A Robust One-Stage Detector for Multiscale Ship Detection With Complex Background in Massive SAR Images

Xi Yang, *Member, IEEE*, Xin Zhang, Nannan Wang, *Member, IEEE*, and Xinbo Gao, *Senior Member, IEEE*

***Abstract*— With the development of synthetic aperture radar (SAR) imaging and deep learning, SAR ship detection based on convolutional neural networks (CNNs) has been extensively applied in the last few years. Nevertheless, there are two main obstacles in SAR ship detection: 1) the SAR images have too much noise, such as the interference from land area, making it difficult to distinguish ship objects from the surrounding background, and 2) due to the multiscale characteristics of ship objects, there are numerous false negatives in the detection results, especially for small objects. To alleviate the above problems, we propose a one-stage ship detector with strong robustness against scale changes and various interferences. First, to mitigate the disturbance from complex background, a coordinate attention module (CoAM) is introduced for obtaining more representative semantic features to accurately locate and distinguish ship objects. Second, a receptive field increased module (RFIM) is devised to capture multiscale contextual information to improve the detection performance for ships with diverse scales. Finally, we verify the robustness of our method on several public SAR datasets, i.e., SAR-Ship-Dataset, high-resolution SAR images dataset (HRSID), and SAR ship detection dataset (SSDD). The experimental results demonstrate that the proposed method has a competitive performance, exceeding other state-of-the-art methods by at least 2.6% AP_{50} on HRSID.**

***Index Terms*— Deep learning, ship detection, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) is a microwave active probing imaging technique with the ability of all-day, all-weather, wide mapping, and high resolution. These advantages make it the most suitable imaging method for object detection and ocean monitoring in the field of space technology [1]–[3].

Manuscript received July 6, 2021; revised September 30, 2021; accepted November 9, 2021. Date of publication November 12, 2021; date of current version February 17, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61976166, Grant 62036007, Grant 62176195, Grant 61922066, and Grant 61876142; in part by the Key Research and Development Program of Shaanxi under Grant 2021GY-030; in part by the Innovation Capacity Support Plan of Shaanxi Province under Grant 2020KJXX-027; in part by the Fundamental Research Funds for the Central Universities under Grant JB210115; and in part by the Open Research Projects of Zhejiang Lab. *(Corresponding author: Xinbo Gao.)*

Xi Yang, Xin Zhang, and Nannan Wang are with the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mail: yangx@xidian.edu.cn; xzhang_98@stu.xidian.edu.cn; nnwang@xidian.edu.cn).

Xinbo Gao is with the State Key Laboratory of Integrated Services Networks, School of Electronic Engineering, Xidian University, Xi'an 710071, China, and also with the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: xbgao@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3128060

Ship detection, including localization and classification, is a momentous application of SAR images and plays an important role in the military and civil fields. At the moment, abundant SAR-imaging technologies provide a large amount of available data, which makes ship detection in SAR images a particularly important task. Therefore, there is an urgent need for accurate and robust ship detection algorithms.

The constant false-alarm rate (CFAR) algorithm [4]–[7], a frequently used traditional approach for SAR ship detection, is based on the statistical distributions of the sea clutter and adaptive thresholding strategy. However, this method is highly dependent on the distributions predefined by humans and greatly affected by the background statistical region. Consequently, it is difficult to settle the problem of ship accurate detection with the traditional approaches for SAR images under the interference of complex background. Recently, due to the powerful feature representations in convolutional neural network (CNN), object detection technology based on deep learning has attracted extensive research and achieved great success. Currently, object detectors based on deep learning can be primarily divided into two families. The one is two-stage detectors such as region-based CNN (R-CNN) [8]–[10], and the most representative one is Faster R-CNN. Another is one-stage detectors, such as SSD [11], YOLO [12]–[14], and RetinaNet [15].

Although the above detection algorithms have superior performance compared with traditional methods [4]–[7], it is difficult to apply them directly to ship detection in SAR images. The main obstacles to its successful application are as follows.

1) Complex background with extensive noise, particularly in the shore area. In virtue of the unique imaging technology of SAR, there is much speckle noise in SAR images [see Fig. 1(a)]. Besides, ship detection in SAR images has many disturbances, including land, islets, and sea clutter, which will produce useless false alarms.
2) Multiscale, especially small object detection. Because of multiresolution imaging modes and various ship shapes, multiscale, especially small objects are the characteristics of SAR images [see Fig. 1(b)]. It is worth noting that when small objects are mapped to the final feature map, there is little information for location refinement and classification, which emerges high false negatives and cuts down the detection performance.

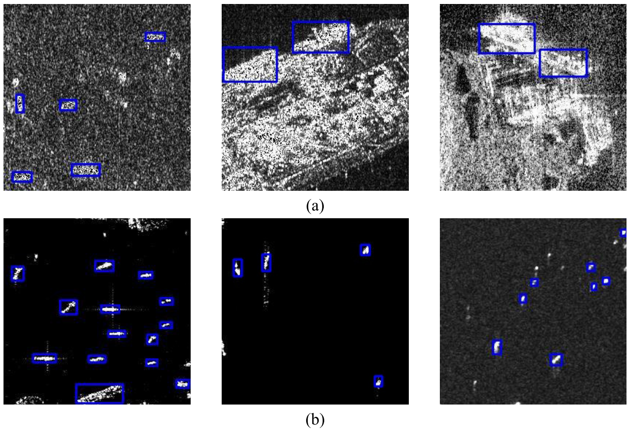

Fig. 1. Examples of SAR images in SAR-Ship-Dataset with distinct scales and backgrounds, showing (a) complex backgrounds and (b) multiscale characteristics. (Ground truths indicated with blue bounding boxes.)

3) Weak generalization ability. Most detection algorithms show weak robustness against different datasets, and specifically, the detection accuracy is high on a specific dataset, but the performance on other similar datasets is unsatisfactory.

Taking the above matters into consideration, we propose a robust detector with superior performance for multiscale ship detection with complex background in massive SAR images. First, to mitigate the disturbance from complex background in SAR images, a coordinate attention module (CoAM) is introduced and embedded into backbone to accurately locate and distinguish ship objects. Second, to obtain an enhanced feature pyramid, receptive field increased module (RFIM) is devised for capturing multiscale contextual information to improve the detection performance for ships with diverse scales, especially small objects. In short, we achieve a detection system that is robust enough to multiscale objects, complex background, as well as datasets.

The main contributions of this article are as follows.

1) For the complicated background, CoAM is introduced and embedded into backbone to obtain stronger semantic features that better represent objects, thereby dropping false alarms.
2) For multiscale ship detection, RFIM is designed to enrich the receptive field and capture multiscale contextual information at the same time, which contributes to reduce false negatives and improve the detection performance.
3) To verify the effectiveness of the proposed method, we make extensive experiments on several SAR image datasets, i.e., SAR-Ship-Dataset, high-resolution SAR images dataset (HRSID), and SAR ship detection dataset (SSDD). Our method reaches the state-of-the-art performance and the detection accuracy up to 96.0%, 92.7%, and 95.6%.

II. Related Work

A. Object Detection

Object detection, which intends to classify and locate objects in an image or video, is a fundamental task in computer vision. It has been widely concerned in recent years due to its wide applications. Object detectors based on deep learning can be primarily divided into two families: two-stage detectors and one-stage detectors.

1) Two-Stage Detectors: As the term suggests, the detection process includes two steps. The algorithms first generate a sparse set of proposals, and then, these proposals of interest are classified and regressed. R-CNN [8] first applies a deep learning method to object detection. Inspired by R-CNN and SPP-Net [16], Fast R-CNN [9] proposes a region of interest (RoI) pooling layer to improve detection accuracy and speed. Later, Faster R-CNN [10] extends Fast R-CNN by replacing selective search with the novel region proposal network (RPN) to extract region proposals. Significantly, it is the first end-to-end detection algorithm. After that, Mask R-CNN [17] is developed to predict an object mask by adding a mask branch to Faster R-CNN. Based on Faster R-CNN, region-based fully convolutional networks (R-FCNs) [18] greatly improve the detection speed through network sharing calculations. In addition, to solve the problem of multiscale change in object detection, feature pyramid networks (FPNs) [19] implement a feature pyramid, which integrates multilayer feature information and it is widely used by the subsequent algorithms. Libra R-CNN [20] solves the imbalanced problems when training through IoU balanced sampling, balanced feature pyramid, and balanced L1 loss. Cascade R-CNN [21] alleviates the problems of overfitting at training and quality mismatch at inference by multistage architecture.

2) One-Stage Detectors: Different from two-stage detectors, one-stage detectors directly classify and regress object on each position of the feature maps without RPN. You only look once (YOLO) series algorithms [12]–[14], [22] regard detection as a regression problem and directly adopt a convolution neural network to implement the entire detection process. Single-shot multibox detector (SSD) [11] detects object of different sizes on the multiscale feature map. RetinaNet [15] overcomes the problem of imbalance between positive and negative samples in one-stage algorithms by focal loss. Recently, anchor-free models emerge endlessly, which do not need to design anchors based on prior knowledge. The representative algorithms include key-point-based algorithms such as CornerNet [23] as well as anchor-point-based algorithms such as FCOS [24] and CenterNet [25].

B. Ship Detection

Due to the requirements of military and civil fields, researchers have made extensive efforts in ship detection. Traditional SAR ship detection methods are more based on CFAR. Wang *et al.* [26] improved the CFAR algorithm by fusing intensity and spatial information. Shi *et al.* [27] extracted object features by directional gradient histograms to separate ship object and background. However, the performance of traditional methods such as the above [26], [27] is far lower than algorithms based on deep learning. Benefiting from deep learning, ship detection based on CNN has once again attracted the attention of researchers. Fan *et al.* [28] applied a fully convolutional network with compact polarimetric SAR images for ship detection. Kang *et al.* [29] used Faster R-CNN combined with CFAR to obtain the final

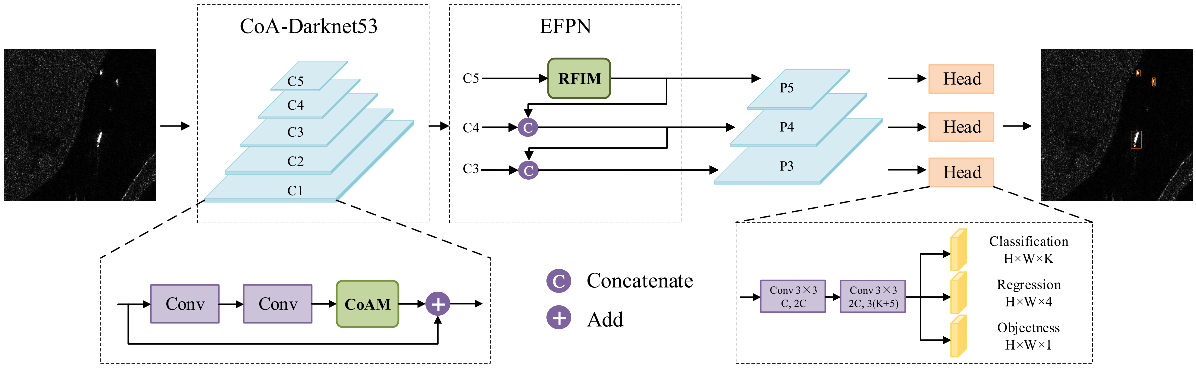

Fig. 2. Overall framework of our method, in which CoAM and RFIM are the proposed modules, and EFPN is enhanced FPNs including RFIM. The input image is fed to the backbone with the designed CoAM to extract features, then the enhanced feature pyramid is obtained through the proposed RFIM, and finally, the detection results are output through the head. Some internal structures of the network are at the bottom of the figure. The block diagram at the bottom left shows the specific location where we insert the CoAM into the residual block of backbone. The block diagram at the bottom right shows the structure of the head, in which the BCE loss is used for classification and objectness and the GIoU loss is used for regression.

ship detection results. To improve detection performance and speed, Wang *et al.* [30] incorporated transfer learning within SSD detector. Facing the challenge of large-scale SAR images, Cui *et al.* [31] introduced the spatial shuffle-group enhance (SSE) attention module into CenterNet to extract stronger semantic features and reduce false positives. Guo *et al.* [32] employed the rotational Libra R-CNN to address three imbalances in feature level, sample level, and objective level. Later, Guo *et al.* [33] continually proposed CenterNet++, which consists of feature refinement module, feature pyramids fusion module, and head enhancement module to resolve the problems of small object detection and complex background. Furthermore, Li *et al.* [34] proposed a new multidimensional deep learning network based on the complementary characteristics of spatial and frequency domains, which improved the detection performance. In recent work, to relief the problem of dense ship detection, Yu *et al.* [35] proposed CR2A-Net, which achieved high-precision detection of ships in any orientation through three parts: data preprocessing module, rotated anchor-aided detection module, and rotated align convolution layer. To solve the problem of multiscale ship detection, Cui *et al.* [36] proposed a novel multiscale ship detection method based on a dense attention pyramid network (DAPN) in SAR images. The salient features refined by convolutional block attention module (CBAM) are integrated with global unblurred features to improve the accuracy effectively in SAR images. Lin *et al.* [37] presented squeeze and excitation rank (SER) Faster R-CNN, which used multiscale feature map concatenation strategy to improve the quality of shared feature maps and suppressed redundant subfeature maps through SE mechanism and rank modification, thereby improving detection performance. Zhang and Zhang [38] exploited a lightweight SAR ship detector ShipDeNet-20 with 20 convolution layers and less than 1-MB model size, which greatly improved the detection speed without reducing accuracy.

The proposed method is a robust one-stage detection algorithm. We introduce innovative modules to improve the network performance and robustness by solving the complex background and multiscale object problems faced by SAR ship detection.

III. Methodology

The overall architecture of our method is shown in Fig. 2. Specifically, first, we integrate CoAM into the Darknet backbone in order to better capture all ship object information in complex backgrounds, so as to obtain more representative feature maps. Then, the multiscale context information is extracted by RFIM to get the enhanced feature pyramids. The details are described in the following subsections.

A. Coordinate Attention Module (CoAM)

The background of SAR images is extraordinarily complex, especially for the berthing ships. For this purpose, we design a lightweight module CoAM as shown in Fig. 3, which decouples the 2-D space into vertical and horizontal directions by two 1-D global pooling operations. It means that the coordinate information of the two directions is embedded into channels correspondingly. In this way, the network can not only retain the accurate location information of one spatial direction but also capture the long-range dependencies of another spatial direction, which is very essential for the object detection task. After that, the module can learn to get the location- and direction-aware feature maps, which can enhance the representations of the ships from complex background. The following will be elaborated.

1) Spatial Directions Decoupling: The famous SENet [39] obtains the global spatial information through 2-D global average pooling, which loses the object location information unfortunately. In contrast, we apply two 1-D average pooling operations to decouple the vertical and horizontal directions so that we can keep the position information in one direction, meanwhile capturing the long-range dependences in the other direction. Specifically, for the input features $I = [i_1, i_2, \ldots, i_C] \in \mathbb{R}^{C\times H\times W}$, we use a pooling kernel size $H \times 1$ to encode vertical coordinate. The element of the cth channel at vertical position h is calculated by

$$x_c(h) = \frac{1}{W}\sum_{j=1}^{W} i_c(h, j). \quad (1)$$

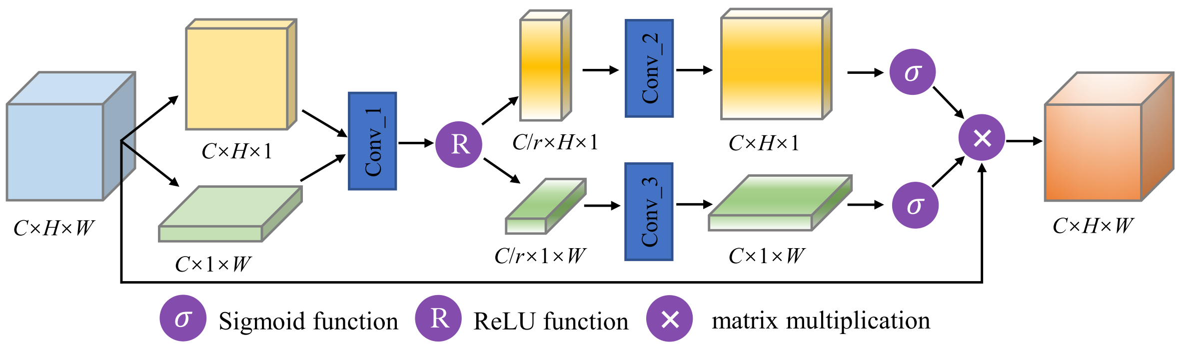

Fig. 3. Architectural details of the proposed CoAM. The modules of Conv_1, Conv_2, and Conv_3 are 1×1 convolutional layer, and it is worth noting that Conv_1 is shared.

Moreover, the pooling kernel size $1 \times W$ is used to encode the horizontal coordinate. The element of the cth channel at the horizontal position w is formulated as

$$x_c(w) = \frac{1}{H}\sum_{j=1}^{H} i_c(j, w). \tag{2}$$

Through the above process, the information of one spatial direction is aggregated to another spatial direction. The features generated in this way have not only direction-awareness but also location-awareness, which can prompt the network to pay more attention to the object RoI.

2) Adaptive Coordinate Attention: After the above-mentioned spatial directions decoupling, we get two tensors with the size of $X^{(h)} \in \mathbb{R}^{C\times H\times 1}$ and $X^{(w)} \in \mathbb{R}^{C\times 1\times W}$, respectively, which have global receptive field and accurate position information. Then, we will carry out convolution operation to make full use of this information, that is to say, to generate adaptive coordinate attention. Specifically, the implementation is defined as

$$\tilde{X}^{(h)} = \mathrm{ReLU}\big(\mathrm{BN}\big(\mathrm{Conv}\big(X^{(h)}\big)\big)\big) \tag{3}$$
$$\tilde{X}^{(w)} = \mathrm{ReLU}\big(\mathrm{BN}\big(\mathrm{Conv}\big(X^{(w)}\big)\big)\big) \tag{4}$$

where $\mathrm{Conv}(\cdot)$ is a shared 1×1 convolutional layer, $\mathrm{BN}(\cdot)$ is the batch normalization, and $\mathrm{ReLU}(\cdot)$ is the ReLU activation function. In addition, the number of channels for tensors $\tilde{X}^{(h)}$ and $\tilde{X}^{(w)}$ is C/r, where hyperparameter r is the reduction ratio. r is introduced to reduce parameter overhead and we set it to 32. Then, the tensors $\tilde{X}^{(h)}$ and $\tilde{X}^{(w)}$ are fed separately to diverse 1×1 convolutional layers to restore the original number of channels, yielding

$$Y^{(h)} = \sigma\left(\mathrm{Conv}\big(\tilde{X}^{(h)}\big)\right) \tag{5}$$
$$Y^{(w)} = \sigma\left(\mathrm{Conv}\big(\tilde{X}^{(w)}\big)\right) \tag{6}$$

where σ is the sigmoid activation function, and the outputs $Y^{(h)}$ and $Y^{(w)}$ are regarded as adaptive coordinate attention weights.

3) Improved Feature Maps: According to the outputs $Y^{(h)}$ and $Y^{(w)}$, we receive attention maps along both the vertical and horizontal directions. Considering them comprehensively, the network can locate and identify the RoI more accurately.

TABLE I
SETTINGS OF MAX-POOLING LAYERS IN RFIM

Layers	Kernel size	Stride	Padding
max pooling_1	5×5	1	same
max pooling_2	9×9	1	same
max pooling_3	13×13	1	same

Consequently, the final output of CoAM O is expressed as

$$o_c(i, j) = i_c(i, j) \times y_c^{(h)}(i) \times y_c^{(w)}(j). \tag{7}$$

Inspired by previous work, we integrate CoAM into the residual block of Darknet53, as shown in the lower left block diagram of Fig. 2. Ultimately, the refined backbone can extract improved semantic features.

B. Receptive Field Increased Module (RFIM)

As we mentioned at the beginning, ship objects in SAR images are multiscale, especially with plenty of small objects, so multiscale information is quite important for ship detection task. For this purpose, we design multiple parallel branches with different parameters. Each branch extracts feature maps at different spatial scales based on the respective receptive field, and then, a spatial pyramid containing rich scale information is constructed. After that, the important scale features are adaptively selected through the convolutional layer to enhance significant information of feature maps at specific scales corresponding to different scales of detected ships. The whole module is called RFIM, which constructs a spatial pyramid through various receptive fields and extracts multiscale spatial contextual information that is conducive to the detection of multiscale, especially small ship objects.

The detailed structure of the module (see Fig. 4) is composed of several pooling layers and convolution layers. In particular, we use three parallel max-pooling layers to enrich the receptive field. The specific parameters are shown in Table I. Unfortunately, although the pooling operation does not add extra parameters, some information is lost. Therefore, an additional branch of dilated convolution layer (kernel size = 3×3, stride = 1, output channels = 512, dilation = 4, and padding = 4) is added to expand the receptive field

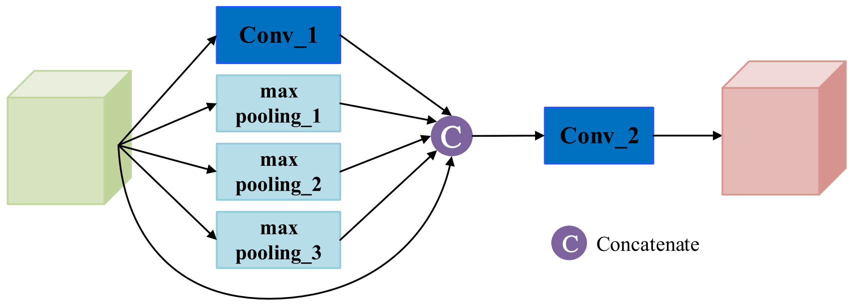

Fig. 4. Flow diagram of the proposed RFIM. Three parallel max-pooling layers have different convolution kernel sizes. Here, module Conv_1 represents the special dilated convolutional layer and module Conv_2 means the 1×1 ordinary convolutional layer.

TABLE II
SETTINGS OF ANCHOR BOXES ON DIFFERENT DATASETS

Datasets	Anchor Boxes (Width, Height)
SAR-Ship-Dataset	(10×11), (13×24), (26×16), (16×43), (25×29), (41×23), (38×36), (29×61), (56×45)
HRSID	(19×9), (12×17), (30×21), (20×46), (63×22), (40×47), (80×43), (88×106), (263×248)
SSDD	(13×23), (29×27), (16×50), (49×43), (30×90), (98×63), (59×162), (175×78), (142×250)

TABLE III
RESULT COMPARISONS UNDER DIFFERENT SETTINGS OF THE REDUCTION RATIO r IN COAM

Reduction ratio r	Params(M)	AP_{50}(%)
8	64.09	95.0
16	62.81	95.1
32	62.16	**95.3**
64	61.87	95.0

TABLE IV
EXPERIMENTAL RESULTS OF EACH PROPOSED MODULE ON SAR-SHIP-DATASET

CoAM	RFIM	AP_{50}(%)	Runtime(ms)
		94.4	7.0
✓		95.3	14.1
	✓	95.1	7.2
✓	✓	96.0	15.0

without losing resolution, so as to compensate for these losses to a certain extent. Finally, to preserve spatial and semantic information in the feature maps with different receptive fields as completely as possible, these features, including input features, are concatenated, followed by a 1×1 convolution layer to merge information, meanwhile adjusting the number of channels. It can be seen that RFIM has achieved a good performance without adding parameters as much as possible, which can effectively extract multiscale context information and is efficient in detecting small ships.

In short, we design a module integrating different receptive fields. It realizes the interaction and transmission of context information through pooling layers with different kernel sizes and dilated convolution layer, which can effectively alleviate the problem of false negatives for small ships.

IV. EXPERIMENT

A. Datasets and Evaluation Metric

We evaluate the proposed method on multiple public datasets, including SAR-Ship-Dataset [40], HRSID [41], and SSDD [42]. These datasets have various scenarios and contain ship objects at different scales. They can be used to develop object detectors for multiscale and small object detection. The SAR-Ship-Dataset labeled by SAR experts was created using Gaofen-3 images and Sentinel-1 images. It consists of 43 819 images of 256×256 pixels. We randomly divide the dataset into training, validation, and testing sets according to the ratio of 7:2:1. For the high-resolution SAR dataset HRSID, these are constructed from Sentinel-1B, TerraSAR-X, and TanDEMX imageries. There are 5604 SAR images with 800×800 pixels and we perform experiments according to the dataset division of the original paper. For SSDD, images in this dataset are from multiple sensors, and it is composed of 1160 images of approximately 500×500 pixels. Image indexes' suffixes 1 and 9 are set as the test set following the author's suggestion.

To evaluate the performance of different methods, AP_{50} is used as the main evaluation metric. Specifically, average precision (AP) is the area under the precision–recall curve. AP_{50} is AP at 0.5 of IoU threshold. In addition, we use precision, recall, F1, inferencing time, models' number of parameters, and FLOPs as an auxiliary evaluation metrics, where precision, recall, and F1 are taken at the confidence score threshold of 0.3. The calculation formula is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

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

TABLE V

COMPARISON OF EVALUATION METRICS OF DIFFERENT METHODS ON SAR-SHIP-DATASET. EACH SECTION SHOWS THE RESULTS OF R-CNN-BASED ALGORITHMS, ANCHOR-BASED SINGLE-STAGE ALGORITHMS, ANCHOR-FREE ALGORITHMS, AND OUR METHOD WITH DIFFERENT BACKBONES. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Method	Backbone	Precision(%)	Recall(%)	F1(%)	AP_{50}(%)	Runtime(ms)	Params(M)	FLOPs(G)
Faster R-CNN [10]	ResNet-101-FPN	91.0	91.0	91.0	91.0	25.1	60.1	31.1
Libra R-CNN [20]	ResNet-101-FPN	87.8	91.4	89.6	91.5	25.5	60.4	31.2
Cascade R-CNN [21]	ResNet-101-FPN	92.0	91.6	91.8	92.0	34.0	87.9	58.9
CR2A-Net [35]	ResNet-101-FPN	91.7	92.2	91.9	90.1	41.7	88.6	59.7
DAPN [36]	ResNet-101-FPN	91.0	91.4	91.2	91.9	27.8	63.8	39.7
SSD512 [11]	SSDVGG	90.9	91.5	91.2	94.2	23.7	24.4	87.7
RetinaNet [15]	ResNet-101-FPN	84.5	93.3	88.7	93.8	25.8	55.1	17.9
YOLOv3 [14]	Darknet-53	91.3	94.3	92.8	94.4	7.0	61.5	12.4
YOLOv4 [22]	CSPDarknet-53	85.7	92.7	89.1	93.2	**6.9**	64.3	11.3
FCOS [24]	ResNet-101-FPN	92.6	93.4	93.0	94.9	22.8	50.8	17.5
CenterNet [25]	DAL-34	84.6	93.5	88.8	95.0	12.8	**20.2**	**6.5**
CenterNet++ [33]	DAL-34	85.4	93.5	89.3	94.9	13.8	20.3	6.7
Ours	Darknet-53	93.7	**95.3**	**94.5**	**96.0**	15.0	65.8	12.6
	CSPDarknet-53	**93.8**	94.6	94.2	95.9	18.1	51.4	9.9
	ResNet-101	**93.8**	94.4	94.1	95.2	22.1	73.5	13.8

TABLE VI

COMPARISON OF EVALUATION METRICS OF DIFFERENT METHODS ON HRSID

Method	Backbone	Precision(%)	Recall(%)	F1(%)	AP_{50}(%)	Runtime(ms)	Params(M)	FLOPs(G)
Faster R-CNN [10]	ResNet-101-FPN	88.8	77.5	82.8	78.2	56.1	60.1	181.9
Libra R-CNN [20]	ResNet-101-FPN	83.1	77.9	80.4	77.5	57.6	60.4	182.6
Cascade R-CNN [21]	ResNet-101-FPN	89.9	79.3	84.3	79.2	64.7	87.9	209.7
CR2A-Net [35]	ResNet-101-FPN	88.5	78.9	83.4	80.9	77.3	88.6	212.5
DAPN [36]	ResNet-101-FPN	88.9	77.6	82.9	79.8	74.9	63.8	266.1
SSD512 [11]	SSDVGG	87.4	85.3	86.3	88.8	44.8	24.4	87.7
RetinaNet [15]	ResNet-101-FPN	69.8	83.8	76.2	82.5	55.0	55.1	175.4
YOLOv3 [14]	Darknet-53	90.6	78.2	84.0	87.2	26.0	61.5	121.0
YOLOv4 [22]	CSPDarknet-53	90.6	84.0	87.2	90.1	**22.4**	64.3	110.5
FCOS [24]	ResNet-101-FPN	91.9	79.5	85.3	86.6	50.9	50.8	170.6
CenterNet [25]	DAL-34	81.8	87.4	84.5	86.3	55.0	**20.2**	**63.3**
CenterNet++ [33]	DAL-34	82.2	87.3	84.7	86.3	54.5	20.3	64.9
Ours	Darknet-53	**92.7**	**88.1**	**90.3**	**92.7**	37.3	65.8	123.5
	CSPDarknet-53	90.1	86.6	88.3	91.3	40.1	51.4	96.6
	ResNet-101	88.4	86.7	87.6	91.3	53.0	73.5	133.9

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

B. Experiment Settings

All experiments are implemented on Ubuntu 16.04 system with PyTorch 1.6, CUDA 10.0, CUDNN 7.4.2, and TITAN RTX GPU with 24-GB memory. The ablation experiments and parameter analysis experiments in this article are all carried out on the SAR-Ship-Dataset. Because of the uniqueness of SAR images, our model is trained from scratch, that is to say, no pretrained weights are loaded. In addition, we apply a stochastic gradient descent (SGD) optimizer and cosine scheduler. The whole training procedure has 300 epochs and the batch size is set to 16. For the above three datasets, the input image size of the network is 256×256, 800×800, and 512×512. Especially, we use k-means clustering to yield bounding box priors and the settings of anchor boxes on different datasets are shown in Table II The total loss is composed of classification loss, regression loss, and objectness loss, where the BCE loss is used for classification and objectness and the GIoU loss is used for regression.

C. Parameter Analysis

As stated in Section III, a hyperparameter reduction ratio r is proposed in CoAM. r has a certain impact on the detection performance of ships. We attempt to change r to find a suitable

TABLE VII
COMPARISON OF EVALUATION METRICS OF DIFFERENT METHODS ON SSDD

Method	Backbone	Precision(%)	Recall(%)	F1(%)	AP_{50}(%)	Runtime(ms)	Params(M)	FLOPs(G)
Faster R-CNN [10]	ResNet-101-FPN	90.9	87.6	89.2	88.3	30.2	60.1	82.7
Libra R-CNN [20]	ResNet-101-FPN	88.6	88.6	88.6	89.9	30.2	60.4	83.0
Cascade R-CNN [21]	ResNet-101-FPN	94.3	89.9	92.0	89.5	38.8	87.9	110.5
CR2A-Net [35]	ResNet-101-FPN	94.0	87.8	90.8	89.8	67.2	88.6	112.0
DAPN [36]	ResNet-101-FPN	87.6	91.4	89.4	90.1	34.5	63.8	117.2
SSD512 [11]	SSDVGG	92.9	88.0	90.4	94.0	30.2	24.4	87.7
RetinaNet [15]	ResNet-101-FPN	81.6	92.3	86.6	89.6	30.2	55.1	71.8
YOLOv3 [14]	Darknet-53	90.7	**94.7**	92.6	95.0	**10.4**	61.5	49.6
YOLOv4 [22]	CSPDarknet-53	93.6	94.0	93.8	96.1	12.9	64.3	45.3
FCOS [24]	ResNet-101-FPN	94.4	85.6	89.8	88.7	25.9	50.8	69.8
CenterNet [25]	DAL-34	93.3	94.5	93.9	93.5	21.5	**20.2**	**25.9**
CenterNet++ [33]	DAL-34	92.6	94.5	93.6	92.7	21.5	20.3	26.6
Ours	Darknet-53	94.4	92.1	93.2	95.6	16.4	65.8	50.6
	CSPDarknet-53	93.1	90.6	91.8	95.3	19.1	51.4	39.6
	ResNet-101	**95.1**	94.5	**94.8**	**96.4**	24.4	73.5	54.9

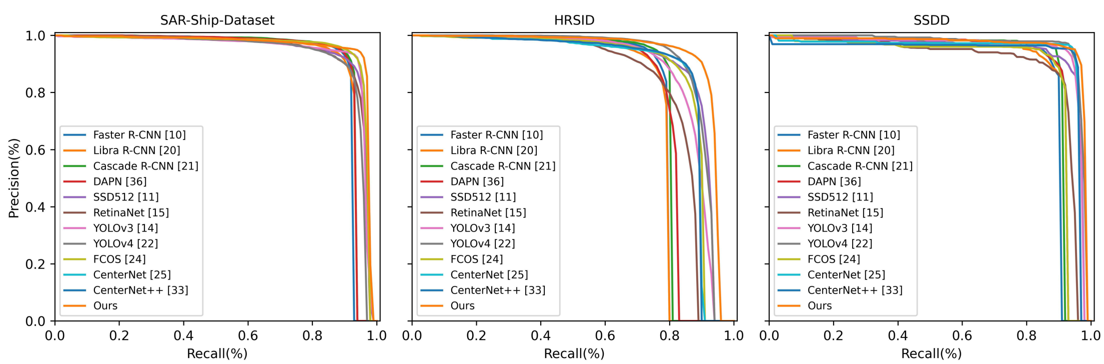

Fig. 5. Precision–recall curves of different methods on SAR-Ship-Dataset, HRSID, and SSDD.

value for superior performance. As shown in Table III, with the increasing of r, the number of parameters will obviously increase, but the performance has a tendency to increase first and then decrease. Considering comprehensively, we set r to 32 to achieve a good performance gain ($AP_{50} = 95.3\%$) while hardly increasing the amount of parameters. The reason is analyzed and, when r is small, the convolution layer will eliminate redundant information in channels, but with the increasing of r, some useful feature information will be lost, which is disadvantageous.

D. Ablation Study

In this article, two novel modules, CoAM and RFIM, are proposed. In order to intuitively contrast the validity of each module, we conducted an ablation study and the results are shown in Table IV.

1) CoAM: Compared with the baseline, the proposed method CoAM can improve AP_{50} from 94.4% to 95.3%. The improvement benefits from the fact that CoAM can extract precise spatial location information and obtain more representative feature maps through adaptive coordinate attention so that the network pays more attention to ship objects in complex background interference. In essence, CoAM can alleviate the influence of complex background on detection performance and the ablation study also proves its effectiveness.

2) RFIM: AP_{50} reaches 95.1% by adding lightweight module RFIM, which is 0.7% higher than baseline. This indicates that this module can capture multiscale contextual information with the inflow of various receptive fields, which is effective for multiscale, especially small ship detection.

Ultimately, by combining the modules we proposed, AP_{50} can reach 96.0%. It can be seen that the proposed CoAM and RFIM are effective in improving the performance of ship detection, and their interaction can further improve the performance of our network.

E. Comparison With the State-of-the-Art Methods

To confirm the feasibility and generalization ability of our method, we conduct experiments on three datasets of SAR-Ship-Dataset, HRSID, and SSDD and compare with existing state-of-the-art methods. As shown in Tables V–VII, it can be seen that while real-time detection is achieved, the detection performance of our method outperforms other comparison methods, reaching state-of-the-art (except performance on

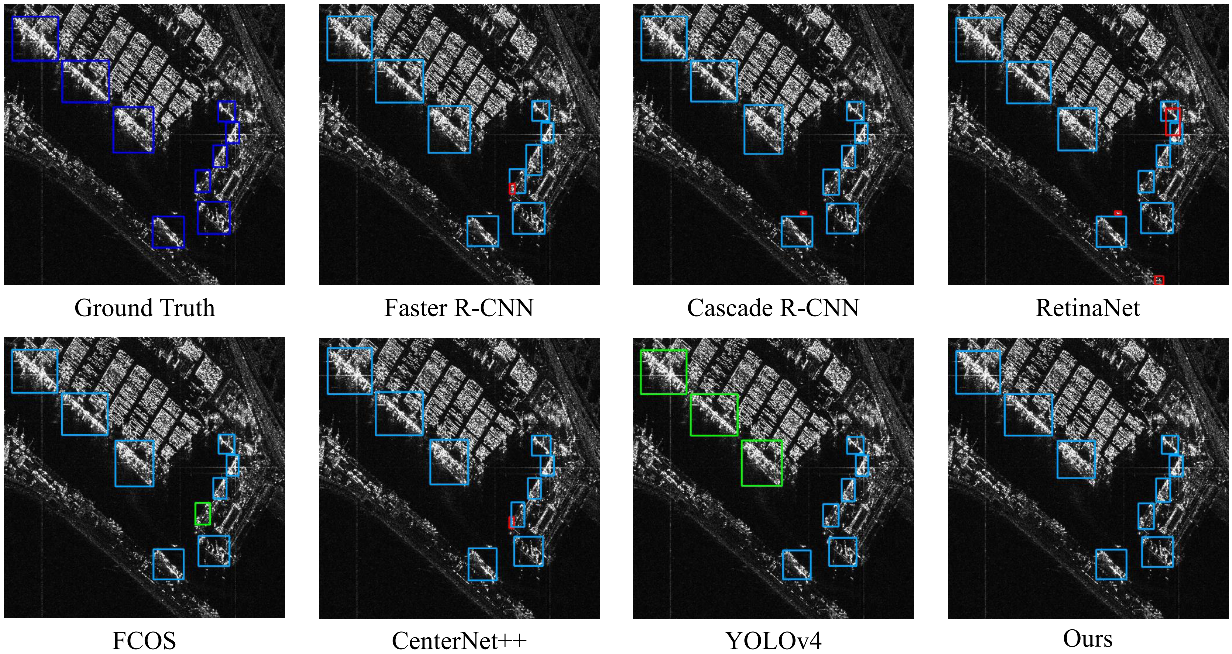

Fig. 6. Comparison of the detection results by different methods under complex background. The navy blue, light blue, green, and red boxes indicate the ground truths, true positives, false negatives, and false positives, respectively. (Unless otherwise specified, the following pictures are the same.)

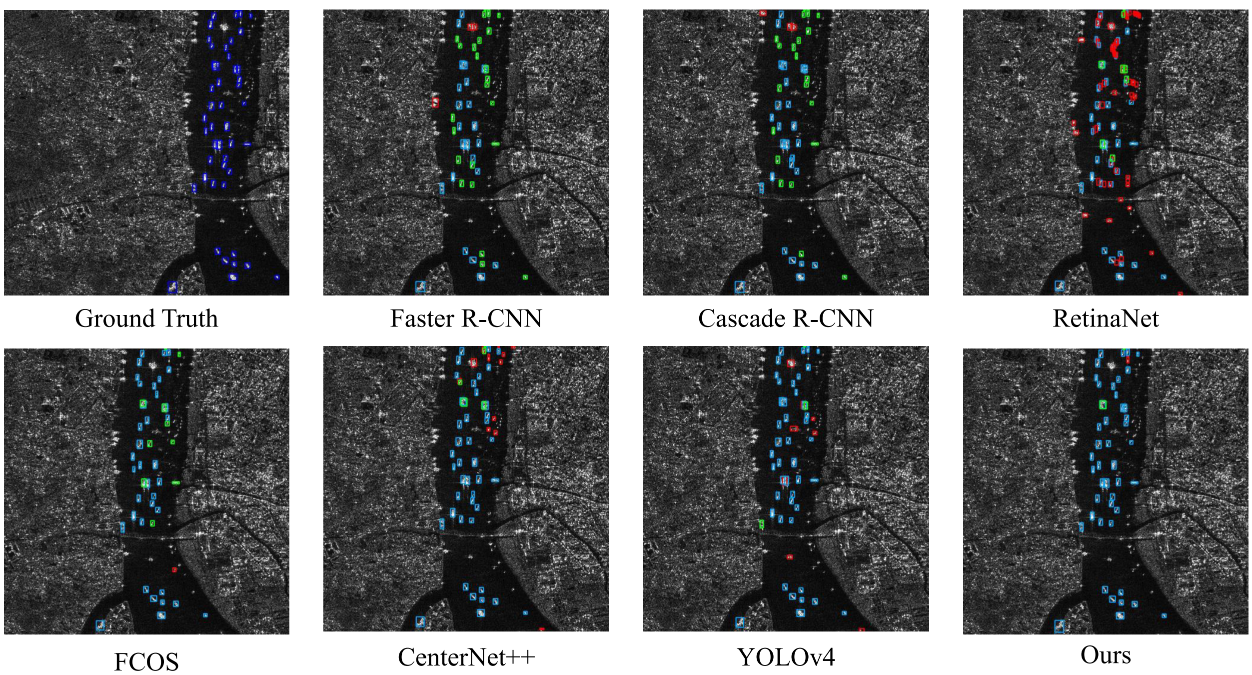

Fig. 7. Comparison of the detection results by different methods for small ships.

SSDD, which is slightly lower than YOLOv4). Not only that, we also draw the precision–recall curves of detection results with different methods (as shown in Fig. 5). The following is a detailed analysis of the experimental results.

1) SAR-Ship-Dataset: In general, the algorithms we compare are divided into three categories: R-CNN-based algorithms including two-stage and multistage, and anchor-based and anchor-free single-stage algorithms. On SAR-Ship-Dataset, the one-stage methods, especially the anchor-free methods, have superior performance. It is worth noting that our method can achieve 96.0% AP_{50}, which is better than all comparison methods. Compared with the classic R-CNN-based methods, i.e., Faster R-CNN, Libra R-CNN, and Cascade R-CNN, our method can improve by about 4%–5%. As for the common one-stage algorithms SSD and RetinaNet, ours can improve 1.8% and 2.2%, respectively. In addition, the proposed method is 1.6% higher than the YOLOv3 and also exceeds the YOLOv4 algorithm by 2.8%. Moreover, ours has 1.1% and 1.0% advantages over anchor-free FCOS and CenterNet separately. Compared with the improved ship detection algorithms CR2A-Net, DAPN, and CenterNet++, our method is competitive as well, which is 5.9%, 4.1%, and

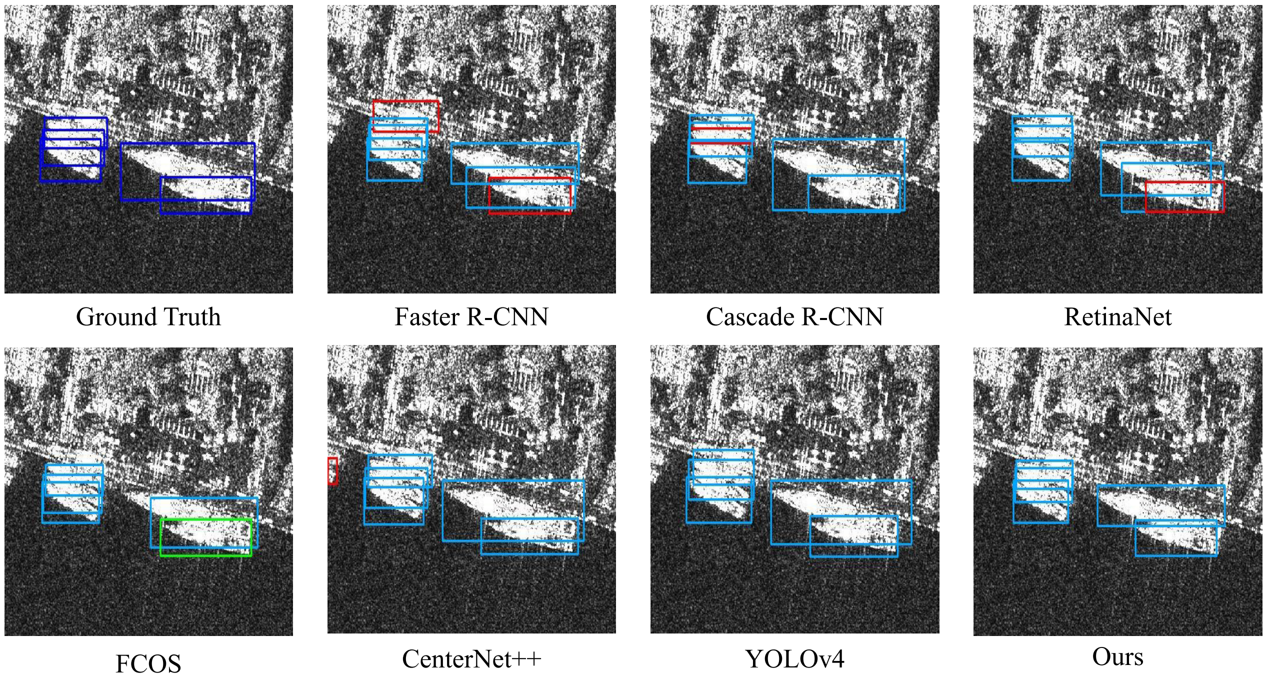

Fig. 8. Comparison of the detection results by different methods for densely arranged ships in SSDD.

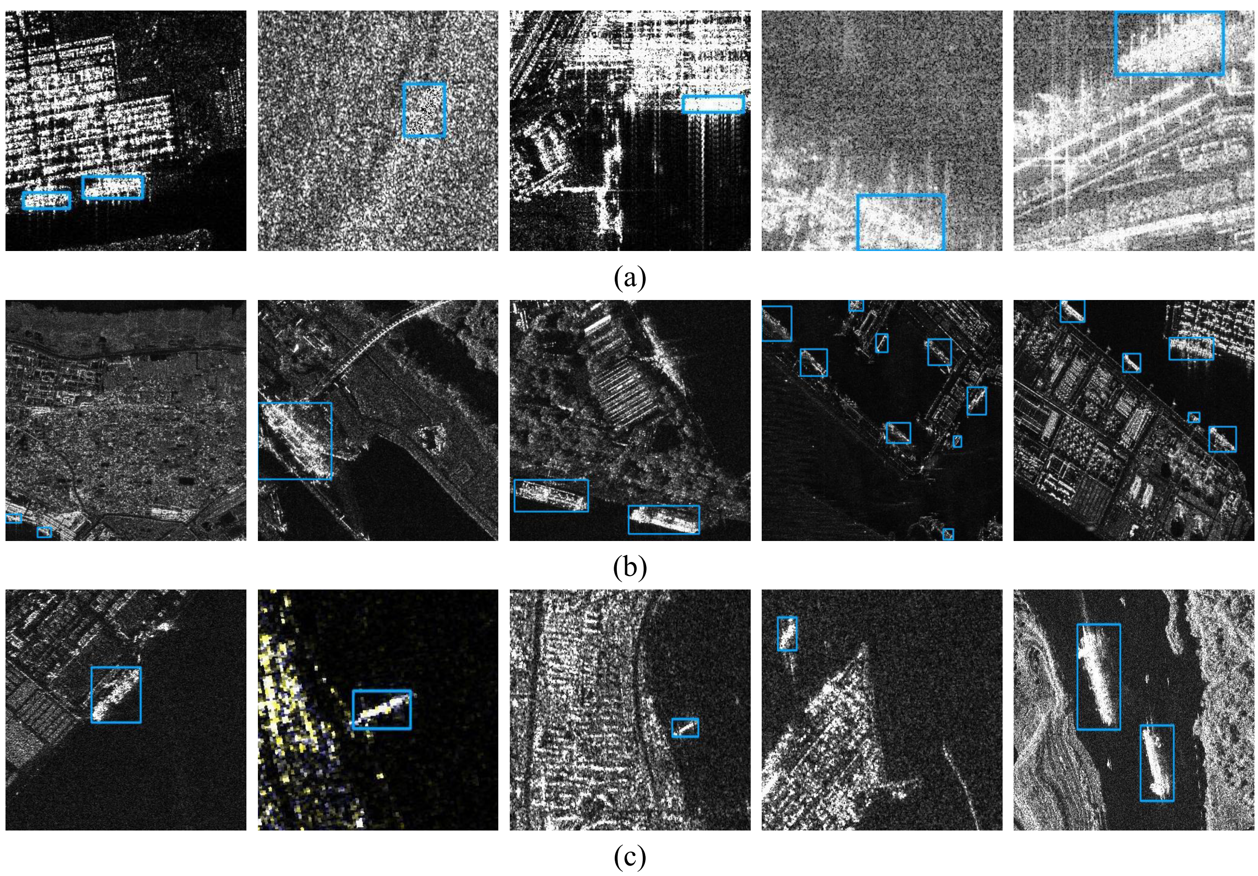

Fig. 9. Part of detection results of the proposed method for objects under complex background on various SAR datasets. All ships are correctly detected. (a) SAR-Ship-Dataset. (b) HRSID. (c) SSDD.

1.1% higher than the above algorithms, respectively. It is worth mentioning that the inferencing time of the method in this article is not as fast as some methods [14], [22], but it can fully meet the real-time detection requirement.

2) HRSID: The image background of the dataset is more complicated, and there are more small ship objects, so it can better reflect the effectiveness of the proposed method. Specifically, compared to the state-of-the-art methods, our

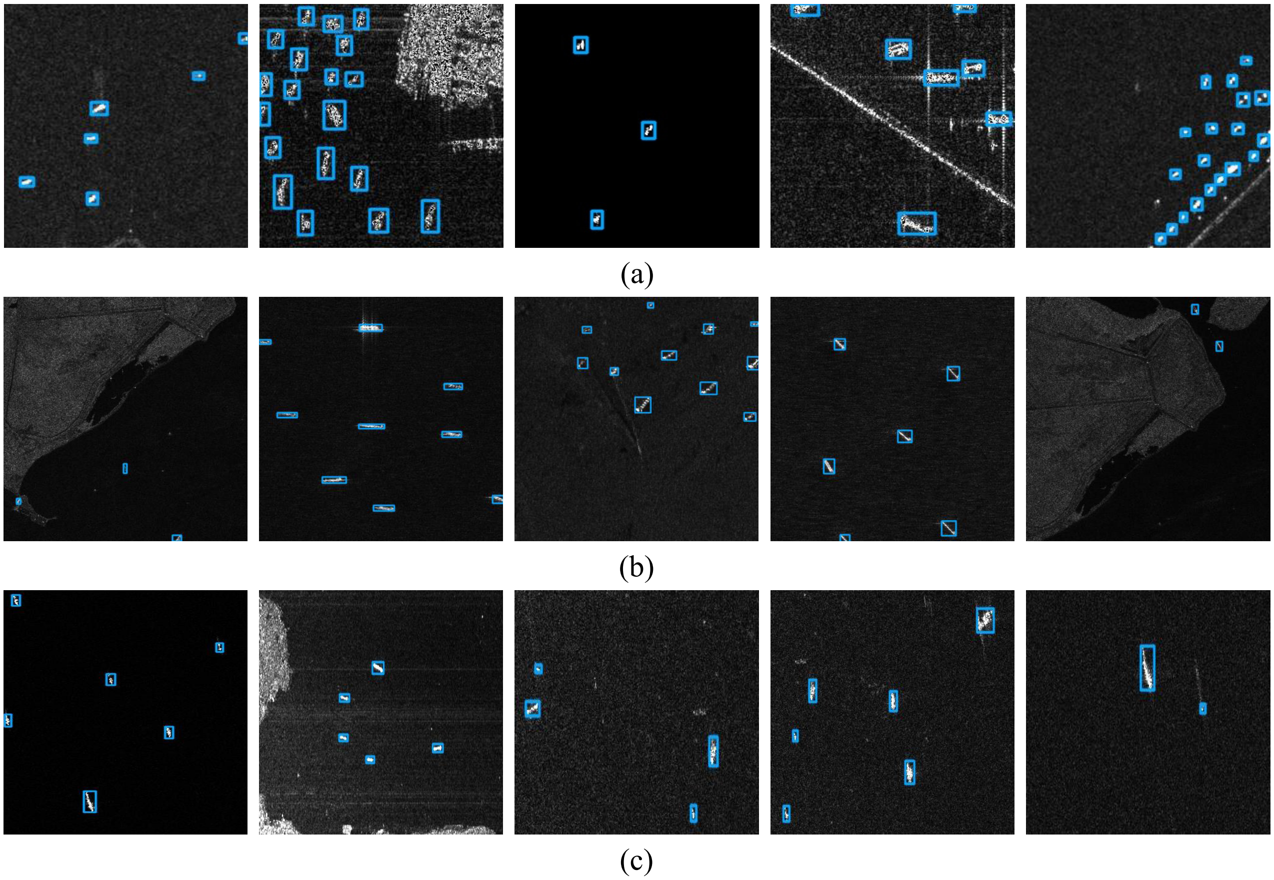

Fig. 10. Some detection results of our method for multiscale, especially small objects on three SAR datasets. All ships are correctly detected. (a) SAR-Ship-Dataset. (b) HRSID. (c) SSDD.

method can improve by about 2.6%–15.2%, benefitting from the proposed CoAM and RFIM. The CoAM obtains coordinate attention by extracting more characteristic features, which can eliminate the influence of complex background on the detection of ship objects to a certain extent. In addition, RFIM alleviates the challenge of multiscale objects, especially small objects by capturing multiscale context information. Even compared with YOLOv4 algorithm, our method is still 2.6% higher than it. Similarly, the proposed method outperforms other methods, achieving improvements of 14.5%, 15.2%, 13.5%, 3.9%, 10.2%, 6.1%, and 6.4% over Faster R-CNN, Libra R-CNN, Cascade R-CNN, SSD, RetinaNet, FCOS, and CenterNet, respectively. Furthermore, the detection performance is also better than CR2A-Net (80.9%), DAPN (79.8%), and CenterNet++ (86.3%).

3) SSDD: Experimental results on this dataset show that our method is competitive, but its performance is slightly inferior to YOLOv4 by about 0.5%. Apart from this, our method excels over other classical and achieves improvements of 7.3%, 5.7%, 6.1%, 5.8%, 5.5%, 1.6%, 6.0%, 0.6%, 6.9%, 2.1%, and 2.9% over Faster R-CNN, Libra R-CNN, Cascade R-CNN, CR2A-Net, DAPN, SSD, RetinaNet, YOLOv3, FCOS, CenterNet, and CenterNet++, respectively. It is worth mentioning that we also test the performance of the proposed method under different backbones, and the experimental results are competitive. Especially on SSDD dataset, using ResNet-101 as the backbone can further improve the overall performance of the detector, which outperforms all the compared algorithms and even YOLOv4.

In summary, our method can achieve remarkable detection accuracy. Moreover, the detection results on multiple datasets also verify the fine generalization ability of the proposed method.

F. Visual Results and Insight

To further demonstrate the superiority of our method, it is necessary to visualize the detection results. Figs. 6–8 show the visualization results of ours and other state-of-the-art methods. Among them, Fig. 6 shows the detection results under complex background, and Fig. 7 shows the detection results of various methods for small objects. It can be seen intuitively that our method is superior to other methods. In addition, the proposed method is also effective for dense ship detection, and the comparison results for densely arranged ships in SSDD are shown in Fig. 8. Our method is comparable with YOLOv4. There is no false negative and false positive in the detection results compared with other methods. The overall visual results of our method are shown in Figs. 9 and 10, where subimages (a), (b), and (c) are the ship detection results on SAR-Ship-Dataset, HRSID, and SSDD, respectively. Specifically, the pictures in Fig. 9 are the results of ship detection under the

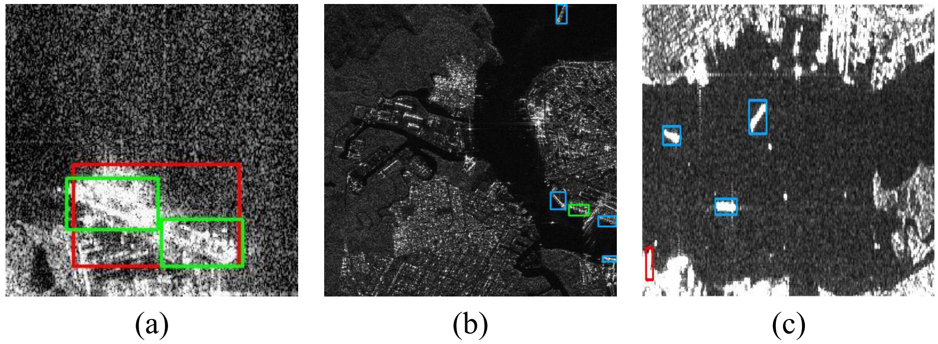

Fig. 11. Some cases of unsuccessful results in (a) SAR-Ship-Dataset, (b) HRSID, and (c) SSDD.

disruption of intricate background. The CoAM can alleviate this problem and realize accurate ship positioning and recognition of complex images. The pictures in Fig. 10 are the results of multiscale and small object detection. It can be seen that our method can obtain satisfactory detection results by capturing multiscale context information. In other words, the visualization results intuitively reflect that our method can effectively alleviate several major problems faced by ship detection in SAR images and shows good performance on multiple datasets.

G. Analysis of Undesirable Results

Although most of the detection results of our method are correct, some unexpected results are inevitable. Fig. 11 shows the results of the failure cases. It can be seen that under complex background, if the head and tail of the ships are closely connected, as shown in Fig. 11(a), our method cannot distinguish them well, thus treating the two real ships as one. Similarly, inshore ships are susceptible to interference from land noise so that sometimes the ship objects cannot be accurately located, resulting in few false positives and false negatives [see Fig. 11(b)]. In response to these abnormal situations, we plan to use an image super-resolution reconstruction network to increase the resolution of ship objects to obtain better discriminate features in the future.

V. Conclusion

In this article, we achieve a robust ship detector against scale changes and various interferences in massive SAR images. Specifically, CoAM is introduced and embedded into backbone to obtain stronger semantic features, which reduces the interference caused by complex background, such as offshore and inland. In addition, RFIM is introduced with the motivation of capturing multiscale contextual information to improve the detection performance of multiscale ship objects, especially small objects. Judging from our adequate experimental results, our method achieves a competitive performance compared with the state-of-the-art works while achieving real time on SAR ship detection. It is worth noting that although the proposed method achieves a superior detection performance, there are some undesirable detection results. We will borrow image super-resolution reconstruction technology to further alleviate these problems in future work.

References

[1] F. Wong and T. S. Yeo, "New applications of nonlinear chirp scaling in SAR data processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 5, pp. 946–953, May 2001.

[2] Z. Li, J. Wu, Y. Huang, Z. Sun, and J. Yang, "Ground-moving target imaging and velocity estimation based on mismatched compression for bistatic forward-looking SAR," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3277–3291, Jun. 2016.

[3] X. Yang *et al.*, "Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, p. 132, 2018.

[4] 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.

[5] H. Rohling, "Radar CFAR thresholding in clutter and multiple target situations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-19, no. 4, pp. 608–621, Jul. 1983.

[6] 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.

[7] J. Ai, X. Qi, W. Yu, Y. Deng, F. Liu, and L. Shi, "A new CFAR ship detection algorithm based on 2-D joint log-normal distribution in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 806–810, Oct. 2010.

[8] 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.*, Jun. 2014, pp. 580–587.

[9] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2015, *arXiv:1506.01497*.

[11] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2016, pp. 21–37.

[12] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

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

[14] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[15] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[16] 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.

[17] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.

[18] J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," 2016, *arXiv:1605.06409*.

[19] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[20] J. Pang, K. Chen, J. Shi, H. Feng, W. Ouyang, and D. Lin, "Libra R-CNN: Towards balanced learning for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 821–830.
[21] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6154–6162.
[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
[23] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis.*, Sep. 2018, pp. 734–750.
[24] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9627–9636.
[25] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, *arXiv:1904.07850*.
[26] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.
[27] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4511–4523, Aug. 2014.
[28] Q. Fan *et al.*, "Ship detection using a fully convolutional network with compact polarimetric SAR images," *Remote Sens.*, vol. 11, no. 18, p. 2171, Sep. 2019.
[29] M. Kang, X. Leng, Z. Lin, and K. Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. Int. Workshop Remote Sens. Intell. Process. (RSIP)*, May 2017, pp. 1–4.
[30] 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, Aug. 2018.
[31] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.
[32] H. Guo, X. Yang, N. Wang, B. Song, and X. Gao, "A rotational Libra R-CNN method for ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5772–5781, Aug. 2020.
[33] H. Guo, X. Yang, N. Wang, and X. Gao, "A CenterNet++ model for ship detection in SAR images," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107787.
[34] D. Li, Q. Liang, H. Liu, Q. Liu, H. Liu, and G. Liao, "A novel multidimensional domain deep learning network for SAR ship detection," *IEEE Trans. Geosci. Remote Sens.*, early access, Mar. 11, 2021, doi: 10.1109/TGRS.2021.3062038.
[35] Y. Yu, X. Yang, J. Li, and X. Gao, "A cascade rotated anchor-aided detector for ship detection in remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, early access, Dec. 7, 2021, doi: 10.1109/TGRS.2020.3040273.
[36] Z. Cui, Q. Li, Z. Cao, and N. Liu, "Dense attention pyramid networks for multi-scale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8983–8997, Nov. 2019.
[37] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.
[38] T. Zhang and X. Zhang, "ShipDeNet-20: An only 20 convolution layers and <1-MB lightweight SAR ship detector," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 7, pp. 1234–1238, Jul. 2021.
[39] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[40] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, Mar. 2019.
[41] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A high-resolution SAR images dataset for ship detection and instance segmentation," *IEEE Access*, vol. 8, pp. 120234–120254, 2020.
[42] 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. (BIGSARDATA)*, Nov. 2017, pp. 1–6.

Xi Yang (Member, IEEE) received the B.Eng. degree in electronic information engineering and the Ph.D. degree in pattern recognition and intelligence system from Xidian University, Xi'an, China, in 2010 and 2015, respectively.

From 2013 to 2014, she was a Visiting Ph.D. Student with the Department of Computer Science, The University of Texas at San Antonio, San Antonio, TX, USA. In 2015, she joined the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, where she is an Associate Professor in communications and information systems. Her research interests include image/video processing, computer vision, and multimedia information retrieval.

Xin Zhang received the B.Eng. degree in communication engineering from Xidian University, Xi'an, China, in 2020, where he is currently pursuing the M.S. degree in electronic and communication engineering.

His research interests include deep learning and object detection.

Nannan Wang (Member, IEEE) received the B.Sc. degree in information and computation science from the Xi'an University of Posts and Telecommunications, Xi'an, China, in 2009, and the Ph.D. degree in information and telecommunications engineering from Xidian University, Xi'an, in 2015.

From September 2011 to September 2013, he was a Visiting Ph.D. Student with the University of Technology Sydney, Sydney, NSW, Australia. He is currently a Professor with the State Key Laboratory of Integrated Services Networks, Xidian University. He has published over 100 articles in refereed journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IJCV, and NeurIPS. His research interests include computer vision, pattern recognition, and machine learning.

Xinbo Gao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electronic engineering, signal, and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Post-Doctoral Research Fellow with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University. He is also a Cheung Kong Professor of the Ministry of Education of China, a Professor of pattern recognition and intelligent system with Xidian University, and a Professor of computer science and technology with the Chongqing University of Posts and Telecommunications, Chongqing, China. He has published six books and over 300 technical articles in refereed journals and proceedings. His research interests include image processing, computer vision, multimedia analysis, machine learning, and pattern recognition.

Prof. Gao is also a fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics. He has served as the general chair/co-chair, the program committee chair/co-chair, or a PC member for over 30 major international conferences. He is also on the Editorial Board of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier).