MLBR-YOLOX: An Efficient SAR Ship Detection Network With Multilevel Background Removing Modules

Jindong Zhang, Weixing Sheng, Hairui Zhu, Shanhong Guo, and Yubing Han

***Abstract*—On the remote sensing images of marine synthetic aperture radar (SAR), ship targets often occupy only a small part of an image, and the rest are all sea and coastal backgrounds. Existing neural networks based on SAR ship detection often directly detect an entire SAR image, which consumes a large amount of computing resources. In this article, a new marine SAR ship detection network, called multilevel background removing–you only look once X (MLBR-YOLOX), is proposed. First, a new plug-and-play module, called standalone spatial patch detector, is proposed to predetect the position of ships and filter out most of the sea backgrounds at the source image level. Second, a deep spatial feature detector is presented for detecting deep semantic features of the output of the backbone module to further reduce the computational cost of the neck and head modules. Finally, the original YOLOX network is adopted to locate and classify the pre-detected results. Experimental results on the SAR ship detection dataset and the high-resolution SAR images dataset indicate that the detection performance of MLBR-YOLOX is close to that of YOLOX, but the computational complexity is merely 23.97% and 12.50% of YOLOX's, respectively. Moreover, the experiment conducted on a large-scene Sentinel-1 SAR image illustrates that the proposed network has good migration application capability.**

***Index Terms*—Convolutional neural network (CNN), multilevel background removing modules, ship detection, synthetic aperture radar (SAR) image.**

I. Introduction

SYNTHETIC aperture radar (SAR) is an active remote sensing device that utilizes microwave imaging to gain large-area and high-resolution images (all day and all weather conditions). By actively transmitting and receiving microwave signals in a specified frequency band, SAR sensors can continuously get geographic information in complex weather environments and have strong adaptability to the changing marine environment [1], [2]. Therefore, SAR has been widely applied in marine ecosystem protection, resource exploration, and military reconnaissance [3], [4], [5]. Furthermore, SAR plays an important role in marine ship detection [6], [7]. Due to the special angular reflectivity of ships, the echo signal intensity of ships obtained by the SAR system is stronger than that of oceans. Thus, ships appear bright in SAR images, while the sea backgrounds often are dark. Nevertheless, SAR ship detection still faces many challenges. For example, some coasts and islands appear reflection characteristics similar to ships in SAR images and they may be misclassified as targets in ship detection. A majority of ships are considered small and densely distributed targets, which makes them harder to accurately detect. Meanwhile, the existence of speckle noise in marine SAR images seriously reduces the interpretability of images [8], [9], [10]. Moreover, most CNN-based ship detection networks only focus on detection accuracy, ignoring the computational complexity of networks, which increases the strain on a computing platform. Therefore, SAR ship detection is still a focus of marine research.

Conventional methods for the analysis of marine SAR images mainly include detection methods based on background clutter statistical distribution [11], [12], polarization characteristics [13], and saliency [14], among which the most common method is ship detection algorithms based on background clutter statistical distribution. The constant false alarm rate (CFAR) algorithm [15], [16] is one of the most widely studied and applied ship detection algorithms, which uses statistical distribution to model background clutter and calculates an adaptive threshold to determine whether a pixel belongs to the target area. On the basis of CFAR, a majority of researchers carried out SAR ship detection research works and proposed many improvement strategies. To improve the ship detection accuracy in the situation of multiscale targets, Dai et al. [11] introduced an object proposal generator based on gradient features and regarded the object proposals as the CFAR guard windows. Li et al. [12] proposed an improved two-stage superpixel-level CFAR detection algorithm, using a global CFAR detector based on the weighted information entropy to find out candidate target superpixels and a local detector to clutter parameter estimation and overcome the multitarget situations. To reduce the interference of SAR ambiguities and sea clutter, Leng et al. [15] presented a kernel density estimation algorithm to estimate the spatial distribution and proposed a bilateral CFAR ship detector combining both the intensity distribution and spatial distribution of SAR images. Nevertheless, these traditional methods are usually designed in a cumbersome way that involves manual steps and heavily relies on background conditions. When multiscale ships appear in a

Manuscript received 2 February 2023; revised 6 March 2023 and 24 April 2023; accepted 23 May 2023. Date of publication 29 May 2023; date of current version 16 June 2023. This work was supported by the National Natural Science Foundation of China under Grant 61971224, Grant 62001227, and Grant 62001232. *(Corresponding author: Weixing Sheng.)*

The authors are with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: zhangjindong@njust.edu.cn; shengwx@njust.edu.cn; hairui.zhu@njust.edu.cn; guosh@njust.edu.cn; hanyb@njust.edu.cn).

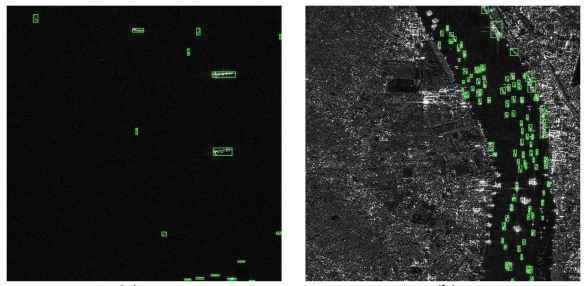
Digital Object Identifier 10.1109/JSTARS.2023.3280741

new complex background, the detection performance becomes not stable and the robustness of these algorithms is poor.

With the development of machine learning technology, target detection methods based on convolutional neural networks (CNN) have become a research priority, and many CNN methods have emerged and achieved cracking performance. The existing CNN-based detectors can be divided into two categories: the two-stage methods and the one-stage methods. For the two-stage detection methods, the first stage is generating some region candidate proposals through region proposal algorithms, such as selective search (SS) [17], region proposal network (RPN) [18], and so on. The second stage is to further locate and classify the target for the region proposals. The two-stage methods, such as the regions with CNN features (RCNN) series [18], [19], [20], spatial pyramid pooling network [21], usually achieve high accuracy in target detection. Nevertheless, these networks consume a large amount of computation, and are difficult to meet the demand for real-time detection. The one-stage detection methods directly predict the category probability and location information of the object through the extracted features. Classic one-stage algorithms include you only look once (YOLO) series [22], [23], [24], [25], single shot multibox detector [26], RetinaNet [27], and so on. Since the one-stage methods directly predict the target classification and regression, the detection speed of the one-stage methods is faster than that of the two-stage algorithms while having a similar detection effect to the two-stage methods.

The majority of target detection networks are developed according to optical images. However, due to the special imaging characteristics of SAR images, the detection networks based on optical images may not be applicable to SAR images. Therefore, more and more scholars have begun to design and propose ship detection networks based on the characteristics of SAR images. For example, Zhang et al. [28] proposed a quad-feature pyramid network (FPN) for complex backgrounds and multiscale features of ships, which improved the ship detection accuracy by cascading four unique FPN modules. Yang et al. [29] introduced a coordinate attention module to mitigate the disturbance from SAR complex background and designed a receptive field increased module for capturing contextual information of multiscale ships. Xiao et al. [30] utilized a power-based convolution block to suppress speckle noise and coasts, and designed a feature alignment guided block to prevent the problem of ship dislocation. To enhance the ship outline feature and weaken the speckle noise, Li et al. [31] proposed a ship feature enhancement category through outline extraction based on side window mean filtering, which improved the ability to extract key features. To suppress scattered spots and noises, Bai et al. [32] constructed an enhanced feature pyramid that employed a spatial attention mechanism. In the meantime, a shallow feature reconstruction module was presented for effectively detecting small-scale ships.

Most of the above CNN-based SAR ship detection networks improved the detection accuracy at the cost of increasing the amounts of parameters and computation of the network. However, in some specific radar application scenarios, the computing resource and power consumption of computing devices are limited. It leads to the difficulty of deploying large-scale and high-computation target detection networks on the hardware platform. How to achieve a better tradeoff between detection accuracy and computational complexity has become a challenge. To reduce the computational cost, most existing ship detection networks tend to design lightweight structures. Nevertheless, few scholars have considered the characteristics of ship distribution in SAR images.



(a) (b)

Fig. 1. Ships marked in green boxes in SAR images. (a) Offshore scene. (b) Inshore scene.

In marine SAR images, sea and coastal regions usually occupy most of the areas of images, whereas ships simply take a small part, as shown in Fig. 1. If the target regions can be selected from the SAR image using a simple algorithm and then put into a complex detection network, the amount of network computation will be reduced at the source image level, which is similar to the two-stage network. Nevertheless, existing algorithms for generating candidate regions, such as SS in [17] and RPN in [18], have some disadvantages. For the SS algorithm, there are a lot of overlapping regions in the proposed candidate boxes, and most of them are redundant boxes, which wastes a lot of computational cost. RPN is designed after the feature extraction module, and it is not conducive to detecting small-scale ships. Moreover, RPN also has the problem of generating plenty of redundant boxes. Thus, this article focuses on designing a new module to select ship regions from SAR image and filter out the backgrounds. Image segmentation methods can be used to generate target regions, which divide the SAR image into several regions according to the features of shape, gray, and semantics [33], [34]. Additionally, texture features such as skewness wavelet energy [35] and kurtosis wavelet energy [36] are efficient and useful for SAR image segmentation. In recent years, many SAR image segmentation algorithms based on deep learning have been proposed and achieved good segmentation results [37]. However, the difference between ships and sea is significant in SAR images, so it is not necessary to separate these two regions by complex algorithms.

In response to the aforementioned problems, this article proposes a new SAR ship detection network, called multilevel background removing-YOLOX (MLBR-YOLOX), based on the one-stage detection network (i.e., YOLOX [38]). First, according to the special angular reflectivity of ships and backgrounds, a standalone spatial patch detector (SSPD) is designed to predetect the SAR image at the source image level. Since SSPD that has simple structure is not suitable for detecting the whole SAR image, the SAR image is first cropped into small-scale patches.

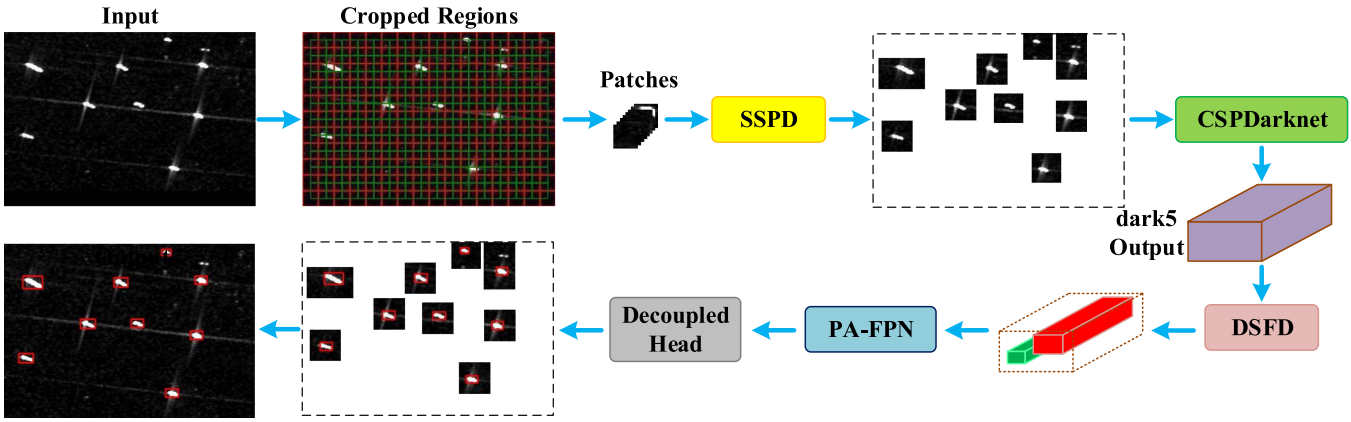


Fig. 2. Architecture of MLBR-YOLOX. Red and green lines represent the cropped marks in cropped regions.

SSPD detects these patches and determines whether the detected patches belong to the ship regions. It filters out the large sea and coastal areas of marine SAR images and only the regions of interest (RoIs) are input to YOLOX, drastically reducing the computational cost of the YOLOX network. After extracting the deep features of the retained RoIs using the backbone module, a deep spatial feature detector (DSFD) is proposed to detect the deep semantic features of the extracted feature blocks. It classifies the features at different spatial locations in the feature map to further filter out the regions of sea and coasts, which reduces the size of the extracted feature blocks and the computational cost of the neck and head modules. Finally, the retained target features are fed into the neck and head modules for target classification and localization. MLBR-YOLOX can obtain accurate detection results with quite lower computation. Experimental results based on the SAR ship detection dataset (SSDD) [39] and the high-resolution SAR images dataset (HRSID) [40] show that the proposed MLBR-YOLOX network achieves much smaller computational cost with similar ship detection accuracy compared with most existing advanced detection networks.

The main contributions of this article can be summarized as follows.

1) To predetect SAR images and filter out the regions of the background, a tiny plug-and-play module, namely standalone spatial patch detector, is proposed. This module selects the target areas from the SAR image and filters out most of the sea and coastal backgrounds with a small amount of computation.
2) A deep spatial feature detector is designed and detects the deep semantic features of different spatial locations from the backbone module and removes most deep features that contain the sea and coast information.
3) A novel SAR ship detection network, MLBR-YOLOX, is proposed based on YOLOX. The SSPD and DSFD modules are inserted into the YOLOX network to filter out the majority of the background regions of SAR images. Experiments on SSDD and HRSID demonstrate that the detection performance of MLBR-YOLOX is close to that of the state-of-the-art CNN networks, while the computational cost is only 23.97% and 12.50% of that of the baseline, YOLOX, respectively. Additionally, based on the well-trained model, the detection results on a large-scene Sentinel-1 SAR image prove that the proposed network has good migration application ability.

II. Proposed Method

Fig. 2 presents the architecture of the proposed MLBR-YOLOX network. MLBR-YOLOX comprises five modules: SSPD, CSPDarknet, DSFD, PA-FPN, and Decoupled Head. The SSPD and DSFD modules are proposed in this article, whereas the other three modules are included in the YOLOX network. The first part of MLBR-YOLOX is the image predetection module, SSPD, which selects the regions of targets and filters out large areas of sea and coastal backgrounds in the SAR image level. The predetected results are fed to the backbone module, CSPDarknet, for extracting the semantic features of SAR images. Then the DSFD module detects the deep semantic features of the dark5 output and further separates the features that contain ships and coasts. The selected target features are input into the neck module, PA-FPN, for fusing the semantics with different scales. Finally, Decoupled Head is employed to generate the prediction results of target location and classification.

YOLOX is introduced as the baseline network, which is the first network that adopts an anchor-free structure in the YOLO series. As ships have multiple scales and sparseness in SAR images, an anchor-free structure is more suitable for predicting ships compared with an anchor-based structure. Additionally, YOLOX presents the SimOTA algorithm to better allocate fuzzy samples, which is conducive to assigning the prediction samples of ship targets in complex backgrounds. Without changing the YOLOX network, we propose the SSPD and DSFD modules to filter out the large regions of sea and coastal backgrounds while preserving ship features.

Section II-A introduces the SSPD module employed to classify the regions of targets and backgrounds in the SAR images. The YOLOX network is presented in Section II-B. In Section II-C, DSFD is described in detail.

A. SSPD

Due to different backscattering features in the echo signals of ships and sea backgrounds obtained by the SAR system, the

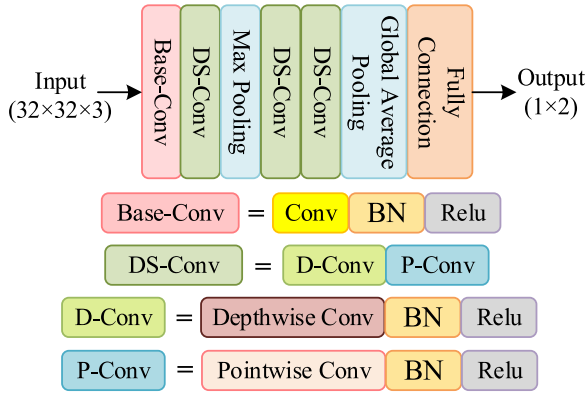


Fig. 3. Architecture of SSPD.

ships are represented as bright pixels in SAR images, whereas most of the sea backgrounds are represented as dark pixels. Considering this characteristic of a marine SAR image, a plug-and-play module, SSPD, is designed to identify the regions of ships and sea backgrounds in SAR images. Independent of the YOLOX network, SSPD detects SAR images to filter out the regions of the backgrounds without missing any ships.

The architecture of SSPD is shown in Fig. 3. The SSPD module is a tiny CNN structure and simply has seven layers. It adopts depthwise separable convolution [41] to substitute for conventional convolution. Fig. 4 shows the architectures of conventional convolution and depthwise separable convolution, where I_1–I_3 are input channels, K_1–K_9, k_1–k_{12} are convolution kernels, and O_1–O_3 are output channels. In conventional convolution, each input channel corresponds to a convolutional kernel, and an output channel is the sum of the outputs of all convolutional kernels. Depthwise separable convolution is divided into depthwise convolution and 1×1 convolution which is called pointwise convolution. Depthwise convolution has the same convolution kernel size as conventional convolution, but each output channel is determined only by a convolution kernel sliding in an input channel. Then pointwise convolution is used to generate a linear combination of the depthwise layer outputs. Similar to conventional convolutions, depthwise separable convolution adds batch normalization (BN) [42] and rectified linear unit function after each convolution. Compared with the conventional convolution, the amounts of parameters and computation for depthwise separable convolution are drastically decreased. In addition, the SSPD module employs a max pooling to expand the receptive field of the network and a global average pooling to extract the global semantic features of the SAR image. Finally, a fully connected layer is adopted and outputs two elements, which predict whether the detection image contains targets.

SSPD is not appropriate to detect large-scale images due to its simple structure. Hence, we crop sequentially the original SAR image into small-scale patches with a size of 32×32, and there is no overlap between patches. To prevent small-scale ships from being segmented and missed, a SAR image is cropped twice. The twice-cropped patches have a 25% overlap, and the cropped regions of Fig. 2 depict the cropped positions on the SAR image. The SSPD module considers these patches as the input, and selects the patches that belong to the pieces of targets. Then, these patches are numbered and mapped back to the original image for classification. The steps of the patch classification are as follows.

1) Assume that the pixel center coordinates of two patches in the original image are (x_1, y_1) and (x_2, y_2), respectively. The distance d between the pixel centers of these two patches is as follows:
$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}. \tag{1}$$
2) If $d < 32$, the corresponding regions of these two patches in the original image belong to a same RoI. The serial numbers of two patches are recorded to form a sequence pair. Find all the sequence pairs. If a patch does not form a sequence pair with all other patches, it is considered as an independent RoI.
3) All sequence pairs are classified again. If two sequence pairs contain a same serial number, it is considered that the patches corresponding to the serial numbers in the two sequence pairs belong to a same RoI. The classification is repeated until all the sequence pairs are classified.
4) The principle of the minimum external rectangle [43] is employed to determine the regions of all the RoIs. Since the width and height of the image inputting to YOLOX should be an integer multiple of 32. If the size of the RoI does not meet it, we expand the RoI region on the original image until both the width and height reach an integer multiple of 32. Then the RoIs are cropped from the original SAR image.

Since SSPD is a binary classification module, the cross entropy loss function [44] is chosen to measure the errors between the prediction results and the actual label, i.e., the ground truth (GT). The loss function is formulated as follows:
$$\text{Loss} = -\left(y \log p + (1 - y) \log (1 - p)\right) \tag{2}$$
where y donates the GT of the detected patch. If the detected patch belongs to the ship region, $y = 1$; otherwise, $y = 0$. The p donates the probability that the patch may contain the ship features. Combined with the adaptive momentum (Adam) optimizer [45], the SSPD module is trained as a classifier for ships and backgrounds.

The SSPD module selects and crops the regions that may be the target from SAR images, which achieves the effect of predetection of the SAR images and effectively reduces the calculation amount of subsequent detection. To retain the characteristics of targets, CSPDarknet is input with the predetected RoIs without resizing.

B. YOLOX

YOLOX is the latest target detection network in the YOLO series that adopts the anchor-free method. The YOLOX network achieves high detection accuracy by combining the anchor-free structure with a new label assignment strategy, SimOTA. To obtain a better balance between parameter amount and detection performance, the YOLOX-s model is chosen as the baseline.

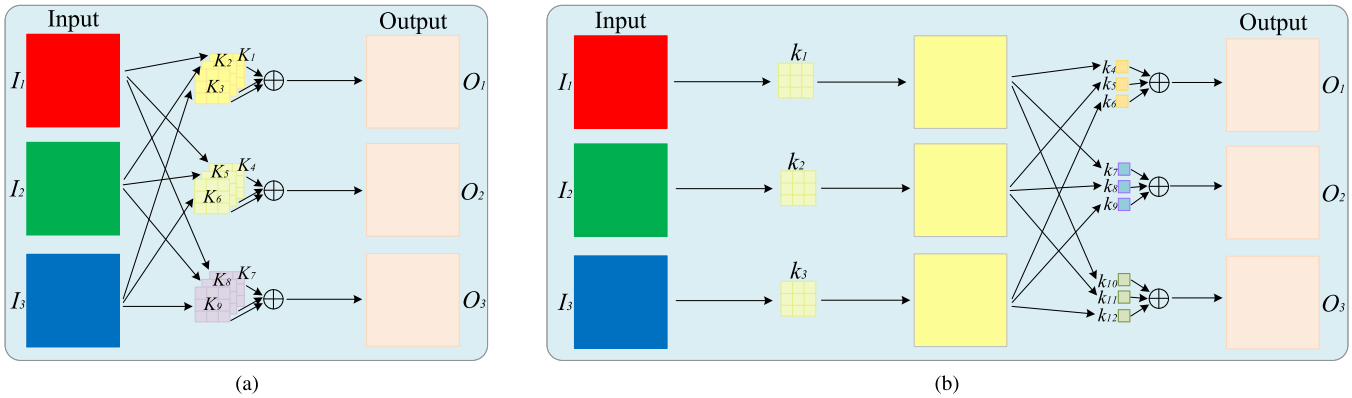


Fig. 4. Architectures of conventional convolution and depthwise separable convolution. (a) Conventional convolution. (b) Depthwise separable convolution.

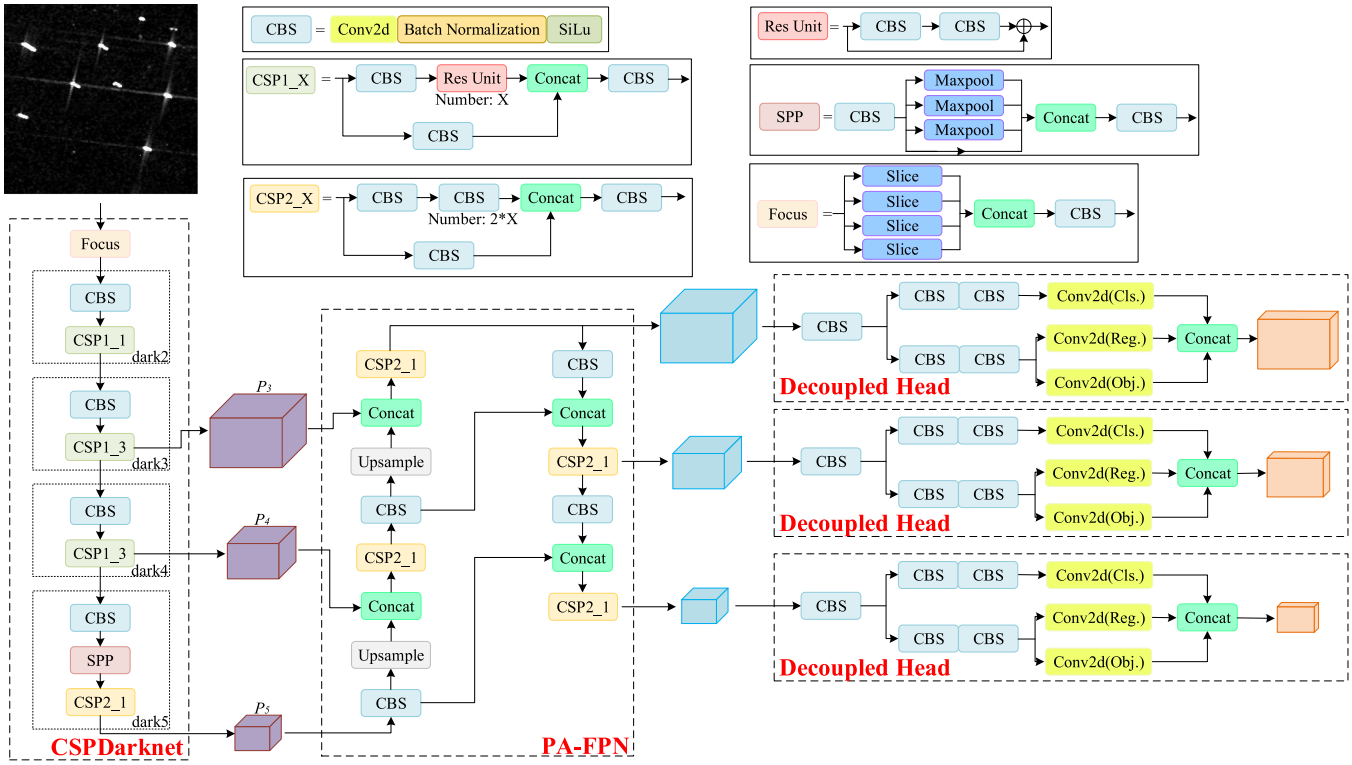


Fig. 5. Architecture of YOLOX-s.

The architecture of YOLOX-s is shown in Fig. 5. It is mainly composed of three parts: CSPDarknet, PA-FPN, and Decoupled Head, which are described below.

CSPDarknet serves as the backbone module to extract the features of a SAR image. Focus structure is added at the beginning of CSPDarknet to split a high-resolution feature map into four low-resolution feature map slices. It converts the width and height features of the SAR image to the channel dimension. Then the features of different channels are extracted by a 3×3 convolution. Compared with the original three-layer convolutions, the Focus structure reduces the amounts of parameters and computation, which can further improve the training speed of the network. The CBS structure consists of conventional convolution, BN, and sigmoid-weighted linear units (SiLu) [46] activation function, as the basic convolutional unit of the CSP-Darknet module. The definition of SiLu is as follows:

$$\mathrm{SiLu}(x) = \frac{x}{1 + e^{-x}}. \tag{3}$$

Meanwhile, CSPDarknet adopts residual block [47], cross stage partial (CSP) [48], and SPP structures [21] to strengthen the effect of feature extraction. Residual block increases the gradient of back-propagation between layers to avoid gradient disappearance. The CSP structure divides the input feature map into two branches for extracting the features of different scales, and then concatenates the two branches. As shown in Fig. 5,

there are two types of CSP structure: CSP1 and CSP2. In CSP1, the residual block is employed in one branch, while it is replaced with two CBS structures in CSP2. The CSP structure enhances the learning ability of different layers. The SPP module fuses the feature map of local and global features, which enriches the expression ability of the feature map and improves the detection performance. CSPDarknet has three output layers, namely dark3, dark4, and dark5, and they output the deep semantic features P_3, P_4, and P_5. P_5 is chosen as the input of the DSFD module.

PA-FPN functions as the neck part and comprises FPN [49] and path aggregation network (PAN) [50]. They are employed to fuse different information of P_3, P_4, and P_5. FPN conveys the semantic features of the SAR images from top to bottom, while PAN conveys the localization features of targets from bottom to top. The PA-FPN module outputs the fused features of three scales that are adopted to predict large, medium, and small-scale targets, respectively.

Decoupled Head serves as the head module and replaces the original coupled head. It adopts two unique detection heads to predict the information of target classification and position regression, respectively. With a few more parameters and computational complexity, Decoupled Head not only improves the detection accuracy, but also accelerates the convergence speed.

For the input images, YOLOX adopts Mosaic and Mixup strategies for data augmentation. Mosaic makes random scaling and cropping of the input image, and then splices them into a new image, which improves the detection performance of small-scale targets and the robustness of the network. Mixup combines two images into a new image in a certain proportion, and uses the new image and label to participate in training, which increases the diversity of targets and backgrounds. In addition, to assign labels better, the SimOTA algorithm is proposed in YOLOX. First, along with obtaining the training loss, the cost matrix of the pairs of GT boxes and pre-screening prediction boxes (prediction-gt pairs) is calculated. Second, the quantity of selecting prediction boxes is determined according to the intersection over union (IoU) of the prediction-gt pairs. Finally, based on the cost matrix, the corresponding quantity of prediction boxes is selected as the final positive samples.

C. DSFD

The SSPD module can filter out most of the sea background of SAR images. Nevertheless, some patches containing coasts have familiar features with those of ships, and SSPD does not have good recognition ability with them. To improve the performance of removing the backgrounds of SAR images, DSFD is designed. Since the comparison between the features of ships and coastal backgrounds becomes clearer after feature extraction of SAR images, the outputs of the backbone module are selected as the input. The DSFD module is employed to detect the deep semantic features of different spatial locations and separate the areas that contain ship and coastal features.

The backbone, CSPDarknet, has three output layers: dark3, dark4, and dark5, and outputs three deep feature blocks: P_3, P_4, and P_5. Compared with the size of the input image, the feature maps' sizes of P_3, P_4, and P_5 are downsampled by 8, 16, and 32 times, respectively. However, the information contained in the corresponding spatial location of their feature maps is the same. Thus, we can merely detect one output of the backbone module, and then the detection results can be mapped to the other two outputs. To reduce the computational cost, P_5 is selected as the input of DSFD.

Fig. 6 shows the processing flowchart of the DSFD module. First, the CSPDarknet module is fed with a SAR image and generates the deep semantic features of three output layers. We assume that the size of an input image is (3, H, W); hence, the size of P_5 is (512, H/32, W/32). Second, P_5 is fed into DSFD. The steps of the DSFD module are as follows.

1) The local textural features of P_5 are fused using an average pooling (kernel size $= 3 \times 3$ and padding $= 1$).
2) The semantic feature block is reshaped to a two-dimensional feature with a size of (HW/1024, 512), and different spatial features whose size is (1, 512) are fed to a classifier with two fully connected layers.
3) The classification results are mapped to the feature map of P_5 and the location that contains ship features is determined according to the principle of the minimum external rectangle. To prevent some ship features from being filtered, we expand two-pixel units around the obtained target region in the feature map.
4) The feature map of P_5 is upsampled four and two times, and the regions of the ship features in P_3 and P_4 are predicted, as well.
5) The features predicted to contain ship information are cropped from three output blocks, and they are input to the neck and head modules for accurate positioning.

The DSFD module is employed to separate the ships and backgrounds in the deep feature block. The background semantic features are further filtered and the size of feature maps of P_3, P_4, and P_5 is reduced. Furthermore, the amount of computation of PA-FPN and Decoupled Head is decreased.

III. Experimental Results

The experiments are conducted on the PyTorch framework. A computer equipped with Intel I7-7700K CPU, 32-GB RAM, and Nvidia GeForce RTX 3060 GPU is used. The operating system of the computer is Ubuntu 20.04.

A. Datasets and Evaluation Criteria

1) Datasets: In this article, we select the SSDD and HRSID for a series of experiments to verify the detection performance of MLBR-YOLOX. SSDD is the first public dataset used for SAR ship detection, which is acquired by RadarSat-2, TerraSAR-X, and Sentinel-1 sensors, with four polarization modes: HH, HV, VV, and VH. It contains 1160 SAR images and 2456 ships, and can be mainly divided into two types of scenes: offshore and inshore. The range resolution of SSDD ranges from 1 to 15 m. According to [51], image indexes' suffixes 1 and 9 are set as the test set, and the rest are set as the training set. As for HRSID, it has 5604 high-resolution SAR images with pixel resolution 800×800, containing a total of 16951 ships. HRSID is mainly

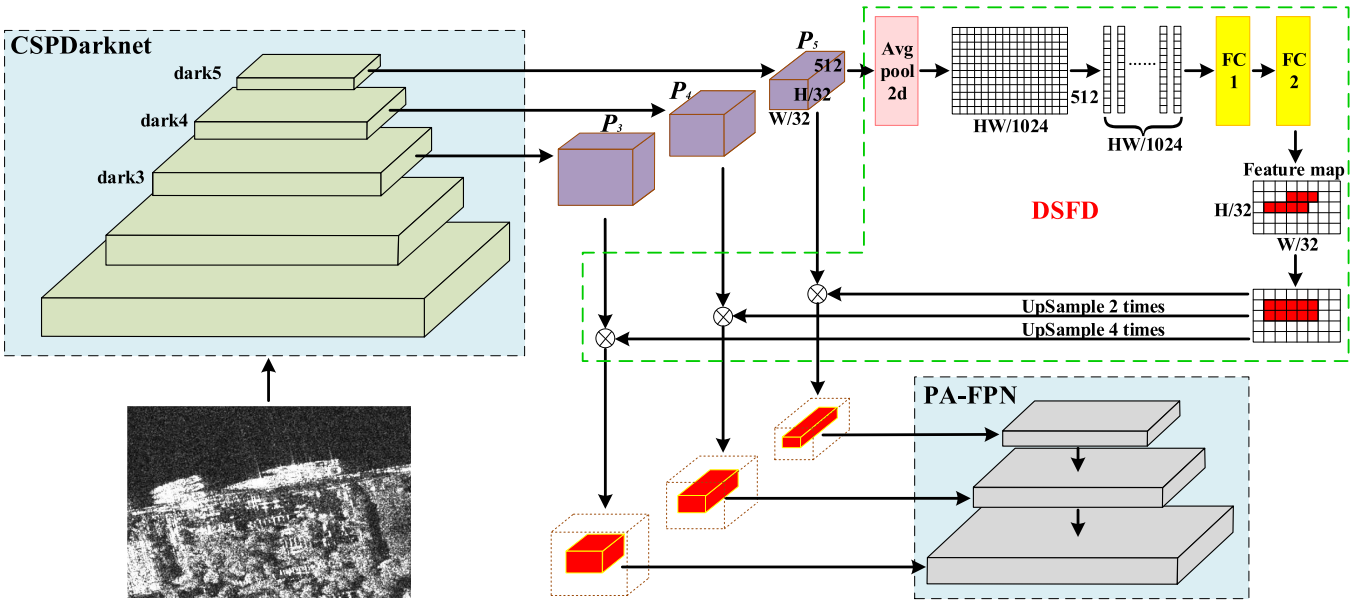


Fig. 6. Processing flowchart of DSFD. FC = Fully connected layer.

used for ship detection and semantic segmentation tasks. The resolution of SAR images varies from 0.5 to 3 m and the polarization methods of the radar remote sensing system include HH, HV, and VV. Following the given dataset configuration, the proportion of SAR images used for training and testing is 65% and 35%, respectively. Both datasets provide SAR images of different sea conditions and scenarios, and their GTs of ship targets.

2) Evaluation Criteria: The computational complexity of the network and detection accuracy are selected as the evaluation indexes in this article, represented by the average floating-point operations (FLOPs) and mean average precision (mAP) respectively. The FLOPs represents the amount of computation of a detection network. The mAP is determined by recall rate R and precision rate P. The definition of R, P, and mAP are respectively

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

$$\mathrm{mAP} = \int_0^1 P(R)dR \tag{6}$$

where TP, FN, and FP are the number of true positive, false negative, and false positive, respectively. In the SAR ship detection, TP represents that the IoU of the ship prediction-gt pairs is higher than 0.5, while FP means false alarms or the IoU of the prediction-gt pairs is lower than 0.5. FN stands for the quantity of missed ships. $P(R)$ denotes the precision-recall curve. The mAP represents the detection accuracy of the network. Both SSDD and HRSID only have one class of target; hence, the mAP is the average precision of ships.

Additionally, the $F1$ score is introduced to weigh the recall and precision, and it is a comprehensive evaluation metric of detection performance, which is calculated by the following:

$$F1 = 2 \times \frac{P \times R}{P + R}. \tag{7}$$

B. Implementation Details

1) Training Strategies: In the part of YOLOX, the YOLOX-s model is selected as the baseline. When training the SSDD, the epoch is set to 500, and the batch-size is set to 8. The SAR images resized to 512 × 512 are used as the input. Other parameters are set according to the default parameters of the YOLOX network [52]. As to HRSID, the size of the input images for training is 800 × 800, and 300 epochs are trained while keeping other parameters unchanged.

For training the SSPD module, the patches containing ships and backgrounds are respectively cropped from SAR images according to the given annotations. The batch-size is 16, and the training epoch is 100. The learning rate of the Adam optimizer is set to 0.001.

2) Test Strategies: When testing the SSDD dataset, to ensure that the features of ships in SAR images remain unchanged, SAR images are no longer resized, but the right and bottom of SAR images are padded with zeros until the image size becomes an integer multiple of 32. The processed image is input to the MLBR-YOLOX network. As for HRSID, the size of the SAR images is 800 × 800, so we feed the original images into the proposed network directly.

When MLBR-YOLOX is employed for ship detection, SSPD can filter out most of the sea backgrounds of SAR images. However, it does not work effectively for recognizing some coastal backgrounds. Additionally, when the DSFD module is adopted for semantic feature detection, some features containing ship information may be lost, especially the features of small-scale ships. After comprehensive consideration, we decide to merely

TABLE I
ABLATION EXPERIMENTS OF MLBR-YOLOX

Model			SSDD			HRSID		
YOLOX	SSPD	DSFD	mAP (%)	Average FLOPs	Rate (%)	mAP (%)	Average FLOPs	Rate (%)
√	×	×	**96.97**	11.2 G	100	**93.79**	41.8 G	100
√	√	×	96.62	3.2 G	28.33	93.34	6.1 G	14.69
√	√	√	96.59	**2.7 G**	23.97	92.16	**5.2 G**	12.50

feed the large-scale P_5 into the DSFD module. The P_5, whose product of the width and height of the feature map is greater than 30, is defined as a large-scale feature block. Otherwise, it is defined as a small-scale feature block. Specifically, small-scale feature blocks skip DSFD and are put into the neck module directly.

C. Ablation Experiments of MLBR-YOLOX

Ablation experiments are conducted on the SSDD and HRSID to verify the feasibility of MLBR-YOLOX. Table I presents the results of the ablation experiments. The average FLOPs represents the average computational complexity of detecting a SAR image on the test set, and the rate represents the proportion of a network's computational cost compared with that of YOLOX. The following conclusions can be drawn from Table I.

1) When using the YOLOX network to detect SAR images of SSDD and HRSID, the mAPs are 96.97% and 93.79%, respectively, which prove that YOLOX has strong ability in detecting SAR ships under various scenarios. Nevertheless, the average FLOPs of detecting a SAR image is quite large for a computing platform, increasing the difficulty of hardware transplantation.
2) The SSPD module is added to predetect SAR images, and the detected results are fed to YOLOX for accurate detection. The amount of computation is rapidly reduced, and the FLOPs are respectively 28.33% and 14.69% that of YOLOX when testing SSDD and HRSID. That is because most of the SAR images in both datasets are taken offshore, and sea regions occupy the majority of the areas of marine SAR images. The SSPD module can easily filter out the sea backgrounds and retain the target regions, whereas the detection accuracy of YOLOX on both SAR datasets is slightly reduced. Nevertheless, as some patches containing coasts have features similar to those of ships, they are predicted as the patches that contain ships, and that produces false alarms.
3) To further filter out the coastal backgrounds and lower computational cost, the SSPD and DSFD modules are inserted into the YOLOX network. On SSDD, the mAP is 96.59% and the gap of the mAP between YOLOX and MLBR-YOLOX is merely 0.38%. However, the average FLOPs of MLBR-YOLOX only requires 23.97% compared with YOLOX. When testing HRSID, the mAP decreased by 1.63% compared with YOLOX. The reduction of the detection accuracy is acceptable because the size of most ships in HRSID is small. Some small-scale ships in SAR images have little texture information and may be missed in DSFD. Moreover, the average FLOPs is only 12.50% of the original YOLOX network, which saves massive computation resources.

TABLE II
COMPARISON OF DIFFERENT METHODS ON SSDD

Network	Input size	mAP (%)	Average FLOPs
Faster R-CNN [18]	512 × 512	93.71	126.6 G
RetinaNet [27]	512 × 512	90.56	104.1 G
YOLO v3 [24]	512 × 512	96.69	98.9 G
YOLO v3-tiny [24]	512 × 512	95.42	8.2 G
YOLO v5-s [25]	512 × 512	96.61	10.1 G
YOLOX [52]	512 × 512	96.84	17.1 G
YOLOX [52]	320 × 256~ 672 × 512	**96.97**	11.2 G
MLBR-YOLOX (ours)	320 × 256~ 672 × 512	96.59	**2.7 G**

TABLE III
COMPARISON OF DIFFERENT METHODS ON HRSID

Network	Input size	mAP (%)	Average FLOPs
Faster R-CNN [18]	800 × 800	88.10	269.0 G
RetinaNet [27]	800 × 800	88.80	254.4 G
YOLO v3 [24]	800 × 800	**94.19**	241.5 G
YOLO v3-tiny [24]	800 × 800	90.11	20.1 G
YOLO v5-s [25]	800 × 800	93.49	24.9 G
YOLOX [52]	800 × 800	93.74	41.8 G
MLBR-YOLOX (ours)	800 × 800	92.16	**5.2 G**

D. Comparison With Different Detection Methods

1) Comparison With Latest Target Detection Methods: To further confirm the detection performance of our proposed method, we conduct experiments on SSDD and HRSID, and compare them with other existing target detectors. These detectors include Faster RCNN, RetinaNet, YOLO v3, YOLO v5, and the baseline, YOLOX. The comparison results of the detection performance and computational complexity are shown in Tables II and III.

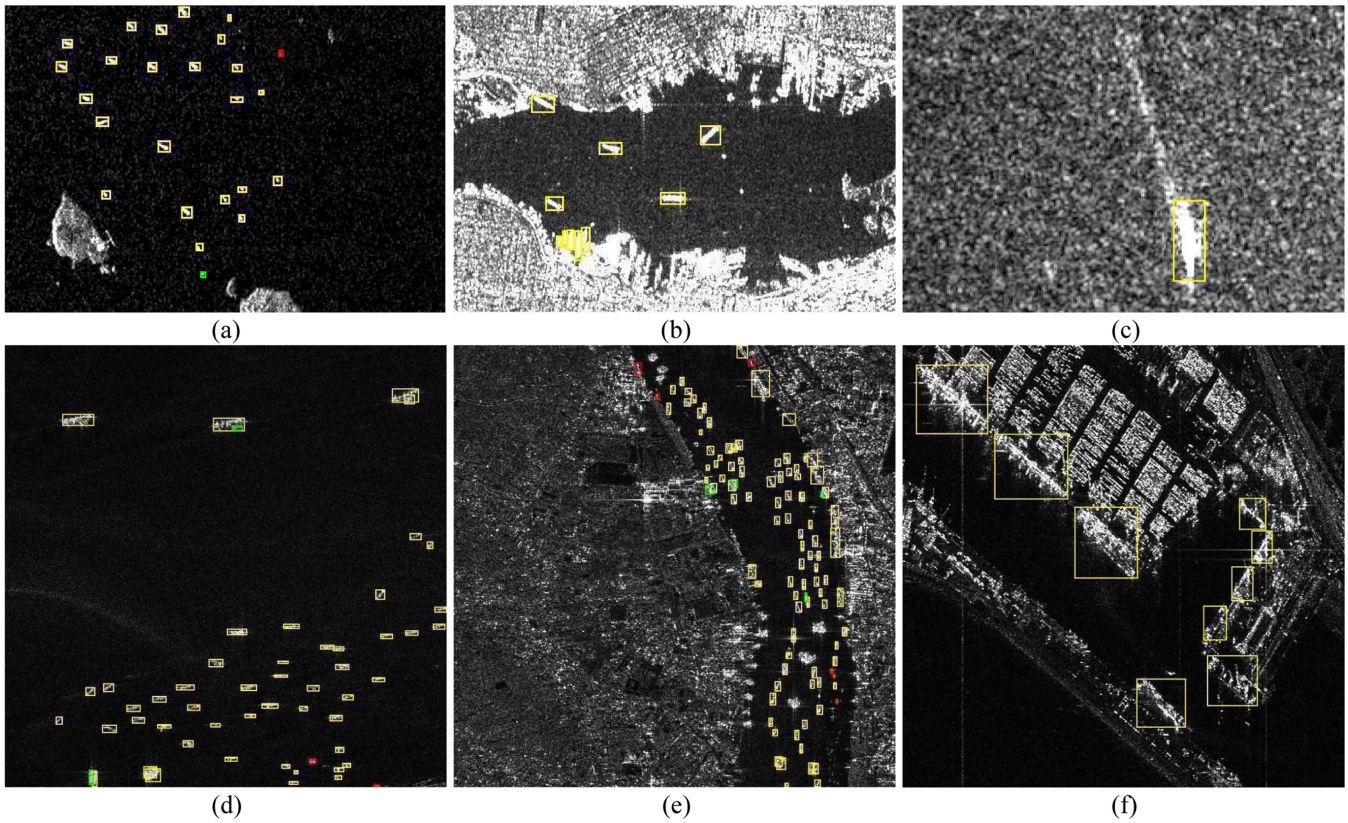


Fig. 7. Visual detection results of MLBR-YOLOX. (a) Small-scale and offshore targets on SSDD. (b) Small-scale, inshore and densely distributed targets under speckle noise on SSDD. (c) Large-scale targets under speckle noise on SSDD. (d) Small-scale and offshore targets on HRSID. (e) Small-scale, inshore and densely distributed targets on HRSID. (f) Large-scale and inshore targets on HRSID.

Table II presents the comparison between MLBR-YOLOX and other advanced detection networks on SSDD. Experimental results indicate that the original YOLOX's mAP of detecting SAR images without resizing reaches 96.97%, and that is higher than those of other detection networks. It further verifies that the baseline network has great detection performance. Nevertheless, its average FLOPs is 11.2G. Although the FLOPs of YOLOX is far less than that of Faster RCNN, RetinaNet, and YOLO v3, it still consumes a lot of computing resources and increases the strain on a computing platform. The mAP of MLBR-YOLOX is 96.59%. Compared with the classic Faster RCNN and RetinaNet, the proposed MLBR-YOLOX can improve 2.88% and 6.03%, respectively. As for YOLO v3 and YOLO v5, the mAP of our detector simply reduce about 0.1%. Meanwhile, the average FLOPs of MLBR-YOLOX is only 2.7 G, which is quite smaller than other detection networks.

Compared with SSDD, the backgrounds of the SAR images on HRSID are more complicated and the quantity of small-scale ships is larger. Thus, it can better reflect the detection effect of different algorithms. As shown in Table III, the detection performance of YOLO v3 outperforms other detection networks, but its FLOPs is 241.5G, which is nearly six times as much as that of YOLOX. The mAP of MLBR-YOLOX reaches 92.16% and is close to that of YOLOX and YOLO v5. Additionally, MLBR-YOLOX is 4.06% higher than Faster RCNN and also exceeds RetinaNet by 3.36%. Moreover, it is worth noting that the FLOPs of the proposed network is merely 5.2G, and it is also far less than other detectors. Tables II and III indicate that the proposed network can greatly reduce the computational burden while having a similar detection effect to the existing advanced networks.

TABLE IV
COMPARISON OF SAR SHIP DETECTION NETWORKS

Network	mAP (%)	Average FLOPs
DSDet [6]	90.70	14.3 G
Pow-FAN [30]	89.74	–
LPEDet [53]	89.70	18.4 G
SAR-Net [54]	87.49	104.2 G
Reference [29]	**92.70**	123.5 G
Reference [55]	88.39	28.5 G
MLBR-YOLOX (ours)	92.16	**5.2 G**

2) Comparison With SAR Ship Detection Networks: We carry out the performance comparison on HRSID with some existing advanced SAR ship detection networks to demonstrate the effectiveness of our proposed network, as shown in Table IV.

The networks and detection results listed in Table IV need to be specially explained. Since all these SAR ship detection networks are not open source, and some parameters and details involved in the experiments are not specifically described, their experimental codes and results are difficult to reproduce

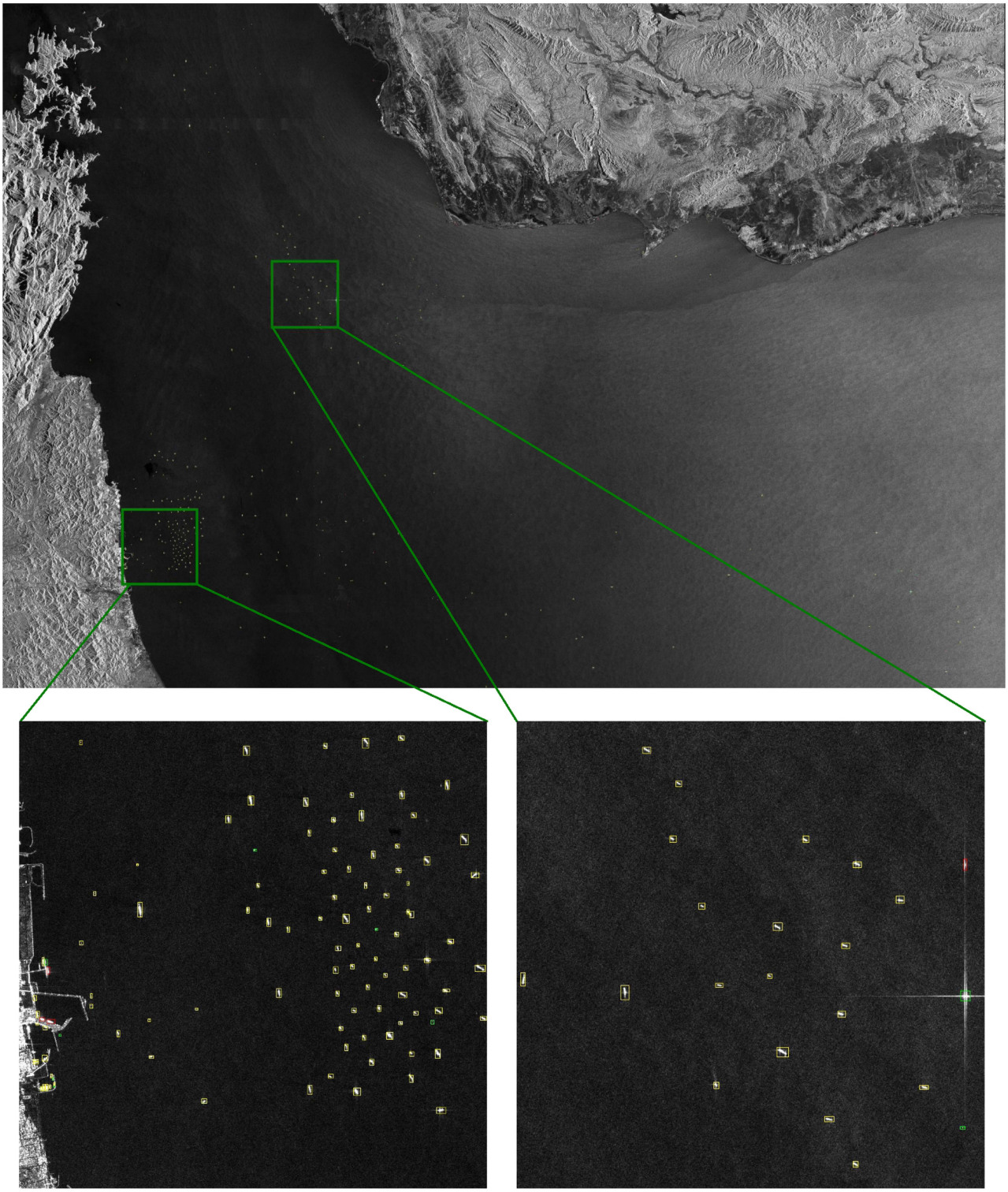


Fig. 8. Visual detection results of the proposed network on a large-scene Sentinel-1 SAR image.

completely. Thus, to fairly compare the detection performance and computational complexity of different networks, we use as many of the same parameter settings as the comparison networks, while citing the best experimental results reported in the references of the comparison methods.

As can be seen from Table IV, the detection accuracy of the proposed MLBR-YOLOX is merely 0.54% lower than that of the network proposed in [29], and the FLOPs of our proposed network is only 4.21% of it. Compared with other state-of-the-art SAR ship detection networks, our proposed network achieves

TABLE V
DETECTION RESULTS OF INSHORE SCENARIOS

Network	GT	TP	FP	FN	*R*(%)	*P*(%)	*F*1(%)
YOLOX	163	159	30	4	**97.5**	84.1	90.3
MLBR-YOLOX	163	156	24	7	95.7	**86.7**	**91.0**

higher detection accuracy and computing efficiency. Furthermore, the proposed network gets a better balance between the detection performance and computational cost compared with other advanced SAR ship detectors.

E. Visual Detection Results

1) SSDD and HRSID: Fig. 7 shows some visual detection results of MLBR-YOLOX on the SSDD and HRSID in which the detected ships, false alarms, and missed ships are respectively marked with yellow, red, and green rectangles. It can be seen from the detection results that a majority of multiscale ships can be detected accurately in various complicated situations, such as inshore scenes, strong speckle noise, and densely distributed targets. However, there are few missed ships and false alarms. For one thing, some ships are so small-scale and densely distributed that the texture information of ships is not obvious. For another, some backgrounds are complex, and some islands and coasts have similar features of ships in SAR images. The visual detection results further demonstrate that MLBR-YOLOX has strong ship detection capability and robustness in SAR images.

2) Large-Scene SAR image: To test the migration capability of MLBR-YOLOX, we select a large-scene Sentinel-1 SAR image derived from the large-scale SAR ship detection dataset-v1.0 (LS-SSDD-v1.0) [56] based on the model trained on the SSDD dataset. The size of the image is 24 000 × 16 000 and the polarization mode is VV. The large-scene image is firstly split into 800 × 800 subimages because of the limitation of GPU memory. Then, the well-trained network is used to detect these images. Finally, the prediction results are integrated into the large-scene SAR image. The visual detection results are shown in Fig. 8. It can be observed that MLBR-YOLOX can accurately detect most ships with few false alarms and missed ships. The satisfactory detection result confirms the good migration application capability of the proposed network.

F. Detection Results of Inshore Scenarios

We also test and compare the detection performance of the proposed network and the basic network, YOLOX, in inshore scenarios. According to [51], 46 inshore SAR images are selected from the SSDD test set and tested on YOLOX and MLBR-YOLOX, respectively. The confidence threshold is set to 0.35. The prediction results are shown in Table V. It can be seen that, compared with YOLOX, although the quantity of missed ships in the MLBR-YOLOX network increases by 3, the false alarms in MLBR-YOLOX are significantly reduced. Meanwhile, the *F*1 score of the proposed network is 91.0%, which is better than YOLOX. Table V shows that our proposed network can effectively reduce the number of false detection of nearshore factories and other buildings in complex scenarios, and it further confirms the strong detection performance of MLBR-YOLOX.

IV. CONCLUSION

To obtain low computational cost in SAR ship detection, we propose a new SAR ship detection network, called MLBR-YOLOX. Without changing the YOLOX network, the SSPD and DSFD modules are proposed to filter out large regions of sea and coastal backgrounds in SAR images from different levels. Only the regions predicted to contain ship features are adopted and input into YOLOX for accurate positioning. Experimental results on the SSDD and HRSID indicate that MLBR-YOLOX has good detection performance close to advanced detection networks with quite lower computational complexity. Meanwhile, based on the trained model, a large-scene SAR image of Sentinel-1 is used to test the migration ability of MLBR-YOLOX. Satisfactory detection results prove that the proposed network has good migration application capability. Furthermore, experimental results based on inshore images show that the proposed network reduces the number of false detection in inshore scenes compared to YOLOX.

REFERENCES

[1] Z. Sun, C. Meng, J. Cheng, Z. Zhang, and S. Chang, "A multi-scale feature pyramid network for detection and instance segmentation of marine ships in SAR images," *Remote Sens.*, vol. 14, 2022, Art. no. 6312, doi: 10.3390/rs14246312.

[2] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "GreyWolfLSM: An accurate oil spill detection method based on level set method from synthetic aperture radar imagery," *Eur. J. Remote Sens.*, vol. 55, pp. 181–198, 2022, doi: 10.1080/22797254.2022.2037468.

[3] S. Yang, W. An, S. Li, G. Wei, and B. Zou, "An improved FCOS method for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8910–8927, 2022.

[4] F. Mahmoudi Ghara, S. B. Shokouhi, and G. Akbarizadeh, "A new technique for segmentation of the oil spills from synthetic-aperture radar images using convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8834–8844, 2022.

[5] N. Aghaei, G. Akbarizadeh, and A. Kosarian, "OSDES_Net: Oil spill detection based on Efficient_Shuffle network using synthetic aperture radar imagery," *Geocarto Int.*, vol. 37, pp. 13539–13560, 2022.

[6] K. Sun, Y. Liang, X. Ma, Y. Huai, and M. Xing, "DSDet: A lightweight densely connected sparsely activated detector for ship target detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 14, Aug. 2017, Art. no. 2743, doi: 10.3390/rs13142743.

[7] H. Wan et al., "AFSar: An anchor-free SAR target detection algorithm based on multiscale enhancement representation learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5219514.

[8] Z. Tirandaz, G. Akbarizadeh, and H. Kaabi, "PolSAR image segmentation based on feature extraction and data compression using weighted neighborhood filter bank and hidden Markov random field-expectation maximization," *Measurement*, vol. 153, 2020, Art. no. 107432.

[9] F. Sharifzadeh, G. Akbarizadeh, and Y. Seifi Kavian, "Ship classification in SAR images using a new hybrid CNN–MLP classifier," *J. Indian Soc. Remote Sens.*, vol. 47, pp. 551–562, 2019.

[10] F. Samadi, G. Akbarizadeh, and H. Kaabi, "Change detection in SAR images using deep belief network: A new training approach based on morphological images," *IET Image Process.*, vol. 13, no. 12, pp. 2255–2264, 2019.

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

[12] T. Li, Z. Liu, R. Xie, and L. Ran, "An improved superpixel-level CFAR detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 184–194, Jan. 2018.

[13] 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.

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

[15] 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.

[16] T. Liu, J. Zhang, G. Gao, J. Yang, and A. Marino, "CFAR ship detection in polarimetric synthetic aperture radar images based on whitening filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 58–81, Jan. 2020, doi: 10.1109/TGRS.2019.2931353.

[17] J. Uijlings, K. van de Sande, T. Gevers, and A. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, pp. 154–171, 2013.

[18] 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.

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

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

[21] 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.

[22] 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.*, 2016, pp. 779–788.

[23] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6517–6525.

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

[25] G. Jocher, "YOLOv5," 2020. [Online]. Available: https://github.com/ultralytics/yolov5

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

[27] 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.*, 2017, pp. 2999–3007.

[28] T. Zhang, X. Zhang, and X. Ke, "Quad-FPN: A novel quad feature pyramid network for SAR ship detection," *Remote Sens.*, vol. 13, no. 14, 2021, Art. no. 2771, doi: 10.3390/rs13142771.

[29] X. Yang, X. Zhang, N. Wang, and X. Gao, "A robust one-stage detector for multiscale ship detection with complex background in massive SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5217712.

[30] M. Xiao, Z. He, X. Li, and A. Lou, "Power transformations and feature alignment guided network for SAR ship detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4509405.

[31] S. Li, X. Fu, and J. Dong, "Improved ship detection algorithm based on YOLOX for SAR outline enhancement image," *Remote Sens.*, vol. 14, 2022, Art. no. 4070, doi: 10.3390/rs14164070.

[32] L. Bai, C. Yao, Z. Ye, D. Xue, X. Lin, and M. Hui, "Feature enhancement pyramid and shallow feature reconstruction network for SAR ship detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 1042–1056, 2023.

[33] P. Yu, A. K. Qin, and D. A. Clausi, "Unsupervised polarimetric SAR image segmentation and classification using region growing with edge penalty," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1302–1317, Apr. 2012.

[34] F. Lang, J. Yang, D. Li, L. Zhao, and L. Shi, "Polarimetric SAR image segmentation using statistical region merging," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 509–513, Feb. 2014.

[35] G. Akbarizadeh, G. A. Rezai-Rad, and S. B. Shokouhi, "A new region-based active contour model with skewness wavelet energy for segmentation of SAR images," *IEICE Trans. Inf. Syst.*, vol. E93-D, no. 7, pp. 1690–1699, Jul. 2010.

[36] G. Akbarizadeh, "A new statistical-based kurtosis wavelet energy feature for texture recognition of SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4358–4368, Nov. 2012.

[37] Z. Yue, F. Gao, Q. Xiong, J. Wang, A. Hussain, and H. Zhou, "A novel attention fully convolutional network method for synthetic aperture radar image segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4585–4598, 2020.

[38] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

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

[40] 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.

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

[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.

[43] D. Chaudhuri and A. Samal, "A simple method for fitting of bounding rectangle to closed regions," *Pattern Recognit.*, vol. 40, pp. 1981–1989, 2007.

[44] L. Li, M. Doroslovački, and M. H. Loew, "Approximating the gradient of cross-entropy loss function," *IEEE Access*, vol. 8, pp. 111626–111635, 2020.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–15.

[46] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Netw.*, vol. 107, pp. 3–11, Nov. 2018.

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

[48] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1571–1580.

[49] 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.*, 2017, pp. 936–944.

[50] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 8759–8768.

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

[52] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX," 2021. [Online]. Available: https://github.com/Megvii-BaseDetection/YOLOX

[53] Y. Feng et al., "A lightweight position-enhanced anchor-free algorithm for SAR ship detection," *Remote Sens.*, vol. 14, 2022, Art. no. 1908, doi: 10.3390/rs14081908.

[54] S. Gao, J. M. Liu, Y. H. Miao, and Z. J. He, "A high- effective implementation of ship detector for SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4019005.

[55] W. Yu, Z. Wang, J. Li, Y. Luo, and Z. Yu, "A lightweight network based on one-level feature for ship detection in SAR images," *Remote Sens.*, vol. 14, no. 14, 2022, Art. no. 3321, doi: 10.3390/rs14143321.

[56] T. Zhang et al., "LS-SSDD-v1.0: A deep learning dataset dedicated to small ship detection from large-scale Sentinel-1 SAR images," *Remote Sens.*, vol. 12, no. 18, 2020, Art. no. 2997.

Jindong Zhang received the B.S. degree in communication engineering from Nanjing University of Science and Technology, Nanjing, China, in 2021, where he is currently working toward the M.S. degree in communication and information system.

His research interests include deep learning and SAR ship detection.

Weixing Sheng received the B.Sc. degree from Shanghai Jiao Tong University, Shanghai, China, in 1988, and the M.S. and Ph.D. degrees from the Nanjing University of Science and Technology, Nanjing, China, in 1991 and 2002, respectively, all in electronic engineering.

Since 1991, he has been with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, where he is currently a Professor. His current research interests include adaptive signal processing.

Hairui Zhu received the B.S. degree in digital media technology from the Nanjing University of Posts and Tele-communications, Nanjing, China, in 2016, and the M.E. degree in multimedia communication from the University of Liverpool, Liverpool, U.K., in 2018. He is currently working toward the Ph.D. degree in electronic science and technology with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China.

His research interests include radar signal processing and image processing.

Shanhong Guo received the B.S. degree in electronic engineering from the National University of Defense Technology, Changsha, China, in 1991, and the M.S. and Ph.D. degrees in communication and information system from the Nanjing University of Science and Technology, Nanjing, China, in 1994 and 2012, respectively.

Since 1994, she has been with Nanjing University of Science and Technology, where she is currently an Assistant Professor. Her current research interests include radar signal processing, and communication and information system.

Yubing Han received the Ph.D. degree in signal and information processing from Southeast University, Nanjing, China, in 2006.

Since 2006, he has been a Faculty Member with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, where he is currently a Professor. His current research interests include SAR imaging, sensor signal processing, radar signal processing, wireless communications, and digital image processing.