# Small Vessel Detection Based on Adaptive Dual-Polarimetric Feature Fusion and Sea–Land Segmentation in SAR Images

Yongsheng Zhou , *Member, IEEE*, Feixiang Zhang, *Student Member, IEEE*, Fei Ma , *Member, IEEE*, Deliang Xiang , *Member, IEEE*, and Fan Zhang , *Senior Member, IEEE*

*Abstract*—Detection of small sea vessels in synthetic aperture radar (SAR) images has received much attention in recent years because the small vessels have weak scattering intensity and few image pixels. The existing detection network structures are not well adapted to small-scale targets, the polarimetric data are not properly utilized, and the sea–land segmentation process to remove land false alarms is time-consuming. Regarding these problems, first, a single low-level path aggregation network is designed specifically for small targets. The structure reduces false alarms at the feature level by finding suitable single-scale feature maps for detection and adding a semantic enhancement module. Second, adaptive dual-polarimetric feature fusion is proposed to filter the multichannel features acquired by dual-polarimetric decomposition to reduce feature redundancy. Third, a segmentation layer is added to the network to shield the land from false alarms. The detection and segmentation layers share the feature extraction and feature fusion modules and are jointly trained by a joint loss. Finally, polarimetric SAR detection and segmentation dataset containing small vessel detection and sea–land segmentation labels is created with reference to the LS-SSDDv1.0 dataset, and experimental results on this dataset verify the improvement of this proposed method over other typical methods.

*Index Terms*—Feature fusion, polarimetric synthetic aperture radar (SAR), sea–land segmentation, small vessel detection.

## I. INTRODUCTION

SYNTHETIC aperture radar (SAR) has an all-day, all-weather observation and morphological detection capability, allowing continuous monitoring of the ocean over long periods of time without interference from clouds, fog, etc. [1]–[3]. With the continuous development of SAR sensors (such as Sentinel-1, Gaofen-3, Radarsat-2, etc.) and the increase of publicly available datasets [4]–[8], a vast amount of researches focus on the detection of sea vessels using SAR images [9]–[14].

One of the most commonly used methods for vessel detection in SAR images is the constant false alarm rate (CFAR) [15]–[20]. CFAR is based on the statistical distribution of clutter and calculates a threshold while maintaining a certain false alarm rate. CFAR is widely used and easily integrated with methods such as superpixels [21]. However, the calculation of the threshold value is usually affected due to the complexity of the background clutter. For polarimetric SAR image, some methods based on polarimetric decomposition have been proposed, and they exploit the difference in polarimetric scattering characteristics between the vessel target and the background clutter [22]–[25]. The polarimetric decomposition is used to find the polarimetric features that have a strong differentiation between background and vessels to complete the detection. The existing vessel detection methods are mostly based on scene-specific statistical properties and are not robust enough in complex scenarios that consist of land or strong sea clutter.

With the recent advancement in convolutional neural network (CNN), many CNN-based detection algorithms have been developed and have achieved better results in various scenarios. They can be generally divided into two broad categories: two-stage and one-stage detection algorithm. The two-stage detection algorithm, as its name suggests, consists of two steps. First, a coarse extraction of the detection boxes is performed using RPN [26]. The extracted detection boxes are then mapped to feature maps and fed into ROI Pooling [26] or ROI Align [27]. Second, the regression and classification of the detection boxes are performed. The representative two-stage algorithms include faster R-CNN [26], cascade R-CNN [28], double-head R-CNN [29], etc. The one-stage detection method performs regression and classification directly, without crude extraction as the first part; so it has an advantage in terms of time consumption. The representative one-stage algorithms include YOLO [30], SSD [31], RetinaNet [32], YOLOv4 [33], YOLOX [34], etc.

To adapt to multiscale targets in optical images, it is a common practice to adopt the FPN [35] with multiscale structure. However, the size of small vessels in SAR images differs significantly

Yongsheng Zhou, Feixiang Zhang, and Fei Ma are with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: zhyosh@mail.buct.edu.cn; 2019200710@mail.buct.edu.cn; mafei@mail.buct.edu.cn).

Deliang Xiang is with the Beijing Advanced Innovation Center for Soft Matter Science and Engineering, Beijing University of Chemical Technology, Beijing 100029, China, and also with the Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: xiangdeliang@gmail.com).

Fan Zhang is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China, and also with the Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: zhangf@mail.buct.edu.cn).
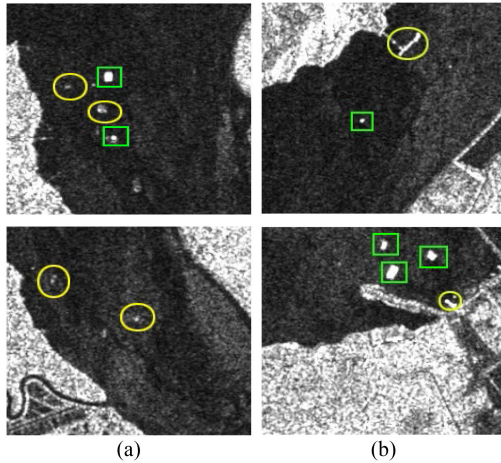
Fig. 1. False alarms (green rectangular boxes are vessel targets, and yellow oval boxes are false alarms). (a) False alarms for clutter. (b) False alarms for strong scatterers on land.

from that of targets in natural optical images, and detection layers with large sensory fields are not suitable for detecting small vessel targets. In addition, the target in an optical image usually occupies the main part of the image and has a large gap with other classes of objects, which does not easily produce more false alarms. In contrast, due to interference from strong land scatterers and sea clutter, SAR amplitude image vessel detection is prone to more false alarms, some of which are shown in Fig. 1.

In order to improve the detection accuracy of small vessel in SAR image and reduce the corresponding false alarm, many methods have been proposed from different aspects.

1) Some scholars have made improvements to the CNN-based detection algorithms from the aspect of network structure. Zhang *et al.* [36] design a separate feature map stitching mechanism instead of FPN for feature fusion to improve the fusion effect. Gao *et al.* [37] use a high-resolution feature map with a downsampling count of 4 times as the detection layer to improve the detection recall of small vessel targets in SAR images. Cui *et al.* [9] also perform detection on a high-resolution feature map with a downsampling count of 4 times. Considering that high-resolution feature maps as detection layers tend to cause more false alarms, the corresponding semantic enhancement modules (SEMs) are added to reduce false alarms. However, the focus on high-resolution feature maps tends to overlook the most suitable feature map scales for small vessel detection in SAR images. High-resolution feature maps are more useful for identifying the texture and edges of small vessel targets, but it is a fairly time-consuming process and more likely to cause false alarms without a proper SEM.

2) Some scholars have made improvements to the CNN-based detection algorithms by introducing multipolarimetric features. Most CNN-based detection methods use amplitude images for training and prediction. Zhang *et al.* [23] demonstrate through extensive experiments that polarimetric decomposition for polarimetric data can well

distinguish small vessel targets from background clutter in some scenarios. Fan *et al.* [38] use polarimetric information to enhance small vessel targets to effectively mitigate the problem of miss detection and false alarms for small vessel targets. However, the polarimetric enhancement process simply synthesizes pseudocolor images and does not make full use of the polarimetric information from multiple channels.

3) In order to effectively reduce false alarms caused by strong scatterers on land, some scholars have introduced CNN-based semantic segmentation into sea–land segmentation of SAR images and achieved better results [39], [40]. Liu *et al.* [41] use a cascaded method in which the segmented result map is first overlaid onto the original image for land masking, and then the masked image is fed into the detection. This process suffers from two problems. On the one hand, the two tasks of sea–land segmentation and target detection are not end-to-end and require separate training and inference, which is time-consuming. On the other hand, the two tasks are intrinsically correlated, and the global semantic information of the sea–land segmentation cannot be fully utilized in the cascaded process to optimize the vessel detection model.

In this article, *first*, a single low-level path aggregation network (S-LPAN) structure is designed for the characteristics of small vessel targets in SAR images. Considering the size of small vessel targets and the resolution of SAR images, a large number of downsampling times is likely to cause the loss of texture and edges of small vessel targets. The S-LPAN utilizes the feature map with fewer downsampling times for detection, and an SEM is designed to enhance the global semantic information of this part of the feature map. *Second*, considering that amplitude images contain less information, dual-polarimetric decomposition is performed on the dual-polarimetric image and a mixture of components is extracted to enhance small vessel targets with polarimetric information. However, feeding all features into the network is bound to have feature redundancy. Therefore, ADPFF is designed to adaptively fuse all input features, and the convolution kernel parameters of ADPFF are updated according to backpropagation to obtain the most suitable fusion parameters for small vessel detection. Third, the cascaded approach to remove land false alarms is time-consuming and does not take full advantage of the global semantic information of sea–land segmentation. Therefore, a segmentation layer is added by sharing the feature extraction and feature fusion parts and implements sea–land segmentation and target detection in one network. A joint loss function is also designed to optimize the vessel detection model by using the high-level semantic information extracted from the sea–land segmentation and to optimize the vessel detection model while reducing the land false alarms.

In all, the main contributions of this article are as follows.

1) Considering the differences between SAR images and conventional optical images, the S-LPAN detector is designed for the characteristics of small vessel targets in SAR images.

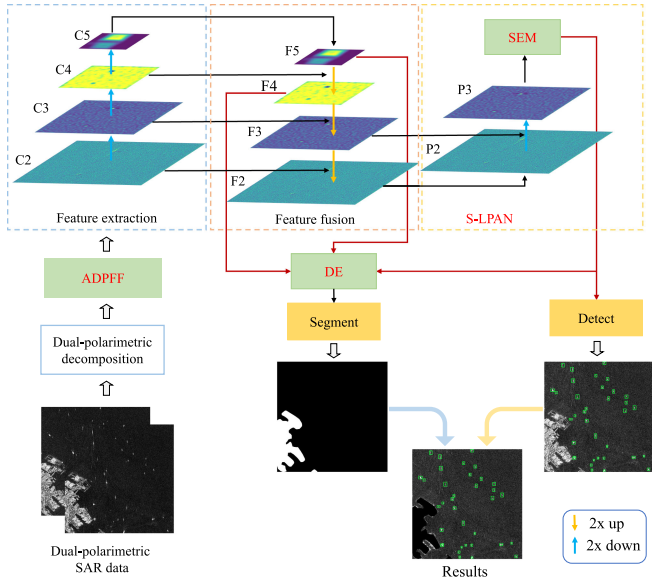2) The Sentinel-1 dual-polarimetric data are polarimetrically decomposed to enhance the features of small vessel

Fig. 2. Overall structure of the method proposed in this article.



Fig. 3. Conventional feature fusion structure.



Fig. 4. Improved feature fusion structure for small vessels in SAR images.

targets, and the polarimetric, amplitude, and mixture components are adaptively fused to extract features, discarding redundant information and reducing miss detections and false alarms of small vessel targets.

3) The sea–land segmentation layer is added to the overall network to share feature extraction and feature fusion. Compared to the cascaded approach, this method not only reduces time consumption but also utilizes the semantic information from the sea–land segmentation to optimize the small vessel detection.

The rest of this article is organized as follows. Section II presents the proposed method. Section III verifies the proposed method by comparing it with other methods. Finally, Section IV concludes this article.

## II. METHODOLOGY

The overall structure of the proposed method is shown in Fig. 2, which consists of an adaptive dual-polarimetric feature fusion (ADPFF) module, an S-LPAN with an SEM, and a multitasking head for sea–land segmentation and vessel detection with a dilate encoder (DE) module. In this section, the S-LPAN structure with SEM designed for the difference between small vessel targets in SAR images and optical images is first described. It then details how ADPFF is used to fuse multichannel polarimetric information for the enhancement of small vessel targets. Finally, it describes how a joint loss function can be designed to optimize the vessel detection model using sea–land segmentation during training.

### A. S-LPAN

The conventional feature extraction and feature fusion parts are shown in Fig. 3. The commonly used feature extraction networks include VGG [42], ResNet [43], Inception [44],
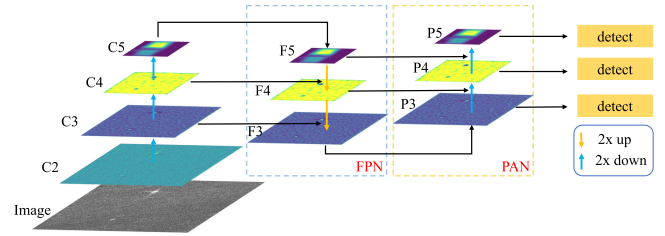
DenseNet [45], etc. Some experiments have proved that CSP-Darknet53 [33] performs better as a feature extraction module in the detection task; so it is used for feature extraction in this article. C2, C3, C4, and C5 are the feature maps extracted by the feature extraction network, and the downsampling times under the feature map scale are 4, 8, 16, and 32. The FPN part is the top-down feature fusion. F3, F4, and F5 are the feature maps after FPN, and the downsampling times are 8, 16, and 32. With the proposed PAN [46] structure, most networks add the bottom-up PAN structure to the FPN to obtain the P3, P4, and P5 feature maps. These three scales of feature maps constitute the three detection layers of multiscale detection, corresponding to the downsampling times of 8, 16, and 32. The relationship between these layers is as follows:

$$C_i = \text{down}\,(C_{i-1}) \qquad i = 3, 4, 5 \qquad (1)$$

$$F_{j-1} = \text{concat}\,(C_{j-1}, \text{up}\,(F_j)) \qquad j = 4, 5, F_5 = C_5 \quad (2)$$

$$P_k = \text{concat}\,(F_k, \text{down}\,(P_{k-1})) \qquad k = 4, 5, P_3 = F_3 \quad (3)$$

where down is 2 times downsampling, up is 2 times upsampling, and concat is feature stitching. However, due to the limitation of SAR image resolution, small vessel targets are much smaller in size than most optical images and are susceptible to interference from clutter and some near-shore strong scatterers; so this structure is not very suitable for small vessel detection in SAR images.

For small vessel detection, some scholars [9], [47] have improved the feature fusion part. As shown in Fig. 4, the resolution of the detection feature map is increased to reduce the miss detections of small vessel targets by adding F2 and P2 layers with a downsampling time of 4. However, there are two problems with increasing the number of high-resolution detection layers. On the one hand, the computational effort increases dramatically. For an image of $800 \times 800$ pixels, the size of the P2, P3, P4, and P5 feature maps are $200 \times 200$, $100 \times 100$, $50 \times 50$, and $25 \times 25$ pixels. With the addition of the P2 layer, the size of the overall detection layer increases to 4.048 times than the original
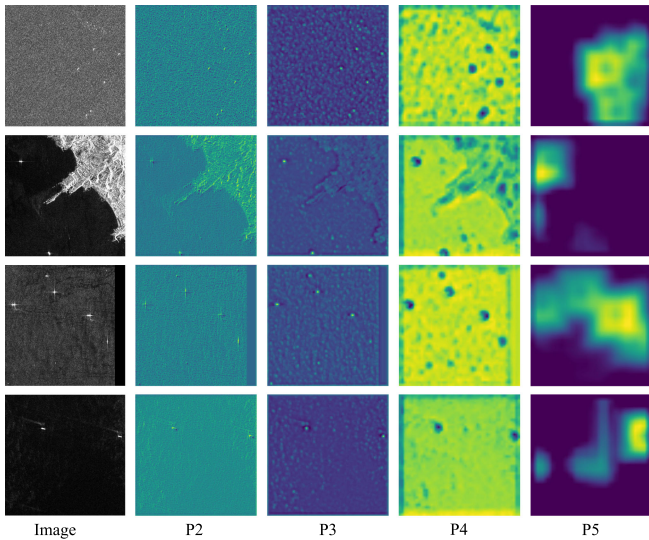
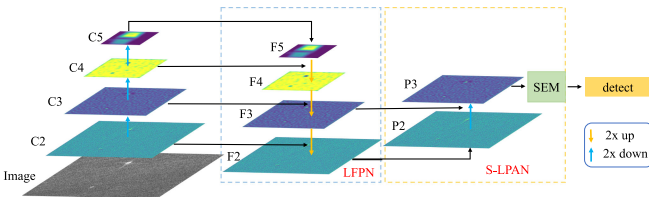Fig. 5. Feature map visualizations at different scales.



Fig. 6. Single low-level path aggregation network structure.

size. On the other hand, the utilization of high-resolution feature maps tends to introduce more redundant information, which can easily cause false alarms.

The visualization of the feature maps P2, P3, P4, and P5 is shown in Fig. 5. The comparison between the original image and the feature image of small vessel targets shows that the texture and edges of the vessels are clearer in P2 than in the others. P3 also shows the position and texture edges of the targets and without flaps and trails. P4 and P5 only show the approximate position of the vessel targets and do not provide information on the size and shape of the vessels.

As shown in Fig. 6, the single-scale detection structure in S-LPAN is designed to reduce redundant computations and better detect small vessels in SAR images. P3 feature map contains texture and position information, which has the effect of suppressing flank and wake trails in SAR images, and is, therefore, selected as the detection layer. Considering that the lower detection layer lacks some semantic information, an SEM is added after P3 to enhance the semantic information in the P3 layer.

The main structure of the SEM is shown in Fig. 7. The spatial pyramid pooling [48] increases the perceptual field of the feature map using maximum pooling. In [33], it is experimentally demonstrated that the maximum pooling with kernels of 5, 9, and 13 can effectively improve the detection accuracy. Therefore, the maximum pooling with kernels 5, 9, and 13 is used in the SEM and the pooled feature maps are spliced with
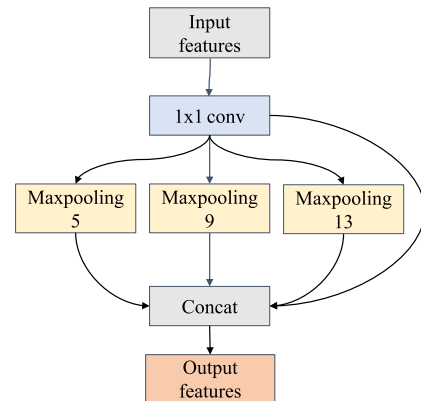


Fig. 7. Structure of semantic enhancement module.



Fig. 8. Polarimetric decomposition.

the original feature maps. The texture and position information of the original feature map is retained while increasing the perceptual field to obtain more global semantic information.

Through subsequent experiments, it is demonstrated that the proposed S-LPAN structure can effectively improve the detection of small vessels in SAR images and reduce the computational effort by combining only a single detection layer with an SEM.

### B. ADPFF

The amplitude image only contains intensity information. For some land-based strong scatterers with similar intensity to small vessel targets and scenes with more complex sea conditions, it is difficult to distinguish them. In order to enhance the feature of small vessel targets, polarimetric decomposition is utilized to obtain multiple features of the data, making it possible to enhance the vessel targets while suppressing the clutter background and various types of noise.

The method of polarimetric decomposition is generally based on quad-polarimetric SAR data. The polarimetric scattering characteristics of the target are decomposed into several combinations of the fundamental scattering mechanisms. Detection is then performed based on the target characteristics obtained from the decomposition [49]–[51].

Compared to quad-polarimetric data, dual-polarimetric data are simpler to obtain and have a larger swath, allowing for monitoring of a larger sea area. The polarimetric information obtained by polarimetric decomposition of dual-polarimetric data can effectively increase the input information for vessel detection [52], [53]. Therefore, the Sentinel-1 dual-polarimetric data in SLC format is used in this article. The process of polarimetric decomposition of Sentinel-1 dual-polarimetric data is shown in Fig. 8.

The Sentinel-1 SLC format data are first multiviewed with a range direction factor of 3 and an azimuth direction factor of 1.

This step is to make the target scale in SAR image closer to the real scene scale distribution. Then the polarimetric matrix conversion is done on the multiview processed data to obtain the $C_2$ matrix, and then the $C_2$ matrix is decomposed. The H/A/$\alpha$ decomposition [54] is a method to decompose the polarization coherence matrix into polarization entropy H, inverse entropy A, and average scattering angle $\alpha$. The decomposition equation is defined as

$$C_2 = [U_2] \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} [U_2]^{*T} \tag{4}$$

where $\lambda_1$ and $\lambda_2$ are the eigenvalues and $U_2$ is the eigenvector. The corresponding polarization channels can be obtained by calculating different polarimetric components, where the polarimetric entropy component is calculated as

$$H = -\sum_{i=1}^{2} P_i \log_2 (P_i) \tag{5}$$

where $P_i$ is the relative scattering intensity of the scattering process $i$. The polarization entropy describes the randomness of the scattering process, which is a reflection of the relationship between the eigenvalues $\lambda_1$ and $\lambda_2$ as follows:

$$P_i = \frac{\lambda_i}{\lambda_1 + \lambda_2}. \tag{6}$$

The anisotropy component is defined as

$$A = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2}. \tag{7}$$

The mean scattering angle component is defined as

$$\alpha = \sum_{i=1}^{2} P_i \alpha_i. \tag{8}$$

The mixing components are obtained from the two decomposition components: H and A. The four mixing components are $H*A$, $(1-H)*A$, $H*(1-A)$, and $(1-H)*(1-A)$, respectively. The $(1-H)*(1-A)$ channel corresponds to the case where only one main scattering mechanism exists (low entropy and low anisotropy degree). The $H*(1-A)$ channel characterizes the random scattering process (high entropy and low anisotropy). The $H*A$ channel corresponds to the case where there are two scattering mechanisms with similar probabilities (high entropy and high anisotropy). The $(1-H)*A$ channel corresponds to the case where there are two scattering mechanisms, one of which is the dominant scattering mechanism (low or medium entropy) and the other with medium probability (high anisotropy).

Examples of the amplitude image, H/A/$\alpha$ polarimetric decomposition components, and mixed components for a total of ten channels are shown in Fig. 9.

There is feature redundancy in the amplitude image, polarimetric decomposition image, and mixed component image with a total of ten channels fed directly into the feature extraction network. In order to extract effective features that are useful for detecting small vessel targets and reduce feature redundancy, ADPFF is proposed, which consists of three components: adaptive fusion of amplitude, adaptive fusion of polarization, and adaptive fusion of mixed components.
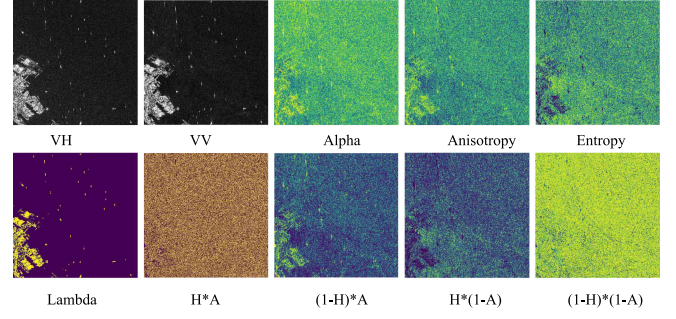


Fig. 9. Sentinel-1 SLC data amplitude, polarimetric decomposition, and mixed component channels.
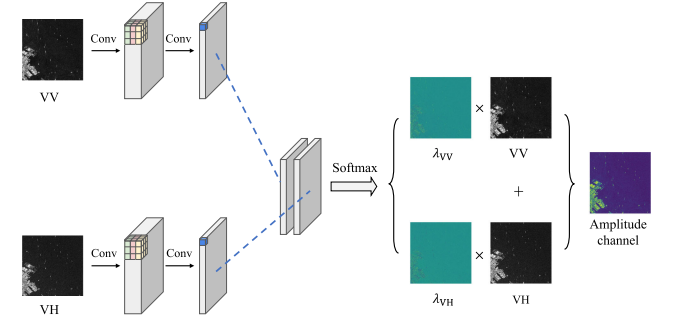


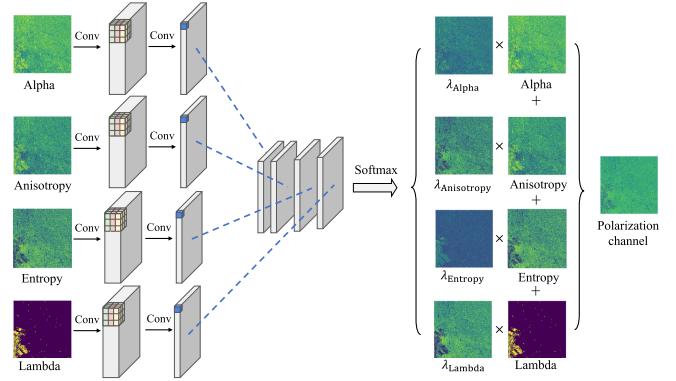Fig. 10. Adaptive fusion of amplitude channels.



Fig. 11. Adaptive fusion of polarization channels.

The amplitude adaptive fusion is shown in Fig. 10. The VH and VV amplitude channels are convoluted separately. The convoluted feature maps are stitched together and passed through Softmax to obtain the fusion coefficient matrix. The fusion coefficient matrices $\lambda_{VV}$ and $\lambda_{VH}$ are then multiplied with the corresponding points of the original channel separately and finally summed to obtain the fused channel. The polarimetric adaptive fusion is performed on the polarization channels Alpha, Anisotropy, Entropy, and Lambda as shown in Fig. 11. Mixed component adaptive fusion is performed on the four channels $H*A$, $(1-H)*A$, $H*(1-A)$, and $(1-H)*(1-A)$ as shown in Fig. 12. The fusion is similar to amplitude adaptive fusion, except that the number of channels fused is increased.

Their fusion equations are defined as

$$F_{Amp} = \lambda_{VV} \times F_{VV} + \lambda_{VH} \times F_{VH} \tag{9}$$
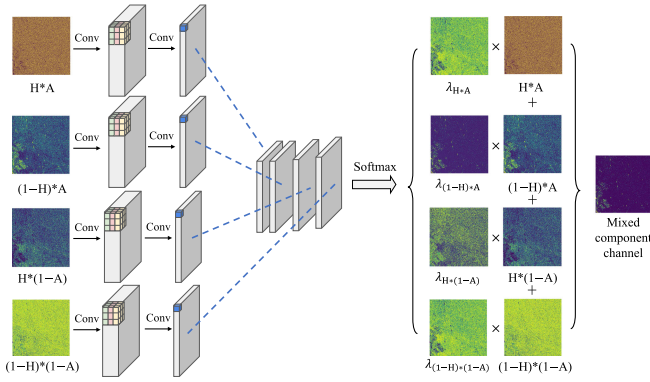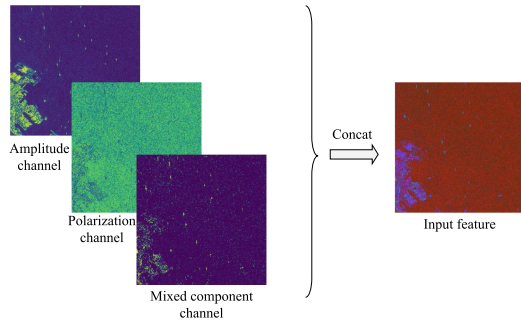
Fig. 12. Adaptive fusion of mixed component channels.



Fig. 13. Amplitude channel, polarization channel, and mixed component channel are stitched together as the input feature.

$$F_{Pol} = \lambda_{Alpha} \times F_{Alpha} + \lambda_{Anisotropy} \times F_{Anisotroey}$$
$$+ \lambda_{Enropy} \times F_{Entropy} + \lambda_{Lambda} \times F_{Lambda} \quad (10)$$

$$F_{Mix} = \lambda_{H*A} \times F_{H*A} + \lambda_{(1-H)*A} \times F_{(1-H)*A} + \lambda_{H*(1-A)}$$
$$\times F_{H*(1-A)} + \lambda_{(1-H)*(1-A)} \times F_{(1-H)*(1-A)} \quad (11)$$

where $F_x$ represents the $x$ channel to be fused, $F_{Amp}$, $F_{Pol}$, and $F_{Mix}$ represent the fused amplitude channel, polarization channel, and mixed component channel, respectively. Assuming $(i, j)$ is the coordinate of the corresponding point on the different channel features, for amplitude adaptive fusion

$$\lambda_{VV}[i,j] + \lambda_{VH}[i,j] = 1. \quad (12)$$

For polarimetric adaptive fusion and mixed component adaptive fusion

$$\lambda_{Alpha}[i,j] + \lambda_{Anisotropy}[i,j] + \lambda_{Entropy}[i,j]$$
$$+ \lambda_{Lambda}[i,j] = 1 \quad (13)$$

$$\lambda_{H*A}[i,j] + \lambda_{(1-H)*A}[i,j] + \lambda_{H*(1-A)}[i,j]$$
$$+ \lambda_{(1-H)*(1-A)}[i,j] = 1. \quad (14)$$

As shown in Fig. 13, the input features are obtained by stitching the amplitude channel, polarization channel, and mixed component channel obtained by ADPFF. A convolution operation exists in the adaptive fusion part. Since the parameters of the convolution can be updated from the backpropagation, the fusion coefficient matrix can be continuously updated during iterative training, resulting in a fusion coefficient matrix that
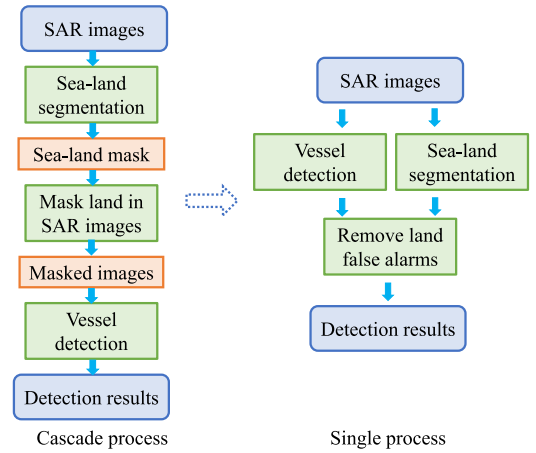


Fig. 14. Simplification of the sea–land segmentation process.

facilitates small vessel detection. The ADPFF can not only reduce feature redundancy but also make full use of the effective features of multiple polarization channels. It improves the degree of differentiation between small vessel targets and sea clutter and near-shore strong scatterers.

### C. Sea–Land Segmentation

In order to reduce the land false alarms, a sea–land segmentation module is usually added to mask the land before the vessel target detection. The conventional processing step of adding a sea–land segmentation module is shown in the cascaded process in Fig. 14. First, the input SAR image is segmented by land and sea, and then the land is blacked out on the original image according to the mask obtained from the sea–land segmentation. Finally, the processed original image is fed into the vessel detection network to obtain the final detection results.

However, there are two problems with the cascaded segmentation and detection process. On the one hand, the two tasks are not end-to-end and require separate training and inference, which is very time-consuming. On the other hand, the intrinsic correlation of the individual tasks is not fully exploited in the cascaded segmentation and detection process.

It is worth noting that by adding a segmentation layer to the vessel detection network, the two tasks of small vessel detection and sea–land segmentation can be achieved simultaneously. In this structure, the sea–land segmentation network and the vessel detection network share two parts: feature extraction and feature fusion. In order to optimize both tasks, a joint loss is proposed in this article. Through joint training, the global semantic information of sea–land segmentation is introduced into the optimization process of the small vessel target detection model, which makes it easier for the detection model to distinguish the land false alarms.

However, this joint training also raises some problems. The feature map should not be downsampled too many times to prevent the loss of small vessel target features [41], [55]. However, the sea–land segmentation task requires more downsampling to obtain global semantic information to optimize the semantic segmentation model, which contradicts the design of the small vessel detection network. It is necessary to design a separate
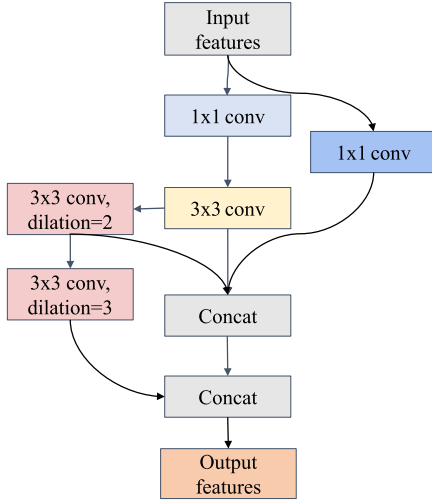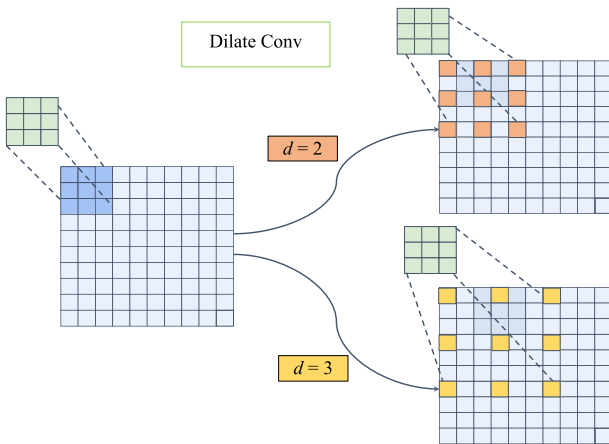
Fig. 15.    Architecture of dilate encoder.



Fig. 16.    Perceptive field for different dilation rates of the dilate convolution.

module that is not shared with vessel detection in order to enhance the global semantic information for sea–land segmentation. Therefore, a DE is added before the segmentation layer to enhance the semantic information of the segmentation layer feature maps, and the overall structure of the DE is shown in Fig. 15.

The DE module takes advantage of the larger perceptive field of the dilate convolution. The dilate convolution can increase the perceptive field without changing the scale of the feature map. The coverage of the dilate convolution can be adjusted by the dilate rate parameter $d$ of the normal convolution. Fig. 16 shows the coverage of the dilate convolution with dilate rate $d$ of 2 and 3, respectively, during the convolution process. Compared with the regular convolution, the addition of the dilate convolution can make the DE have a larger perceptual field, which is beneficial to enhance the global semantic information of the sea–land segmentation.

### D.  Joint Loss

In order to optimize both sea–land segmentation and small vessel detection models in a single network structure, a joint

loss function is designed in this article. Compared with the two tasks trained and optimized separately, the joint loss of sea–land segmentation and small vessel detection allows to introduce global semantic information from sea–land segmentation into the optimization process of small vessel detection.

Considering the small size of the vessel targets contained in the dataset, the GIoU [56] loss is used as a regression loss for vessel detection. It can effectively filter out those high-quality detection results that are closer to the labeled boxes. The GIoU loss is calculated as

$$\text{loss}_{\text{GIoU}} = 1 - \text{IoU} + \frac{C - (A \cup B)}{C} \tag{15}$$

where $C$ is the smallest convex set of $A$ and $B$. The IoU is defined as

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}. \tag{16}$$

For the classification part, only small vessel targets need to be distinguished from the background; hence, a binary cross-entropy function is used

$$\text{loss}_{\text{cls}} = -\frac{1}{n} \sum_{i=1}^{n} [t \log u + (1 - t) \log(1 - u)] \tag{17}$$

where $t$ is the value of the detection label: 0 for the background and 1 for the vessel. $u$ is the output of the detection layer network via the Sigmoid function, while $n$ is the number of samples detected.

The cross-entropy function is often used as a loss function for semantic segmentation. However, when performing sea–land segmentation, there are only two categories: sea and land; so the binary cross-entropy is also used for the classification loss of sea–land segmented pixels with the following equation:

$$\text{loss}_{\text{seg}} = -\frac{1}{n} \sum_{i=1}^{n} [y \log s + (1 - y) \log(1 - s)] \tag{18}$$

where $y$ is the true value of the sea–land segmentation: 0 for sea and 1 for land, $s$ is the output of the segmentation layer through the Sigmoid function, and $n$ is the number of segmented samples.

The joint loss of sea–land segmentation and target detection is shown below, with the values of $\text{loss}_{\text{total}}$ involved in the final parameter update

$$\text{loss}_{\text{total}} = \text{loss}_{\text{GIoU}} + \text{loss}_{\text{cls}} + \text{loss}_{\text{seg}}. \tag{19}$$

### III.  Experimental Results and Evaluation

In order to assess the effectiveness of the proposed method, a polarimetric detection and segmentation dataset (P-DSD) are produced. The required environment and hyperparameter settings for the experiments are then presented, and ablation experiments are carried out for each module. Finally, the experiments are compared with other well-known detection algorithms.

### A.  Dataset Description and Parameter Settings

The P-DSD is annotated by 15 large scene Sentinel-1 SLC images with a resolution of 10 m and contains small vessel detection
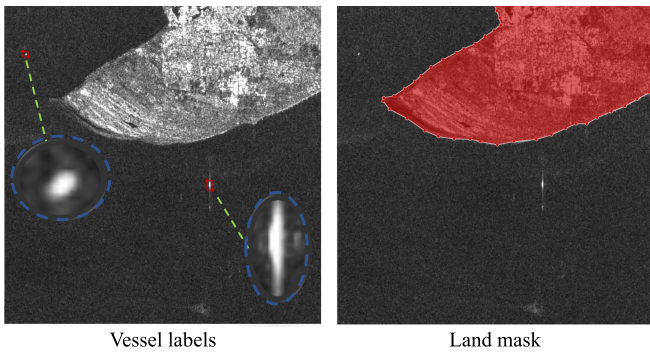
Fig. 17.    P-DSD labels.



Fig. 18.    Vessel size distribution.

labels with sea–land segmentation labels. The dataset detection labels are annotated with reference to LS-SSDDv1.0 [57]. First, the VV amplitude images are extracted from the SLC format data, and then the vessel detection frames are labeled according to the GRDH data labels. The coordinates of the left-top and right-bottom corners of the vessel are stored in the detection box labels. Finally, the vessel slices containing the ground truth of the labeled box are obtained by updating the coordinates of the left-top and right-bottom corners of the vessel according to the location of the slice while cropping the large image.

The differences between the P-DSD dataset and LS-SSDDv1.0 dataset are as follows.

1) The P-DSD dataset is annotated with SLC format data, which allows polarization matrix conversion and polarimetric decomposition. The LS-SSDDv1.0 dataset is annotated with GRDH format data, which only contains amplitude information and cannot extract polarization information.

2) The P-DSD dataset contains both vessel detection labels and sea–land segmentation labels. The LS-SSDDv1.0 dataset contains only vessel detection labels.

3) The P-DSD dataset eliminates the background during the cropping process, and each image contains the vessel target. Nine thousand images in LS-SSDDv1.0 dataset contain a large number of background images.

The definition of small vessels in this article refers to the definition in LS-SSDDv1.0 dataset, where $M$ is defined as the number of pixels of an image, and then the vessel with a label box area less than $M \times 0.37\%$ is defined as small vessel target. The number of pixels of the images in P-DSD dataset is $800 \times 800$; so the vessel with a label box area less than $800 \times 800 \times 0.37\% = 2368$ is defined as small vessel targets.

The small vessel detection boxes and the sea–land segmentation mask labels are shown in Fig. 17. The dataset contains a total of 5026 annotated small vessel targets. The aspect ratio statistics and area statistics of the annotated boxes are shown in Figs. 18 and 19. It can be seen from the figures that most of the vessel targets in this dataset have annotated boxes with an area smaller than 2368 pixels, and, thus, the vessel targets in the P-DSD dataset can all be defined as small vessels. In order to meet the input scale of the network, a sliding window crop is required. The final cropping results in 902 training slices
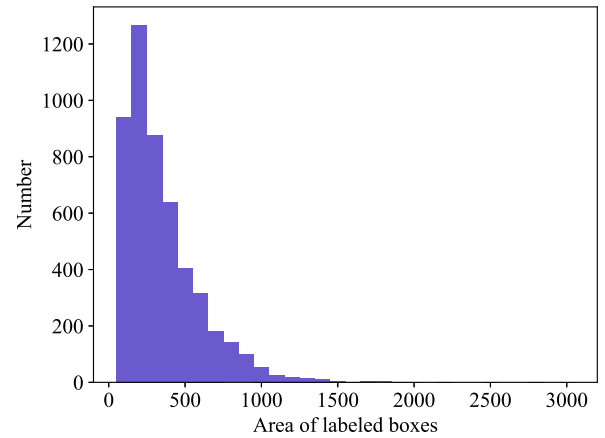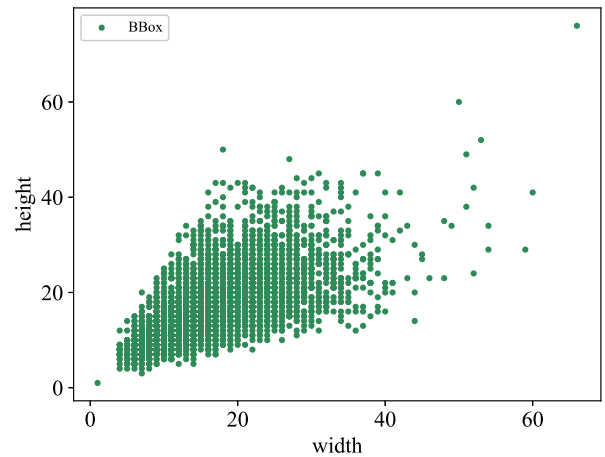


Fig. 19.    Length and width distribution of vessels.

and 595 testing slices. Considering the small number of valid training sets containing vessel targets, the test set is used as the validation set during the training process.

The proposed method is based on the Pytorch deep learning framework. The parameters are set empirically as in[1], where they are tested and proved to be effective. The optimizer is stochastic gradient descent with momentum. The initial learning rate is set to 1e-3. Momentum is set to 0.9 and the weight decay is set to 5e-4. The scatter plot in Fig. 19 shows that the aspect ratio of the small vessel target is balanced and there are no extreme aspect ratios. Therefore, the anchors are set to [1:1, 1:2, 2:1] for three aspect ratios, the anchor size is set to 15, the epoch is set to 100, and the learning rate is reduced to 0.0005 at the 50th epoch to facilitate the search for the optimal model. The training process is shown in Algorithm 1.

As the labels of sea–land segmentation and vessel detection are different, the losses need to be calculated separately during the training process. Since there is no difference between summing the losses when derived separately and when added

---

[1][Online]. Available: https://github.com/ultralytics/yolov5

**Algorithm 1:** Update Parameters During Training.

1 **Initialization:**
2 accumulate = 0;
3 batch-size = 4;
4 **while** accumulate $<= 64$ **do**
5    **if** accumulate$<64$ **then**
6       $loss_{seg}$.backward();
7       $loss_{det}$.backward();
8       accumulate = accumulate+batch-size;
9    **else**
10       optimizer.step();
11       optimizer.zero_grad();
12       accmulate = 0;
13    **end**
14 **end**

together and then derived. The segmentation loss is first back-propagated, and then the vessel detection loss is backpropagated. The gradients are accumulated to a preset value of 64, and then the parameters are updated. The advantage of gradient accumulation training is that it enables large batches to be trained even on machines with small memory. This helps to mitigate oscillations in losses during training and allows for faster acquisition of the best model.

## B. Evaluation Metric

Precision, recall, and average precision (AP) are used to assess the performance of vessel detection models, which are calculated as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{20}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{21}$$

$$\text{AP} = \int_0^1 P(R)dR \times 100\% \tag{22}$$

where TP (true positives), FP (false positives), and FN (false negatives) refer to the number of vessel targets correctly predicted, the number of vessel targets incorrectly predicted, and the number of vessel targets predicted to be nonvessel targets, respectively. AP describes the area under the precision–recall (PR) curve. It is a tradeoff between the two metrics and shows the performance of the different methods.

For the sea–land segmentation, pixel accuracy (PixAcc) and mean intersection over union (MIoU) are used and calculated as

$$\text{PixAcc} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{23}$$

$$\text{MIoU} = \left[ \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} + \frac{\text{TN}}{\text{TN} + \text{FN} + \text{FP}} \right] \Big/ 2 \tag{24}$$

where TP is the number of pixels of land segmented as land, FP is the number of pixels of sea segmented as land, TN is the number of pixels of sea segmented as sea, and FN is the number of pixels of land segmented as sea.

TABLE I
COMPARISON OF DETECTION RESULTS FOR DIFFERENT FEATURE MAPS

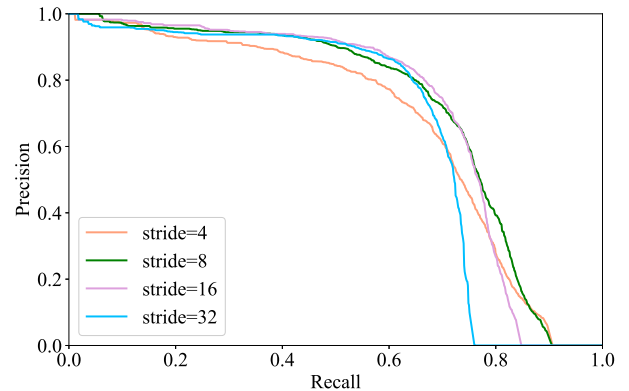| Method | Precision/% | Recall/% | AP/% |
|---|---|---|---|
| P2 (stride=4) | 71.58 | 64.64 | 67.36 |
| P3 (stride=8) | 78.74 | 66.12 | **71.90** |
| P4 (stride=16) | 80.6 | 66.47 | 71.32 |
| P5 (stride=32) | 84.05 | 62.66 | 66.19 |



Fig. 20. Comparison of precision–recall curves with different downsampling times.
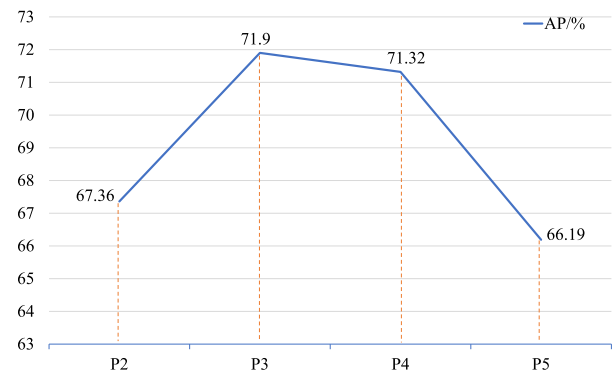


Fig. 21. Comparison of AP for detection at different feature scales.

## C. S-LPAN

As the feature map is continuously downsampled, the semantic information is continuously enhanced and the position and texture information is continuously diminished. Usually, the detection algorithms use multiscale feature maps as the detection layer and finally combine the detection results. The small vessel targets in SAR images are small in size and do not vary greatly in scale; so a suitable feature map scale can be selected for single-scale detection to reduce calculations. Table I shows the results obtained by using different scales of feature maps as a single detection layer, where stride is the downsampling time compared to the original image.

From the results in Table I and the PR curve in Fig. 20, it can be seen that using the P3 feature map with a downsampling time of 8 as the detection layer improves the AP more than other features maps. It is also clear from the curves in Fig. 21 that P3 is most suitable for small vessel detection in SAR images.

TABLE II
COMPARISON OF SEM MODULE EXPERIMENTS

| Method | Precision/% | Recall/% | AP/% |
|---|---|---|---|
| PAN | 82.39 | 63.81 | 71.36 |
| P3 | 78.74 | 66.12 | 71.90 |
| P3+SEM | 80.79 | 65.74 | **72.91** |

TABLE III
COMPARISON OF EXPERIMENTAL RESULTS ON SSDD DATASET

| Method | Precision/% | Recall/% | AP/% |
|---|---|---|---|
| PAN | 97.79 | 97.07 | 98.88 |
| S-LPAN (Ours) | 95.66 | 96.89 | **99.06** |
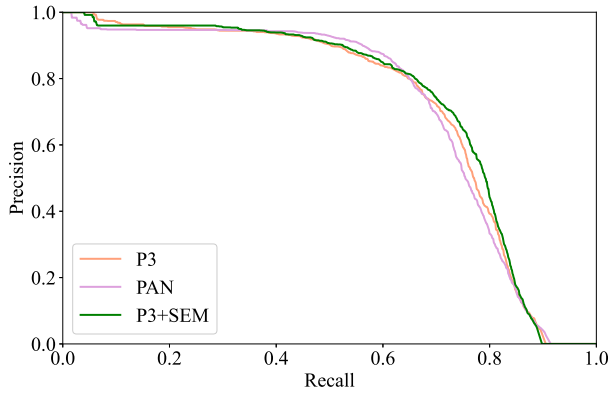


Fig. 22. Comparison of precision–recall curves of P3, PAN, and P3+SEM.



Fig. 23. Comparison of inshore detection results. (a) Detection results of conventional PAN method. (b) Detection results of S-LPAN method.



Fig. 24. Comparison of offshore detection results. (a) Detection results of conventional PAN method. (b) Detection results of S-LPAN method.

Compared to the results of the P2 feature map, P3 has higher precision accuracy, indicating that false alarms are effectively reduced.

The global semantic information of the high-resolution single-layer feature map is insufficient. To alleviate this problem, SEM is proposed for the semantic enhancement of P3. The comparison results of the SEM module are shown in Table II.

In Table II, PAN is the conventional multiscale feature map detection, P3 is the single-scale feature map detection, and P3+SEM is the single-scale feature map detection with the addition of the SEM module. The results show that the recall of P3 is higher, but the precision is lower. The addition of the SEM module increases the precision by 2.05%. SEM can effectively reduce false alarms by enhancing the semantic information in the single detection layer and is more effective than multiscale detection of PAN.

The PR curve in Fig. 22 shows that P3+SEM maintains the high recall of P3, while the gap with PAN in precision gradually decreases, which can balance the precision and recall well.

In order to verify that the S-LPAN is not only better for the detection on small vessels but also applicable to the detection of other sizes of vessels, the public dataset, SAR ship detection dataset (SSDD), [58] is employed for experimental validation. The SSDD dataset images are mainly from RadarSat-2, TerraSAR-X, and Sentinel-1 sensors. It contains HH, HV, VV, and VH polarizations and has a resolution of 1–15 m. There are 1160 images, including 928 images in the training set and 232 images in the test set. According to the definition in the SSDD dataset, Area is the number of pixels of a vessel detection box, and the vessel category is determined based on the value of Area:

1) small vessel targets: Area $< 1000$;
2) medium vessel targets: $1000 \leq$ Area $\leq 8000$;
3) large vessel targets: Area $> 8000$.
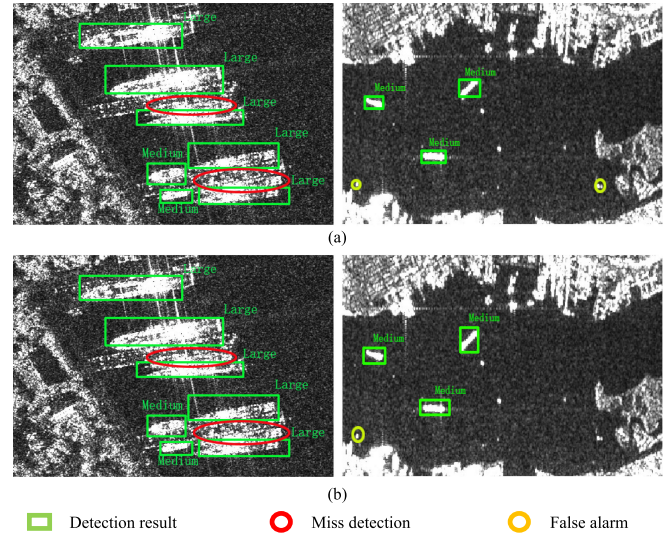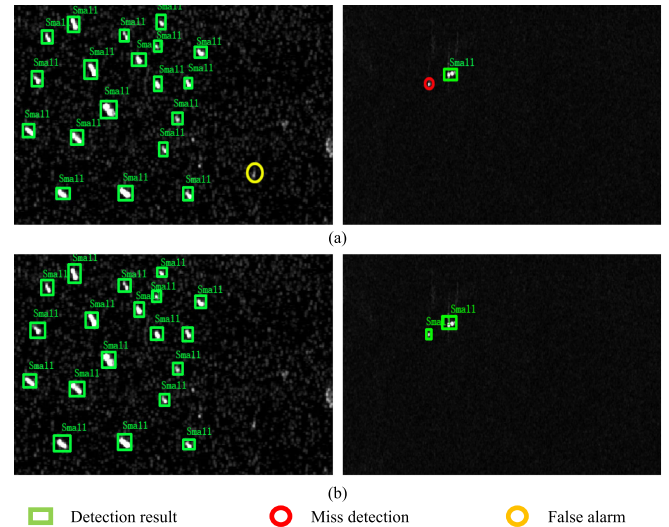
The experimental results are shown in Table III. It can be seen that the S-LPAN structure designed in this article outperforms the conventional PAN structure in terms of AP metrics in the SSDD vessel detection dataset containing multiple scales.

Some test result images are shown in Figs. 23 and 24, where "Small" represents small vessels, "Medium" represents medium vessels, and "Large" represents large vessels. The false alarms and miss detections for medium and large vessels are comparable to those of the traditional method. The effectiveness

TABLE IV
ABLATION EXPERIMENTS OF ADPFF

| Method | Precision/% | Recall/% | AP/% |
|---|---|---|---|
| Amplitude | 80.79 | 65.74 | 72.91 |
| 10channel | 82.04 | 72.40 | 75.99 |
| ADPFF (Ours) | 80.64 | 72.78 | **76.74** |



Fig. 25. Comparison of precision–recall curves of amplitude training, all polarization channel training and ADPFF training.



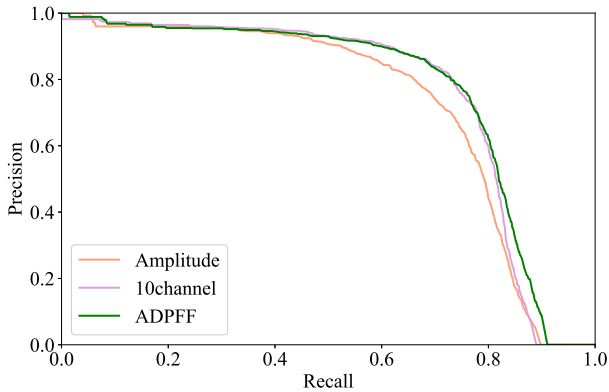Fig. 26. Comparison of sea–land segmentation results (where the red line is the result without DE and the green outline is the result with DE in contours).

of the S-LPAN structure for small vessel detection is further demonstrated.

### D. ADPFF

One contribution of this article is ADPFF. To analyze the contribution of the ADPFF module to small vessel detection, a series of comparative experiments are carried out as shown in Table IV.

In Table IV and Fig. 25, "Amplitude" means that VV channels are used for training and testing, and "10channel" means that all amplitude channels, polarization channels, and mixed component channels are overlaid for training and testing. The experimental results show that the addition of polarimetric and mixed component channels can effectively improve the precision and recall of small vessel detection, and the AP is increased by 3.08%. This further shows that the polarimetric information can enhance the characteristics of small vessel targets and reduce missed detections and false alarms. Moreover, the combined information of ten channels contains more redundant information. The input features of the three channels are obtained by ADPFF and the AP is improved by 3.83%, further demonstrating the effectiveness of the proposed ADPFF module.

### E. Sea–Land Segmentation

Implementing both vessel detection and sea–land segmentation tasks in one network can save the time of training and inference. In addition, the global semantic information of sea–land segmentation can be introduced to optimize vessel detection by combining sea–land segmentation losses with small vessel detection losses through backpropagation.

Table V shows the vessel detection results for the single loss training model and the joint loss training model. The comparison
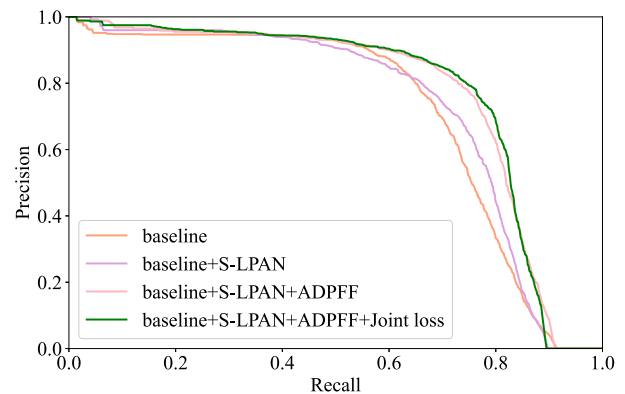


Fig. 27. Comparison of precision–recall curves for ablation experiments.

TABLE V
COMPARISON OF SINGLE AND JOINT TRAINING

| Method | Precision/% | Recall/% | AP/% |
|---|---|---|---|
| Single loss | 82.52 | 62.57 | 72.42 |
| Joint loss | **83.02** | **69.36** | **75.75** |

of the results shows that the joint loss optimization improves in both precision and recall, with recall improved by 6.79%, effectively alleviating the problem of missed detection. AP is improved by 3.33%, and the overall performance of the model is also effectively improved.

Considering that the semantic segmentation task requires deeper global semantic information compared to the small vessel detection, the DE module is designed to enhance the global semantic information of the sea–land segmentation part. Table VI shows the ablation experiments of the DE module. It can be seen
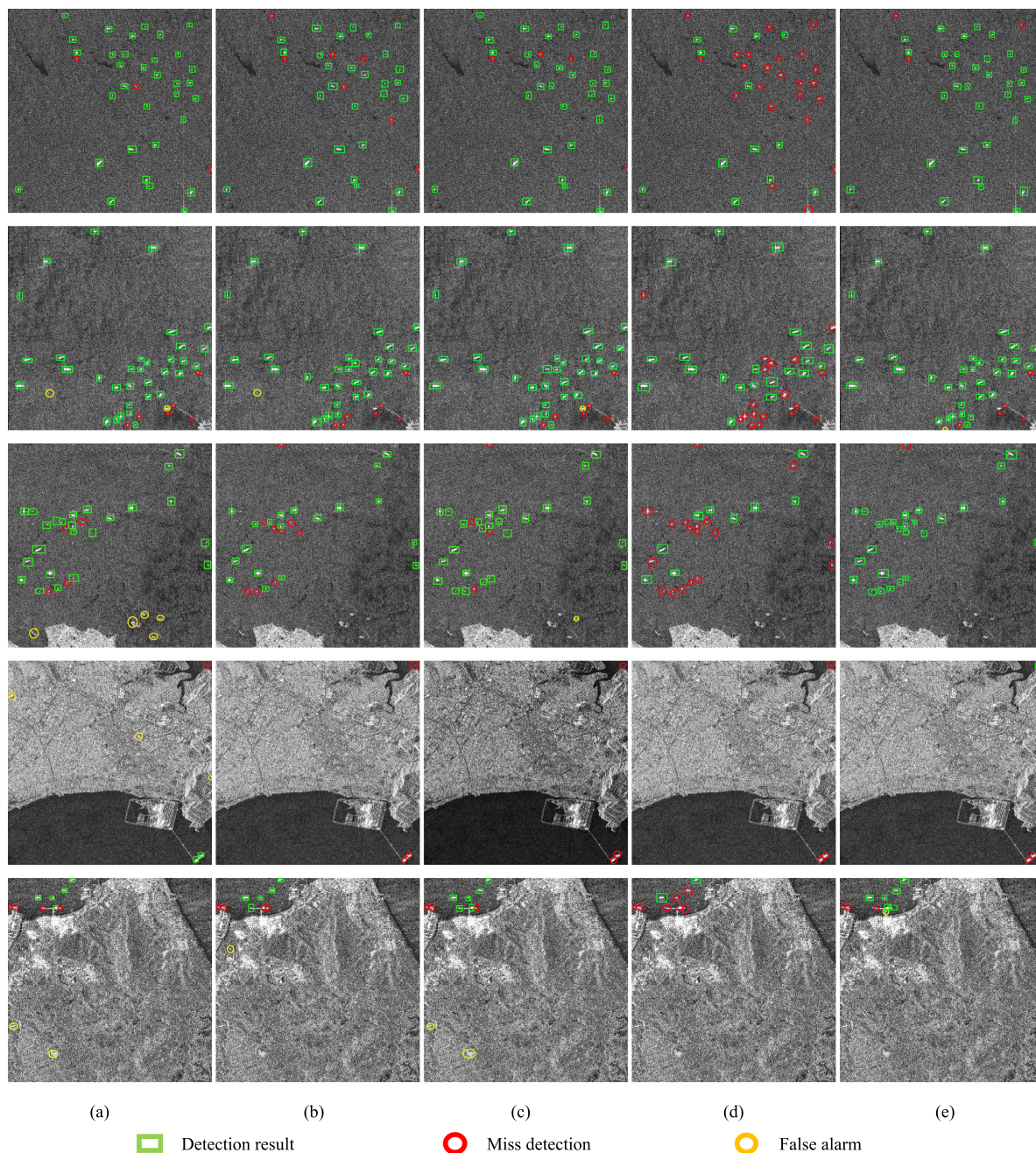
Fig. 28. Comparison with some detection results of relevant benchmark methods. (a) Faster R-CNN. (b) Cascade R-CNN. (c) Double-head R-CNN. (d)YOLOv3. (e) Our methods.

TABLE VI
ABLATION EXPERIMENTS OF DE

| Method | PixAcc/% | mIoU/% |
|---|---|---|
| Without DE | 99.02 | 91.30 |
| DE | **99.40** | **94.69** |

that the inclusion of DE can effectively improve the PixAcc and mIoU of the model.

In the real scene detection process, there is no vessel target in the central part of the land without rivers. The comparison of the partial segmentation results in Fig. 26 shows that the segmentation result in the central part of the scene is better with addition of DE than without DE, indicating that the global semantic information of the model is enhanced after the addition of DE, and the global information of the land is learned better. Thus, the central part of the land is all correctly segmented.

### F. Overall Results and Method Comparison

The results of the ablation experiments for each module proposed in this article are shown in Table VII.

|   (f)   |   (g)   |   (h)   |   (i)   |   (e)   |
|---------|---------|---------|---------|---------|

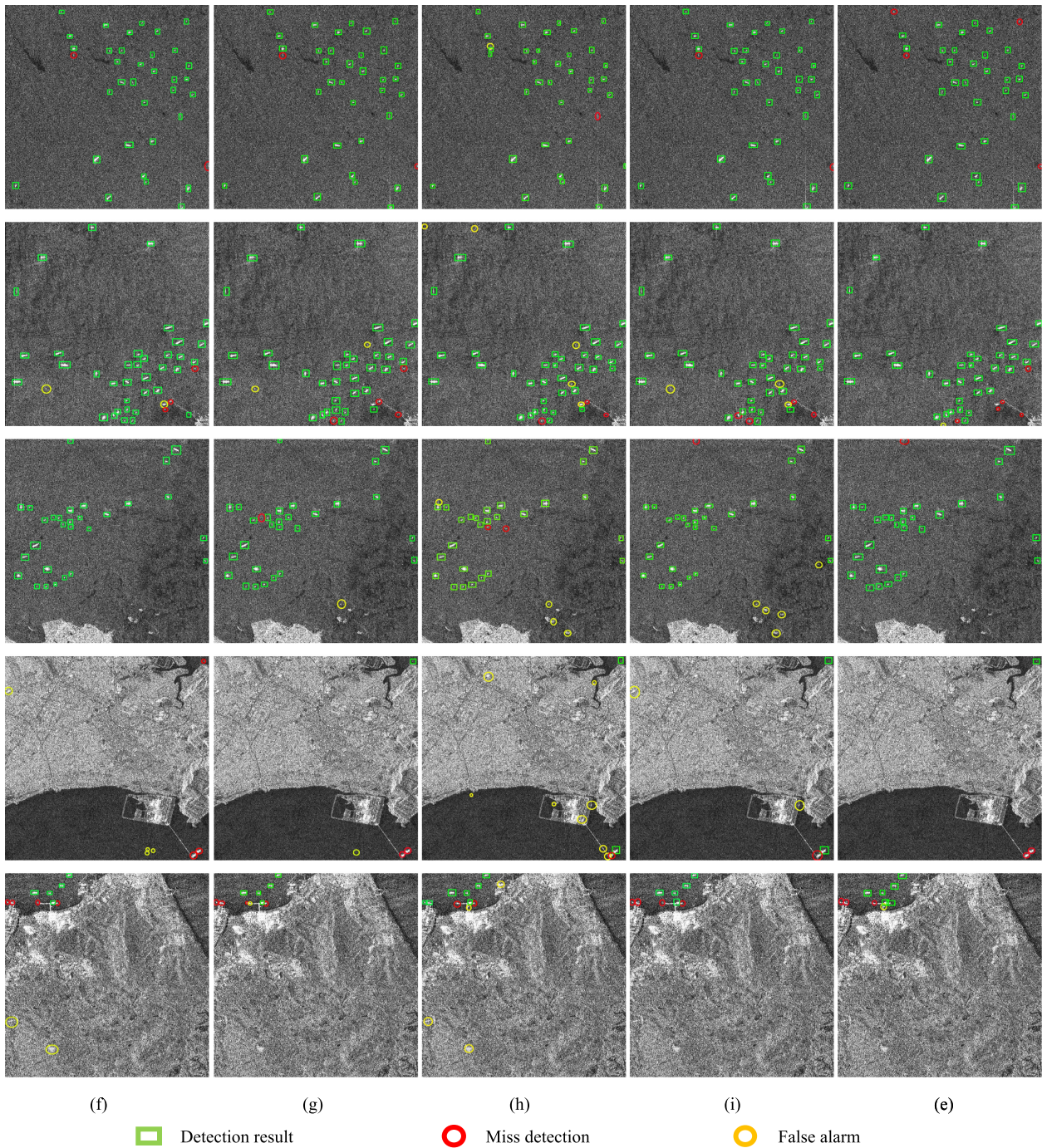□ Detection result     ○ Miss detection     ○ False alarm

Fig. 29. Comparison with some detection results of relevant benchmark methods. (f) YOLOv4. (g) YOLOX. (h) PP-YOLO. (i) TPH-YOLOv5. (e) Our methods.

Through the superposition of multiple modules, the final AP reaches 77.30%. A comparison of the PR curves for the ablation experiment is shown in Fig. 27. The green curve indicates the effect of all modules in this article, and it is clear from the curves that the area enclosed by the green curve is larger than the other curves. The effectiveness of the improved modules proposed is further verified.

To demonstrate the superiority of the proposed method, a series of experiments are conducted on the P-DSD dataset using several of the best-known algorithms, including faster R-CNN [26], cascade R-CNN [28], double-head R-CNN [29], YOLOv3 [59], YOLOX [34], YOLOv4 [33], PP-YOLO [60], TPH-YOLOv5 [61]. Among them, PP-YOLO and TPH-YOLOv5 are newer algorithms that perform well on small targets.

The results are shown in Table VIII. The proposed method outperforms the current well-known algorithms in terms of recall and AP on the P-DSD dataset. Some visual results are shown in Figs. 28 and 29. It is also clear from the figures that the proposed method has a significantly lower false alarm on land compared to

TABLE VII
ABLATION EXPERIMENTS OF THE PROPOSED METHOD

| S-LPAN | ADPFF | Joint loss | AP/% |
|--------|-------|-----------|------|
| – | – | – | 71.36 |
| ✓ | – | – | 72.91 |
| ✓ | ✓ | – | 76.74 |
| ✓ | ✓ | ✓ | **77.30** |

TABLE VIII
COMPARISON WITH BEST-KNOWN METHODS

| Method | Precision/% | Recall/% | AP/% |
|--------|-------------|----------|------|
| Faster R-CNN [26] | 48.07 | 64.57 | 57.21 |
| Cascade R-CNN [28] | 54.49 | 69.09 | 58.67 |
| Double-Head R-CNN [29] | 12.71 | 64.39 | 57.46 |
| YOLOv3 [59] | 75.70 | 54.48 | 56.32 |
| YOLOX [34] | 87.17 | 54.74 | 67.34 |
| YOLOv4 [33] | 76.25 | 60.27 | 64.85 |
| PP-YOLO [60] | 81.25 | 48.94 | 61.49 |
| TPH-YOLOv5 [61] | 80.28 | 64.91 | 72.07 |
| Ours | 82.33 | **73.00** | **77.30** |

other algorithms and the missed detection of small vessel targets is also mitigated.

## IV. CONCLUSION

In this article, a small vessel detection method based on adaptive bipolar feature fusion and sea–land segmentation in SAR images is proposed. An S-LPAN module for small vessel detection in SAR images is designed to address the differences between natural optical images and SAR images. The complexity of the network is reduced by visualizing the feature map and experimentally finding the most suitable feature map for single-scale prediction, and the SEM is used for semantic enhancement. In addition, to enhance the features of small vessel targets, the ADPFF module is designed to fuse the amplitude, polarization, and mixed component channels of the dual-polarization decomposition to obtain effective input features. Finally, a sea–land segmentation module is introduced. Through joint training, the global semantic information extracted from the sea–land segmentation is used to optimize the small vessel detection model and improve the discrimination of land false alarms. Experimental results on the detection and segmentation dataset P-DSD constructed in this article demonstrate the effectiveness of the proposed method in reducing false alarms and missed alarms.

## ACKNOWLEDGMENT

## REFERENCES

[1] H. Heiselberg and A. Stateczny, "Remote sensing in vessel detection and navigation," *Sensors*, vol. 20, no. 20, pp. 1–5, 2020.

[2] F. Ma, F. Zhang, D. Xiang, Q. Yin, and Y. Zhou, "Fast task-specific region merging for SAR image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, Jan. 2022.

[3] R. Pelich et al., "Advancements for Sentinel-1 based vessel monitoring: Dual-polarization detection and SAR-based coastline detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 5030–5032.

[4] 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.

[5] F. Zhang, X. Yao, H. Tang, Q. Yin, Y. Hu, and B. Lei, "Multiple mode SAR raw data simulation and parallel acceleration for Gaofen-3 mission," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 2115–2126, Jun. 2018.

[6] X. Sun, Z. Wang, Y. Sun, W. Diao, Y. Zhang, and K. Fu, "AIR-SARShip-1.0: High-resolution SAR ship detection dataset," *J. Radars*, vol. 8, no. 6, pp. 852–862, 2019.

[7] 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.

[8] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 765.

[9] Z. Cui, X. Wang, N. Liu, Z. Cao, and J. Yang, "Ship detection in large-scale SAR images via spatial shuffle-group enhance attention," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 379–391, Jan. 2021.

[10] 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, pp. 1–12, Nov. 2022.

[11] X. Wang, G. Li, A. Plaza, and Y. He, "Ship detection in SAR images via enhanced nonnegative sparse locality-representation of Fisher vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9424–9438, Nov. 2021.

[12] J. Fu, X. Sun, Z. Wang, and K. Fu, "An anchor-free method based on feature balancing and refinement network for multiscale ship detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1331–1344, Feb. 2021.

[13] F. Zhang, Y. Zhou, F. Zhang, Q. Yin, and F. Ma, "Small vessel detection based on adaptive dual-polarimetric SAR feature fusion and attention-enhanced feature pyramid network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp*, 2021, pp. 2218–2221.

[14] Y. Mao, X. Li, H. Su, Y. Zhou, and J. Li, "Ship detection for SAR imagery based on deep learning: A benchmark," in *Proc. IEEE Joint Int. Inf. Technol. Artif. Intell. Conf.*, vol. 9, 2020, pp. 1934–1940.

[15] J. Ai, Z. Cao, Y. Mao, Z. Wang, F. Wang, and J. Jin, "An improved bilateral CFAR ship detection algorithm for SAR image in complex environment," *J. Radars*, vol. 10, no. 4, pp. 499–515, 2021.

[16] J. Karvonen, A. Gegiuc, T. Niskanen, A. Montonen, J. Buus-Hinkler, and E. Rinne, "Iceberg detection in dual-polarized C-band SAR imagery by segmentation and nonparametric CFAR (SnP-CFAR)," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, Apr. 2022.

[17] J. Ai, Z. Cao, and M. Xing, "An adaptive-trimming-depth based CFAR detector of heterogeneous environment in SAR imagery," *Remote Sens. Lett.*, vol. 11, no. 8, pp. 730–738, 2020.

[18] A. Safa et al., "A low-complexity radar detector outperforming OS-CFAR for indoor drone obstacle avoidance," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9162–9175, Aug. 2021.

[19] W. An, M. Lin, and H. Yang, "Stationary marine target detection with time-series SAR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6406–6413, Jun. 2021.

[20] S.-W. Chen, X.-C. Cui, X.-S. Wang, and S.-P. Xiao, "Speckle-free SAR image ship detection," *IEEE Trans. Image Process.*, vol. 30, pp. 5969–5983, Jun. 2021.

[21] M.-D. Li, X.-C. Cui, and S.-W. Chen, "Adaptive superpixel-level CFAR detector for SAR inshore dense ship detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, Mar. 2021.

[22] T. Zhang, L. Jiang, D. Xiang, Y. Ban, L. Pei, and H. Xiong, "Ship detection from PolSAR imagery using the ambiguity removal polarimetric notch filter," *ISPRS J. Photogramm. Remote Sens.*, vol. 157, pp. 41–58, 2019.

[23] T. Zhang, Z. Yang, H. Gan, D. Xiang, S. Zhu, and J. Yang, "PolSAR ship detection using the joint polarimetric information," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 8225–8241, Nov. 2020.

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

[25] X.-C. Cui, C.-S. Tao, Y. Su, and S.-W. Chen, "PolSAR ship detection based on polarimetric correlation pattern," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 471–475, Mar. 2020.

[26] 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.

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

[28] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.

[29] Y. Wu *et al.*, "Rethinking classification and localization for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10186–10195.

[30] 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.

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

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

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

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

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

[36] T. Zhang, X. Zhang, J. Shi, and S. Wei, "Depthwise separable convolution neural network for high-speed SAR ship detection," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2483.

[37] F. Gao, Y. He, J. Wang, A. Hussain, and H. Zhou, "Anchor-free convolutional network with dense attention feature aggregation for ship detection in SAR images," *Remote Sens.*, vol. 12, no. 16, 2020, Art. no. 2619.

[38] W. Fan, F. Zhou, X. Bai, M. Tao, and T. Tian, "Ship detection using deep convolutional neural networks for PolSAR images," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2862.

[39] R. Li *et al.*, "DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018.

[40] P. Shamsolmoali, M. Zareapoor, R. Wang, H. Zhou, and J. Yang, "A novel deep structure U-net for sea-land segmentation in remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3219–3232, Sep. 2019.

[41] L. Liu, G. Chen, Z. Pan, B. Lei, and Q. An, "Inshore ship detection in SAR images based on deep neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 25–28.

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

[43] 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.

[44] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.

[45] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[46] 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.

[47] J. Shao, Q. Yang, C. Luo, R. Li, Y. Zhou, and F. Zhang, "Vessel detection from nighttime remote sensing imagery based on deep learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 12536–12544, Nov. 2021.

[48] 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.

[49] A. Kumar, R. K. Panigrahi, and A. Das, "Three-component decomposition technique for hybrid-pol SAR data," *IET Radar, Sonar Navig.*, vol. 10, no. 9, pp. 1569–1574, 2016.

[50] J. Engelbrecht, A. Theron, L. Vhengani, and J. Kemp, "A simple normalized difference approach to burnt area mapping using multi-polarisation C-band SAR," *Remote Sens.*, vol. 9, no. 8, 2017, Art. no. 764.

[51] K. Ji and Y. Wu, "Scattering mechanism extraction by a modified Cloude-Pottier decomposition for dual polarization SAR," *Remote Sens.*, vol. 7, no. 6, pp. 7447–7470, 2015.

[52] G. A. Russell, U. D. Castillo, A. S. Elgueta, and J. C. Sierralta, "Automated fish cages inventoryng and monitoring using H/A/$\alpha$ unsupervised Wishart classification in sentinel 1 dual polarization data," in *Proc. IEEE Latin Amer. GRSS ISPRS Remote Sens. Conf.*, 2020, pp. 471–476.

[53] L. Xie, H. Zhang, C. Wang, and Z. Shan, "Similarity analysis of entropy/alpha decomposition between HH/VV dual-and quad-polarization SAR data," *Remote Sens. Lett.*, vol. 6, no. 3, pp. 228–237, 2015.

[54] S. Cloude, "The dual polarization entropy/alpha decomposition: A PAL-SAR case study," *Sci. Appl. SAR Polarimetry Polarimetric Interferometry*, vol. 644, pp. 2–8, Mar. 2007.

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

[56] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 658–666.

[57] 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.

[58] 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.

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

[60] X. Long *et al.*, "PP-YOLO: An effective and efficient implementation of object detector," 2020, *arXiv:2007.12099*.

[61] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 2778–2788.

**Yongsheng Zhou** (Member, IEEE) received the B.E. degree in communication engineering from Beijing Information Science and Technology University, Beijing, China, in 2005, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2010.

He was with the Academy of Opto-Electronics, Chinese Academy of Sciences during 2010 and 2019, and is currently a Professor of Electronic and Information Engineering with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His general research interests include target detection and recognition from microwave remotely sensed image, digital signal, and image processing.

**Feixiang Zhang** (Student Member, IEEE) is currently working toward the master's degree, majoring in information and communication engineering, with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China.

His research interests include image processing and deep learning based small target detection.

**Fei Ma** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronic and information engineering from the Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China, in 2013, 2016, and 2020, respectively.

He is currently with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China, as an Associate Professor. His research interests include radar signal processing, image processing, machine learning, and target detection.

**Deliang Xiang** (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2010, the M.S. degree in photogrammetry and remote sensing from the National University of Defense Technology, Changsha, China, in 2012, and the Ph.D. degree in geoinformatics from the KTH Royal Institute of Technology, Stockholm, Sweden, in 2016.

Since 2020, he has been a Full Professor with the Interdisciplinary Research Center for Artificial Intelligence, Beijing University of Chemical Technology, Beijing, China. His research interests include urban remote sensing, synthetic aperture radar (SAR)/polarimetric SAR image processing, artificial intelligence, and pattern recognition.

Dr. Xiang currently serves as a reviewer for the *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *IEEE Geoscience and Remote Sensing Letters*, and several other international journals in the remote sensing field. In 2019, he was awarded a Humboldt Research Fellowship.

**Fan Zhang** (Senior Member, IEEE) received the B.E. degree in communication engineering from the Civil Aviation University of China, Tianjin, China, in 2002, the M.S. degree in signal and information processing from Beihang University, Beijing, China, in 2005, and the Ph.D. degree in signal and information processing from the Institute of Electronics, Chinese Academy of Sciences, Beijing, China, in 2008.

He is currently a Full Professor of Electronic and Information Engineering with the Beijing University of Chemical Technology, Beijing, China. His research interests include remote sensing image processing, high-performance computing, and artificial intelligence.

Dr. Zhang is currently an Associate Editor for *IEEE Access* and a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *IEEE Geoscience and Remote Sensing Letters*, and *Journal of Radars*.