A Fast Threshold Neural Network for Ship Detection in Large-Scene SAR Images

Jingyu Cui , *Student Member, IEEE*, Hecheng Jia, *Student Member, IEEE*, Haipeng Wang , *Senior Member, IEEE*, and Feng Xu , *Senior Member, IEEE*

Abstract—Multiscale ship detection in large-scene offshore synthetic aperture radar (SAR) images is of great significance in civil and military fields, such as maritime management and wartime reconnaissance. Methods based on deep learning apply a deep neural network to extract multiscale information from SAR images, which improves detection performance. However, deep neural networks are computationally complex, and even with GPU acceleration, the timeliness of ship detection in large-scene SAR images is still constrained. Methods based on threshold segmentation, in contrast, are efficient and straightforward, but they are less robust and need to be adjusted with complex and changing scenes. This article combines two methods and proposes a lightweight framework based on a threshold neural network (TNN) to achieve fast detection. Specifically, the TNN is carefully designed to extract the grayscale features of the SAR image, which predicts the optimal detection threshold within the sliding window and separates the targets adaptively. In addition, a false alarm rejection network is used to discriminate candidate targets and improve detection accuracy. Experiments are carried out on the public SSDD offshore dataset and the FUSAR-Ship-Detection dataset. The results show that the proposed framework performs 14.43% better than the Multi-CFAR for the SSDD offshore dataset and 7.36% better for the FUSAR-Ship-Detection dataset when using F1 as the metric. Furthermore, the floating point operations of the proposed framework are only 1/240 of those of YOLO-v4 with comparable performance.

Index Terms—False alarm suppression network (FSN), FUSAR-Ship-Detection dataset, lightweight ship detection framework, synthetic aperture radar (SAR), threshold neural network (TNN).

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has high application value in civil and military fields due to its all-day all-weather imaging capability and superior penetrating ability. Ship detection is an essential branch of SAR image interpretation, as the ship is the critical target for maritime management and wartime attacks. SAR ship detection has been widely researched, and the detection performance has substantially improved with methods ranging from traditional detectors to deep learning

target detection algorithms. However, the observation scene of SAR images is generally large. It is still challenging to detect multiscale ship targets fast with limited computing power in the large-scene SAR images.

Many ship detection methods for high-resolution SAR images have been proposed in recent years, which can be mainly divided into methods based on threshold segmentation and methods based on deep learning.

Threshold-based methods segment targets from the background by comparing gray values or other features with the threshold. Among them, the constant false alarm rate (CFAR) detector is one of the most prominent algorithms. The traditional CFAR detector focuses on the gray features of SAR images, which adaptively calculates the detection threshold by analyzing the distribution characteristics of sea clutter. The improved CFAR detectors proposed in recent years mainly include two aspects: more accurate clutter statistical models and more intelligent clutter sample selection. Qin *et al.* [1] introduced the generalized Gamma model into the SAR ocean clutter parameter estimation. The model can degrade to many common distributions in some cases, which improves the fitting ability for different sea clutter. The good fitting performance of the Fisher distribution and the generalized Gamma distribution was demonstrated in [2] with different research cases. Both the distributions are complementary and have some limitations, but the robustness of the generalized Gamma distribution is often preferable. To model SAR images, Leng *et al.* [3] proposed a fast estimation method for complex generalized Gaussian distribution, which efficiently estimated the shape parameter through the complex signal kurtosis (CSK). Another challenge for traditional CFAR detectors is to select clutter sample adaptively and exclude target interference inside the estimation window, especially for dense multiscale targets. Tao *et al.* [4] used truncation to eliminate possible targets within the sliding window, and the remaining clutter was modeled and computed by a truncated statistical model. The outlier-robust CFAR detector proposed by Ai *et al.* [5] used an adaptive threshold to exclude high-intensity clutter outliers and select samples from the reference window, followed by a truncated maximum-likelihood (ML) estimator to calculate the sea parameters. In [6], an object proposal generator was introduced, and the fixed guard window was replaced by the generated target proposal, thus realizing the variable guard window. In addition, researchers have attempted to employ superpixel as the detection unit, segmenting targets using the spatial relationships and statistical characteristics between

pixels [7], [8]. In addition to the grayscale difference caused by the scattering intensity, many other features are also extracted, and targets are separated by the threshold in feature spaces, such as the polarization cross-entropy in [9], the reflection symmetry in [10], the CSK in [11], the variance weighted information entropy in [12], and the modified Fisher metric (lognormal ρ-metric) in [13]. These algorithms consider more physical characteristics and show good performance in polarization SAR data.

Methods based on deep learning are inherited from target detection in optical images, which are mainly divided into networks based on anchor boxes and networks based on key points. Anchor-based networks further include two-stage networks and one-stage networks. Faster-RCNN [14] is one of the most representative two-stage detection networks, where the anchor boxes are first regressed by the region proposal network (RPN) to generate region proposals, and then, more refined regression and classification are performed for the final bounding boxes. After that, the RPN [15] introduced the feature pyramid to fuse multi-scale features, which increases the ability to characterize small targets. Mask-RCNN [16] proposed the ROI Align to replace ROI Pooling, using bilinear interpolation to reduce feature misalignment, and a segmentation branch was added to the prediction header. Without generating proposals, one-stage detection networks remove the RPN and directly regress the anchor boxes. YOLO [17], SSD [18], and RetinaNet [19] are typical one-stage detection networks. Many target detection networks have recently abandoned anchor boxes in favor of directly regressing key points. For example, ConerNet [20] predicts the top-left corners and bottom-right corners, ExtremNet [21] predicts four extreme points and one center point, and CenterNet [22] predicts the center point and the bounding box size without grouping pairs of key points. In the field of SAR, Fu *et al.* [23] resolved the problem of SAR-optical reciprocal translation with a cascaded-residual adversarial network to assist SAR image interpretation. Chen *et al.* [24] realized the SAR automatic target recognition by all-convolutional networks with only sparsely connected layers. To solve the multiscale ship detection in multiscene SAR images, Jiao *et al.* [25] introduced a densely connected network with multiscale feature fusion and used the focal loss to reduce the weight of easy examples in the loss function. Cui *et al.* [26] refined the multiscale features in the FPN using the convolutional block attention module, which highlights salient information at the channel and spatial scales. Sun *et al.* [27] added the guidance vector to the classification branch of the FCOS network, and the proposed category-position module optimized the position regression. Yang *et al.* [28] improved the RetinaNet to predict rotatable bounding boxes, employing various strategies to solve the problems such as feature scale mismatches, contradictions between different learning tasks, and unbalanced distributions of positive samples in SAR ship detection.

In SAR ocean images, ship targets are relatively dispersed with multiple scales, and the background is complex and variable. Ship detection in these scenes is primarily focused on timeliness and achieving a good balance of precision and recall, which necessitates a fast detection algorithm with a low missing rate and fewer false alarms. Deep neural networks, which are computationally complex, are used in methods based on deep learning to extract features of SAR images. As a result, the requirements for the runtime environment are high, and the timeliness is constrained. Threshold-based methods, in contrast, frequently rely on certain assumptions and hyperparameters that must be adjusted to fit complicated and varied SAR scenes. In order to achieve fast detection for large-scene offshore SAR images, we combine two methods and design a lightweight framework based on threshold segmentation. First, a threshold neural network (TNN) is proposed to predict the optimal detection threshold within the sliding window and adaptively segment ships. Then, a false alarm suppression network (FSN) is adopted to discriminate candidate targets and improve the detection accuracy.

The main contributions of this article are summarized as follows.

1) A lightweight TNN with optimal segmentation ideas is proposed, which achieves fast ship detection in large-scene SAR images.
2) A pixel-level target segmentation metric is introduced to calculate the optimal detection threshold according to labeled bounding boxes, and the loss function of the TNN is developed based on it.

The rest of this article is organized as follows. Section II introduces the proposed lightweight framework with details about the TNN and the FSN. In Section III, the experiments on SSDD and FUSAR-Ship-Detection datasets are presented, and the analysis of the results is exhibited. Finally, Section IV concludes this article.

II. PROPOSED METHOD

A. Physical Background

The SAR imaging of sea surface and ships can be described as a composite scattering problem of the target and rough surface. The backscatter intensity of objects determines the gray values of pixels in the SAR amplitude images.

When under low-sea state or the sea surface is relatively calm, the sea is dominated by specular scattering, and the energy of electromagnetic echo is small. When under poor sea conditions or when the wind speed is high, the wave movements cause the sea surface to be rough, and the sea is reflected mainly by diffuse scattering. Compared with the natural distributed scenarios, such as sea surface, man-made metallic ship targets consist of planes, dihedral, and more complicated structures [29]. Hence, ships have more complex scattering mechanisms than the background, including direct reflection from areas perpendicular to the radar beam, corner reflections, and multiple reflections from the ship and the sea surface, resulting in ships appearing brighter in SAR images. What cannot be ignored is that backscatter from ships is also determined by construction material and characteristics of the radar instrument, such as incidence angle, frequency, polarization, and resolution [30].

From an electromagnetic perspective, a ship can be considered as a dominant scatterer characterized by a strong and coherent backscattered signal [30]. The difference in brightness between
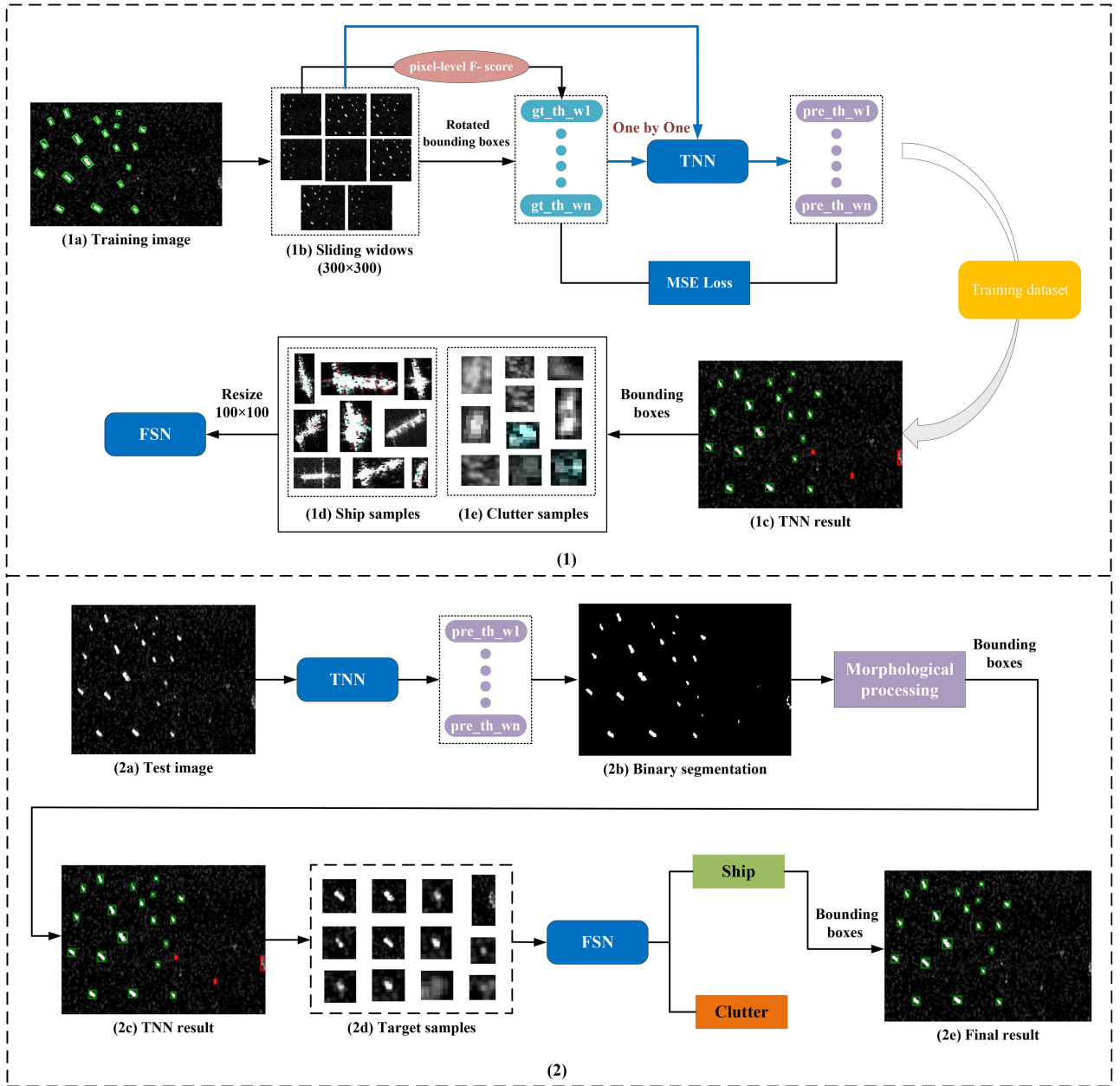
Fig. 1. Flowchart of the lightweight framework. (1) Overall training flowchart. (2) Overall test flowchart.

the ship targets and the background allows us to segment them using a detection threshold, which is the reason why CFAR is one of the most frequently used ship detection methods for SAR images.

The proposed two-stage lightweight framework aims to realize fully automatic threshold segmentation and discrimination. As shown in Fig. 1, the algorithm consists of two stages: 1) TNN and 2) FSN. The SAR image is first sent into the TNN, which predicts the optimal detection threshold of each sliding window in sequence. Then, the ship targets are segmented by comparing the SAR image with the thresholds. After simple morphological processing for the binary segmentation image, the bounding boxes of the targets can be obtained, as shown

in (2c). Finally, the FSN is applied to identify the candidate targets and get the final detection result, as shown in (2e). When training the TNN, SAR images are cropped into sliding windows with the size of 300×300, and the optimal detection threshold (gt_th) of each input window is calculated according to truth rotated bounding boxes. Then, the ship and clutter samples detected by the TNN are resized to 100×100 to train the FSN.

The details of the two-stage lightweight framework are introduced in the rest of this section. First, the loss function and the architecture of the TNN are described to show how they work. Then, the structure of the FSN is briefly introduced.
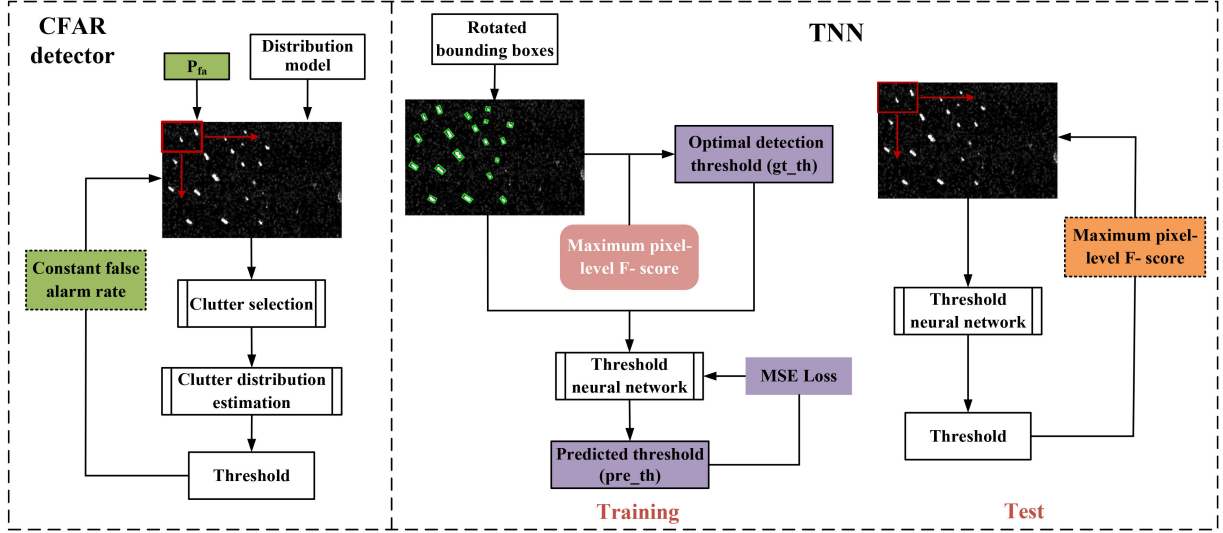
Fig. 2. Comparison of the CFAR detector and the TNN. The left is the CFAR detector, and the right is the TNN.

B. Lightweight TNN

Inspired by the traditional CFAR detector, the TNN extracts the multiscale features from the SAR image and predicts the detection threshold adaptively.

The workflows of the traditional CFAR detector and the TNN are compared in Fig. 2. The CFAR detector relies on the given false alarm probability (P_{fa}) and the sea clutter distribution model to determine the detection threshold. It can theoretically obtain the threshold for a specific P_{fa} by estimating the parameters of the clutter. However, in SAR ship detection, the false alarm rate is not the only metric considered, and the recall rate should also be focused on. Missed detections, especially for multiscale ship detection in large scenes, are often more fatal than false alarms. Therefore, the TNN intends to segment ship targets with the maximum harmonic average of precision and recall.

As shown in Fig. 2, the whole SAR image is fed into the trained TNN with no additional variables; then, the detection threshold for each sliding window is predicted in sequence. In high-resolution SAR images, assigning a detection threshold to each pixel is unnecessary, which leads to repeated calculations for homogeneous regions and low computational efficiency in large scenes. Therefore, setting a uniform threshold within a small window can often better balance the detection speed and accuracy. Moreover, overlapping sliding windows are used in the TNN to ensure the integrity of the detected targets.

1) Loss Function: It is assumed that pre_th is the detection threshold predicted by the TNN, and gt_th is the optimal threshold determined from the truth labels. Then, the simple mean square error (MSE) is used to measure the loss between pre_th and gt_th since the output of the TNN is only one parameter

$$Loss = \mathrm{MSE}(pre\_th, gt\_th). \tag{1}$$

The target detection dataset commonly uses bounding boxes as truth labels, so the optimal detection threshold (gt_th) should be defined to train the TNN. From the labeled rotated bounding boxes, the positions of the targets can be obtained, and the gt_th should minimize the clutter pixels and reserve the most target pixels. Therefore, a pixel-level F-score is introduced for target segmentation, similar to the F-score metric in target detection. The pixel-level F-score is defined as the harmonic average of the pixel-level precision and the pixel-level recall, which ranges from 0 to 1, and a higher value denotes better segmentation performance.

As displayed in Fig. 3, the pixel-level F-score of a threshold is calculated according to the truth rotated bounding boxes, with intermediate variables of the pixel-level precision and the pixel-level recall. The pixel-level precision is defined as the ratio of correctly detected target pixels to all segmented pixels, and the pixel-level recall is defined as the ratio of correctly detected target pixels to all target pixels. The segmentation result can be obtained by a threshold such as (b), and segmented pixels can be classified into green and red pixels. The green pixels inside the bounding boxes are correctly detected target pixels, whereas the red pixels outside the bounding boxes are false alarms. Thus, the theoretical calculation of the pixel-level F-score of the threshold is as follows:

$$F_{\beta_p} = (1 + \beta^2) \cdot \frac{\mathrm{precision}_p \cdot \mathrm{recall}_p}{(\beta^2 \cdot \mathrm{precision}_p) + \mathrm{recall}_p} \tag{2}$$

$$\mathrm{precision}_p = \frac{\mathrm{TP}_p}{\mathrm{TP}_p + \mathrm{FP}_p} \tag{3}$$

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

where F_{β_p} is the pixel-level F-score, $\mathrm{precison}_p$ is the pixel-level precision, and recall_p is the pixel-level recall. These can be obtained by true positive pixels (TP_p), false negative pixels (FN_p) and false positive pixels (FP_p). As shown in Fig. 3(c), true positive pixels are green pixels, and false negative pixels are red pixels. The target pixels are simplified as the area of the rotated
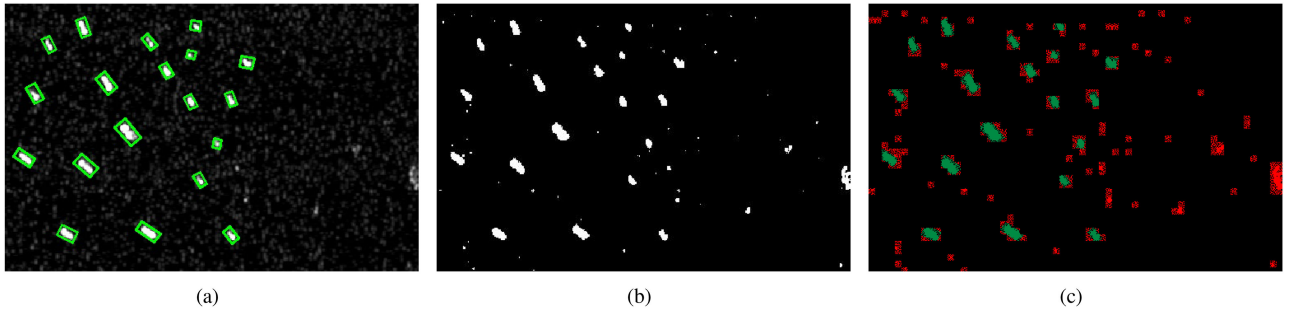
Fig. 3. Calculation of the pixel-level F-score of a threshold. (a) Truth rotated bounding boxes of targets. (b) Segmentation result by a threshold. (c) Green pixels within the boxes are detected ship targets, whereas red pixels outside the boxes are false alarms. The areas of green and red pixels are used to calculate the pixel-level F-score.

Algorithm 1: Optimal Detection Threshold (gt_th).

Input:
SAR image I;
Truth rotated bounding boxes $\boldsymbol{gtboxes}$;

1   $T_0 \leftarrow \frac{1}{m} \sum_{i=1}^{m} x_i$, x_i not in gtboxes
    // mean value of background;
2   $gtship \leftarrow Area(gtboxes)$
    // approximate number of truth ship pixels;
3   **for** $T \leftarrow T_0$ **to** 255 **do**
4     $\mathrm{TP}_p \leftarrow Area(I > T \text{ in gtboxes})$;
5     $\mathrm{FP}_p \leftarrow Area(I > T \text{ not in gtboxes})$;
6     $precision_p \leftarrow \mathrm{TP}_p / (\mathrm{FP}_p + \mathrm{TP}_p)$;
7     $recall_p \leftarrow \mathrm{TP}_p / gtship$;
8     $F_{\beta_P} \leftarrow (1 + \beta^2) \cdot (precision_p \cdot recall_p) / \{(\beta^2 \cdot precision_p) + recall_p\}$;
9   **end**

Output:
Optimal detection threshold $\boldsymbol{gt\_th}$;
$\boldsymbol{gt\_th} \leftarrow \operatorname{argmax}(F_{\beta_p}(T))$

Fig. 4. Calculation of the gt_th and the corresponding segmentation results. (a)–(c) demonstrate the variation of the $recall_p$, $precison_p$, and $F_{0.4_p}$ with different thresholds. (d) is the testing SAR image from the SSDD offshore dataset. (e) is the segmentation result using the gt_th. (f) can be obtained by morphological operations such as dilation, region filling, and removing small objects.

bounding boxes to facilitate the calculation. β is the harmonic average factor, and when $\beta = 1$, $precision_p$ and $recall_p$ have equal weight. The larger β denotes the higher weight of $recall_p$.

The optimal detection threshold (gt_th) is obtained as follows.

1) Compute the mean gray value of the background according to the truth rotated bounding boxes.
2) Increase the detection threshold from the mean value to 255 to calculate F_{β_p}.
3) As a result, the threshold with the largest F_{β_p} is the gt_th.

It is clear that the gt_th and the loss of the TNN will be affected by β. In addition, it was found that the threshold segmentation generally results in more false alarm pixels than missing pixels in SAR images, as the unavoidable sea clutter highlights will be segmented. Thus, a smaller β with a higher $precison_p$ weight is more suitable for the TNN. $F_{0.4_p}$ is chosen as the criterion of the optimal detection threshold, and the related analysis will be introduced in the experiments section.

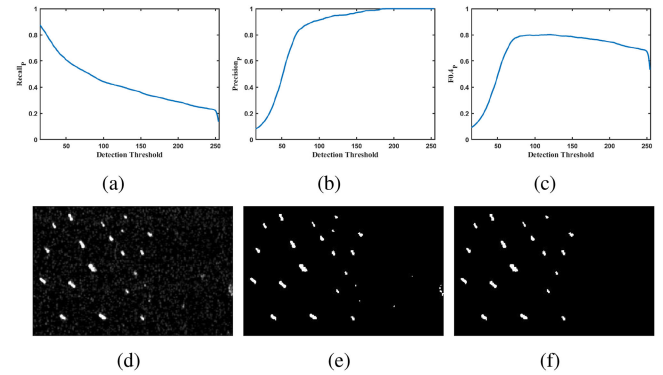With the detection threshold increasing, as demonstrated in Fig. 4(a)–(c), $recall_p$ and $precison_p$ exhibit the trend of monotonic decrease and monotonic increase, respectively. In contrast, the optimal threshold criterion $F_{0.4_p}$ grows first and subsequently decreases, with a highest peak in the middle. As illustrated in Fig. 4(d)–(f), the calculated gt_th can minimize the false alarm pixels while retaining the targets integrally. It should be noticed that some tiny background highlights will be segmented as target pixels, which can be eliminated by the subsequent FSN.

2) Architecture: The TNN differs from traditional target detection networks in that the output is the detection threshold rather than the bounding boxes, and the information extracted is limited to gray features without other complex features. As a result, the structure of the TNN can be greatly simplified, and better segmentation results can be obtained with fewer computing resources.

The lightweight TNN is built on the improved DenseNet, containing three dense blocks. As shown in Fig. 5, each layer in the dense block gets its input from all preceding layers by channel concatenation. In this way, the DenseNet achieves feature reuse, which allows maximal information transmission in each block and improves the effectiveness of the learned features. Since each layer in the block contains the information of all previous layers, the training purpose of each layer is to learn
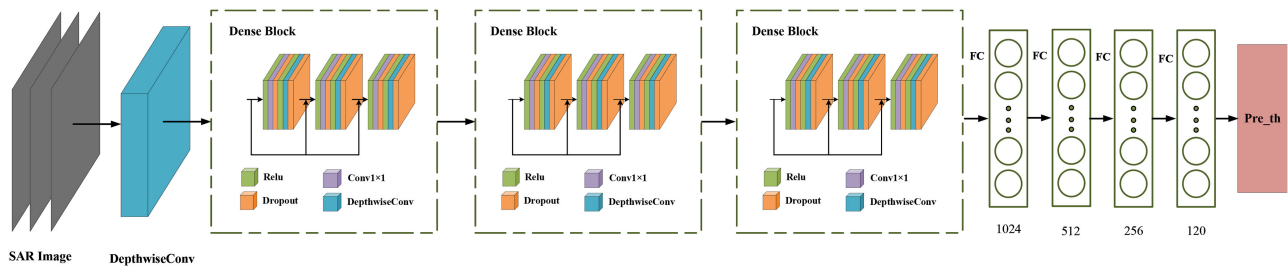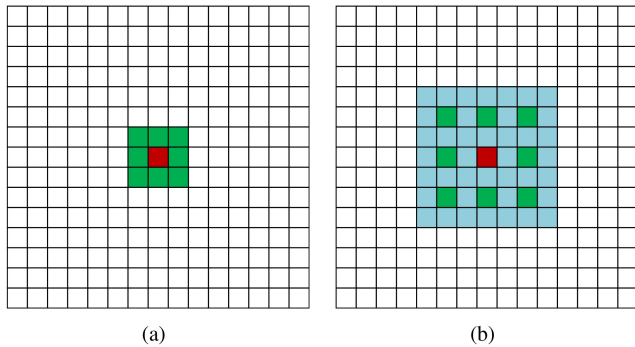
Fig. 5. Framework of the proposed lightweight TNN.

Fig. 6. (a) 3×3 standard convolution; each element has a receptive field of 3×3. (b) 3×3 2-dilated convolution [31]; each element has a receptive field of 7×7.

more features to add to the global features, which drastically reduces the parameters of the TNN.

The nonlinear transformation of each layer in the original dense block consists of batch normalization (BN), ReLU, Conv(1×1), BN, ReLU, and Conv(3×3) in sequence. Conv(1×1) reduces the output dimension of the previous layer, preventing network parameters from spiking due to the considerable increase in feature dimension during channel concatenation. Considering the practical requirements of target segmentation in SAR images, the transfer connection is optimized by removing the BN layer and replacing the Conv(3×3) with depthwise separable dilated Conv(3×3). Thus, the nonlinear transformation of each layer in the improved block is ReLU–Conv(1×1)–DropOut–ReLU–Depthwise Separable Dilated Conv(3×3)–DropOut.

Since the TNN only extracts gray features from SAR images, the BN is removed in hopes of extracting more specific information on the limited gray values. As a result, the hidden layer is not normalized to maintain its diversity, increasing the predicted threshold's sensitivity to gray characteristics in the input window.

As shown in Fig. 6, the 3×3 2-dilated convolution is utilized to replace the standard convolution, resulting in a receptive field expansion from 3×3 to 7×7 with the same parameters. Thus, the gray characteristics of the farther background can be fused in space. Meanwhile, the dilated convolution is operated by a separable form, with depthwise convolution for each channel first and then pointwise convolution to fuse the information between channels. Parameters of standard convolution and depthwise

separable convolution in generating the same output dimension are as follows:

$$N\_std = C_i \times K \times K \times C_o \qquad (5)$$

$$N\_separable = N\_depthwise + N\_pointwise$$
$$= C_i \times K \times K \times 1 + C_i \times 1 \times 1 \times C_o \quad (6)$$

$$\frac{N\_separable}{N\_std} = \frac{C_i \times (K \times K + C_o)}{C_i \times K \times K \times C_o} \qquad (7)$$

where C_i is the input channel, K is the kernel size, H and W are output feature sizes, and C_o is the output channel. Depthwise convolution can be perceived as standard convolution with an output channel of 1, while pointwise convolution is actually standard convolution with a kernel of 1×1. Compared with the standard convolution, the computation of depthwise separable dilated convolution is much lower when having the same large receptive field and the same output dimension.

3) Morphological Processing: The TNN aims to segment targets and obtain bounding boxes adaptively, where some morphological operations are critical for extracting bounding boxes from the binary segmentation. The flowchart of morphological processing in the TNN is shown in Fig. 7, which contains dilation, closing, hole filling, connected component labeling, and removing small objects.

1) *Dilation:* Use a structure element (SE) to stride the binary segmentation, replacing the original pixel with the maximum value in the SE. It enlarges the overall contour of the target and fills in holes smaller than the SE.
2) *Closing:* Dilation first and then erosion (the inverse of dilation). It can fill in the holes and cracks of the target.
3) *Hole filling:* Use the SE to continuously dilate the binary segmentation until convergence, with the complement of the original image as a mask to limit the boundary of the target from expanding. It fills in all the holes inside the objects.
4) *Connected component labeling:* Different connected components are extracted by marking all the neighboring foreground pixels with the same value, which can be achieved by constrained dilation in the connected regions until convergence. The minimum enclosing rectangles of the connected regions are the bounding boxes we need.
5) *Removing small objects:* Each connected region is regarded as a target, and some tiny clutter bright spots are substantially smaller than the actual ships. Therefore,
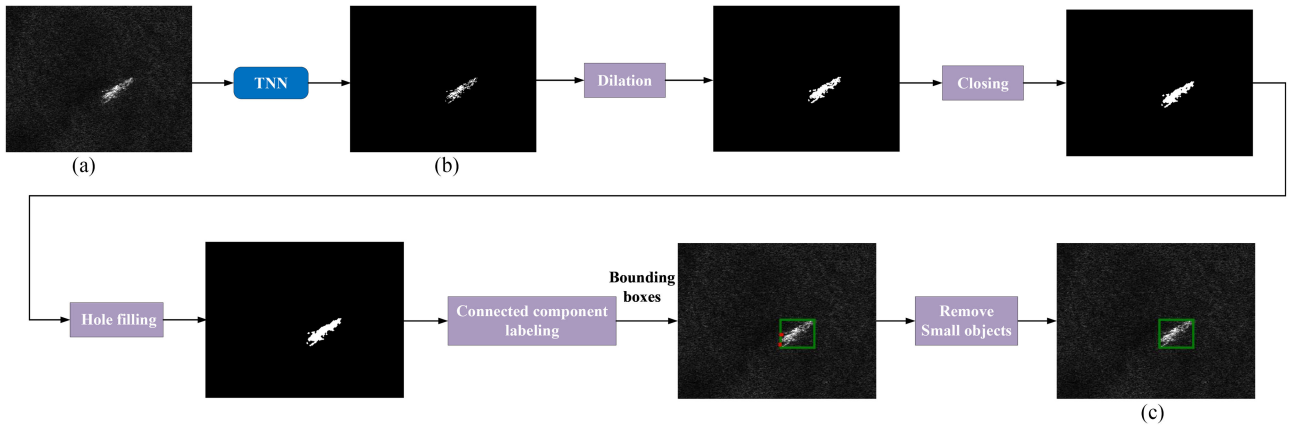
Fig. 7. Flowchart of morphological processing in the TNN. (a) SAR image (b) Binary segmentation. (c) TNN result.
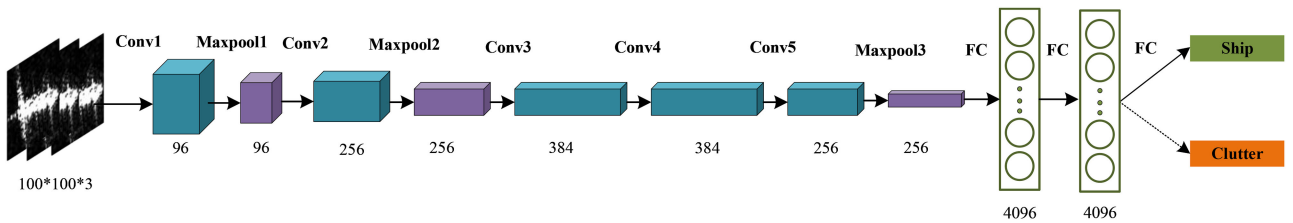

Fig. 8. Framework of the lightweight FSN.

these objects can be eliminated by setting the value of the connected regions to 0 to reduce the false alarms.

C. False Alarm Suppression Network

The unique phenomena in SAR images, such as speckle and Doppler range ambiguity, and the presence of natural phenomena, such as atmospheric fronts and changes in ocean backscattering [32], are all very likely to result in clutter highlights, making it challenging to distinguish from ships relying solely on the detection threshold. Compared with clutter, ships have distinct outlines and specific component structures, which are more evident in high-resolution SAR images. Therefore, a binary classification network is designed to discriminate the candidate targets segmented by the TNN.

Specifically, as shown in Fig. 1, the candidate targets segmented by the TNN can be cropped according to their bounding boxes. These samples will be classified into ships and clutter by the FSN, with the help of deeper features, such as texture and structure, adaptively extracted by the neural network.

To reduce the network parameters, AlexNet is selected as the backbone of the lightweight FSN. AlexNet consists of five convolutional layers and three fully connected layers with ReLU as the activation function. The Dropout layer deactivates some neurons with a certain probability during each iteration, generating different network structures to reduce overfitting. These structures share weights, thus reducing the complex coadaptation relationships of neurons in the fully connected layer. The architecture of the FSN is shown in Fig. 8.

III. EXPERIMENTS

In this section, experiments are set up to evaluate the performance of the proposed lightweight ship detection framework. First, the experimental datasets and settings are introduced. Then, the proposed method is compared with the Multi-CFAR detector and three deep learning detection networks on two datasets to verify its effectiveness. Next, the validity of the backbone and the optimal threshold criterion of the TNN is evaluated. Finally, the proposed method is combined with the sea–land segmentation module to achieve an end-to-end ship detection method in large-scene SAR images.

A. Experiment Dataset and Settings

Two datasets from offshore SAR images and large-scene SAR images are chosen for evaluation. The details are as follows.

1) SSDD Offshore [33]: The public SSDD dataset was constructed by Li *et al.* [33], which consists of images from multiple SAR satellites, including RadarSat-2, TerraSAR-X, and Sentinel-1. In 2021, the SSDD dataset was officially revised as three annotation types [34] with standardized training and test sets that include offshore and inshore scenarios. The SSDD offshore dataset is composed of 476 images with 1520 targets in the training set and 186 images with 374 targets in the test set. Since gt_th is calculated by truth rotated bounding boxes, the RBox-SSDD offshore training dataset is selected to train the TNN. Moreover, the proposed framework is applied to multiscale ship detection in offshore scenarios, and the detection results are minimum enclosing rectangles. Thus, the BBox-SSDD offshore test dataset is selected to evaluate the performance.
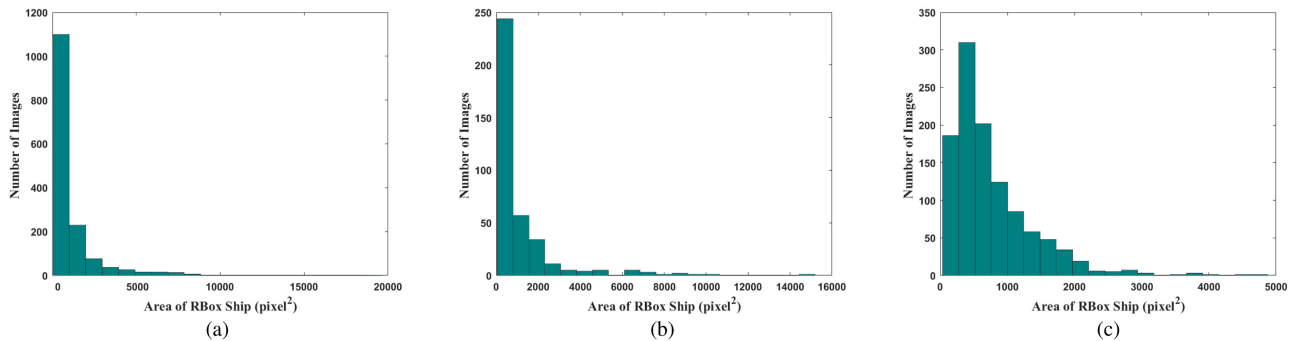
Fig. 9. Distribution of ship area in two datasets. (a) SSDD offshore training set. (b) SSDD offshore test set. (c) FUSAR-Ship-Detection dataset.

TABLE I
INFORMATION OF THE EXPERIMENT DATASETS

	SSDD offshore train	SSDD offshore test	FUSAR-Ship-Detection
Number	476	186	105
Mean size	483×332	477×330	5000×5000
Ship num	1520	374	1094

2) FUSAR-Ship-Detection: The FUSAR-Ship proposed in [35] is a SAR ship identification dataset containing slices of key ship targets and various false alarms. In this article, a FUSAR-Ship Detection dataset for ship detection is constructed to validate the algorithm's robustness, which contains 105 SAR images with 1094 ship targets. First, 70 GF-3 images with 1.2-m resolution in both azimuth and range resolutions are downsampled to 4.6 m and cut into large-scene images of 5000×5000 pixel size (compared to the average 481×331 of SSDD). Then, the offshore scenes are selected and labeled.

Table I shows the information of the datasets used in this article. The distribution of ship areas in two datasets is shown in Fig. 9, using pixels as the statistical unit. Small-scale targets account for the vast majority in the SSDD offshore training and SSDD offshore test datasets, and there are extreme values with distributing nonuniformly. In contrast, the vast majority of the targets in the FUSAR-Ship-Detection dataset are small and medium sized, and the target area distribution is relatively uniform.

With an input window size of 300×300 pixels, the TNN is trained on 3504 image slices cropped from 476 SAR images in the SSDD offshore training dataset. For a total of 20k training steps, we set a 32-batch training size and a 1e-5 initial learning rate. At 12k and 16k steps, the learning rate drops by tenfold.

The positive and negative samples from the SSDD offshore training dataset segmented by the TNN are used to train the FSN with an input size of 100×100. In order to improve the discrimination performance, seven transformations are utilized to augment the samples: rotated 90°, rotated 180°, rotate 270°, added salt-pepper noise, added Gaussian noise, increased brightness, and decreased brightness. Then, all the samples are randomly divided into the training set and the validation set in the ratio of 9:1. The training batch size is 32, and the initial learning rate is 1e-5. When the loss of the validation set decreases for

five consecutive epochs, the learning rate drops to 1/10 of the previous, and the training ends when the loss of the validation set does not decrease for ten consecutive epochs.

Specifically, the TNN and the FSN are implemented in the TensorFlow framework, and three deep learning networks in the PyTorch framework. All the experiments are performed on a NVIDIA Corporation GP102 GPU card with CUDA 10.1.

B. Evaluation Criteria

The proposed framework is evaluated quantitatively in terms of both accuracy and timeliness. Accuracy criteria used in this article are precision, recall, AP, and F1, which are as follows:

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

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

$$\text{F1} = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \tag{10}$$

where TP is the correctly detected ship, FP is the clutter mistakenly detected as a ship, and NP is the missing ship. AP is the average precision, which is calculated by the area under the precision–recall curve.

The floating point operations (FLOPs) of a CNN model represent the amount of computation required for forward propagation, which is often used to measure the complexity of the neural network, for example, for the convolution operation. When containing the bias, its FLOPs can be calculated as follows:

$$\text{FLOPs}_{\text{conv}} = 2 \times C_i \times K^2 \times H \times W \times C_o \tag{11}$$

where C_i is the input channel, K is the kernel size, H and W are output feature sizes, and C_o is the output channel. We use GFLOPs in the subsequent subsection, and 1 GFLOPs $= 10^9$ FLOPs.

C. Comparisons With Other Methods

In this subsection, the Multi-CFAR detector and three deep learning methods are used to compare with the proposed framework. Experiments are conducted on the SSDD offshore test dataset and the FUSAR-Ship-Detection dataset to verify the

TABLE II
DETECTION PERFORMANCE OF SSDD OFFSHORE TEST DATASET COMPARISON BETWEEN DIFFERENT METHODS

Detection Method	Runtime (s)	FLOPs (G)	Recall	Precision	AP	F1
Multi-CFAR	7.386	–	0.9011	0.7539	0.7328	0.8210
RetinaNet	28.875	48.96	0.9492	0.9621	0.9483	0.9556
YOLO-v3	13.905	32.76	0.9786	**0.9786**	0.9772	**0.9786**
YOLO-v4	16.867	33.15	**0.9866**	0.9685	**0.9845**	0.9775
Multi-CFAR+FSN	15.722	–	0.8930	0.9598	0.8897	0.9252
TNN	**6.053**	**0.127**	**0.9866**	0.6685	0.7205	0.7970
TNN+FSN (Ours)	13.469	0.138	0.9679	0.9628	0.9620	0.9653

TABLE III
DETECTION PERFORMANCE OF FUSAR-SHIP-DETECTION DATASET COMPARISON BETWEEN DIFFERENT METHODS

Detection Method	Runtime (s)	FLOPs (G)	Recall	Precision	AP	F1
Multi-CFAR	421.82	–	0.9644	0.7891	0.8430	0.8680
RetinaNet	671.83	48.96	0.9397	0.6461	0.8363	0.7657
YOLO-v3	481.12	32.76	0.9689	0.7157	0.8589	0.8233
YOLO-v4	510.63	33.15	0.9655	0.7727	0.8671	0.8584
Multi-CFAR+FSN	579.38	–	0.9489	0.8764	0.9297	0.9112
TNN	**134.86**	**0.127**	**0.9963**	0.6012	0.7833	0.7499
TNN+FSN (Ours)	290.73	0.138	0.9653	**0.9191**	**0.9516**	**0.9416**

effectiveness of our method. Furthermore, the proposed framework and the deep learning methods are trained on the SSDD offshore training dataset and then transferred to the FUSAR-Ship-Detection dataset without any fine-tuning.

The Multi-CFAR proposed in [36] includes global CFAR, large-scale CFAR, and small-scale sliding CFAR. Their estimation regions are the whole SAR image, the sliding windows, and the neighborhood of the candidate targets, respectively. Then, eigenellipse discrimination and ML discrimination are carried out in Multi-CFAR to reduce false alarms. RetinaNet [19], YOLO-v3 [37], and YOLO-v4 [38] are all one-stage target detection networks with high inference speeds. RetinaNet dynamically increases the weight of the hard examples by focal loss, surpassing the two-stage detection networks both in speed and accuracy for the first time. YOLO-v3 contains excellent structures, such as Darknet-53 and FPN, which balance detection speed and accuracy and improve the detection performance of small targets. Some latest tricks, such as Mosaic augmentation, SPP module, Mish activation, CIOU Loss, and DIOU-NMS, are used in YOLO-v4 to provide more efficient detection. The detection performance of different methods is shown in Tables II and III.

As shown in Table II, the Multi-CFAR detector can guarantee a low missed detection rate for the SSDD offshore test dataset, but it results in more false alarms. With a neural network and multiscale feature extraction, the detection performance of three deep learning methods is considerably improved compared to the traditional detector. The RetinaNet performs more poorly than the YOLO, while the YOLO-v3 and the YOLO-v4 dominate in precision and recall, respectively. The highest recall rate can be obtained when employing the proposed TNN to segment the ship targets, but the false alarms are increased at the same time. After applying the FSN, the precision of both the TNN and the Multi-CFAR improves substantially. However, the final

recall of Multi-CFAR+FSN is 7.49% lower than TNN+FSN (ours), which demonstrates the superiority of the TNN over the traditional CFAR. The TNN+FSN outperforms the RetinaNet and achieves comparable performance to the YOLO. In addition, deep learning methods use more complex models to generate prior regions and regress targets, resulting in much larger FLOPs than ours. The FLOPs of the TNN+FSN represent the amount of computation required for forward propagation in two neural networks. However, we cannot directly equate the FLOPs with the detection runtime. The remarkable thing is that each candidate target detected by the TNN requests a forward propagation of the FSN, which means that more targets result in a longer discrimination time. Even though the detection time of our method varies with the number of targets, it is still a fast ship detection method.

The detection results of the Multi-CFAR detector, YOLO-v4, Multi-CFAR+FSN, and the proposed framework for the SSDD offshore test dataset are displayed in Fig. 10 from left to right. We compared the detection results with the manually labeled true ships and counted the correct detection ships (TP), false alarms (FP), and missing ships (NP) in each image. Many clutter highlights are presented in the poorly focused SAR images, such as Fig. 10(1a)–(1d), and the Multi-CFAR detector generates many false alarms since it relies solely on gray information. Since both the YOLO-v4 and the FSN in our framework extract more characteristics, such as texture and structure, the false alarms can be reduced substantially. The difference in brightness between the target and clutter is relatively small in Fig. 10(2a)–(2d) and Fig. 10(3a)–(3d), and both the Multi-CFAR detector and the YOLO-v3 have more missed detections than our method. In addition, our framework also obtains the best performance in SAR images with dense targets, such as Fig. 10(5a)–(5d).

Given sufficient data, deep learning methods outperform traditional CFAR and the proposed framework. However, when
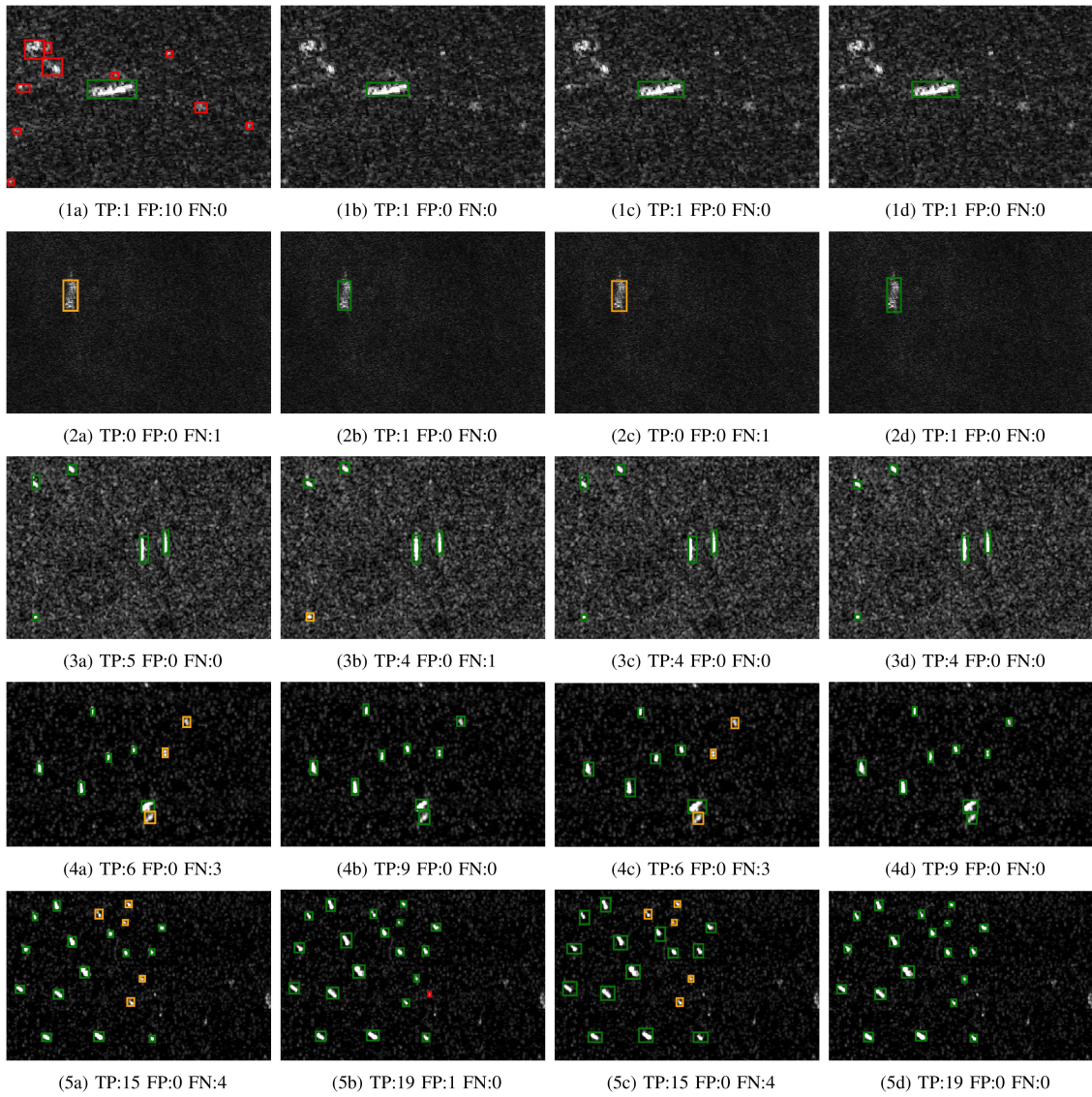
Fig. 10. Comparison results of Multi-CFAR, YOLO-v4, Multi-CFAR+FSN, and proposed framework for the SSDD offshore test dataset. (a)–(d) are results of Multi-CFAR, YOLO-v4, Multi-CFAR+FSN, and our method, respectively. The green rectangles are correctly detected targets (TP), the red rectangles are false alarms (FP), and the orange rectangles are missed targets (FN).

TABLE IV
DETECTION PERFORMANCE OF FUSAR-SHIP-DETECTION DATASET COMPARISON BETWEEN DIFFERENT BACKBONES IN THE TNN

Backbone	RunTime (s)	FLOPs (G)	Recall	Precision	AP	F1
ResNet18	271.04	6.81	0.9840	**0.6740**	0.7341	**0.8000**
MobileNet-v2	213.45	1.32	0.9813	0.6360	0.6696	0.7718
DenseNet	146.15	0.23	0.9854	0.3506	0.5855	0.5172
DenseNet-Dconv	**134.86**	**0.12**	**0.9963**	0.6012	**0.7833**	0.7499

TABLE V
DETECTION PERFORMANCE OF THE MULTI-CFAR AND THE PROPOSED FRAMEWORK

Algorithm	Time	Detected targets	Truth	False alarms	Missing ships	Recall	Precision	F1
Multi-CFAR	56s	205	188	30	13	0.9309	0.8537	0.8906
TNN+FSN (Ours)	5s	206	188	25	7	0.9628	0.8786	0.9188

(1a) TP:7 FP:11 FN:0   (1b) TP:5 FP:3 FN:2   (1c) TP:7 FP:0 FN:0

(2a) TP:22 FP:3 FN:0   (2b) TP:22 FP:10 FN:0   (2c) TP:22 FP:0 FN:0

(3a) TP:63 FP:3 FN:2   (3b) TP:62 FP:14 FN:1   (3c) TP:65 FP:0 FN:0

(4a) TP:14 FP:6 FN:0   (4b) TP:14 FP:2 FN:0   (4c) TP:14 FP:0 FN:0
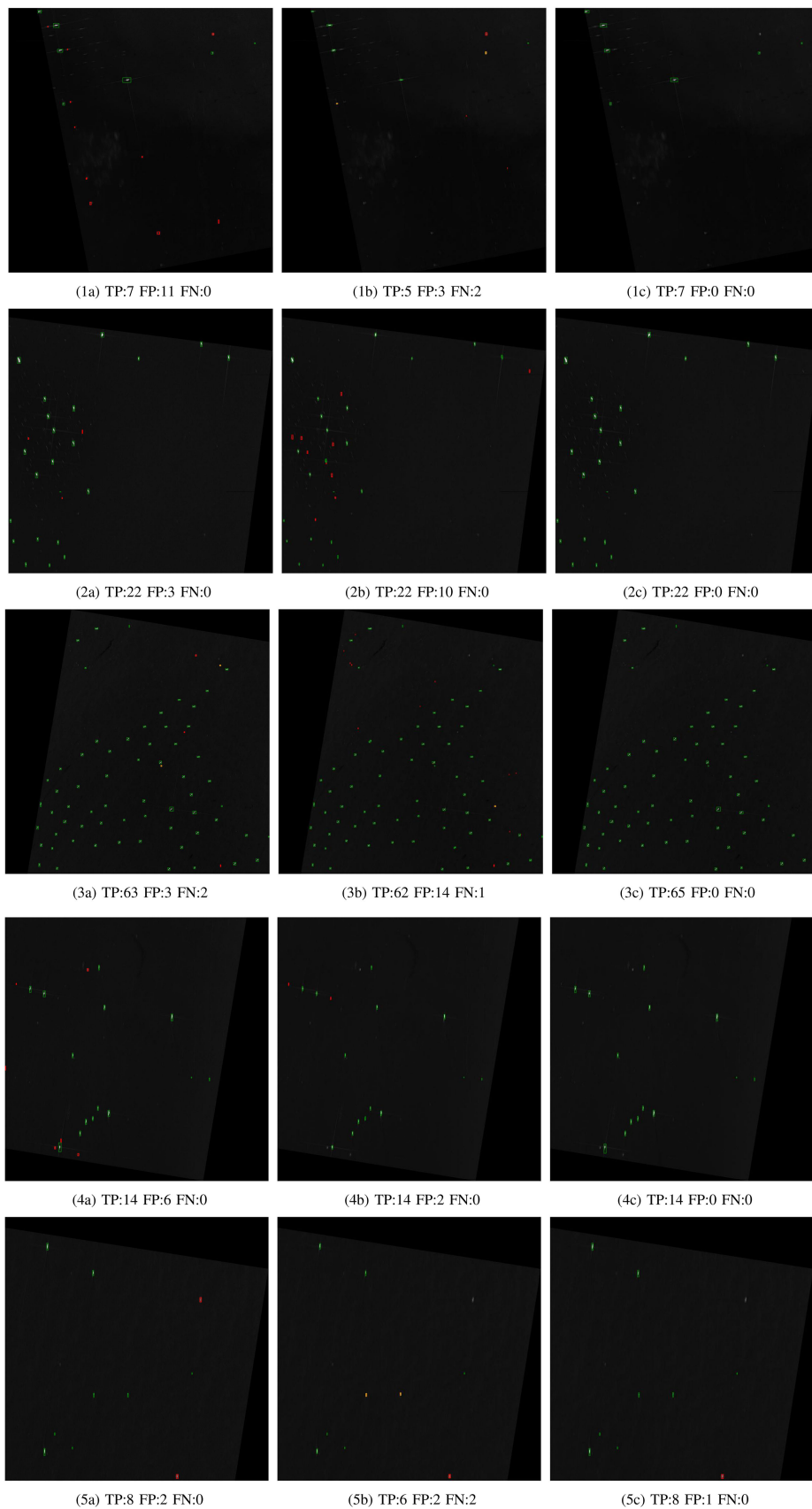
(5a) TP:8 FP:2 FN:0   (5b) TP:6 FP:2 FN:2   (5c) TP:8 FP:1 FN:0

Fig. 11. Comparison results of Multi-CFAR, YOLO-v3 and the proposed framework for FUSAR-Ship-Detection dataset. (a)–(c) are results of Multi-CFAR, YOLO-v3, and our method, respectively. The green rectangles are correct detected targets (TP), the red rectangles are false alarms (FP), and the orange rectangles are missed targets (FN). These are all counted manually by comparing with ground truth.
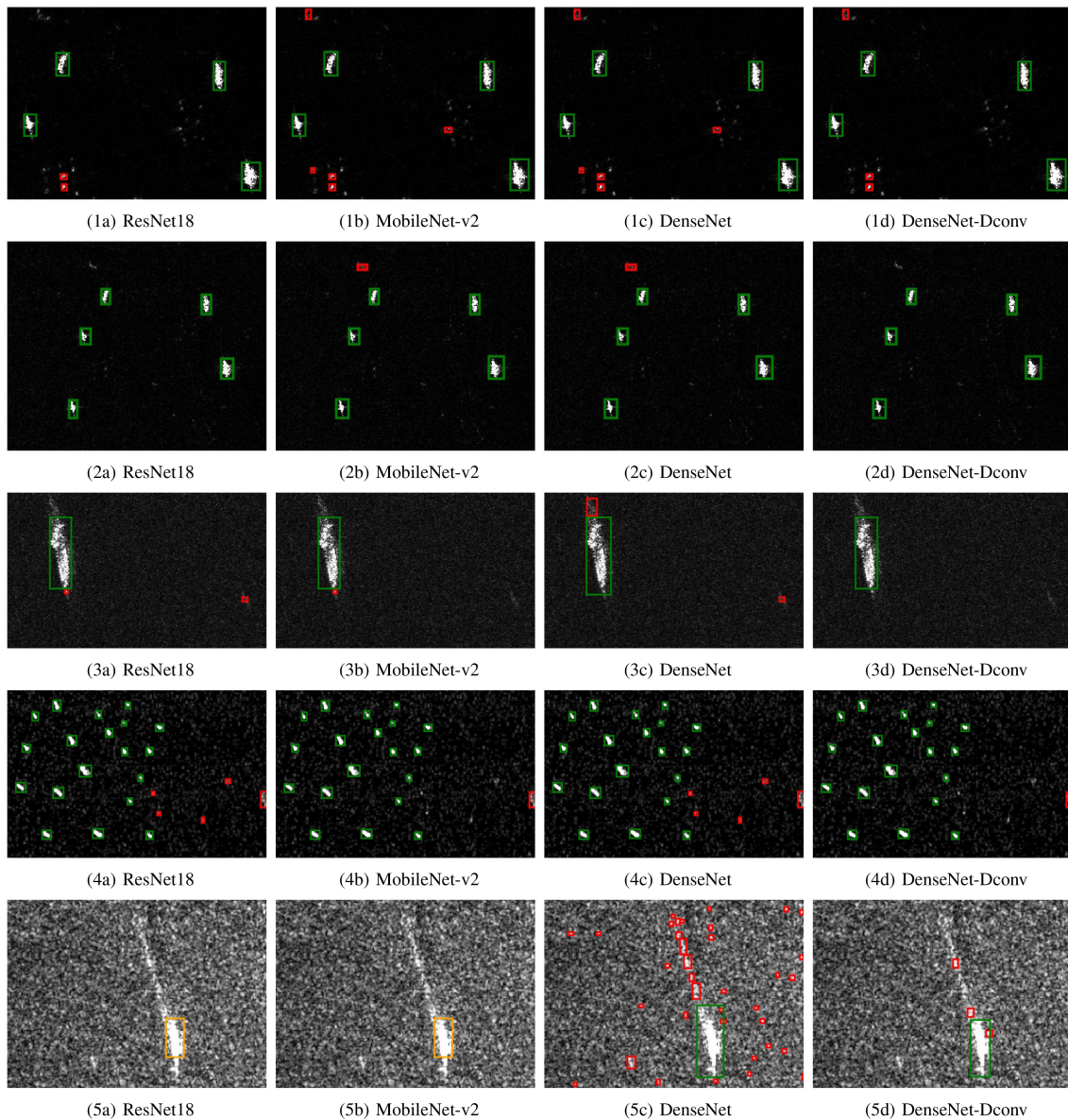
Fig. 12. Comparison results of the TNN with different backbones for the SSDD offshore dataset. (a)–(d) are results of TNN with ResNet18, MobileNet-v2, DenseNet, and DenseNet-Dconv, respectively. The green rectangles are correctly detected targets (TP), the red rectangles are false alarms (FP), and the orange rectangles are missed targets (FN). These are all counted manually by comparing with the ground truth.

applying to new scenes with large differences, the drawbacks of deep neural networks will become apparent. Their generalization ability is limited, and the detection speed will be greatly reduced. In contrast, the proposed framework can better balance detection accuracy and speed in the absence of supervision. In order to validate this inference, we constructed the FUSAR-Ship-Detection dataset and used the models trained on the SSDD offshore training dataset to perform detection without any fine-tuning.

Table III shows the detection performance of the FUSAR-Ship-Detection dataset by different methods. The Multi-CFAR detector can obtain a high recall rate due to the large brightness difference between the targets and the background in the FUSAR-Ship-Detection dataset. In contrast, the detection performance of three deep learning methods on the untrained new dataset is degraded, with no considerable improvement compared to the Multi-CFAR detector. For example, the precision of the YOLO-v3 and the YOLO-v4 are reduced by 26.29% and 19.58%. The proposed TNN can nearly detect all the targets in the FUSAR-Ship-Detection dataset, and the precision improves significantly after adding the FSN. In the meantime, our framework obtains the highest F1-score and AP with a little lower recall than the YOLO-v3 and the YOLO-v4. And the recall and precision are reduced by only 0.26% and 4.37% compared to the SSDD dataset. Moreover, the TNN+FSN has the quickest detection speed of these methods except the TNN, and the FLOPs are substantially smaller than those of the deep learning methods. These results show that the proposed framework can achieve more stable performance with little runtime in new SAR scenarios.
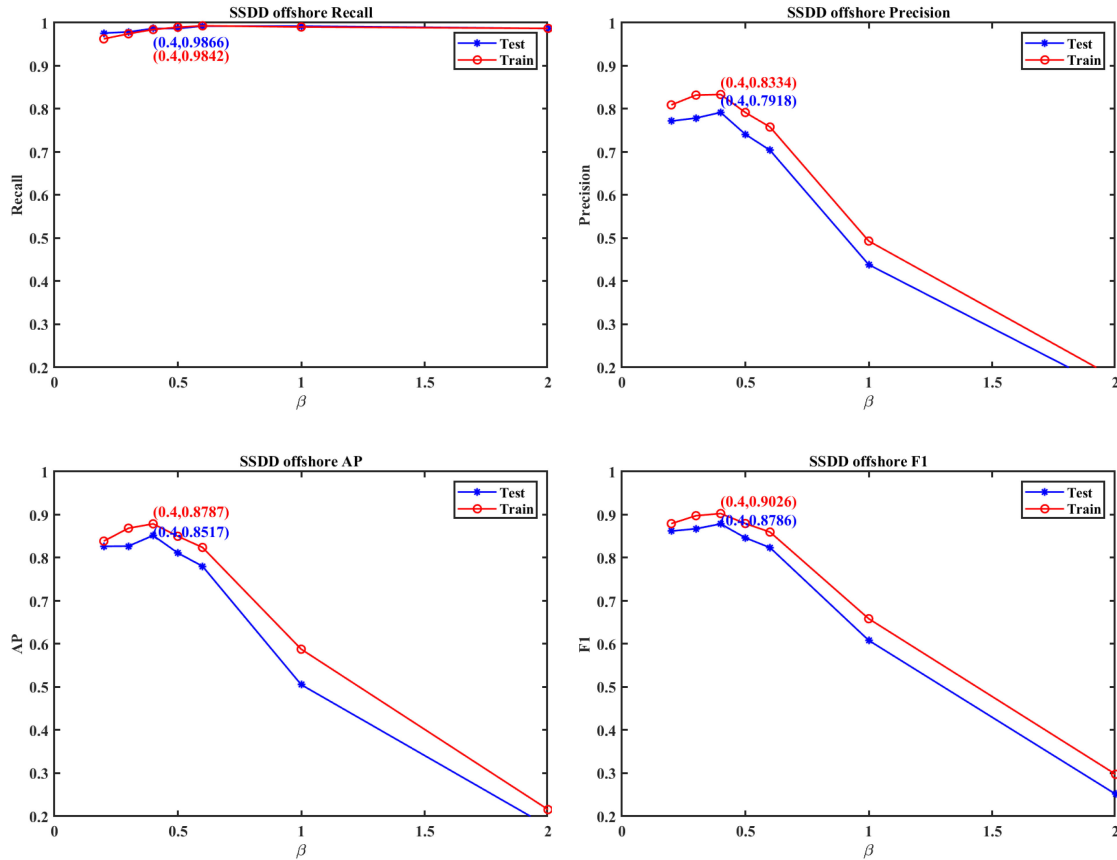
Fig. 13.   Segmentation results obtained by gt_th with different values of β. The red curve is the segmentation results of the SSDD offshore training dataset. The blue curve is the segmentation results of the SSDD offshore test dataset.

Undeniably, the performance of the deep learning methods will be enhanced when using transfer learning techniques to fine-tune the models. However, in some scenarios where training is not possible, such as on-satellite processing and deployed embedded devices, the advantages of the proposed method can be well represented.

Fig. 11 shows the detection results of the FUSAR-Ship-Detection dataset by three methods. The green rectangles are correctly detected targets (TP), the red rectangles are false alarms (FP), and the orange rectangles are missed targets (FN). These are all counted manually by comparing with the ground truth. The Multi-CFAR detector and the YOLO-v3 detection network produce more false alarms than our method in the large-scene SAR images. Furthermore, the Multi-CFAR detector and the YOLO-v3 detection network tend to miss dense targets, such as Fig. 11(3a)–(3c). In addition, the YOLO-v3 detection network is most prone to missed detection in multiscale target scenarios, such as Fig. 11(5a)–(5c).

D. Comparisons With Different Lightweight Backbones in the TNN

In this subsection, different lightweight backbones are applied in the proposed TNN to segment ships from the FUSAR-Ship-Detection dataset, and models of the TNN with different backbones are still trained on the SSDD offshore training dataset.

ResNet18, MobileNet-v2, and DenseNet are selected as backbones of the TNN. ResNet solves the degradation

problem of the deep networks by residual learning for the first time, where ResNet18 has the smallest depth. Mobilenet-v2 introduces the inverted residuals, linear bottlenecks, and depthwise separable convolutions to compress the network. The DenseNet we use contains only three dense blocks, and the feature dimension is maintained at a small value to ensure fewer parameters.

As shown in Table IV, the recall rate in the FUSAR-Ship-Detection dataset maintains above 0.98 as backbone parameters decrease, indicating that the TNN can effectively segment the targets with a lightweight backbone. The MobileNet-v2 has fewer parameters than ResNet18, and the AP and F1 are slightly lower. The Original DenseNet has not learned enough information due to the sharp reduction of parameters, and its precision rate is only half that of ResNet18, implying that a small receptive field with few feature dimensions will result in improper predicted thresholds. In contrast to this, the dilated convolution of the improved DenseNet can increase the receptive field and considerably improve the detection performance, and the depthwise separable convolution can further reduce the parameters of the model. Furthermore, the FLOPs of DenseNet-Dconv are only 1/57 of ResNet18, and the recall rate and AP achieve the highest, while the precision rate and F1 are slightly lower than the highest. In conclusion, the DenseNet-Dconv is chosen as the backbone of the TNN. The detection results of the TNN with different backbones for the SSDD offshore dataset are shown in Fig. 12.
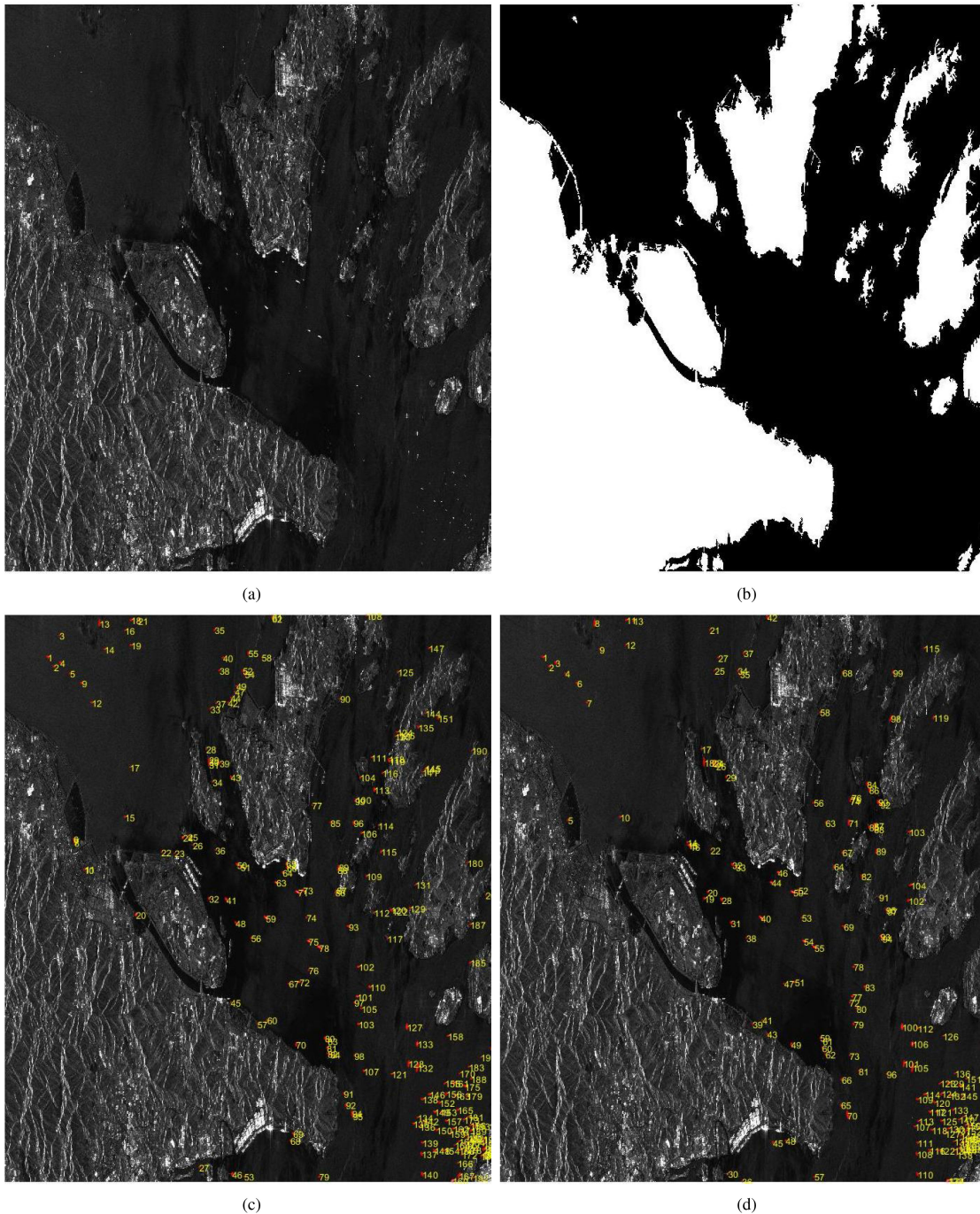
Fig. 14. Comparison results of the Multi-CFAR and the proposed framework for the ALOS-2 SAR image. (a) ALOS-2 SAR image. (b) Refined land–sea segmentation using the method proposed in [36]. (c) Ship detection result of the Multi-CFAR proposed in [36]. (d) Ship detection result of our method.

E. Variants of Optimal Threshold (gt_th) Criterion

In this subsection, experiments are performed on the SSDD offshore dataset using the calculated optimal detection threshold (gt_th) with β taken as 0.2, 0.3, 0.4, 0.5, 0.6, 1, and 2, respectively. It should be noted that the minimum ship area is taken as 30 pixels, and the smaller detected targets are removed.

As demonstrated in Fig. 13, the recall of both the training and test sets can be maintained close to 1 when β ranges from 0.2 to 2. While the precision remains large values as $\beta < 0.5$ with the highest peak at 0.4, it decreases rapidly as β increases. Furthermore, both the AP and the F1-score exhibit similar trends to the precision. In order to ensure a larger recall with a larger precision, $F_{0.4_p}$ is chosen as the criterion of gt_th.

When using gt_th calculated by $F_{0.4_p}$ to detect the SSDD dataset, the AP and the F1-score achieve their maximum with the largest precision in both the training and test sets, and the recall is slightly lower than the highest value.
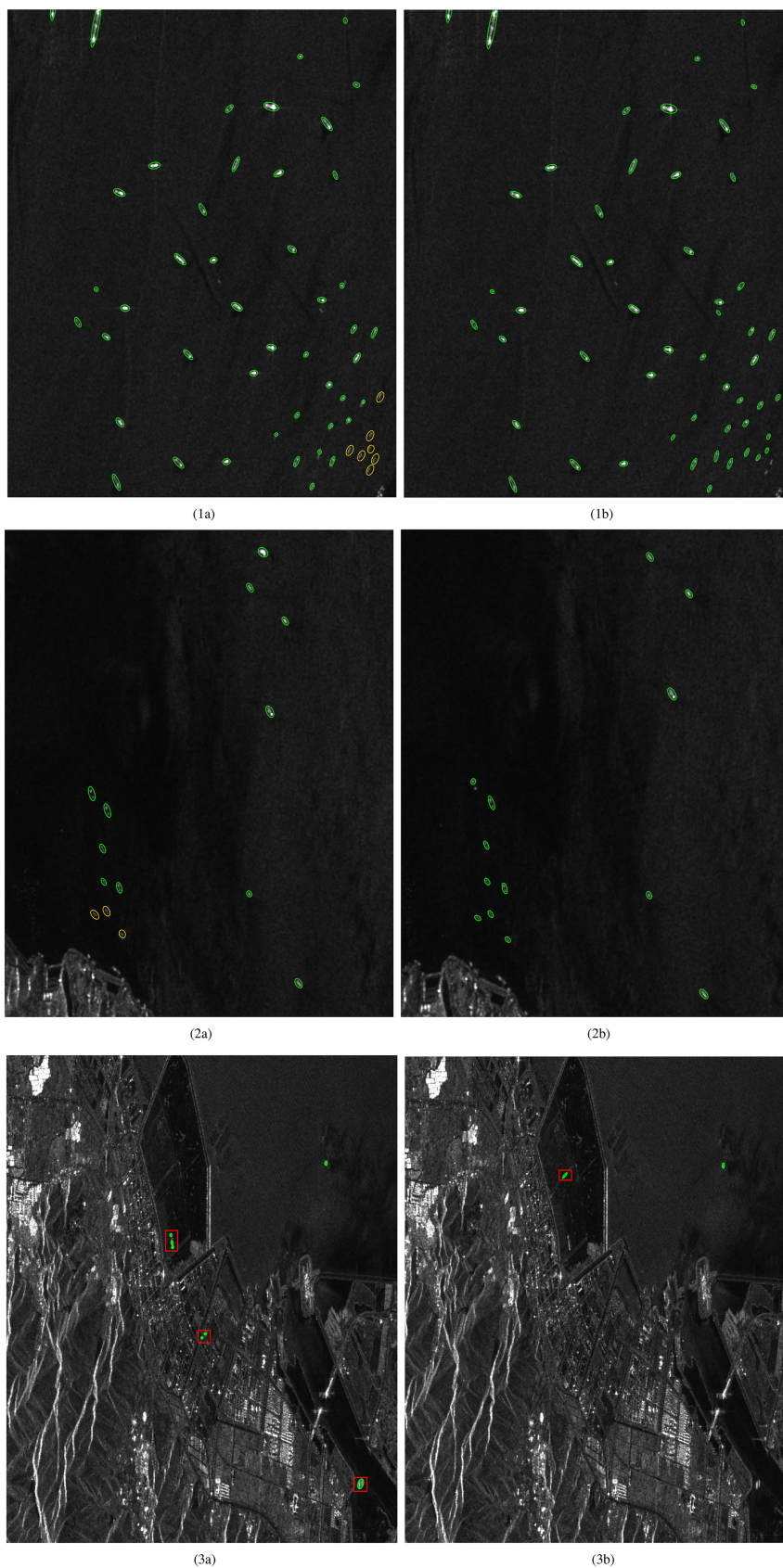
Fig. 15.   Three areas of ship detection results of the Multi-CFAR and the proposed framework for the ALOS-2 SAR image. Detected ships are marked by green ellipses, missing ships are marked by yellow ellipses, and detected targets in red rectangles are false alarms. (1a)–(3a) are the results of the Multi-CFAR, (1b)–(3b) are the results of our method.

F. End-to-End Ship Detection in Large-Scale SAR Images

The proposed framework can be integrated with the sea–land segmentation module to construct an end-to-end ship detection method. In this subsection, experiments are performed using the refined sea–land segmentation module in [36]. The test data is the L-band HH-polarimetric ALOS-2 SAR image of the Zhoushan Islands in Zhejiang Province, China, as shown in Fig. 14(a). The image is 4844×4180 pixels in size, with range and azimuth resolutions of 5.722 and 5.562 m, respectively. The refined sea–land segmentation algorithm is first used to obtain the pixel-level sea–land mask, as shown in Fig. 14(b). On this basis, the Multi-CFAR detector in [36] and the proposed framework are applied to achieve the detection results, as shown in Fig. 14(c) and (d).

As exhibited in Table V, the number of false alarms and missing ships detected by two methods are manually counted. The proposed framework detects almost the same number of ships as the Multi-CFAR detector but fewer false alarms and missing detections.

Three regions are intercepted in this SAR image to compare the detection results of the two methods in a more intuitive way. As shown in Fig. 15, detected ships and missing ships are marked by green ellipses and yellow ellipses, and the targets in red rectangles are false alarms. The Multi-CFAR detector results in more missed detections, while the proposed framework performs better in scenarios where the target scattering intensity is weak and densely distributed. Moreover, even with a refined sea–land mask, some land areas remain, and these land highlights affect the parameter estimation of the sea clutter. Consequently, the Multi-CFAR results in more false alarms in the sea–land mask boundary region. In comparison to the eigenellipse discrimination and ML discrimination [36] used in the Multi-CFAR, the FSN in our method can reduce more false alarms by extracting the deep features of candidate targets.

IV. CONCLUSION

In this article, a lightweight TNN with optimal segmentation ideas was proposed for multiscale ship detection in large-scene offshore SAR images. First, the framework focused on ship segmentation, predicted the detection threshold using the TNN, and performed fast detection for the whole image by sliding windows. Then, the FSN was designed to reduce false alarms with a lightweight model.

Experiments were conducted on the public SSDD offshore dataset and the FUSAR-Ship-Detection dataset. Compared with the Multi-CFAR detector and several deep learning detection networks, the proposed framework can achieve better results with a faster detection speed. Meanwhile, an end-to-end fast detection method was built combined with the sea–land segmentation module for ship detection in a large-scale ALOS-2 SAR image.

The proposed lightweight framework is particularly effective for fast ship detection in large-scene offshore SAR images. Our method obtains the detection threshold by a neural network, which has higher detection accuracy and stronger robustness than other traditional threshold-based detectors. Compared with deep learning detection networks, the lightweight model of

the proposed framework can achieve a good balance between detection efficiency and performance. As a result, our method is more suitable for engineering integration.

REFERENCES

[1] X. Qin, H. Zou, S. Zhou, and Y. Ren, "A generalized gamma distributed CFAR algorithm for layover and shadow detection in InSAR images," in Proc. Chin. Intell. Autom. Conf., 2013, pp. 173–180.

[2] H. Sportouche, J.-M. Nicolas, and F. Tupin, "Mimic capacity of Fisher and generalized gamma distributions for high-resolution SAR image statistical modeling," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 12, pp. 5695–5711, Dec. 2017.

[3] X. Leng, K. Ji, S. Zhou, and X. Xing, "Fast shape parameter estimation of the complex generalized Gaussian distribution in SAR images," IEEE Geosci. Remote Sens. Lett., vol. 17, no. 11, pp. 1933–1937, Nov. 2020.

[4] D. Tao, S. N. Anfinsen, and C. Brekke, "Robust CFAR detector based on truncated statistics in multiple-target situations," IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 117–134, Jan. 2016.

[5] J. Ai, R. Tian, Q. Luo, J. Jin, and B. Tang, "Multi-scale rotation-invariant haar-like feature integrated CNN-based ship detection algorithm of multiple-target environment in SAR imagery," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 12, pp. 10070–10087, Dec. 2019.

[6] 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.

[7] W. Yu, Y. Wang, H. Liu, and J. He, "Superpixel-based CFAR target detection for high-resolution SAR images," IEEE Geosci. Remote Sens. Lett., vol. 13, no. 5, pp. 730–734, May 2016.

[8] 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.

[9] 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.

[10] F. Nunziata, M. Migliaccio, and C. E. Brown, "Reflection symmetry for polarimetric observation of man-made metallic targets at sea," IEEE J. Ocean. Eng., vol. 37, no. 3, pp. 384–394, Jul. 2012.

[11] X. Leng, K. Ji, S. Zhou, and X. Xing, "Ship detection based on complex signal kurtosis in single-channel SAR imagery," IEEE Trans. Geosci. Remote Sens., vol. 57, no. 9, pp. 6447–6461, Sep. 2019.

[12] X. Wang and C. Chen, "Ship detection for complex background SAR images based on a multiscale variance weighted image entropy method," IEEE Geosci. Remote Sens. Lett., vol. 14, no. 2, pp. 184–187, Feb. 2017.

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

[14] 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.

[15] 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.

[16] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 2, pp. 386–397, Feb. 2020.

[17] 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.

[18] W. Liu et al., "SSD: Single shot multibox detector," in Computer Vision. New York, NY, USA: Springer, 2016, vol. 9905, pp. 21–37.

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

[20] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," Int. J. Comput. Vis., vol. 128, no. 3, pp. 642–656, Jan. 2020.

[21] X. Zhou, J. Zhuo, and P. Krähenbühl, "Bottom-up object detection by grouping extreme and center points," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 850–859.

[22] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," 2019, arXiv:1904.07850.

[23] S. Fu, F. Xu, and Y.-Q. Jin, "Reciprocal translation between SAR and optical remote sensing images with cascaded-residual adversarial networks," Sci. China Inf. Sci., vol. 64, no. 2, pp. 1869–1919, 2021.

[24] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

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

[26] 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.

[27] Z. Sun et al., "An anchor-free detection method for ship targets in high-resolution SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 7799–7816, 2021.

[28] R. Yang, Z. Pan, X. Jia, L. Zhang, and Y. Deng, "A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1938–1958, 2021.

[29] R. Guo, J. Cui, G. Jing, S. Zhang, and M. Xing, "Validating GEV model for reflection symmetry-based ocean ship detection with Gaofen-3 dual-polarimetric data," *Remote Sens.*, vol. 12, no. 7, Apr. 2020, Art. no. 1148.

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

[31] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Representations*, 2016.

[32] D. Velotto, F. Nunziata, M. Migliaccio, and S. Lehner, "Dual-polarimetric terraSAR-X SAR data for target at sea observation," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 5, pp. 1114–1118, Sep. 2013.

[33] 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.

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

[35] X. Hou, W. Ao, Q. Song, J. Lai, H. Wang, and F. Xu, "FUSAR-Ship: Building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Sci. China Inf. Sci.*, vol. 63, no. 4, Apr. 2020, Art. no. 140303.

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

[37] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018, *arXiv:1804.02767*.

[38] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
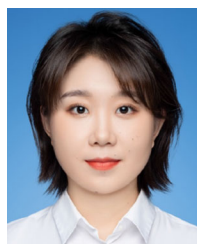
Jingyu Cui (Student Member, IEEE) received the B.S. degree in transportation equipment and control engineering from Northwestern Polytechnical University, Xi'an, China, in 2020. She is currently working toward the M.S. degree in electromagnetic field and microwave technology with the Key Laboratory of Information Science of Electromagnetic Waves, Fudan University, Shanghai, China.

Her research interests include synthetic aperture radar image interpretation and deep learning in remote sensing applications.
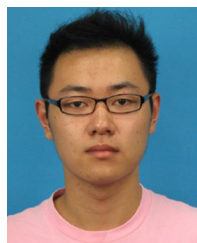
Hecheng Jia (Student Member, IEEE) received the B.E. degree in computer science and technology from the Zhejiang University of Science and Technology, Hangzhou, China, in 2012, and the M.S. degree in software engineering from the University of Science and Technology of China, Hefei, China, in 2015. He is currently working toward the Ph.D. degree in electronics and information with the Key Laboratory of Information Science of Electromagnetic Waves, Fudan University, Shanghai, China.

His research interests include deep learning in remote sensing applications and interactive machine learning.

Haipeng Wang (Senior Member, IEEE) received the B.S. and M.S. degrees in mechanical and electronic engineering from the Harbin Institute of Technology, Harbin, China, in 2001 and 2003, respectively, and the Ph.D. degree in environmental system engineering from the Kochi University of Technology, Kochi, Japan, in 2006.

He was a Visiting Researcher with the Graduate School of Information, Production and Systems, Waseda University, Fukuoka, Japan, in 2008. He is currently a Professor with the Key Laboratory of Electromagnetic Wave Information Science, Department of Communication Science and Engineering, School of Information Science and Engineering, Fudan University, Shanghai, China. His research interests include signal processing, synthetic aperture radar (SAR) image processing and analysis, speckle statistics, applications to forestry and oceanography, and machine learning and its applications to SAR images.

Dr. Wang was the recipient of the Dean Prize of the School of Information Science and Engineering, Fudan University, in 2009, 2017, and 2021. He has been a Member of the Technical Program Committee of IEEE Geoscience and Remote Sensing Symposium since 2011. He is an Associate Editor for IEEE Geoscience and Remote Sensing Letters.

Feng Xu (Senior Member, IEEE) received the B.E. (Hons.) degree in information engineering from Southeast University, Nanjing, China, in 2003, and the Ph.D. (Hons.) degree in electronic engineering from Fudan University, Shanghai, China, in 2008.

From 2008 to 2010, he was a Postdoctoral Fellow with the NOAA Center for Satellite Application and Research, Camp Springs, MD, USA. From 2010 to 2013, he was a Research Scientist with Intelligent Automation, Inc., Rockville, MD, USA. Since 2013, he has been a Professor with the School of Information Science and Technology. He is also the Vice-Dean of the School of Information Science and Technology and the Director of the Key Laboratory for Information Science of Electromagnetic Waves (Ministry of Education). He has authored or coauthored more than 70 articles in peer-reviewed journals and coauthored three books, among many conference papers and patents. His research interests include electromagnetic scattering theory, synthetic aperture radar information retrieval, and intelligent radar systems.

Dr. Xu was a recipient of the Early Career Award of the IEEE Geoscience and Remote Sensing Society (GRSS) in 2014 and the SUMMA Graduate Fellowship in the advanced electromagnetics area in 2007. He received the second-class National Nature Science Award of China in 2011. He is the Founding Chair of the IEEE GRSS Shanghai Chapter and an AdCom Member of the IEEE GRSS. He was an Associate Editor for IEEE Geoscience and Remote Sensing Letters from 2014 to 2020.