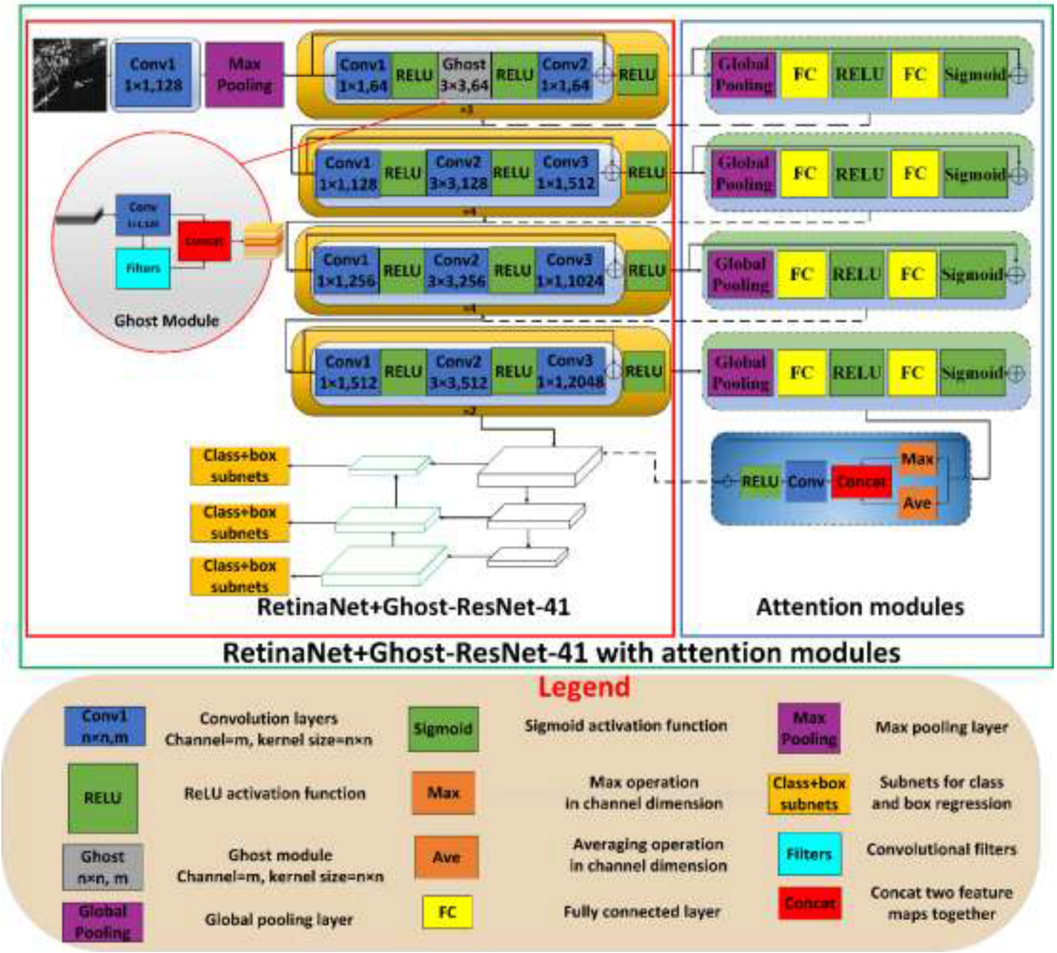


Fig. 7. Structure of the proposed model.

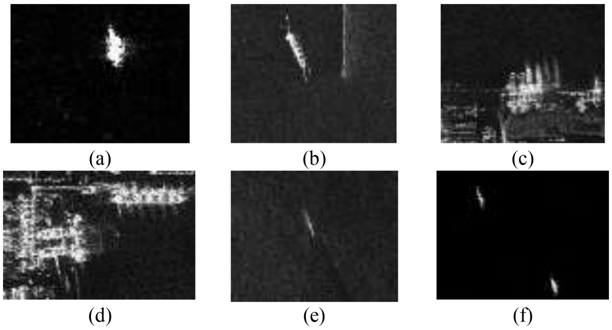


Fig. 8. Some slices of SSDD.

of deep convolution layers is reduced. A series of experiments were carried out to confirm the specific number of layers. We reduce the number of convolutional layers in the last two residual blocks of ResNet-50, and test them in the SSDD dataset. The parameters of the model are determined according to the size and AP of the model.

According to Table IV, the fourth structure gets the highest AP, its FLOPs have reduced from 4.1G to 3.9G, the sixth

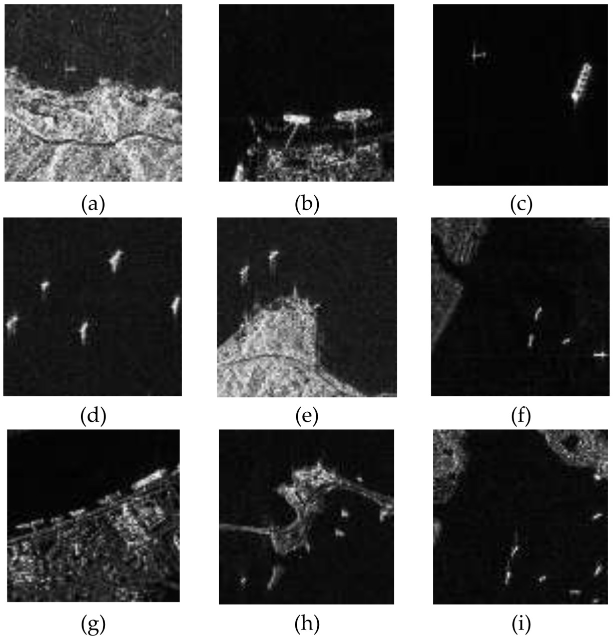


Fig. 9. Some slices of Gaofen-3 SAR ship dataset.

TABLE III
INFLUENCE OF α AND β ON TEST RESULTS

	α	β	mAP
1	0.2	1	91.3
2		2	90.8
3		5	91.4
4	0.25	1	91.5
5		2	**93.4**
6		5	91.1
7	0.3	1	91.9
8		2	93.2
9		5	92.4

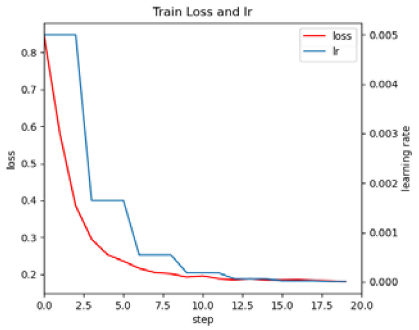


Fig. 10. Training loss and learning rate for RetinaNet+Ghost-ResNet-41 on SSDD dataset.

structure also gets high AP as 58.7, only 0.2 lower than the fourth structure, but its FLOPs has reduced from 4.1G to 3.5G. Though the eighth structure has the lowest FLOPs, its AP is also the lowest in the experiment. The fourth structure achieves high AP while realizing apparent lightweight, which reaches good balance in accuracy and size. As a result, the sixth structure is applied in the proposed model.

D. Experiment With SSDD

The proposed method is trained and tested on the SSDD dataset. We set the training epoch as 20, and the learning rate was reduced at equal intervals. The training loss and learning rate are shown in Fig. 10.

We divided the SSDD dataset into training set, validation set, and test set. The training set contains 812 slices, the validation set contains 174 slices, and the training set includes 174 slices. The trained model is tested on the test set, and some results are shown in Fig. 11.

As shown in Fig. 11, the model shows good detect ability for ships in image slices of the SSDD dataset. Accurate detection can be achieved for a single target in a simple background; for dense targets and inshore targets, most targets can be detected. To comprehensively evaluate the detection ability of the model, we select some standard detect models and compare them with the proposed models on the SSDD dataset. As shown in Table V, Faster-RCNN is a famous two-stage detection network which is proved to have high detection accuracy; SSD and YOLO-v3-spp are common single-stage detection networks, and their detection takes less time than two-stage models. We use the COCO detection evaluation method to evaluate the effects of different models [37], [38]. The results are shown in Table V.

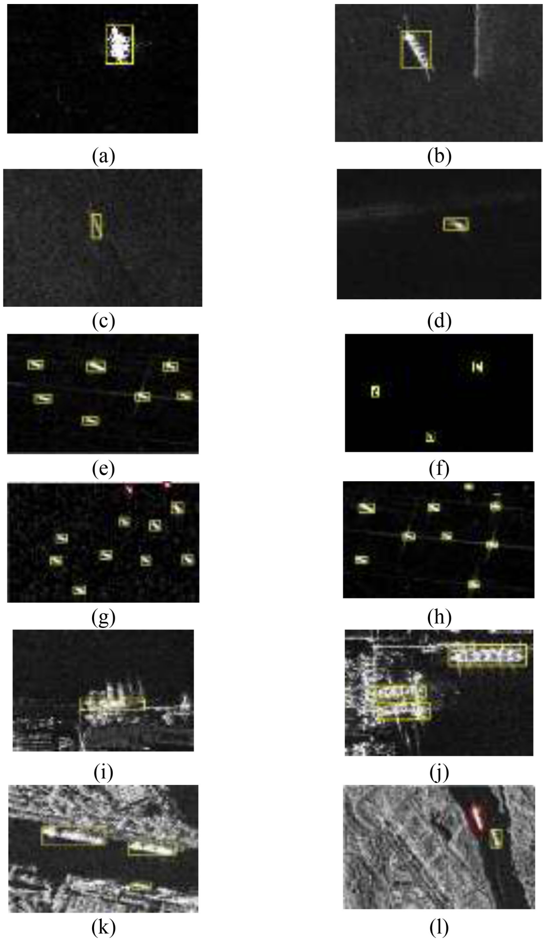


Fig. 11. Detection results of the proposed method on SSDD dataset. In which yellow boxes represent correctly detected targets, and red box represents missed target.

As the evaluated results listed in Table V, RetinaNet+Ghost-ResNet-41 achieves higher precision and recall than other one-stage models in the experiment. Though Faster-RCNN+ResNet-50/SE-ResNet-50 is higher than Reti-naNet+Ghost-ResNet-41 in AP, its recall is closed to the proposed model, and since Faster-RCNN is a two-stage model, the detect time is higher compared with the proposed model. The results also show that embedding attention modules into the proposed model can enhance its detectability. RetinaNet+Ghost-ResNet-41 with attention modules get the highest AP and AR in the experiment. Proposed models have good performance in large targets detection, while RetinaNet+Ghost-ResNet-41 is used as the backbone for RetinaNet, the AR of large target has increased from 50.1 to 52.9, after embedding attention modules, the AR of the large target has increased to 55.1, which is higher than other model's AR_{L} in the experiment. In addition, the AR of small targets, the AP of small and middle targets has also improved. Though we found that the recall rate of large-size targets decreased, the precision and the recall rate of medium- and small-size targets of the model has increased. Considering the low computational cost of additional modules, it is beneficial and acceptable to add additional modules.

TABLE IV
TEST RESULTS AND FLOPS OF DIFFERENT STRUCTURES

No.	**FLOPs**	**Conv3_x**	**Conv4_x**	**AP**	**$AP_{0.5}$**	**$AP_{0.75}$**
1	4.1G	6	3	56.9	91.0	65.2
2	3.9G	5	3	57.8	92.0	64.1
3	3.7G	4	3	56.8	91.1	65.5
4	3.9G	6	2	58.9	93.2	66.8
5	3.7G	5	2	57.8	92.4	63.4
6	3.5G	4	2	58.7	92.8	67.6
7	3.2G	3	2	57.5	91.4	66.2
8	3.0G	3	1	55.1	89.7	60.3

TABLE V
COCO DETECTION EVALUATION FOR SSDD DATASET

Model's type	**Method**	**AP**	**$AP_{0.5}$**	**$AP_{0.75}$**	**AP_S**	**AP_M**	**AP_L**	**AR_S**	**AR_M**	**AR_L**
Two-stage model	Faster-RCNN+VGG16	54.0	90.0	59.9	50.7	61.5	39.4	56.0	68.5	45.5
	Faster-RCNN+MobileNet-v2	47.2	87.5	45.9	41.0	58.4	38.4	47.8	65.1	49.0
	Faster-RCNN+ResNet-50	**58.7**	93.6	68.0	54.2	**66.7**	51.6	60.6	**72.8**	58.0
	Faster-RCNN+SE-ResNet-50	58.4	93.3	66.3	54.1	65.9	50.4	60.6	72.0	60.0
One-stage model	SSD+ResNet-50	52.4	89.4	54.9	47.3	62.2	50.2	54.8	69.1	**65.0**
	SSD+SE-ResNet-50	52.4	88.9	56.4	48.0	61.2	42.9	55.0	68.8	62.0
	YOLOv3-SPP	39.4	80.1	29.4	40.2	39.9	21.7	52.7	58.8	33.5
	RetinaNet+MobileNet-v2	46.9	87.0	45.3	41.3	57.3	37.8	48.3	64.6	46.5
	RetinaNet+ResNet-18	37.2	73.8	30.9	37.7	37.1	18.4	47.2	56.9	46.0
	RetinaNet+ResNet-50	56.9	91.0	65.2	52.8	64.5	50.1	59.7	72.0	65.0
	RetinaNet+Ghost-ResNet-41	57.7	93.2	66.4	54.3	64.7	52.9	61.5	72.1	63.0
	RetinaNet+Ghost-ResNet-41 with attention modules	**58.7**	**93.7**	**68.1**	**54.4**	66.4	**55.1**	**61.9**	**72.8**	61.5

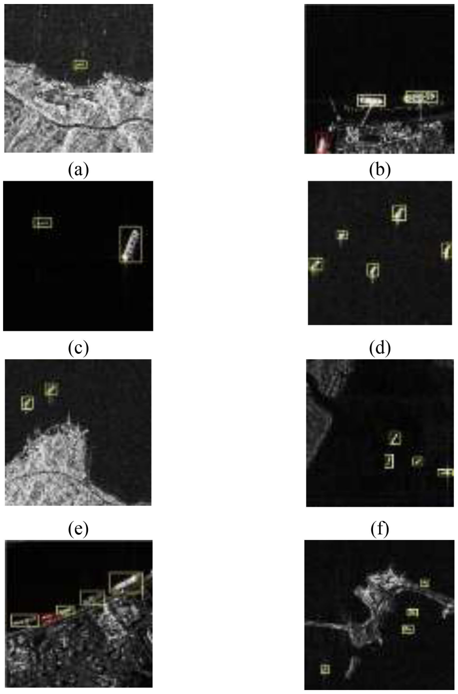
TABLE VI
MAP FOR GAOFEN-3 MINI DATASET

Model's type	**Method**	**mAP**
Two-stage model	Faster-RCNN+VGG16	84.6
	Faster-RCNN+MobileNet-v2	67.9
	Faster-RCNN+ResNet-50	89.8
	Faster-RCNN+SE-ResNet-50	90.4
One-stage model	SSD+ResNet-50	84.4
	SSD+SE-ResNet-50	85.9
	YOLO-v3-SPP	57.2
	RetinaNet+MobileNet-v2	62.5
	RetinaNet+ResNet-18	16.6
	RetinaNet+ResNet-50	88.6
	RetinaNet+Ghost-ResNet-41	91.6
	RetinaNet+Ghost-ResNet-41 with attention modules	**91.8**

E. Experiment With Small Sample Dataset

To test the model's detectability on few samples, a small dataset of Gaofen-3 was used for evaluation. In order to accurately describe the performance of the model, we evaluate models through five-fold cross validation, mAP is used as the evaluation index. The test results are shown in Fig. 12 and Table VI.

The detection results show that the proposed model has good performance. The mAP of RetinaNet+Ghost-ResNet-41 reaches 91.6, which is higher than other models in experiment. After embedding attention modules, the performance of the model has been slightly improved, the mAP has increased from 91.6 to 91.8. The experiment results demonstrate the proposed model's detection ability in small datasets.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 12. Detection results of the proposed method on Gaofen-3 mini dataset. In which yellow boxes represent correctly detected targets, and red box represents missed target.

F. Experiment on Big Scale SAR Image of Hisea−1

To test the robustness of the proposed model, a large-scale SAR image of Hisea-1 was used for detection. Hisea-1 is a commercial satellite equipped with a synthetic aperture radar,

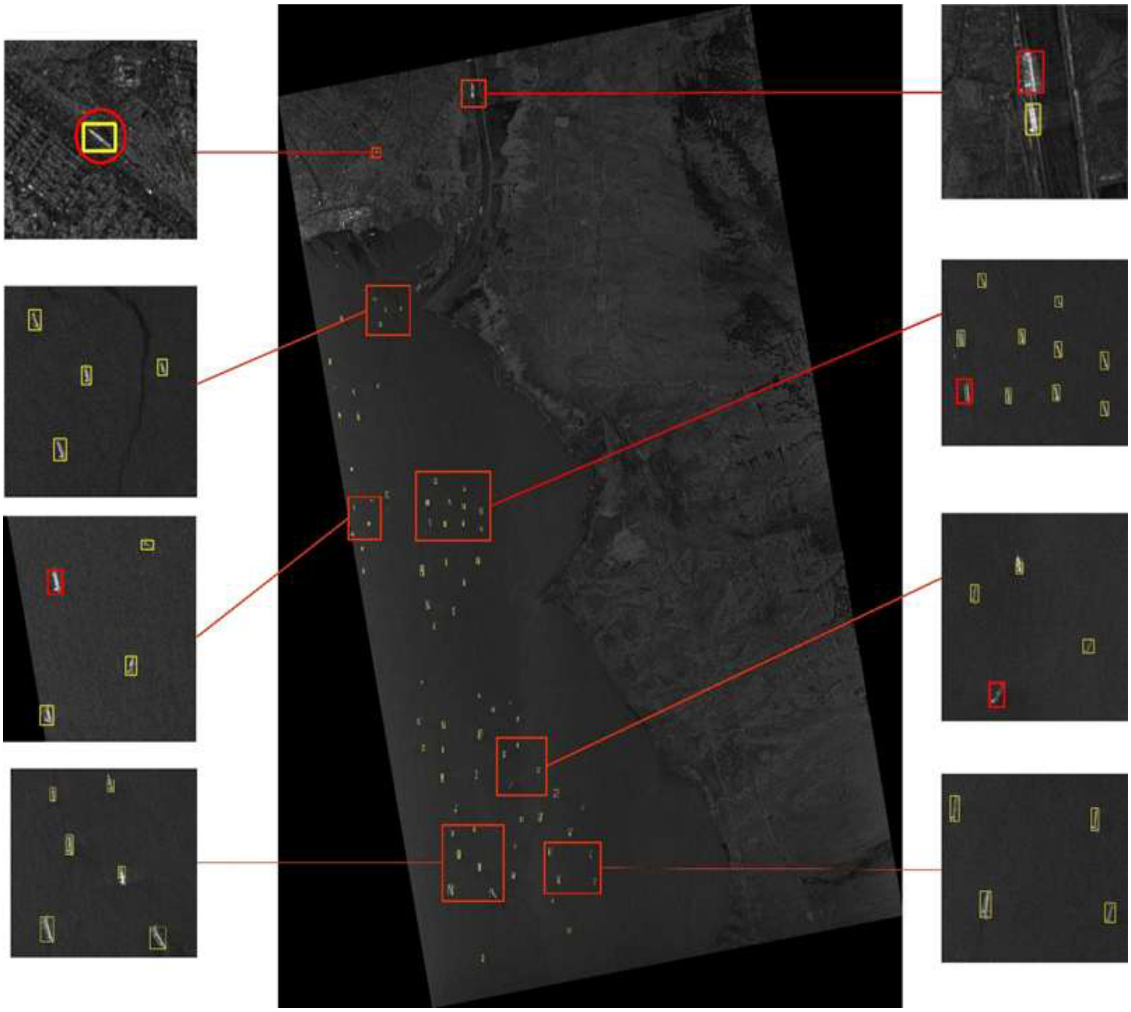


Fig. 13. Detection results of the proposed method on Hisea-1 big-scale SAR image. In which yellow boxes represent correctly detected targets, red box represents missed target and red circle represents the false alarm target.

TABLE VII
INFORMATION OF DETECTED HISEA-1 SAR IMAGE

Time	coordinates	Resolution	Size
2021-03-24	29.86°N, 32.64°E	3m	15436×30440

which works in the C band, and the highest resolution can reach 1 m. Hisea-1 can work with three standard imaging modes including Spotlight (SP), Strip-Map (SM), and ScanSAR (NS/ES) [39]. Hisea-1 SAR image in the Suez Canal area is selected for testing. The information of the image to be detected is shown in Table VII.

To detect the large-scale image of Hisea-1, the weights trained on the Gaofen-3 dataset are used to detect this image. The detection result is shown in Fig. 13 and Table VIII. The test result shows that the proposed model has good robustness and detection ability. After training on the Gaofen-3 mini dataset, the model can detect most of the targets on the SAR image of Hisea-1. The precision has reached 94.5%, and the recall has gained 80.2%. According to the test result, most missed targets are inshore ships and small vessels. The model has also detected some false ships, which are usually bright targets on the ground. As shown in the small picture on the upper left, a bright spot on the ground was incorrectly detected as a boat in the river. This part of false targets can be suppressed after using sea and land segmentation.

G. Ablation Experiment

To study the effect of each module on the improvement of model detection ability, we performed ablation experiments on SSDD dataset. In this part, we mainly carry out experiments on channel attention module, spatial attention module, and K-means clustering method. The results are shown in Table IX.

According to Table IX, these three modules have different degrees of positive impact on the model. After adding K-means algorithm, spatial attention module, and channel attention module to the basic backbone, respectively, the detection ability of the model will be slightly improved. Taking AP as an example, when three modules are used, AP is improved by 0.1, 0.5, and 0.2, respectively. However, when they are combined in pairs, AP will achieve great improvement. For example, after embedding the channel and spatial attention modules into the basic model, the AP has increased from 57.6 to 58.4. For other indicators, most

TABLE VIII
TEST RESULT OF HISEA-1 LARGE-SCALE SAR IMAGE

Detected Ships	True Ships	False Ships	Missed Ships	Precision	Recall
73	69	4	17	94.5%	80.2%

TABLE IX
ABLATION STUDY FOR SSDD DATASET

	Channel Attention Modules	Spatial Attention Modules	K-Means Algorithm	AP	$AP_{0.5}$	$AP_{0.75}$	AP_S	AP_M	AP_L	AR_S	AR_M	AR_L
1	×	×	×	57.6	91.9	65.9	53.1	65.8	48.9	60.0	71.9	63.0
2	×	×	√	57.7	93.2	66.4	54.3	64.7	52.9	61.5	72.1	63.0
3	×	√	×	58.1	93.2	66.0	53.9	65.5	53.0	61.2	71.7	64.5
4	×	√	√	58.2	93.6	65.5	53.9	65.8	51.7	61.0	71.6	66.0
5	√	×	×	57.8	92.3	66.5	53.7	65.7	52.0	61.1	73.2	63.0
6	√	×	√	58.0	92.2	67.3	53.7	66.0	53.8	61.3	73.2	63.5
7	√	√	×	58.4	93.6	65.6	54.2	65.9	50.1	61.5	72.0	62.0
8	√	√	√	**58.7**	**93.7**	**68.1**	**54.4**	**66.4**	**55.1**	**61.9**	72.8	61.5

TABLE X
SIZE OF DIFFERENT BACKBONES

Backbone	Parameters/M	MAcc/G	FLOPs/G
VGG11	132.9	15.2	7.6
VGG13	133.0	22.6	11.3
VGG16	138.4	31.0	15.5
VGG19	143.7	39.3	19.7
DenseNet161	28.7	15.6	7.8
DenseNet169	14.1	6.8	3.4
DenseNet201	20.0	8.7	4.4
MobileNet-v2	3.4	0.3	0.2
ResNet-18	11.7	3.6	1.8
ResNet-34	21.8	7.3	3.7
ResNet-50	25.6	8.2	4.1
ResNet-101	44.5	15.7	7.8
Ghost-ResNet-41	**18.8**	**6.6**	**3.3**

have been improved. Although some indicators have declined (such as AR_M, AR_L), the detection ability of the model has been improved in general.

H. Model Analysis for Different Backbones

To measure and compare the complexity of the models, some common backbones are selected and tested. We measure the complexity of the models by the number of parameters, multiply-accumulate operations (MAcc), and FLOPs. The analysis results are as follows.

As shown in Table X, compared with standard ResNet-50, Ghost ResNet-41 has fewer parameters and requires fewer computing resources. It has 18.8G parameters, 26.56% less than ResNet-50. The FLOPs of Ghost-ResNet-41 is 3.3G, which is 20% less than ResNet-50. Compared with other mainstream backbones like VGG and DenseNet, the proposed backbone has advantages in lightweight.

V. CONCLUSION

An improved ship detection network for SAR images based on RetinaNet is proposed. To reduce the parameter of the model and realize the lightweight design, the shallow 3 × 3 convolutional kernels in the backbone are replaced by ghost modules in the improved RetinaNet, in which the group convolutional filters are performed to generate new feature maps based on the initial feature maps produced by the normal convolutional layers. Then, the number of deep convolutional layers has been reduced to avoid overfitting in complex networks for SAR images. To further improve the detection ability of the model, channel attention modules are added after each residual block, and a spatial attention module is embedded at the end of the backbone. Besides, the *K*-means clustering algorithm is applied to the training set to make the model adapt to the aspect ratio of the target in advance. Compared with common backbones, the backbone of the proposed model is lighter, and FLOPs and the number of parameters is 25% lower than ResNet-50.

SSDD dataset and a small dataset of Gaofen-3 are used to test the effectiveness of the proposed method, the experimental results show that the proposed method can improve the detection ability of the model while reducing the parameter of the model. A large-scale SAR image of Hisea-1 is used to test the robustness of the model. The proposed model achieves good detection results using the weight trained on the Gaofen-3 dataset, which shows the robustness and the potential for portability of the model. In the future, the model can be trained with existing data and transplanted to the satellite for real-time detection, which can eliminate the training on the satellite.

Even though the proposed model shows good detection ability for ships by SAR images in experiments, it limitations exists since the proposed method pursues the lightweight of the model and does not segment the sea and land before detection to suppress the land false alarm, the model may mistakenly capture bright nonship targets on land. Besides, due to the need for a large number of IO operations, although it can reach a good standard on flops, the speed advantage of depth wise convolution on GPU has not been reflected. In GPU platform, lightweight is mainly reflected in the decrease of parameters.

In the future, studies are continued to focus on the lightweight of the model, to overcome the shortcomings of the existing model. The next step of follow-up research directions are as follows.

1) To improve the model's ability detecting ships in rivers or inshore vessels and avoid false detection of targets on the shore, sea and land segmentation will be considered before detection. The traditional threshold segmentation method and the methods based on deep learning will be further studies separately. Multisource information will also be considered to help to extract water body.
2) Lightweight design will be continued to carried out. We will avoid unnecessary computing resource consumption, improve and replace the standard convolutional layers, and reduce the number of model parameters without reducing the detection ability.
3) The existing backbone will be further improved to enhance its feature extraction and generalization capabilities, texture information and reflection in-formation of SAR images will also be considered to improve the model.

Acknowledgment

The authors would like to thank Prof. M. Migliaccio and Prof. A. Buono from Università di Napoli Parthenope for their precious revisions and suggestions for this article.

References

[1] I. Cumming and F. Wong, *Digital Processing of Synthetic Aperture Radar Data*. Norwood, MA, USA: Artech House, 2005, pp. 108–110.

[2] A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, "A tutorial on synthetic aperture radar," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 1, pp. 6–43, Mar. 2013.

[3] A. Gambardella, F. Nunziata, and M. Migliaccio, "A physical full-resolution SAR ship detection filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 760–763, Oct. 2008.

[4] G. Ferrara, M. Migliaccio, F. Nunziata, and A. Sorrentino, "GK-based observation of metallic targets at sea in full-resolution SAR data: A multipolarization study," *IEEE J. Ocean. Eng.*, vol. 36, no. 2, pp. 195–204, May 2011.

[5] X. Cui, Y. Su, and S. Chen, "A saliency detector for polarimetric SAR ship detection using similarity test," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3423–3433, Sep. 2019.

[6] F. Nunziata, A. Montuori, and M. Migliaccio, "Dual-polarized COSMO skymed SAR data to observe metallic targets at sea," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2011, pp. 2270–2273.

[7] M. di Bisceglie and C. Galdi, "CFAR detection of extended objects in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 833–843, Apr. 2005.

[8] W. An, C. Xie, and X. Yuan, "An improved iterative censoring scheme for CFAR ship detection with SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4585–4595, Aug. 2014.

[9] X. Leng, K. Ji, K. Yang, and H. Zou, "A bilateral CFAR algorithm for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1536–1540, Jul. 2015.

[10] H. Dai, L. Du, Y. Wang, and Z. Wang, "A modified CFAR algorithm based on object proposals for ship target detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1925–1929, Dec. 2016.

[11] W. Ao, F. Xu, Y. Li, and H. Wang, "Detection and discrimination of ship targets in complex background from spaceborne ALOS-2 SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 2, pp. 536–550, Feb. 2018.

[12] L. Zhai, Y. Li, and Y. Su, "Inshore ship detection via saliency and context information in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1870–1874, Dec. 2016.

[13] H. Zhao, Q. Wang, J. Huang, W. Wu, and N. Yuan, "Method for inshore ship detection based on feature recognition and adaptive background window," *J. Appl. Remote Sens.*, vol. 8, no. 1, 2014, Art. no. 083608.1.

[14] M. Liao, C. Wang, Y. Wang, and L. Jiang, "Using SAR images to detect ships from sea clutter," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 194–198, Apr. 2008.

[15] F. Lang, J. Yang, S. Yan, and F. Qin, "Superpixel segmentation of polarimetric synthetic aperture radar (SAR) images based on generalized mean shift," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1592.

[16] R. Jin, J. Yin, W. Zhou, and J. Yang, "Level set segmentation algorithm for high-resolution polarimetric SAR images based on a heterogeneous clutter model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 10, pp. 4565–4579, Oct. 2017.

[17] M. Ciecholewski, "River channel segmentation in polarimetric SAR images: Watershed transform combined with average contrast maximization," *Expert Syst. Appl.*, vol. 82, pp. 196–215, 2017.

[18] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 580–587.

[19] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. SAR Big Data Era: Models, Methods Appl.*, 2017, pp. 1–6.

[20] X. Hou and F. Xu, "Inshore ship detection based on multi-aspect information in high-resolution SAR images," in *Proc. Asia-Pac. Conf. Synthetic Aperture Radar*, 2019, pp. 1–4.

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

[22] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "Automatic ship detection based on retinanet using multi-resolution Gaofen-3 imagery," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 531.

[23] T. Zhang, X. Zhang, and X. Ke, "Quad-FPN: A novel quad feature pyramid network for SAR ship detection," *Remote Sens.*,vol. 13, no. 14, 2021, Art. no. 2771.

[24] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 1577–1586, doi: 10.1109/CVPR42600.2020.00165.

[25] A. Howard *et al.*, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv 1704.04861*.

[26] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[27] M. Zhu *et al.*, "Arbitrary-oriented ship detection based on retinanet for remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens .*, vol. 14, pp. 6694–6706, 2021.

[28] H. Su, S. Wei, M. Wang, L. Zhou, J. Shi, and X. Zhang, "Ship detection based on retinanet-Plus for high-resolution SAR imagery," in *Proc. Asia-Pac. Conf. Synthetic Aperture Radar*, 2019, pp. 1–5.

[29] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[30] J. Yang, Y. Zhu, B. Jiang, L. Gao, L. Xiao, and Z. Zheng, "Aircraft detection in remote sensing images based on a deep residual network and super-vector coding," *Remote Sens. Lett.*, vol. 9, no. 3, pp. 228–236, 2018.

[31] A. Bochkovskiy, C. Wang, and H. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020, *arXiv 2004.10934*.

[32] L. Chen *et al.*, "A new framework for automatic airports extraction from SAR images using multi-level dual attention mechanism," *Remote Sens.*, vol. 12, no. 3, 2020, Art. no. 560.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv: 1409.1556*.

[34] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[35] S. Woo, J. Park, J. Lee, and I. Kweon, "Cbam: Convolutional block attention module," in *Proc. Eur. Conf. Comput.*, 2018, pp. 3–19.

[36] A. Farhadi and J. Redmon, "Yolov3: An incremental improvement," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, *arXiv,1804.02767*.

[37] [Online]. Available: https://cocodataset.org/#detection-eval

[38] Z. Hong *et al.*, "Multi-scale ship detection from SAR and optical imagery via a more accurate YOLOv3," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6083–6101, 2021.

[39] P. Xu *et al.*, "On-board real-time ship detection in HISEA-1 SAR images based on CFAR and lightweight deep learning," *Remote Sens.*, vol. 13, no. 10, 2021, Art. no. 1995.

Tian Miao was born in 1997. He received the bachelor's degree in information and communication engineering from Beihang University, Beijing, China, in 2020. He is currently working toward the master's degree with Beihang University, majoring in information and signal processing.

His present research interests include SAR image processing.

HongCheng Zeng (Member, IEEE) was born in 1989. He received the Ph.D. degree in signal and information processing from Beihang University, Beijing, China, in 2016.

Since 2019, he has been an Assistant Professor with the School of Electronics and Information Engineering, Beihang University. He was a Visiting Researcher with the School of Mathematics and Statistics, University of Sheffield, Sheffield, U.K., from 2017 to 2018. He has published more than 20 journal and conference papers, His research interests include high-resolution spaceborne SAR image formation, passive radar signal processing, and moving target detection.

Wei Yang was born in 1983. He received the M.S. and Ph.D. degrees in signal and information processing from Beihang University (BUAA), Beijing, China, in 2008 and 2011, respectively.

From 2011 to 2013, he held a Postdoctoral position with the School of Electronics and Information Engineering, Beihang University. Since July 2013, he has been with the School of Electronics and Information Engineering, BUAA, as a Lecturer. From 2016 to 2017, he researched as a Visiting Researcher with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K. He has been an Associate Professor with the School of Electronics and Information Engineering, BUAA, since 2018. He has authored or coauthored more than 60 journal and conference publications. His research interests include moving target detection, high-resolution spaceborne SAR image formation, SAR image quality improvement, and 3-D imaging.

Boce Chu was born in Xingtai, China, in 1991. He received the master's degree in information and communication engineering from the Beijing University of Posts and Telecommunications, Beijing, China, in 2016. He is currently working toward the Ph.D. degree in Beihang University, Beijing, majoring in information and signal processing.

His present research interests include algorithms for remote sensing image processing based on artificial intelligence.

Fei Zou received the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2012.

Since 2017, she has been an Associate Researcher with the Beijing Institute of Remote Sensing Informatin, Beijing, China. She has published more than ten journal and conference papers. Her research interests include inverse SAR imaging, SAR image quality improving, and remote sensing application.

Weijia Ren received the graduate degree from the Department of Mechanical Engineering, Tsinghua University, Beijing, China, and the Ph.D. degree in space physics from the Graduate School of the Chinese Academy of Science, Beijing.

He conducted research at the University of Southampton, U.K., as a Visiting Scholar for one year. He held positions of Senior Engineer and Research Scientist at the Technology and Engineering Center for Space Utilization, Chinese Academy of Science. He has been the CTO of Spacety Company, Ltd., Changsha, China, since 2016. Under his leadership, Spicery has developed and successfully launched and operated 21 satellites, including "Xiaoxiang-1," the first satellite developed by a private space company in China, and "HiSea-1," China's first commercial miniaturized foldable SAR satellite.

Jie Chen (Senior Member, IEEE) was born in 1973. He received the B.S. and Ph.D. degrees in information and communication engineering from Beishan University, Beijing, China, in 1996 and 2002, respectively.

Since 2004, he has been an Associate Professor with the School of Electronics and Information Engineering, Beijing University. He was a Visiting Researcher with the School of Mathematics and Statistics, University of Sheffield, Sheffield, U.K., from 2009 to 2010, working on ionospheric effects on low-frequency space radars that measure forest biomass and ionospheric electron densities. Since July 2011, he has been a Professor with the School of Electronics and Information Engineering, Beishan University. His research interests include multimodal remote sensing data fusion, and high-resolution spaceborne synthetic aperture radar (SAR) image formation and SAR image quality enhancement.