Dense Attention Pyramid Networks for Multi-Scale Ship Detection in SAR Images

Zongyong Cui, *Member, IEEE*, Qi Li, *Student Member, IEEE*, Zongjie Cao, *Member, IEEE*, and Nengyuan Liu, *Student Member, IEEE*

***Abstract*— Synthetic aperture radar (SAR) is an active microwave imaging sensor with the capability of working in all-weather, all-day to provide high-resolution SAR images. Recently, SAR images have been widely used in civilian and military fields, such as ship detection. The scales of different ships vary in SAR images, especially for small-scale ships, which only occupy few pixels and have lower contrast. Compared with large-scale ships, the current ship detection methods are insensitive to small-scale ships. Therefore, the ship detection methods are facing difficulties with multi-scale ship detection in SAR images. A novel multi-scale ship detection method based on a dense attention pyramid network (DAPN) in SAR images is proposed in this paper. The DAPN adopts a pyramid structure, which densely connects convolutional block attention module (CBAM) to each concatenated feature map from top to bottom of the pyramid network. In this way, abundant features containing resolution and semantic information are extracted for multi-scale ship detection while refining concatenated feature maps to highlight salient features for specific scales by CBAM. Then, the salient features are integrated with global unblurred features to improve accuracy effectively in SAR images. Finally, the fused feature maps are fed to the detection network to obtain the final detection results. Experiments on the data set of SAR ship detection data set (SSDD) including multi-scale ships in various SAR images show that the proposed method can detect multi-scale ships in different scenes of SAR images with extremely high accuracy and outperforms other ship detection methods implemented on SSDD.**

***Index Terms*— Dense attention pyramid network (DAPN), multi-scale feature maps, ship detection, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) is a high-resolution imaging radar. As an active microwave imaging sensor, an SAR imaging process is less affected by environmental factors, such as weather, light, and clouds, than the traditional passive imaging sensors, such as infrared and optical sensors. It has the capability to detect the hidden objects and work in all-weather and all-day conditions. SAR has been extensively used in civilian and military fields with the rapid development of spaceborne and airborne SAR, for instance, TerraSAR-X, RADARSAT-2, Sentinel-1, and Gaofen-3. Studies on ship detection are vital in many aspects [1], such as ocean monitoring, maritime management, and military intelligence acquisition. In recent years, ship detection in SAR images has become a hot spot all over the world [2], [3]. However, there are rare methods suitable for multi-scale ship detection currently, and meanwhile, detecting offshore ships is still a tough task. Therefore, this paper focuses on a novel multi-scale ship detection method for both inshore and offshore scenes of SAR images.

Traditional ship detection methods in SAR images mainly focus on the detection of ship targets and ship wakes. Since ship wakes do not exist all the time, and their features are not as obvious as ship targets, the research on the detection of ship wakes [4] is not extensive. For ship target detection, there are three main methods, including detection algorithms based on statistical characteristics [5]–[7], threshold [8], and transformation [9], [10]. Among the above-mentioned detection methods, a constant false alarm rate (CFAR) [11] detection algorithm and its improved algorithms [12]–[14] are the most widely studied and applied. The previous ship detection methods mainly paid attention to the strong scattering targets. However, the target regions occupy fewer pixels and contain some weak scatters in the large-scene SAR images with increasing resolution. It is difficult for conventional methods of single-polarization SAR to achieve real-time and accurate ship detection.

Compared with single-polarization SAR, polarimetric SAR (PolSAR) can provide more information from the scattering matrix, including target structure and regions of motion [15], thus improving the detection accuracy for SAR images. Moreover, there is a shift toward research on algorithms that make use of PolSAR data [16]. Common ship detection algorithms for PolSAR images mainly consist of polarization target decomposition [17], [18], statistical theory method [19], and machine learning method [20]. In recent years, many state-of-the-art ship detection methods have been proposed. Gao *et al.* [21] put forward a novel decomposition approach and combine the CFAR detector based on the generalized gamma distribution to improve the

Manuscript received February 28, 2019; revised May 16, 2019; accepted June 14, 2019. Date of publication July 15, 2019; date of current version October 31, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61801098, in part by the Fundamental Research Funds for the Central Universities under Grant 2672018ZYGX2018J013, and in part by the Shanghai Aerospace Science and Technology Innovation Fund under Grant SAST2018-079. *(Corresponding author: Zongjie Cao.)*

Z. Cui, Q. Li, and N. Liu are with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zycui@uestc.edu.cn; lucialee0103@gmail.com; nengyuanliu@outlook.com).

Z. Cao is with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Center for Information Geoscience, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: zjcao@uestc.edu.cn).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2923988

signal-to-clutter ratio. In [22], variational Bayesian inference is adopted to detect ships in congested sea areas, which has good detection performance and shape-preserving ability. He *et al.* [23] come up with an approach using three superpixel-level dissimilarity measures to detect ships automatically. It is anticipated that the ship detection algorithms of PolSAR will become more unsupervised and the false alarm targets will reduce. Although the ship detection methods of PolSAR may have better performance than single-polarization SAR, it is also important to explore how to use the limited information of single-polarization SAR to achieve higher detection accuracy. Therefore, the method proposed in this paper focuses on ship detection in single-polarization SAR images.

Recently, the convolutional neural network (CNN) has achieved remarkable success in the field of computer vision [24]–[26] by powerful representation ability and the characteristics of automatic feature extraction. The object detectors based on CNNs can be divided into two kinds of methods. The one is the one-stage method, such as you only look once (YOLO) [27] and single-shot multi-box detector (SSD) [28], which are fast and simple to train. The other is the two-stage method, such as region CNN (R-CNN) [29] and fast R-CNN [30], having high accuracy but heavy computational cost. Ren *et al.* [31] put forward faster R-CNN, which adopts the trainable region proposal network (RPN) to improve detection efficiency while maintaining high accuracy. In general, the two-stage algorithms have higher accuracy, as the calculational cost is also larger. On the contrary, the one-stage algorithms have the characteristics of fast and simple training with the end-to-end mode, but the accuracy is lower than the two-stage algorithms. However, these methods only use the final-level feature map to predict objects while losing features of objects from lower layers, thus making it difficult to detect small targets. In order to make full use of multi-level features, feature pyramid network (FPN) [32] is proposed to extract features from different layers for individual prediction to achieve a better detection performance. Woo *et al.* [33] came up with an attention module that combined the spatial and channel attentions to refine the feature map adaptively.

Ship detection methods in SAR images based on CNNs [34], [35] have also shown superior detection performance. In the early stage, CNN was adopted in various parts of ship detection, such as land masking [36], region of interest (ROI) extraction, and ship targets identification. At the present stage, most of the ship detection algorithms based on CNNs utilize end-to-end mode to send SAR images into CNNs for detection and output ship detection results directly, omitting operations of preprocessing and sea-land segmentation, which improves efficiency and accuracy of detection. For instance, SSD-combined transfer learning [37] was proposed to solve the problem of insufficient SAR training set. Li *et al.* [38] put forward an improved faster R-CNN method and provided a data set called SAR ship detection data set (SSDD) to train and test a model. Jiao *et al.* [39] combined FPN and faster R-CNN to detect multi-scale ships. Wang *et al.* [40] improved SSD with semantic aggregation module and attention module to detect ship and estimate orientation simultaneously. The fusion of ship context information in the feature maps of SSD [41] can improve the accuracy of inshore ships detection.

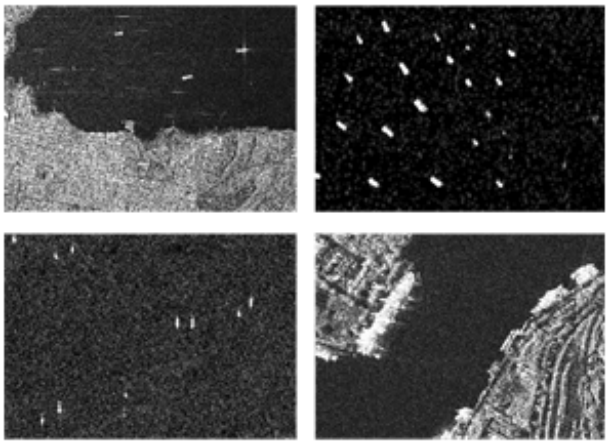


Fig. 1. Examples of multi-scale ships in multi-scene SAR images.

However, these methods pay more attention to salient objects, such as large-scale ships in the offshore scene of SAR images. Moreover, due to the inference of complex background, multi-scale ships in the inshore scene of SAR images are more difficult to detect. As shown in Fig. 1, the prominent objects are those large-scale ships and complex background, and therefore, small-scale ships are easy to be ignored.

Low-level feature maps with higher resolution but shallow features are suitable for detecting small-scale ships, whereas high-level feature maps, with more semantic information but relatively lower resolution, are fit to large-scale ships. In this paper, the ship detection method based on a dense attention pyramid network (DAPN) is proposed.

First, a novel pyramid structure is adopted to extract multi-scale feature maps with higher resolution and more semantic information for multi-scale ship detection.

Next, the abundant multi-scale feature maps are refined adaptively to obtain salient features for specific scales by densely connecting convolutional block attention module (CBAM) to each concatenated feature map from top to bottom of the pyramid network. By this means, multi-scale feature maps are fused by CBAM, adaptively selecting significant scale features and paying more attention to specific scales for multi-scale ship detection. Since up-sampling and CBAM operations can blur target features, only using salient features for detection will result in missed detection.

Finally, the fused feature maps are obtained by integrating salient features from layers higher than the current layer in the top-down network with unblurred features from the corresponding layer in the bottom-up network for subsequent detection. In this way, the proposed method improves the accuracy of multi-scale ship detection in different scenes of SAR images, avoiding the inference of complex backgrounds. Experiments on SSDD demonstrate that the proposed method achieves a better detection performance with extremely high accuracy for multi-scale ship detection in SAR images.

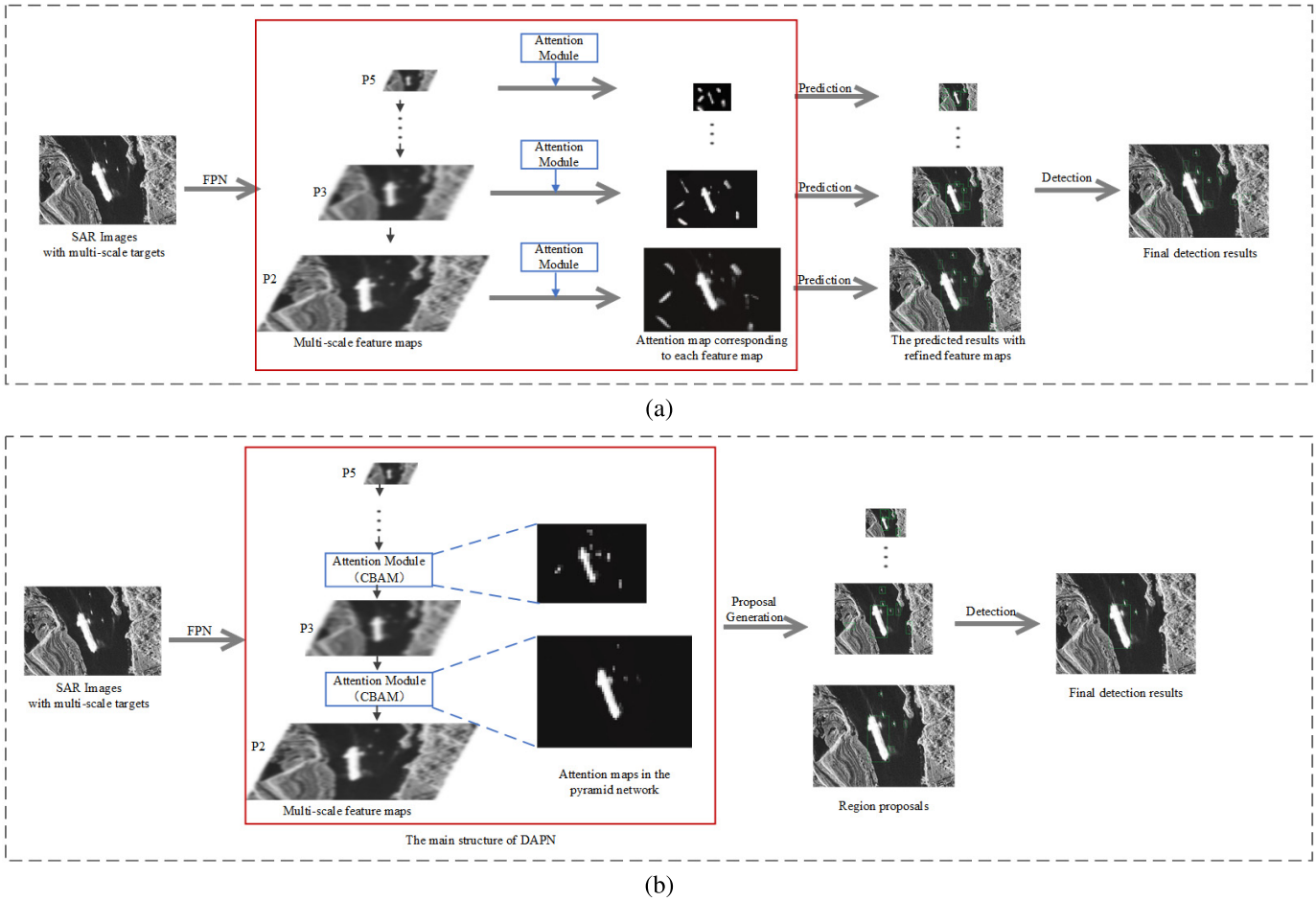


Fig. 2. Flow charts of [40] and the proposed method. (a) Core idea of [40], which adopts the attention module to each feature map separately. (b) Core idea of the proposed method, which embeds CBAM into the FPN.

The rest of this paper is organized as follows. Section II introduces the proposed method based on the DAPN. In Section III, the details of experiments on SSDD and the analysis of the results are exhibited. Finally, the conclusion is given in Section IV.

II. Methodology

In this part, the details of the proposed method will be introduced. At first, by analyzing the defects of existing methods, the key points and the overall architecture of the proposed method are derived. Next, every branch of the proposed architecture is described in detail to show how it works, and the introduction of the loss function is given at the end of this section.

A. Ideas of the Proposed Method

It is vital for multi-scale ship detection to make full use of features from different layers to obtain more spatial and semantic information. Therefore, a pyramid structure with dense connections in the top-down network is adopted in this paper. Compared with dense connection with element-wise addition in [39], the dense connections of the proposed method use the operation of concatenation, which preserves the features of multi-scale fused feature maps to a greater extent. In addition, although the method in [39] utilizes dense connections to fuse feature maps of different scales for multi-scale ship detection, it lacks the process of adaptive feature selection at specific scales. Since some targets have disappeared at different scales, it will interfere with detection results. CBAM is utilized to weight feature maps of different scales, highlighting significant features of specific scales. In this way, it is more targeted for ships of a specific scale.

In [33], CBAM is proposed for feature maps at the same scales, while CBAM is applied to the fusion of multi-scale feature maps in this paper, paying more attention to scale features of targets. CBAM of the proposed method utilizes channel and spatial attentions to weight in dimensions of scale and spatial, respectively. Channel attention focuses on selecting feature maps of suitable scales, and spatial attention concentrates on salient regions containing targets. Therefore, multi-scale ships are detected more effectively with CBAM.

A similar attention module is adopted in [40], which is also for feature maps at the same scales. However, the method in [40] adopted the attention module to each feature map separately, as shown in Fig. 2(a), which caused false alarms generated by the feature map of the higher layer not to be eliminated, resulting in false alarms being transmitted

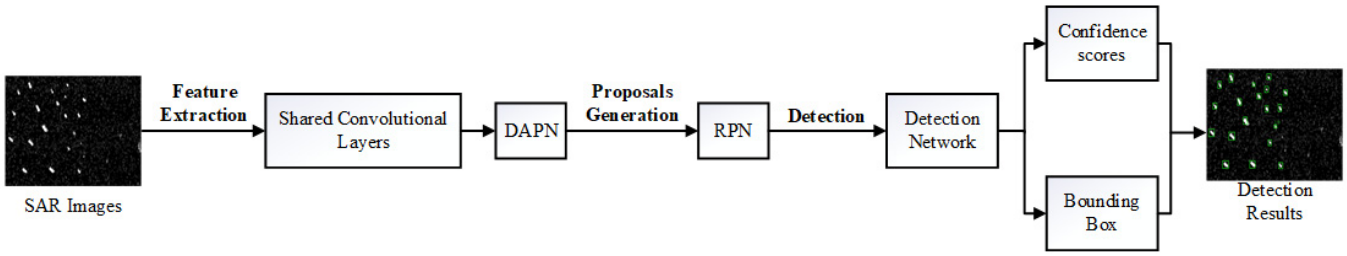


Fig. 3. Flowchart of the proposed method.

layer by layer from the higher level feature map. Therefore, there may be more false alarms and lower accuracy obtained.

In contrast, the proposed method embeds CBAM into the FPN to construct DAPN, as shown in Fig. 2(b). In this way, the attention module can refine feature maps layer by layer so that false alarms of the higher level feature map will be reduced in the lower level feature maps through the attention module. Finally, false alarms will be gradually eliminated, greatly improving the accuracy of multi-scale ship detection.

B. Detailed Description of Network Architecture

According to the above-mentioned ideas, the overall flowchart of the proposed method is shown in Fig. 3, which is mainly composed of DAPN, RPN, and detection network. DAPN is utilized to extract multi-scale fused feature maps for subsequent proposals generation and detection. RPN is adopted to generate proposals, which may contain targets, and then, false alarms are eliminated to obtain final detection results in the detection network.

At first, CBAM, the key component of DAPN used to refine multi-scale feature maps, will be illustrated. Then, each part of the proposed method in Fig. 3 will be described detailed.

1) Convolutional Block Attention Module: There are two main purposes for putting CBAM in the process of feature extraction. First, the refined feature maps by CBAM from higher layers in the top-down network are densely connected to each other to obtain concatenated feature maps of lower layers, so it plays a role in feature enhancement. Second, global unblurred features are merged with local salient features of different scales by lateral connections to optimize fused feature maps for multi-scale targets detection in different scenes.

In this paper, CBAM is adopted to highlight the significant features of specific scales from a huge amount of spatial and channel aggregation information of multi-level concatenated feature maps in the top-down network of DAPN. In this way, it can refine the multi-scale feature maps adaptively, which benefits to reduce false alarms and improve the accuracy of multi-scale ship detection.

It is composed of channel attention and spatial attention, which obtains attention maps along two dimensions separately. As shown in Fig. 4, the input feature map $\mathbf{F} \in \mathrm{R}^{\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$ is multiplied by the channel attention map $\mathbf{A_C} \in \mathrm{R}^{\mathrm{C}\times 1\times 1}$ and the spatial attention map $\mathbf{A_S} \in \mathrm{R}^{1\times\mathrm{H}\times\mathrm{W}}$ to obtain a salient feature map $\mathbf{F_R} \in \mathrm{R}^{\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$. The overall process of CBAM is as follows:

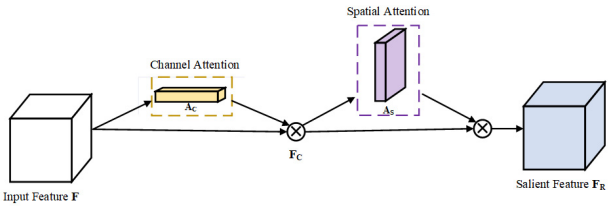


Fig. 4. Structure of CBAM.

$$\mathbf{F_C} = \mathbf{A_C}(\mathbf{F}) \otimes \mathbf{F} \tag{1}$$

$$\mathbf{F_R} = \mathbf{A_S}(\mathbf{F_C}) \otimes \mathbf{F_C} \tag{2}$$

where $\mathbf{F_C} \in \mathrm{R}^{\mathrm{C}\times\mathrm{H}\times\mathrm{W}}$ is the feature map obtained by channel attention and $\otimes$ represents the element-wise multiplication.

The channel attention module mainly focuses on the specific significant contents of input feature maps. It aggregates the spatial information through max-pooling operation and average-pooling operation simultaneously and then reduces the parameters by multi-layer perceptron (MLP). The computation process is summarized as follows:

$$\mathbf{A_C}(\mathbf{F}) = \sigma\left(MLP(AvgPool(\mathbf{F})) + MLP(MaxPool(\mathbf{F}))\right) \tag{3}$$

where AvgPool and MaxPool represent the operations of average pooling and max pooling, respectively, and σ is the sigmoid function.

The spatial attention module pays more attention to the specific locations of salient features. It first applies max-pooling operation and average-pooling operation along the channel axis, respectively, and concatenates the results. Then, a 7×7 convolutional layer is utilized to the concatenated feature maps to generate the spatial attention map. The spatial attention is computed as follows:

$$\mathbf{A_S}(\mathbf{F_C}) = \sigma\left(\mathrm{Conv}_{7\times 7}([\mathrm{AvgPool}(\mathbf{F_C}); \mathrm{MaxPool}(\mathbf{F_C})])\right) \tag{4}$$

where $\mathrm{Conv}_{7\times 7}$ represents the 7×7 convolutional layer.

CBAM improves the representation capability of input feature maps by channel and spatial attentions sequentially. Channel attention is applied globally, and spatial attention concentrates on local information. Therefore, the combination of these two attention modules will extract salient features effectively.

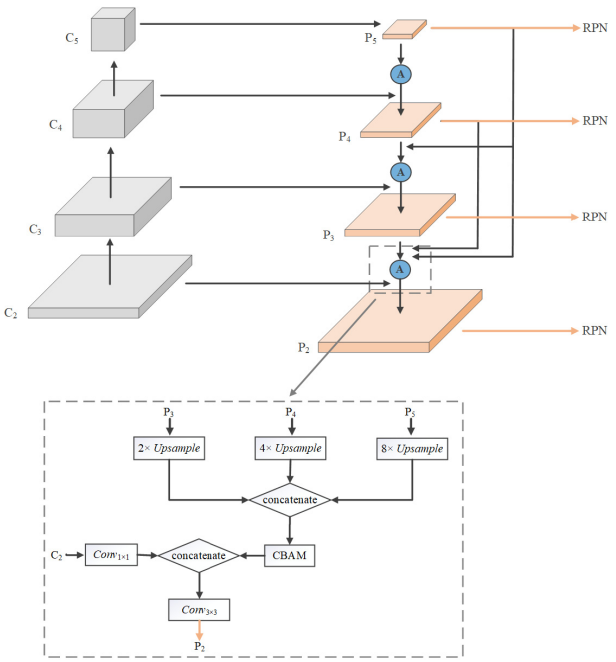


Fig. 5. Architecture of DAPN.

For multi-scale ship detection, CBAM can highlight salient feature maps of specific scales and extract suitable scale features adaptively according to the scales of ships to be detected by weighting multi-scale feature maps. In this paper, CBAM is densely connected to multi-level concatenated feature maps in the top-down network of DAPN, which emphasizes significant information of feature maps at specific scales corresponding to different scales of detected ships. This can effectively eliminate false alarms in various scenes of SAR images, such as ground objects in inshore scenes and speckle noise similar to small-scale ships in offshore scenes.

2) Dense Attention Pyramid Network: High-level feature maps have large receptive fields with more semantic information, and less spatial information is suitable for detecting large-scale ships. On the contrary, the low-level feature maps, with more spatial information but relatively shallow features, possessing small receptive fields are fit to small-scale ships detection. Hence, it is critical for the multi-scale ship detection to integrate the semantic and spatial information effectively. In addition, it is necessary to extract salient information for specific scales in massive features to highlight ship targets of different scales and eliminate the inference of complex backgrounds, which achieves better detection for multi-scale ship detection in various scenes of SAR images. Inspired by [23], the pyramid architecture is adopted in DAPN to extract multi-scale and salient features containing both semantic and spatial information to detect multi-scale ships in SAR images.

As shown in Fig. 5, the DAPN consists of the bottom-up feedforward network, lateral connections, and the top-down network densely connecting CBAM. ResNet101 is utilized as the backbone, which is a 101-layer residual network that adopts shortcut connections to learn residual functions with reference to the input layer instead of learning unreferenced functions [42]. According to the different sizes of output feature maps, the convolutional layers in ResNet101 are divided into five stages, which are defined as conv1, conv2, conv3, conv4, and conv5. In these five stages, only conv1 is a 7×7 convolutional layer, and the other stages are composed of numerous 3×3 and 1×1 convolutional layers. In general, the deepest layer of each stage has the strongest features. Therefore, the activation outputs of the last residual block in conv2, conv3, conv4, and conv5 stages are adopted as $\{C_2, C_3, C_4, C_5\}$ in the bottom-up feedforward network. Especially, each feature map in the bottom-up network reduces the channel dimension by a 1×1 convolutional layer.

In the top-down network, the feature maps with more semantic information and higher resolution are obtained by up-sampling the high-level feature maps. Since some redundant information needs to be suppressed in the huge number of features, while target features are emphasized, the CBAM is densely connected to feature maps concatenated by up-sampled higher level feature maps in the top-down network to highlight the prominent features of specific scales. The feature maps processed by CBAM are defined as $\{A_2, A_3, A_4, A_5\}$, and the computational process refers to (6). In this way, massive features containing more semantic and spatial information in different layers are refined to acquire significant features of different scales for subsequent detection.

Then, the feature maps with the same spatial size in the bottom-up pathway and the top-down pathway are merged by lateral connections. By this means, it integrates the unblurred global features of the bottom-up network and the salient local features of the top-down network. Moreover, a 3×3 convolutional layer is attached to each final fused feature map to alleviate aliasing effect of up-sampling while reducing the dimensions of channel to 256. Especially, the set of final fused feature maps are defined as $\{P_2, P_3, P_4, P_5\}$. The multi-scale fused feature maps for multi-scale ship detection in multi-scene SAR images are obtained by lateral connections and dense connections with CBAM. Furthermore, both lateral connections and dense connections in DAPN are implemented by concatenation. In short, the overall process of DAPN is as follows:

$$P_5 = \mathrm{Conv}_{1\times1}(C_5) \tag{5}$$

$$A_{\mathrm{i}} = \begin{cases} A[U\mathrm{psample}(P_5)] & i=4 \\ A[U\mathrm{psample}(P_{\mathrm{i}+1}) \oplus \cdots \oplus U\mathrm{psample}(P_5)] & i=2,3 \end{cases} \tag{6}$$

$$P_{\mathrm{i}} = \mathrm{Conv}_{3\times3}[\mathrm{Conv}_{1\times1}(C_i) \oplus A_{\mathrm{i}}], i=2,3,4 \tag{7}$$

where A and Upsample represent the operations of CBAM and up-sample, respectively, and $\oplus$ is the operation of concatenation.

DAPN can extract multi-scale fused feature maps containing abundant spatial and semantic information for multi-scale ship detection, as well as unblurred global features and salient local features contained in final fused feature maps adapted to different scenes of SAR images. Then, the final fused feature maps are fed to RPN to generate proposals.

3) Region Proposal Network: Although DAPN generates high-resolution feature maps with more semantic information, multi-scale feature maps of different layers have various resolutions and receptive fields. High-level feature maps with

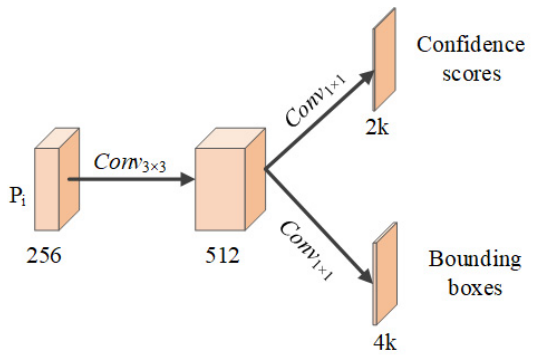


Fig. 6. Structure of RPN

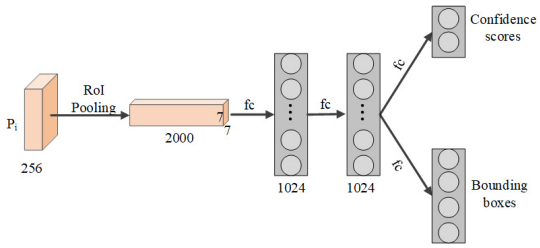


Fig. 7. Structure of the detection network.

low resolution and large receptive fields are fit to detect large-scale ships, and on the contrary, low-level feature maps are suitable for the small-scale ship detection. Therefore, the final fused feature maps $\{P_2, P_3, P_4, P_5, P_6\}$ are fed to RPN, respectively, for multi-scale ship detection, and P_6 is obtained by max-pooling operation of P_5.

As shown in Fig. 6, RPN is composed of a 3×3 convolutional layer and two 1×1 convolutional layers to obtain confidence scores and bounding boxes simultaneously. Anchors are adopted as reference boxes for classification and regression, generating confidence scores and bounding boxes. All anchors are assigned five scales $\{16^2, 24^2, 40^2, 60^2, 80^2\}$ to $\{P_2, P_3, P_4, P_5, P_6\}$, respectively. Considering the different scales of ships, each anchor is designed to have various ratios {1:1, 1:2, 1:3, 2:1, 2:3, 3:1, 3:2}. Therefore, there are $k = 35$ different anchors at each location of each fused feature map, and every proposal has 2k confidence scores and 4k parameters encoding bounding boxes. Besides, positive samples and negative samples are chosen at a ratio of 1:1 to train the network based on intersection-over-union (IoU) with ground truths. The label of an anchor is positive if IoU with the ground truth is over 0.7, and the label is negative if IoU with the ground truth is lower than 0.3. Finally, 2000 RoIs are obtained for each image by top-N and non-maximum suppression (NMS) operations on all proposals. All RoIs from different levels obtained by RPN are attached to faster R-CNN for the final multi-scale ship detection results.

4) Detection Network: The detection network mainly consists of the RoI pooling layer and fully connected (fc) layers to achieve final ship detection with confidence scores and bounding boxes, and the structure of detection network is shown in Fig. 7. The RoI pooling layer is adopted to generate a fixed size of 7×7 features, which maps RoIs of different layers extracted by RPN to the fused feature map of the corresponding layer, respectively. Then, all the 7×7 features are flattened to send to fc layers for the final detection results. In this way, the global unblurred features and local salient features containing rich semantic and spatial information are completely preserved and utilized to detect multi-scale ship in SAR images.

C. Loss Function

Multi-task loss function as faster R-CNN [31] is used in this paper. For an image, the loss function is as follows:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{\text{cls}}}\sum_i L_{\text{cls}}(p_i, p_i^*) + \lambda \frac{1}{N_{\text{reg}}}\sum_i p_i^* L_{\text{reg}}(t_i, t_i^*) \tag{8}$$

where i is the index of an anchor in a minibatch. In addition, p_i and $p_i{}^*$ are predicted probability and ground-truth label, respectively. Similarly, t_i and $t_i{}^*$ are the predicted bounding box and ground-truth box, respectively. The classification loss and the regression loss are represented by L_{cls} and L_{reg}, and the two terms are normalized by N_{cls} and N_{reg}. Besides, they are weighted by λ.

The classification loss L_{cls} adopts the cross-entropy loss to measure the difference between two probability distributions. The definition is as follows:

$$L_{\text{cls}}(p_i, p_i{}^*) = -[p_i \log(p_i{}^*) + (1 - p_i)\log(1 - p_i{}^*)]. \tag{9}$$

The regression loss L_{reg} is defined as

$$L_{\text{reg}}(t_i, t_i^*) = \text{smooth}_{L1}(t_i - t_i^*) \tag{10}$$

where

$$\text{smooth}_{L1}(x) = \begin{cases} 0.5x^2 & if\ |x| < 1 \\ |x| - 0.5 & \text{otherwise.} \end{cases} \tag{11}$$

III. Experiments

In this section, the experiments are implemented to evaluate the detection performance of the proposed method. At first, the introduction of the data set called SSDD, and the detailed settings of related experiments will be illustrated. Then, the evaluation criteria used in this paper are described. The next part mainly concentrates on the exploration of DAPN, including the effect of concatenation, the significance of CBAM, and the importance of the fusion of unblurred and salient features. Finally, the comparison with other methods on SSDD implies the effectiveness of the proposed method.

A. Data Set and Settings

The SSDD data set is constructed by Li *et al.* [38], containing multi-scale ships in various environments, including different scenes, sensor types, polarization modes, and image resolutions. In SSDD, there are multi-scale ships in offshore and inshore scenes, and the SAR images in the data set are mainly from RadarSat-2, TerraSAR-X, and Sentinel-1 sensors. There are four polarization modes, including HH, HV, VV, and

VH. Furthermore, the resolution of SAR images in the data set ranges from 1 to 15 m.

There are totally 1160 images in SSDD, which are divided into three parts (training, validation, and test set) with the proportion of 7:2:1. Besides, the data augmentation methods as horizontal flip and random crop are adopted to increase the data set and make the proposed method more robust.

The pre-trained ResNet101 on the ImageNet data set is utilized as the backbone network. Especially, the weight decay and the momentum are 0.0001 and 0.9, respectively. Besides, the learning rate is 0.001 with a decay ratio of 0.1. The max iteration is 50 000.

Especially, all experiments are implemented in Tensorflow framework and carried out on a computer with an Intel i7-6700HQ processor and a Nvidia Geforce GTX 1080Ti GPU card. The operating system is Linux with CUDNN v8.

B. Evaluation Criteria

In order to evaluate the detection performance quantitatively, some evaluation criteria are adopted to contrast different ship detection methods on SSDD, including the precision, recall, mean average precision (mAP), and F1 score. These criteria are obtained by four well-established components in information retrieval, true positive (TP), true negative (TN), false positive (FP), and false negative (FN). In this paper, TP and TN indicate the number of correct detected ships and correct detected backgrounds, respectively. FP represents the number of false alarms, and FN is the number of undetected ships. Moreover, if the IoU between a bounding box and a single ground truth is higher than 0.5, the target detected by this bounding box is a correct ship.

Recall is the ratio of correct detected ships in all ground truths, which represents the coverage of ground truths, and it is defined as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{12}$$

The precision is the ratio of correct detected ships in all detected objects, which represents the correctness of detected objects. The definition is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{13}$$

Since recall and precision affect each other, F1 score and mAP are adopted to evaluate the overall performance of detection methods. F1 score is computed as

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{14}$$

and mAP is as follows:

$$\text{mAP} = \int_0^1 P(R)\text{dR} \tag{15}$$

where P and R represent the single-point value of precision and recall.

TABLE I
EFFECT OF CONCATENATION IN MULTI-SCENE SAR IMAGES

Scenes	Methods	Recall	Precision	mAP	F1
Inshore	FPN	**71.04**%	45.77%	39.46%	55.67%
	DFPN-EA	67.78%	46.54%	40.02%	55.19%
	DFPN-CON	64.26%	**51.17**%	**41.28**%	**56.98**%
Offshore	FPN	**90.17**%	74.51%	67.29%	81.60%
	DFPN-EA	83.68%	71.73%	61.37%	77.24%
	DFPN-CON	82.99%	**82.13**%	**69.93**%	**82.56**%

C. Evaluation of DAPN

In this part, the effect of DAPN is evaluated through a series of comparative experiments in three aspects comprising concatenation, CBAM, and the fusion of unblurred and salient features. In order to explore the multi-scale ship detection performance of the proposed method in different scenes of SAR images, experiments are conducted in inshore and offshore scenes, respectively.

1) Effect of Concatenation: In DAPN, lateral connections and dense connections between up-sampled feature maps have adopted the operation of concatenation rather than element-wise addition in [32]. To evaluate the effect of dense connection with concatenation, we conduct comparison experiments on FPN, dense connection with element-wise addition, and dense connection with concatenation and they are denoted as FPN, DFPN-EA, and DFPN-CON, respectively. In the framework of FPN, the feature maps $\{P_2, P_3, P_4, P_5\}$ are directly connected to RPN, rather than densely connecting feature maps of higher layers. In this way, there is less semantic information from different deeper layers than methods with dense connections.

Fig. 8(a) shows the ground truth, and Fig. 8(b)–(d) shows the detection results of FPN, DFPN-EA, and DFPN-CON, respectively. In Fig. 8(a), red rectangles represent correct ship targets, and in Fig. 8(b)–(d), green rectangles are the detected targets of these three methods. Moreover, one rectangle represents a single-ship target, whether in the ground truth or detection results. With Fig. 8(a) as a reference, the false alarms and the missed targets of these methods can be figured out. FPN has more false alarms whether it is inshore or offshore scenes, whereas DFPN-EA exhibits the missed detection problem with some false alarms. Compared with these two methods, the detection performance of DFPN-CON is better, which has less false alarms and missed targets. The comparison results demonstrate the effect of dense connection with concatenation. For offshore scenes, DFPN-CON has already achieved pretty good performance for multi-scale ship detection. However, DFPN-CON can only detect a single ship accurately in the inshore scenes, rather than multiple ships.

The comparison results of detection performance evaluation are shown quantitatively in Table I. As can be observed from the results, DFPN-EA has the worst performance whether inshore or offshore scenes and its indices are even lower than FPN, which is without dense connection. It seems that dense connection with element-wise addition is almost pointless for performance improvements, which does not achieve the

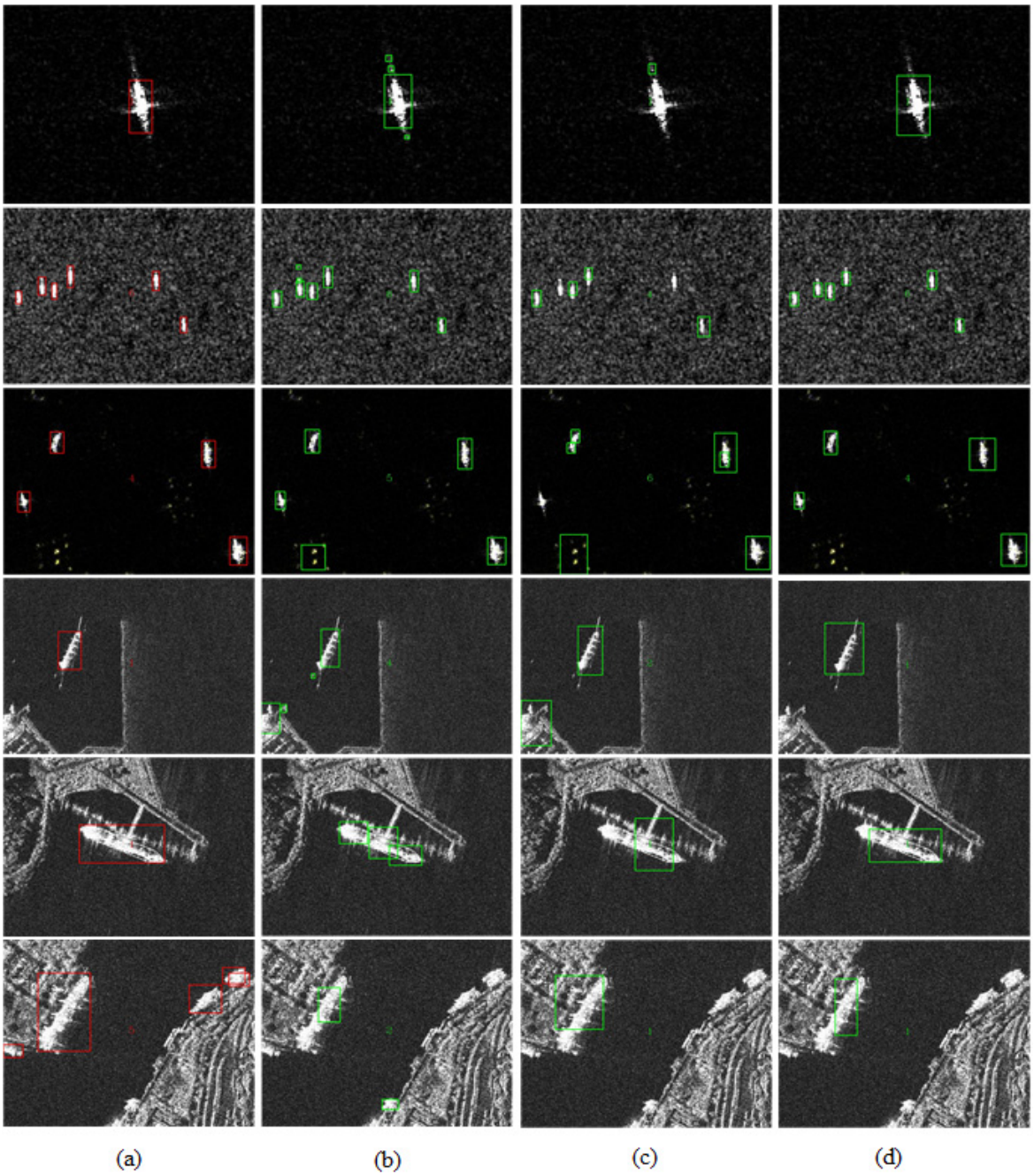


Fig. 8. Comparison results of FPN, DFPN-EA, and DFPN-CON. (a) Ground truth. (b)–(d) Results of FPN, DFPN-EA, and DFPN-CON, respectively. The red rectangles in (a) are correct ship targets. The green rectangles in (b)–(d) indicate detected targets.

desired effect of merging semantic information from different layers. In addition, the overall performance of DFPN-CON is better than FPN. Although its recall is slightly lower than FPN, the precision of DFPN-CON is nearly 8% higher than FPN. The reason is that DFPN-CON utilizes dense connection but no adaptive feature selection process, resulting in the accumulation of features in higher level feature maps. Since high-level feature maps are suitable for detecting large-scale ships, DFPN-CON pays more attention to large-scale targets, ignoring small-scale targets. Therefore, DFPN-CON has a little lower recall value, higher precision value, and best overall detection performance. Consequently, dense connection with concatenation makes full use of the semantic information of all layers deeper than a certain layer, obtaining high-resolution feature maps with more semantic information, which can enhance detection performance effectively.

2) Effect of CBAM: A large number of high-resolution feature maps with more semantic information for multi-scale ship detection are obtained through dense connections with concatenation. In addition, the proposed DAPN adopts CBAM to concatenated feature maps comprising up-sampled feature maps of all layers higher than the certain layer, fusing feature

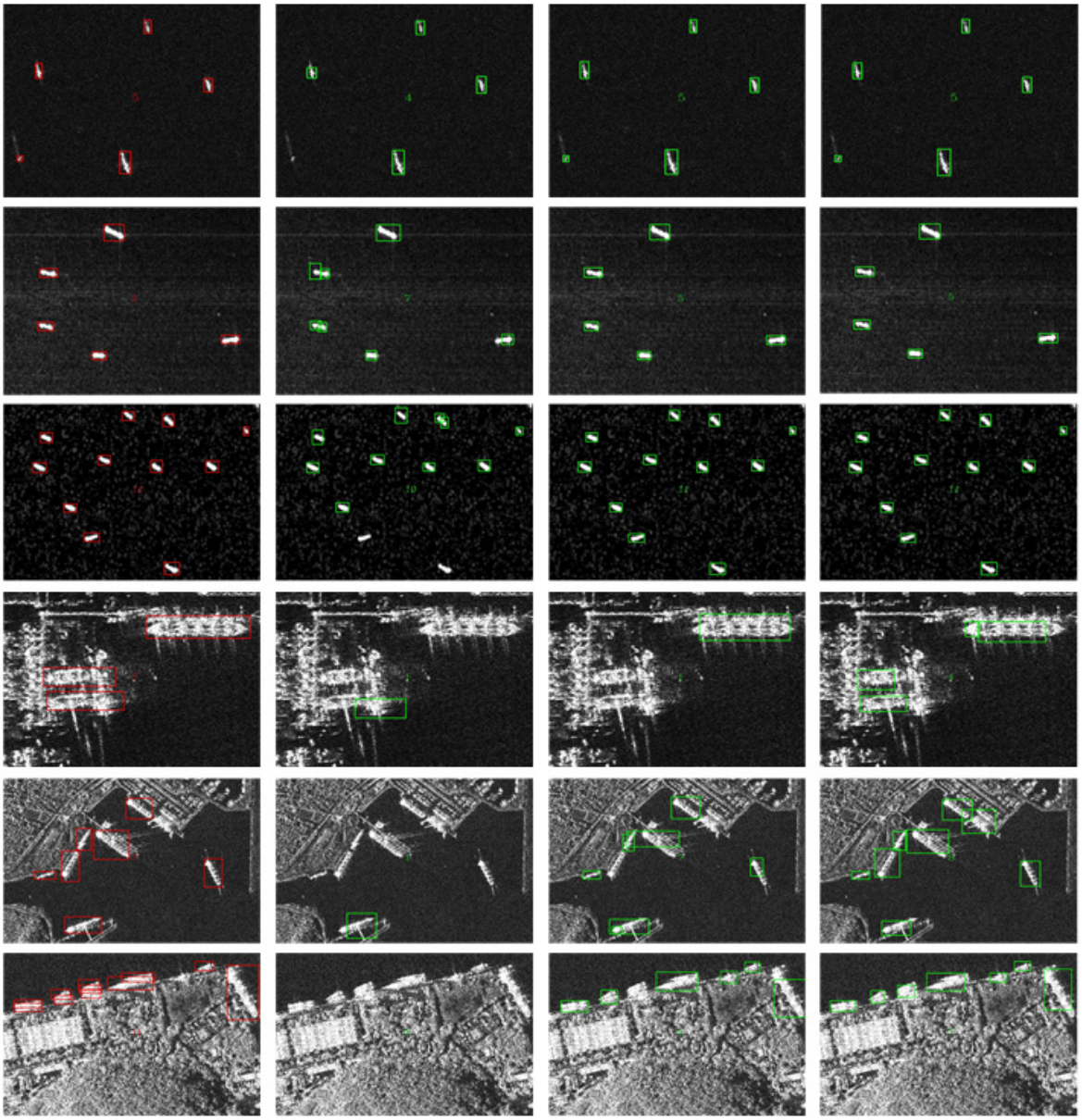


Fig. 9. Effect of CBAM. (a) Ground truth. (b) and (c) Method with dense connections or CBAM, respectively. (d) Method with both. The red rectangles in (a) are correct ship targets. The green rectangles in (b)–(d) indicate detected targets.

maps of different scales and extracting salient information of specific scales from massive fused features. The refined feature maps emphasize significant features containing spatial and semantic for multi-scale ships detection while suppressing redundant information to reduce false alarms. To evaluate the effect of CBAM in the proposed method, comparison experiments are carried out. In order to distinguish the influence of dense connections, we conduct experiments with or without CBAM on the basis of dense connections. Moreover, the method, which only has CBAM without dense connections, is implemented to show the effect of CBAM.

The comparison results of these three methods are shown in Fig. 9. In offshore scenes of SAR images, the method only with dense connections has some false alarms and missed detection problems, especially for small-scale ships, while the other two methods with CBAM can detect multi-scale ships well. As mentioned before, the method by dense connections with concatenation has difficulties in detecting multiple ships in the inshore scenes, which usually detects single ship correctly. Compared with the method with dense connections, the method connecting CBAM improves performance for multiple ships detection in the inshore scenes. Since semantic information is not enough without dense connections, there are still some missed detection problems in this method, as well as the bounding boxes are not very accurate for multi-scale ships. Both dense connections and CBAM are adopted by the

TABLE II
EFFECT OF CBAM

Scenes	Dense	CBAM	Recall	Precision	mAP	F1
Inshore	√		64.26%	51.17%	41.28%	56.98%
		√	89.85%	61.25%	58.45%	72.75%
	√	√	**90.98**%	**71.18**%	**67.55**%	**79.87**%
Offshore	√		82.99%	82.13%	69.93%	82.56%
		√	98.98%	95.42%	95.39%	97.17%
	√	√	**99.66**%	**95.77**%	**95.93**%	**97.67**%

TABLE III
EFFECT OF THE FUSION OF UNBLURRED FEATURES AND SALIENT FEATURES

Scenes	unblurred	salient	Recall	Precision	mAP	F1
Inshore	√		88.72%	58.56%	54.51%	70.55%
		√	74.13%	53.26%	47.38%	61.98%
	√	√	**90.98**%	**71.18**%	**67.55**%	**79.87**%
Offshore	√		91.32%	84.25%	78.18%	87.63%
		√	91.62%	81.38%	76.95%	86.20%
	√	√	**99.66**%	**95.77**%	**95.93**%	**97.67**%

proposed method in this paper, which has superior detection performance for multi-scale ships whether inshore or offshore scenes. Although there are still a few false alarms for the inshore ships detection, the proposed method achieves a better detection performance than other methods on SSDD.

The evaluation indicators of the detection performance of these three methods are displayed in Table II. It can be seen whether it is used in the offshore or inshore scenes, and every evaluation criterion of the method that only has dense connections without CBAM is significantly lower than the other two methods which both connect CBAM. Therefore, connecting CBAM to the pyramid network is effective for detecting multi-scale ships in various scenes of SAR images. For the offshore scenes, the two methods connecting CBAM achieve almost the same excellent detection performance for multi-scale ships. As for the inshore scenes, the overall detection performance of the proposed method is nearly 7% higher than the method that only connects CBAM without dense connections, especially in terms of precision, the index is 10% higher. It is confirmed that the proposed method adopting dense connections with concatenations and CBAM achieves a better detection performance for multi-scale ships in SAR images.

3) Effect of the Fusion of Global Unblurred Features and Local Salient Features: As demonstrated in the first two parts, dense connections with concatenation integrate semantic information with spatial information for multi-scale ship detection. Besides, CBAM is utilized to emphasize significant information on massive features containing both spatial and semantic information and suppress redundant features to reduce false alarms of complex backgrounds. In this paper, CBAM is connected to the concatenated feature maps comprising up-sampled feature maps of all layers than the certain layer to obtain salient features of this layer. By lateral connections, global unblurred features generated from the bottom-up network and local salient features are fused to adapt ship detection in various scenes of SAR images. Comparative experiments are carried out to evaluate the effect of the fusion of unblurred and salient features. We conduct experiments on three methods, including only using blurred features or salient features to detect and adopting the fusion of these two types of features for multi-scale ship detection in different scenes of SAR images.

The effectiveness of the fusion of global unblurred features and local salient features can be seen in Table III. It can be observed that there is a huge difference between these three methods for inshore ship detection. Among them, the method only with local salient features has the worst detection performance, whose recall and precision are lower than the other two methods. This is due to the absence of original global features to refine the fuzzy features caused by up-sampling and CBAM, which causes missed targets detection. In addition, the recall of the method with unblurred features is only 2% lower than the method with fused features, but there are much more false alarms that make the precision far lower. The proposed method, fusing global unblurred features and local salient features, has the highest recall and precision, as well as the best overall detection performance that achieves nearly 80%. With regard to offshore ship detection, the methods only using unblurred features or salient features obtain similar detection performance that is nearly 87%, while the proposed method achieves 97%, which is almost 10% higher than the other two methods. Consequently, it demonstrates the effect of the proposed method, which fuses global unblurred features in the bottom-up network and local salient features in the top-down network, effectively avoiding false alarms and missed targets detection.

The comparison results are shown in Fig. 10. For offshore scenes, there are a few false alarms and missed targets in the methods that only utilize unblurred features of salient features to detect multi-scale ships. Meanwhile, the proposed method has supreme detection performance for multi-scale ships in the offshore scenes. As for inshore ship detection, the methods only with global unblurred features or local salient features have worse performance than offshore ship detection. For the method only with unblurred features, more false alarms are the main problems because of the lack of salient features to identify targets and inference. Although there are a few false alarms in the method only with salient features, missed targets detection due to blur caused by up-sampling and CBAM is the main problem. Therefore, it is necessary to fuse global unblurred features and local salient features to refine final fused feature maps used for multi-scale ship detection in SAR images.

D. Comparison With CFAR-Based Methods

CFAR and its improved methods are the most widely studied and applied ship detection methods in SAR images. In this part, some typical and advanced CFAR-based methods are selected to do comparative experiments with the proposed method, including the gamma- and G^0-based global CFAR [12], [14] and the two-stage CFAR based on G^0 distribution. The sliding window of global CFAR is the entire image, and

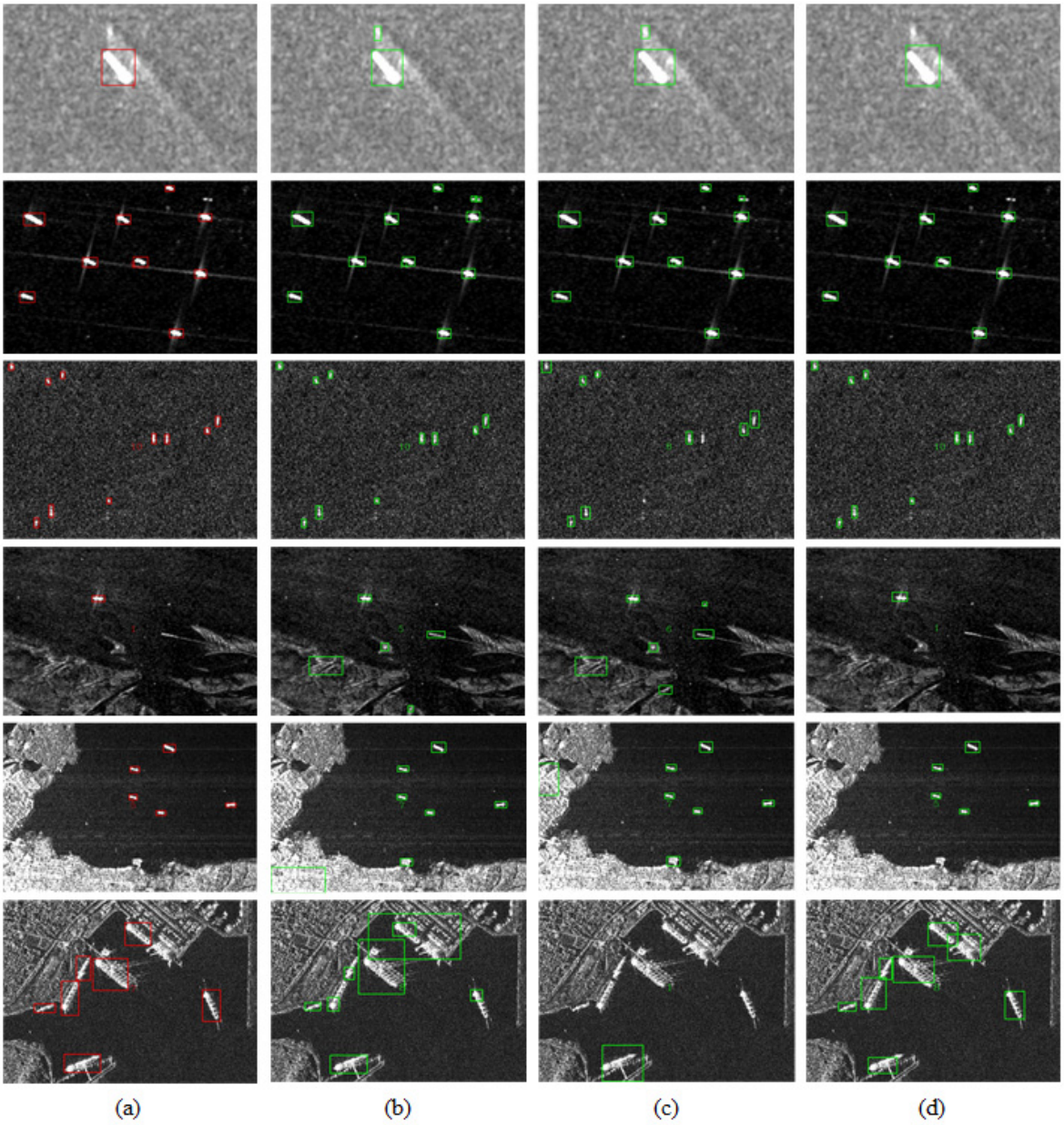


Fig. 10. Effect of fusion of unblurred and salient features. (a) Ground truth. (b) and (c) Only with unblurred or salient features, respectively. (d) Both features. The red rectangles in (a) are correct ship targets. The green rectangles in (b)–(d) indicate detected targets.

it is easy to implement. The two-stage CFAR uses global CFAR first and then eliminates false alarms by local CFAR. All the three methods adopt image processing and subsequent identification process to refine the detection results, and the shape of the detected ship is estimated by fitting an ellipse.

The comparison results of detection performance evaluation are shown quantitatively in Table IV. As can be seen from the results, global gamma-based CFAR has the worst detection performance in both scenes, especially in the inshore scenes. The global CFAR and the two-stage CFAR based on G^0 distribution obtain a similar detection performance. The difference between the two methods is that the precision of the two-stage CFAR is a bit higher because it cascades the local CFAR to eliminate false alarms, but the overall performance of these two methods is almost the same. The overall performance of the proposed method is only 2% higher than these CFAR-based methods in offshore scenes, while it is almost 20%

TABLE IV
COMPARISON WITH CFAR-BASED METHODS

Scenes	Methods	Recall	Precision	F1
Inshore	Global Gamma-based CFAR	65.88%	39.23%	49.18%
	Global G^0-based CFAR	75.59%	48.71%	59.24%
	Two-stage CFAR	73.65%	50.53%	59.94%
	the proposed method	**90.98**%	**71.18**%	**79.87**%
Offshore	Global Gamma-based CFAR	90.92%	84.38%	87.53%
	Global G^0-based CFAR	98.08%	92.73%	95.33%
	Two-stage CFAR	96.15%	94.23%	95.18%
	the proposed method	**99.66**%	**95.77**%	**97.67**%

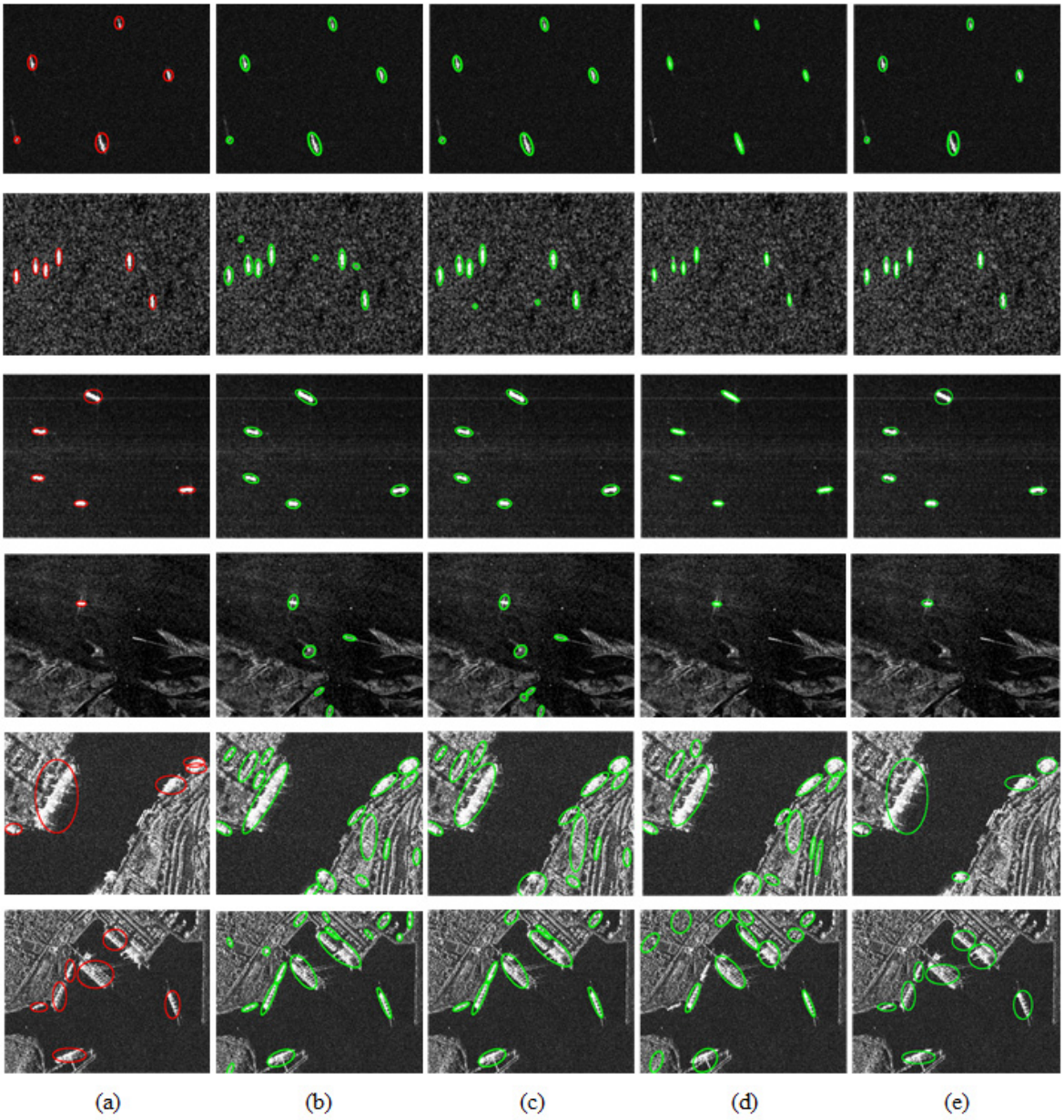


Fig. 11. Comparison with CFAR-based methods. (a) Ground truth. (b)–(d) Detection results of global gamma-based CFAR, global G^0-based CFAR, and two-stage CFAR. (e) Detection result of the proposed method. The red ellipses in (a) are correct ship targets. The green ellipses in (b)–(e) are the detected ships of global gamma-based CFAR, global G^0-based CFAR, two-stage CFAR, and the proposed method.

higher in the inshore scenes. Hence, the proposed method achieves a significantly better detection performance than these CFAR-based methods in the inshore scenes. Furthermore, the proposed method is highly effective in the multi-scale ship detection of SAR images.

The comparison results are shown in Fig. 11. In offshore scenes, global gamma- and G^0-based CFARs have more false alarms than other methods, and the two-stage CFAR misses a small-scale ship in the first image of Fig. 11(d). Compared with these three methods, the proposed method obtains a little better detection performance. However, the proposed method presents the prominent advantages in the inshore scenes. Meanwhile, the main problem of these three CFAR-based methods is the false alarm problem. The number of false alarms generated by the global gamma-based method is the most. The two-stage CFAR has relatively less false alarms, but it still exhibits the missed detection problem. By contrast, the proposed method achieves a supreme detection performance with very few false alarms and missed ships. Therefore, the proposed method significantly improves the CFAR-based methods mainly in the inshore scenes of SAR images.

E. Comparison With CNNs on SSDD

In this part, some CNN-based ship detection methods on SSDD are contrasted with the proposed method in this paper, such as the detection methods based on SSD and faster R-CNN. Besides, an improved faster R-CNN adopting feature fusion, hard example mining, and transfer learning proposed

TABLE V
PERFORMANCE OF SSD, FASTER RCNN, IMPROVED FASTER RCNN, AND THE PROPOSED METHOD

Methods	mAP
Faster R-CNN	70.1%
Improved Faster R-CNN	78.8%
SSD	78.5%
the proposed method	**89.8**%

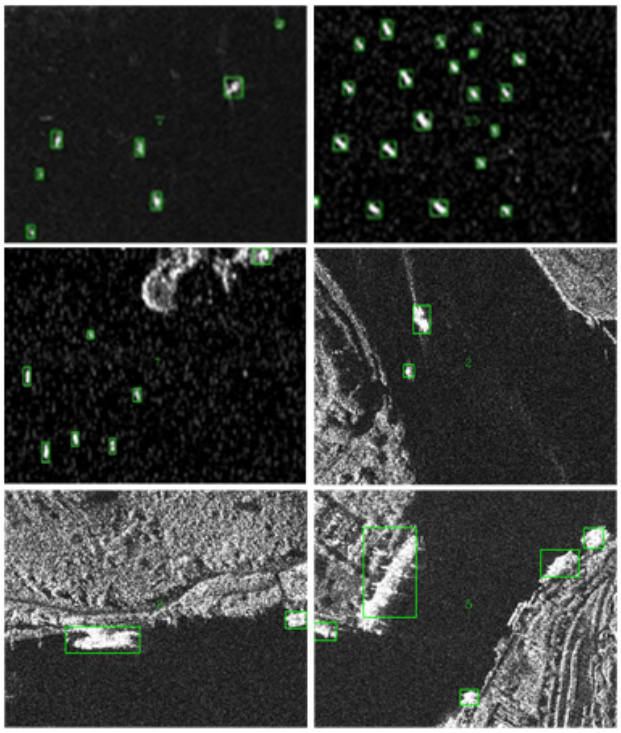


Fig. 12. Detection results of the proposed method.

in [38] is also used for comparison. Since these methods did not carry out special research on ship detection in different scenes of SAR images, respectively, we randomly selected images in SSDD at a ratio of 7:2:1 as training, verification, and test set to evaluate the detection performance of the proposed method. To evaluate the overall detection performance of these methods quantitatively, mAP is utilized in this part, and the result is displayed in Table V.

As can be seen from the results in Table V, the mAP of the proposed method in this paper has achieved 89.8%, which is 11% higher than SSD and the improved faster R-CNN, and nearly 20% higher than faster R-CNN. As a consequence, the much higher mAP value of the proposed method confirms a better detection performance than other ship detection methods on SSDD. It implies the effectiveness of DAPN to extract multi-scale fused feature maps containing more spatial and semantic information for multi-scale ship detection, and the integration of global unblurred features improves the accuracy of ship detection in SAR images. Some detection results of the proposed method for multi-scale ships in the different scenes of SAR images are shown in Fig. 12.

It is clearly observed from Fig. 12 that the proposed method obtains a supreme detection performance for multi-scale ships in various environments, whether they are in offshore or inshore scenes of SAR images. For offshore ship detection, even if the sizes of small-scale ships are so tiny that they are similar to the speckle noise, which can hardly be distinguished by the naked eye, the proposed method can still detect well. Furthermore, the proposed method can eliminate the interference of artificial ground objects in complex inshore scenes, so as to identify multi-scale ships docked on the shore. Hence, the proposed method can be effectively applied to multi-scale ship detection in SAR images.

IV. CONCLUSION

In this paper, the method based on DAPN is proposed for multi-scale ship detection in SAR images. A series of contrast experiments on CFAR-based and CNNs demonstrates the effectiveness of the proposed method based on DAPN for multi-scale ship detection in various scenes of SAR images.

First, it is known from this paper that spatial information in feature maps is beneficial to detect small-scale ships, while semantic information is suitable for large-scale ships. Therefore, the method adopts a pyramid structure for multi-scale ship detection.

Besides, in order to preserve spatial and semantic information in the fused feature maps of different layers as completely as possible, the operation of concatenation is utilized in lateral and dense connections rather than element-wise addition.

Next, to highlight significant information and suppress redundant information in the huge amount of fused feature maps, CBAM is adopted to fuse the feature maps of different scales and extract salient features from the concatenated feature maps comprising up-sampled feature maps of different layers higher than the certain layer. In this way, the feature maps of different scales are weighted to focus on the salient features of the specific scales for detecting ships of suitable scales. Besides, salient features are vital to prevent false alarms of complex backgrounds.

However, only using salient features for ship detection can result in missed ship problems due to severe feature blur caused by up-sampling and CBAM. Thus, for the sake of solving this problem, local salient features extracted in the top-down network are merged with global unblurred features in the corresponding layer of the bottom-up network to refine feature maps used for final ship detection. By this means, the proposed method is flexibly adapted to ship detection in different scenes of SAR images.

In summary, the proposed method based on DAPN extracts high-resolution fused feature maps with more semantic information for multi-scale ship detection; meanwhile, the integration of global unblurred features with local salient features is beneficial to detect ships in various scenes of SAR images with extremely high accuracy. The experimental results verify that it can detect multi-scale ships in different scenes accurately, including inshore and offshore ships. Besides, the proposed method achieves a better performance than other state-of-the-art ship detection methods.

For ships docked side by side on the shore, they can be detected well by changing the threshold of NMS, but the

adaptability of this approach to different scenes is a little bit worse. Therefore, it is the focus of future research to find a universal method to detect ships side by side in the inshore scenes of SAR images.

Furthermore, the proposed method in this paper only utilizes the single-polarization information of SAR images to detect multi-scale ships, and therefore, the research on how to integrate polarization information into deep networks for ship detection in SAR images will be a key direction of the further research in the future.

REFERENCES

[1] S. Brusch, S. Lehner, T. Fritz, M. Soccorsi, A. Soloviev, and B. van Schie, "Ship surveillance with TerraSAR-X," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1092–1103, Mar. 2011.

[2] P. W. Vachon, S. J. Thomas, J. Cranton, H. R. Edel, and M. D. Henschel, "Validation of ship detection by the RADARSAT synthetic aperture radar and the ocean monitoring workstation," *Can. J. Remote Sens.*, vol. 26, no. 3, pp. 200–212, 2000.

[3] C. C. Wackerman, K. S. Friedman, W. G. Pichel, P. Clemente-Colón, and X. Li, "Automatic detection of ships in RADARSAT-1 SAR imagery," *Can. J. Remote Sens.*, vol. 27, no. 5, pp. 568–577, 2001.

[4] A. C. Copeland, G. Ravichandran, and M. M. Trivedi, "Localized radon transform-based detection of ship wakes in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 33, no. 1, pp. 35–45, Jan. 1995.

[5] X. Wei, X. Wang, and J. Chong, "Local region power spectrum-based unfocused ship detection method in synthetic aperture radar images," *Proc. SPIE*, vol. 12, no. 1, 2018, Art. no. 016026.

[6] W. Huo, Y. Huang, J. Pei, Q. Zhang, Q. Gu, and J. Yang, "Ship detection from ocean SAR image based on local contrast variance weighted information entropy," *Sensors*, vol. 18, no. 4, p. 1196, 2018.

[7] M. Yang and C. Guo, "Ship detection in SAR images based on lognormal ρ-metric," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1372–1376, Sep. 2018.

[8] K. Eldhuset, "An automatic ship and ship wake detection system for spaceborne SAR images in coastal regions," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 4, pp. 1010–1019, Jul. 1996.

[9] M. Tello, C. Lopez-Martinez, and J. J. Mallorqui, "A novel algorithm for ship detection in SAR imagery based on the wavelet transform," *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 2, pp. 201–205, Apr. 2005.

[10] H. Shi *et al.*, "A novel ship detection method based on gradient and integral feature for single-polarization synthetic aperture radar imagery," *Sensors*, vol. 18, no. 2, pp. 563–587, 2018.

[11] 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.

[12] A. C. Frery, H.-J. Müller, 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.

[13] 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.

[14] 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.

[15] M. Yeremy, J. W. M. Campbell, K. Mattar, and T. Potter, "Ocean surveillance with polarimetric SAR," *Can. J. Remote Sens.*, vol. 27, no. 4, pp. 328–344, 2001.

[16] D. J. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," DSTO, Dept. Defense, Australian Government, Canberra, IC, Australia, Tech. Rep. DRDC-TM-2005-243, 2004.

[17] J. Chen, Y. Chen, and J. Yang, "Ship detection using polarization cross-entropy," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 723–727, Oct. 2009.

[18] R. Touzi and F. Charbonneau, "Characterization of target symmetric scattering using polarimetric SARs," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2507–2516, Nov. 2002.

[19] M. Migliaccio, F. Nunziata, A. Montuori, and R. L. Paes, "Single-look complex COSMO-SkyMed SAR data to observe metallic targets at sea," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 893–901, Jun. 2012.

[20] Z. Zou and Z. Shi, "Ship detection in spaceborne optical image with SVD networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5832–5845, Oct. 2016.

[21] G. Gao, S. Gao, J. He, and G. Li, "Adaptive ship detection in hybrid-polarimetric SAR images based on the power–entropy decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5394–5407, Sep. 2018.

[22] S. Song, B. Xu, and J. Yang, "Ship detection in polarimetric SAR images via variational Bayesian inference," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 2819–2829, Jun. 2017.

[23] J. He, Y. Wang, H. Liu, N. Wang, and J. Wang, "A novel automatic PolSAR ship detection method based on superpixel-level local information measurement," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 384–388, Mar. 2018.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[25] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[26] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[27] W. Liu, L. Ma, J. Wang, and H. Chen, "Detection of multiclass objects in optical remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 791–795, May 2019.

[28] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

[29] 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.

[30] R. Girshick, "Fast R-CNN," in *Proc. IEEE Conf. Comput. Vis. (ICCV)*, Boston, MA, USA, Jun. 2015, pp. 1440–1448.

[31] 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.

[32] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 936–944.

[33] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Munich, German, Sep. 2018, pp. 8–14.

[34] L. I. Perlovsky, W. H. Schoendorf, B. J. Burdick, and D. M. Tye, "Model-based neural network for target detection in SAR images," *IEEE Trans. Image Process.*, vol. 6, no. 1, pp. 203–216, Jan. 1997.

[35] R. Wang, J. Li, Y. Duan, H. Cao, and Y. Zhao, "Study on the combined application of CFAR and deep learning in ship detection," *J. Indian Soc. Remote Sens.*, vol. 46, no. 9, pp. 1413–1421, 2018.

[36] Q. An, Z. Pan, and H. You, "Ship detection in Gaofen-3 SAR images based on sea clutter distribution analysis and deep convolutional neural network," *Sensors*, vol. 18, no. 2, p. 334, 2018.

[37] Y. Wang, C. Wang, and H. Zhang, "Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images," *Remote Sens. Lett.*, vol. 9, no. 8, pp. 780–788, 2018.

[38] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. BIGSARDATA*, Beijing, China, Nov. 2017, pp. 1–6.

[39] 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.

[40] J. Wang, C. Lu, and W. Jiang, "Simultaneous ship detection and orientation estimation in SAR images based on attention module and angle regression," *Sensors*, vol. 18, no. 9, p. 2851, 2018.

[41] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, no. 8, p. 860, 2017.

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

Zongyong Cui (S'13–M'15) received the B.E. and Ph.D. degrees in signal and information processing from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2007 and 2015, respectively.

From 2013 to 2014, he was a Visiting Student with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore. He is currently a Lecturer with the School of Information and Communication Engineering, UESTC. His research interests include radar image processing, target recognition, and machine learning.

Dr. Cui is a Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the *IET Radar, Sonar & Navigation*, and the *Journal of Electronic Imaging*.

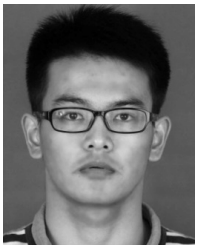
Qi Li (S'19) received the B.E. degree in electronic information engineering from the Harbin Institute of Technology (HIT), Weihai, China, in 2017. She is currently pursuing the M.S. degree in signal and information processing with the University of Electronic Science and Technology of China (UESTC), Chengdu, China.

Her research interests include object detection, image processing, and deep learning.

Zongjie Cao (M'10) received the B.E. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1999 and 2005, respectively.

From 2006 to 2008, he was a Post-Doctoral Researcher with the Communication and Information System Postdoctoral Center, University of Electronic Science and Technology of China (UESTC), Chengdu, China. In 2008, he joined the School of Electronic Engineering, UESTC, where he is currently a Professor with the School of Information and Communication Engineering. He has authored or coauthored over 50 papers. His research interests include radar signal processing, synthetic aperture radar imaging, and target recognition.

Dr. Cao is a Reviewer for several international journals and conferences, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the *IET Radar, Sonar & Navigation*, and *Remote Sensing*.



Nengyuan Liu (S'18) received the B.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering.

His research interests include SAR image processing and target detection.