Received March 8, 2018, accepted April 3, 2018, date of publication April 10, 2018, date of current version May 2, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2825376

A Densely Connected End-to-End Neural Network for Multiscale and Multiscene SAR Ship Detection

JIAO JIAO[1,2], (Student Member, IEEE), YUE ZHANG, (Member, IEEE)[1], HAO SUN[1], XUE YANG[1,2], (Student Member, IEEE), XUN GAO[1,2], (Student Member, IEEE), WEN HONG[1], (Senior Member, IEEE), KUN FU[1,2], AND XIAN SUN[1]

[1]Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing 100190, China

[2]School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

Corresponding author: Xian Sun (sunxian@mail.ie.ac.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 41501485, Grant 41701508, and Grant 61725105.

ABSTRACT Synthetic aperture radar (SAR) images have been widely used for ship monitoring. The traditional methods of SAR ship detection are difficult to detect small scale ships and avoid the interference of inshore complex background. Deep learning detection methods have shown great performance on various object detection tasks recently but using deep learning methods for SAR ship detection does not show an excellent performance it should have. One of the important reasons is that there is no effective model to handle the detection of multiscale ships in multiresolution SAR images. Another important reason is it is difficult to handle multiscene SAR ship detection including offshore and inshore, especially it cannot effectively distinguish between inshore complex background and ships. In this paper, we propose a densely connected multiscale neural network based on faster-RCNN framework to solve multiscale and multiscene SAR ship detection. Instead of using a single feature map to generate proposals, we densely connect one feature map to every other feature maps from top to down and generate proposals from each fused feature map. In addition, we propose a training strategy to reduce the weight of easy examples in the loss function, so that the training process more focus on the hard examples to reduce false alarm. Experiments on expanded public SAR ship detection dataset, verify the proposed method can achieve an excellent performance on multiscale SAR ship detection in multiscene.

INDEX TERMS Ship detection, multiscale, neural network, synthetic aperture radar (SAR).

I. INTRODUCTION

Ship detection has been playing an increasingly essential role in marine monitoring and maritime traffic supervision [1]–[4]. Due to its independence on the solar illumination and all-weather capability, Synthetic Aperture Radar (SAR) such as TerraSAR-X, RADARSAT-2, and Sentinel-1 has developed rapidly recently and it greatly promotes SAR ship detection [5]–[7]. Many algorithms for ship detection in SAR images among which constant false alarm rate (CFAR) and its variations are widely used [8], [9]. They can automatically adapt the threshold to the varying sea background while maintaining the expected performance. However, it is difficult for CFARs to exhibit good performance for small scale ships and inshore complex scenes. To deal with the inshore SAR ship detection, Zhao *et al.* [10] proposed a method through feature recognition and adaptive background window to detect inshore ships in SAR images. Liang *et al.* [11] presented an approach via saliency and context information to deal with inshore SAR ship detection. However, these methods of inshore SAR ship detection require post-processing to deal with many false alarms, they are not end-to-end [12], [13].

With the development of computer hardware and deep learning, convolutional neural network (CNN) has become the dominate approach for object detection, classification, and segmentation [14]–[17]. Features extracted by neural network have great performance than those by hand [18]–[20]. In recent years, many detection algorithms based on deep learning have developed rapidly [21]–[31]. RCNN uses CNN to extract a fixed length feature vector from each region proposal before a set of class specific linear SVMs [21]. Then the object detection methods based on region with CNN are intensively investigated. Fast-RCNN [22] processes the

whole image with CNN to produce a feature map, each object proposal is mapped to the feature map before (region of interesting) RoI pooling then generates two output vectors: softmax probabilities and bounding-box regression offsets with sharing computation. Faster-RCNN [23] introduces a Region Proposal Network (RPN) that shares convolutional features of full image with detection network to generate high quality region proposals, which are used by Fast-RCNN to refine the result of detection. Inspired by the developments of deep leaning, SAR ship detection based on deep learning is an inevitable trend in the future [32]–[35]. However, it does not show the great performance as expected when applying deep learning methods in SAR ship detection. One of the important reasons is that the current deep learning methods used in SAR ship detection have weak ability to deal with multiscale SAR ship detection. It is mainly because they generate region and classification score from a single high-level convolutional feature map. The high-level feature map has more semantic information but lower resolution which is insufficient to show the characteristic of multiscale objects. However, ships always have different scales in different resolution images as shown in Fig. 1. Moreover, SAR images of different resolutions have different effects on ship detection, low resolution SAR images are suitable for wide range detection, high resolution SAR images can achieve more accurate position. Therefore, solving the multiscale ship detection in multiresolution SAR images is of great significance. This paper proposes a network which generates RoIs from each fused feature maps with dense connection to solve multiscale SAR ship detection.

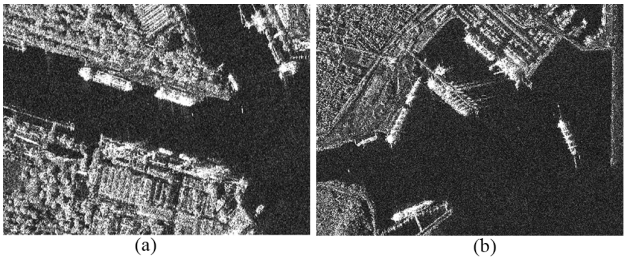


FIGURE 1. Showing two clips of SAR images.

The other important problem of using deep learning to detect ships in SAR images is the inference of inshore complex background which leads to a high false alarm. It is mainly due to the imbalance of easy and hard examples during the training of deep learning. The class imbalance of object detection always addressed via sampling heuristic [36], bootstrapping [37] and online hard example mining (OHEM) [38]. Sample heuristic usually guarantees a fixed foreground-to-background ratio. Bootstrapping in MSCNN [29] ranks the negative examples according to their classification scores and then collects top-N negative examples. In OHEM examples of training are sampled according to the current loss of each example under consideration. However, when aforementioned methods of solving class imbalance are used in SAR ship detection, they will have problems. Sample heuristic [36] is only used to balance positive and negative examples rather than easy and hard examples. Bootstrapping and OHEM are no guarantee that the negative examples of choice include the inshore complex objects which are highly like ships completely. To avoid the inference of inshore complex background, we propose a mechanism to reduce the weight of easy examples in the loss function and focus on the hard examples during training.

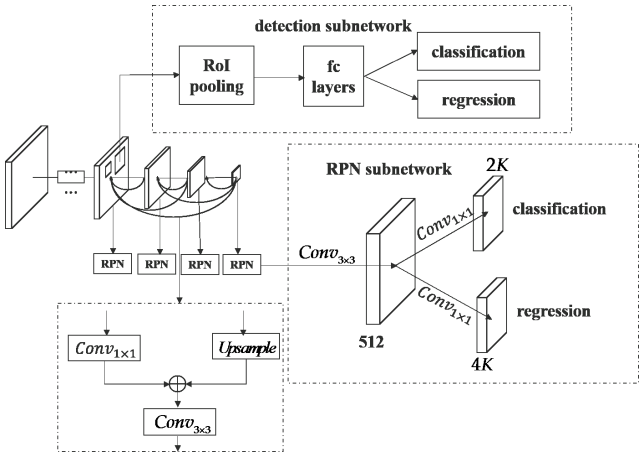


FIGURE 2. The architecture of DCMSNN which consists of RPN subnetwork and detection subnetwork.

This paper proposed a densely connected multiscale neural network (DCMSNN) based on Faster-RCNN as shown in Fig. 2 to achieve multiscale and multiscene SAR ship detection. Due to different layers of CNN feature maps have different spatial resolution and semantic information. Specifically, the lower feature maps of CNN have a higher spatial resolution but more coarse features which are suitable for small-scale object detection. The higher features maps have more semantic information but they are more abstract which are suitable for large-scale object detection. To address multiscale SAR ship detection both inshore and offshore, firstly, we densely connect feature map with other feature maps from top to down and generate proposals from each fused feature map. Secondly, we propose an improved loss function during training to avoid the inference of inshore complex background. Finally, we map RoIs to the lowest fused feature maps with most abundant information to get RoI pooling features which imported to the detection subnetwork to refine the detection results. Experiments on expanded public multiresolution SAR Ship Detection Dataset (SSDD) verify that the proposed method can achieve an excellent performance on multiscale and multiscene SAR ship detection.

The rest of this paper is organized as follows. Section 2 states our proposed network DCMSNN. Section 3 introduces the dataset used by our experiments. Section 4 describes the experimental results. Section 5 gives our discussion.

II. METHODS

Our proposed network consists of two subnetworks as shown in Fig. 2, one is the region proposal subnetwork (RPN) and the other is the detection subnetwork. Two subnetworks share the convolutional features of the images. The RPN is used to generate proposals which are used by detection subnetwork to achieve more refined detection results. In this paper, we use ResNet101 as backbone, feature maps have the same size in ResNet101 are called a stage. Due to the deepest layer of each stage has the strongest features, we use the feature activations output by the last residual block of each stage as our reference set of feature maps. We define the outputs of the last residual block in each stage conv2, conv3, conv4, conv5 as C_2, C_3, C_4, C_5. The rest of this section will introduce the details of the proposed method.

A. FUSING FEATURE MAP

For the outputs of the last residual block at each stage, from top to down there is fewer and fewer semantic information but there is more location information as they are subsampled fewer times. Although deeper feature maps have more semantic information, their resolution are lower and small-scale objects hardly have response on the deeper layers, so deeper feature maps are not adapted to small-scale object detection. However, lower feature maps have higher resolution but semantic information are rare. To make low-level high-resolution feature maps have more semantic information, we densely connect feature maps from top to down as shown in Fig. 2. Specifically, from top to down the feature map undergoes a 1×1 convolutional layer to reduce channel dimension and we use nearest neighbor up sampling to up sample the fused feature maps higher than it to its size. Then the up sampled feature maps are merged to the corresponding feature map. Finally, to reduce the aliasing effect of up sampling, we append a 3×3 convolution on each merged feature maps to generate the final fused feature maps. Specially, there is no higher feature map than C_5, we simply attach a 1×1 convolutional layer on C_5 to produce the coarsest resolution map. In order to be more intuitive, the above process is summarized as the following formula:

$$P_i = Conv_{3\times3}[\sum_{j=i+1}^{5} Upsample(P_j) + Conv_{1\times1}(C_i)] \quad (1)$$

$$P_5 = Conv_{1\times1}(C_5), \quad i = 4, 3, 2 \quad (2)$$

The formula is an iteration calculation until the lowest fused feature map is generated where P is the fused feature map corresponding to C.$Conv_{1\times1}(.)$ is a convolutional to reduce the dimensions of channels to 256, $Upsample(.)$ is nearest neighbor up sampling to up sample P to the same scale as C before merging. Finally, a convolutional $Conv_{3\times3}(.)$ on each fused feature maps to generate the final feature map P and reduce the dimensions of channels to 256. The fused feature maps will naturally provide more detailed information for the following bounding box prediction and classification, it is more conducive to multiscale and multiscene SAR ship detection.

B. RPN SUBNETWORK

We design a RPN subnetwork is realized by a 3×3 convolutional layer followed by two siblings 1×1 convolutional for classification and regression as shown in RPN subnetwork of Fig. 2. According to the characteristics of low-level feature maps have high resolution, high-level feature maps have more semantic information, feature maps of different layers adapt to object detection of different scales. low-level feature maps are adapted to small-scale object detection, whereas high-level feature maps are adapted to large-scale object detection, the RPN subnetwork is attached to each fused feature map to achieve multiscale SAR ship detection. The criterion of object or non-object and the bounding box regression of objects are defined with respect to a set of reference boxes called anchors. The anchors are of multiple predefined scales and aspect ratios to cover objects of different scales. We assign anchors of a single scale to each level, formally, we assign five scales $\{32^2, 64^2, 128^2, 256^2, 512^2\}$ anchors to $\{P_2, P_3, P_4, P_5, P_6\}$ respectively (P_6 is a stride two max-pooling of P_5). Anchors of each level have tree aspect ratios $\{1{:}1, 1{:}2, 2{:}1\}$, so in total there are $k = 3$ anchors at each sliding position for each P. The classification layer outputs $2K$ scores that estimate probability of object or not object for each proposal, the regression layer has $4K$ outputs encoding the coordinates of boxes. While generating the proposals, we assign label to each anchor based on their Intersection-over-Union (IoU) with ground-truth. Specifically, an anchor is labeled positive if it has the highest IoU with a ground-truth box or an IoU over 0.5 with any ground-truth box and labeled negative if it has IoU lower than 0.4 with all ground-truth box, the remaining anchors are ignored [23]. For an image we sample 512 anchors to train where the sampled positive and negative anchors have a ratio of 1:1, whereas the ignored anchors do not be sampled to train. In the experimental part, we discuss the impact of sampling heuristic.

C. TRAINING

This paper minimizes multi-task loss function as Faster-RCNN [23], for an anchor box i, its loss function is defined:

$$L(p_i, t_i) = L_{cls}(p_i, p_i^*) + \lambda p_i^* L_{reg}(t_i, t_i^*) \quad (3)$$

The classification loss L_{cls} is the softmax loss of two classes (object or not object). The L_{reg} is the loss of bounding box regression. In the Faster-RCNN L_{cls} is the cross-entropy loss for binary classification:

$$L_{cls} = -\log(p_t) \quad (4)$$

where:

$$p_t = \begin{cases} p & if\ y = 1 \\ 1 - p & otherwise \end{cases} \quad (5)$$

p is the model's estimated probability for the class with label $y = 1$. The cross-entropy loss is shown as the blue curves

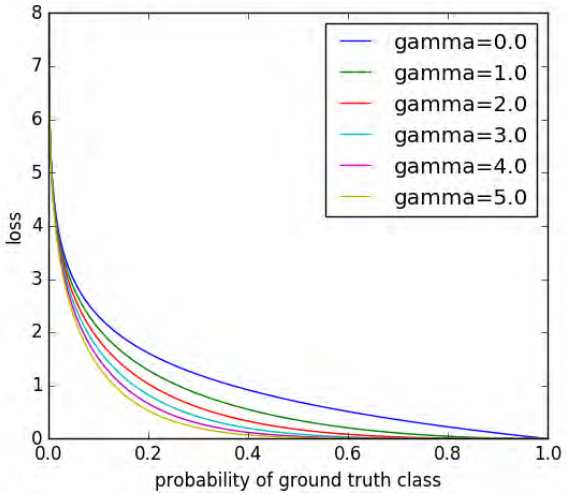


FIGURE 3. The modified loss function is the cross-entropy multiplied a modulating factor $(1-p_t)^\gamma$, when $\gamma = 0$ the improved loss is degenerated into cross-entropy. With the modified loss function the weight of easy examples is reduced a ratio during training.

in Fig. 3. As can be seen, even examples that are easily classified will have a relatively large loss. When summed over many easy examples, the training process will be dominated by the easy examples [30]. The model will not distinguish hard examples well. In the SAR ship detection, there are many inshore complex background that are highly like the ships. To make the model have good performance on multiscene SAR ship detection, we need to solve the problem of the training is dominated by easy examples. In this paper, we multiply a modulating factor $(1-p_t)^\gamma$ to the cross-entropy. We define the modified loss as:

$$L_{cls} = -(1-p_t)^\gamma \log(p_t) \tag{6}$$

Fig. 3 shows the modified loss curves for different values of γ. For easy examples, p_t tends to 1, $(1-p_t)$ goes to 0, the loss is reduced. For hard examples, p_t tends to 0, $(1-p_t)$ goes to 1, the loss is unaffected. The γ is an adjustable parameter that adjusts the ratio of reduced weight for easy examples. When $\gamma = 0$ the improved loss degenerated into cross-entropy loss function.

D. DETECTION SUBNETWORK

The detection subnetwork as shown in detection subnetwork of Fig. 2 consists of two hidden 1024-d fully-connected layers followed by final classification and bounding box regression layers. The RoIs generated by RPN highly overlap with each other. We take top-N (in this paper N is 12000) RoIs according to the classification scores. After that, we apply non-maximum suppression (NMS) threshold of 0.5 on the top-N RoIs based on their classification scores, which leaves us about 1000 high quality RoIs for each image. Due to the lowest fused feature maps has the highest resolution and merge the most semantic information, we use RoI pooling to extract 7×7 feature from p_2. We attach the detection subnetwork to all feature maps extracted by RoI pooling to achieve the refined detection result. For an image we sample 256 RoIs where the sampled positive and negative RoIs have a ratio of 1:1 to train the detection subnetwork. To illustrate that RoIs mapped to p_2 is best, in the experiment section, we will map RoIs to other fused feature maps for comparison. We also use the modified loss to train the detection subnetwork.

III. DATESET

The public SSDD dataset have a similar procedure as PASCAL VOC are provided by [35]. It includes SAR images of resolution from 1m to 15m which are collected from RadarSat-2, TerraSAR-X and Sentinel-1. In addition, it includes the ships of inshore and offshore, ships of different sizes so on. The specific information of ships in SSDD is shown in Table 1. As some small ships only have very few pixels in low resolution, sometimes it is hard to decide whether it is a ship or not, the SSDD dataset only annotates ships which the number of pixels are more than three. In addition, [35] utilize feature fusion, transfer learning, hard negative mining, and other implementation details to improve the AP from 70.1% to 78.8% compared to Faster in SSDD.To make the neural network of detection more robust, we expand SSDD dataset with 20 annotated multiresolution SAR images. We cut the 20 SAR images into 512×512 sized subimages without overlap and the coordinates of the annotated bounding boxes were transformed into the location of the corresponding subimages.

The supplemental dataset is also in the PASCAL VOC format. We add the subimages have annotated ships into SSDD. In the expanded SSDD dataset, there are totally 2246 subimages which were divided into train, test sets with ratio (7:3).

TABLE 1. The details of SSDD.

Sensors	Polarization	Resolution	Position
RadarSat-2 TerraSAR-X Sentinel-1	HH , VV VH , HV	1m-15m	inshore offshore

IV. EXPERIMENTS

In this section experiments are carried out to evaluate the performance of proposed method. Firstly, four experiments are designed to explore the effect of dense connection, the importance of RPN attached to each fused feature maps, the influence of RoIs mapped to different fused feature maps and the significance of modified loss. Then, we illustrate the role of sample heuristic and detection subnetwork. Besides, the comparison with other methods indicates the outperformance of the proposed method.

A. SETTING

All experiments are implemented in the Tensorflow framework and executed on a NVIDIA K80 GPU. The architecture in Fig. 2 is trained end to end. As is common practice,

we use the pre-trained ResNet101 on the ImageNet dataset to initialize the model. The adjustable parameter of modified loss γ changes between 0 and 5. We adopt synchronized SGD to train model. A mini-batch involves 1 images, 512 anchors, and 256 RoIs per images on GPU. We use a weight decay of 0.0001 and a momentum of 0.9. The iterations of training are 50k. Initial learning rate is 0.001 every 20k decrease 10 times.

B. THE STANDARD OF EVALUATION

To evaluate the quality of the model we apply the evaluation criteria mentioned below. We defined the target detection accuracy as:

$$p = \frac{N_{tp}}{N_{total\ \text{target}}} \tag{7}$$

recall as:

$$\mathrm{r} = \frac{N_{tp}}{N_{\text{ground truth}}} \tag{8}$$

To evaluate the overall performance of detector, F1 score which is defined as:

$$F1 = \frac{2 \times p \times r}{p + r} \tag{9}$$

where N_{tp} is the number of correct detected objects, $N_{total\ \text{target}}$ denotes the number of detected ships, $N_{\text{ground truth}}$ is the number of ground-truth. We define the bounding box is correct when it has IoU greater 0.5 with a single ground-truth.

C. EXPERIMENTS ON SSDD

1) THE EFFECT OF DENSE CONNECTION

As mentioned above, different feature maps have different characteristic, the low-level feature maps have high resolution but less semantic information, whereas the high-level feature maps have low resolution but more semantic information. To make feature maps with high resolution have more semantic information, we densely connect feature maps from top to down. To identify the effect of dense connection, comparison experiments with dense connection and without dense connection in the proposed network are conducted in this section. In the network without dense connection, RPN generates RoIs from each feature maps C, mapping RoIs to the lowest feature map C_2. In the network with dense connection, RPN generates RoIs from each fused feature maps P, mapping RoIs to the lowest fused feature map P_2. The two models are the same except for the different of connection, the modulating factor γ of modified loss is set to 0 which degenerates into cross-entropy loss function. The confidence score of all models is set to 0.9. Table 2 displays the detection accuracy, detection recall and F1 scores of networks with dense connection and without dense connection.

It can be seen from the Table 2 that the model with dense connection and the model without dense connection are similar in accuracy, but the model with dense connection has higher recall and F1 score. So, the model with dense connection has the better performance. As shown in Fig. 4, (a) is the ground truth, (b) is the detection result of model without dense connection, (c) is the detection result of model with dense connection. It is clear the model with dense connection can detect SAR ships which the model without dense connection cannot detect whether large or small ship. It is important to use dense connection to fuse the feature maps with different characteristic.

TABLE 2. Detection performance of model with dense connection and without dense connection.

Method	Accuracy	Recall	F1
no-dense	92.7%	76.8%	84.1%
dense	**92.8%**	**83.4%**	**87.9%**

2) THE INFLUENCE OF RPN ATTACHED TO EACH FUSED FEATURE MAPS

To adapt to multiscale ships detection, in our proposed method, we attach RPN to each fused feature map P, which has different resolution and semantic information from different layer. Feature maps from different layers adapt to objects of different scales, to cover objects of different scales the anchors have multiple predefined scales and aspect ratios. The anchors with single scale has three aspect ratios are assigned to the corresponding level. To prove RPN attached to different fused feature maps is more suitable to multiscale SAR ship detection, in this section, in addition to attach PRN to each fused feature maps, we also attach RPN to only one fused feature map from $\{P_2, P_3, P_4, P_5\}$ respectively in different experiments. When attaching RPN to a single fused feature map, to ensure the same number of anchors, we use five scales anchors of $\{32^2, 64^2, 128^2, 256^2, 512^2\}$ with three ratios $\{1{:}1, 1{:}2, 2{:}1\}$ to a single feature maps. The same detection network is attached to the feature maps of RoI pooling the modulating factor γ of modified loss is set to 0 which degenerates into cross entropy loss function. The confidence score of all models is set to 0.9. The detection performance of abovementioned experiments is displayed in the Table 3. Attaching RPN to multi layers has the best performance in accuracy, recall and F1. Attaching RPN to $\{P_2, P_3, P_4\}$ have the similar performance. Attaching RPN to P_5 has the higher accuracy than attaching RPN to $\{P_2, P_3, P_4\}$, but has the lowest recall.

TABLE 3. Detection performance of model with RPN is attached to different fused feature maps.

Method	Accuracy	Recall	F1
P_2	88.7%	75.9%	81.8%
P_3	87.9%	75.1%	81.0%
P_4	87.0%	75.3%	80.8%
P_5	90.1%	70.3%	80.0%
multi	**92.8%**	**83.4%**	**87.9%**

Fig. 5 shows the test result of RPN is attached to different fused feature maps in SSDD. (a) is the ground truth. (b) to (e)

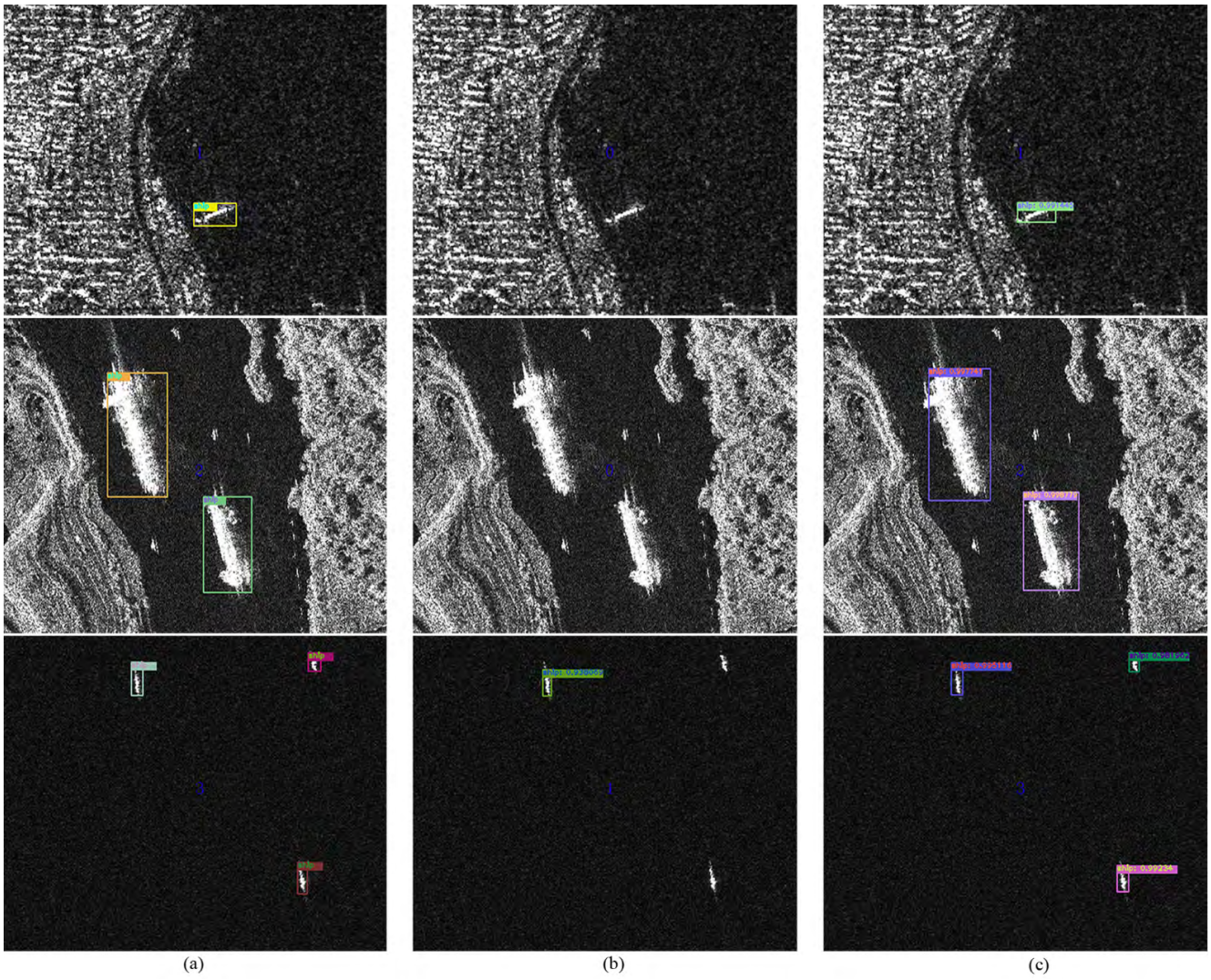


FIGURE 4. (a) is the ground truth, (b) is the detection result of model without dense connection, (c) is the detection result of model with dense connection. The model with dense connection can detect SAR ships which the model without densely connection cannot detect whether large or small ship.

is the RPN is attached to the P_2, P_3, P_4, P_5 respectively. (f) is the PRN is attached to multi fused feature maps. Attaching RPN to lower fused feature maps such as P_2 is suitable to detect small-scale ships, but the large-scale ships will be missed as shown in Fig. 6(b). Attaching RPN to higher fused feature maps such as P_5 is suitable to detect large-scale ships, but the small-scale ships will be missed as shown in Fig. 5(e). Attaching RPN to each fused feature maps can achieve the best test result as shown in Fig. 5(f).

3) THE INFLUENCE OF ROIS ARE MAPPED TO DIFFERENT FUSED FEATURE MAPS

As mentioned before, fused feature maps from different layers have different resolution and semantic information. Mapping RoIs to different fused feature maps to get the RoI features for the same detection network will have different performance. In this section, four models with mapping RoIs to P_2, P_3, P_4, P_5 respectively are trained for exploring the influence of mapping RoIs to different fused feature maps. All models use dense connection with the same detection network, RPN generates RoIs from each fused feature maps, the modulating factor γ of modified loss is set to 0 which degenerates into cross entropy. The confidence score of all models is set to 0.9. Table 4 displays the detection accuracy,

TABLE 4. Detection performance of model with ROIS are mapped to different fused feature maps.

Method	Accuracy	Recall	F1
P_2	**92.8%**	**83.4%**	**87.9%**
P_3	89.3%	73.9%	80.9%
P_4	89.9%	75.4%	82.0%
P_5	86.8%	69.1%	77.0%

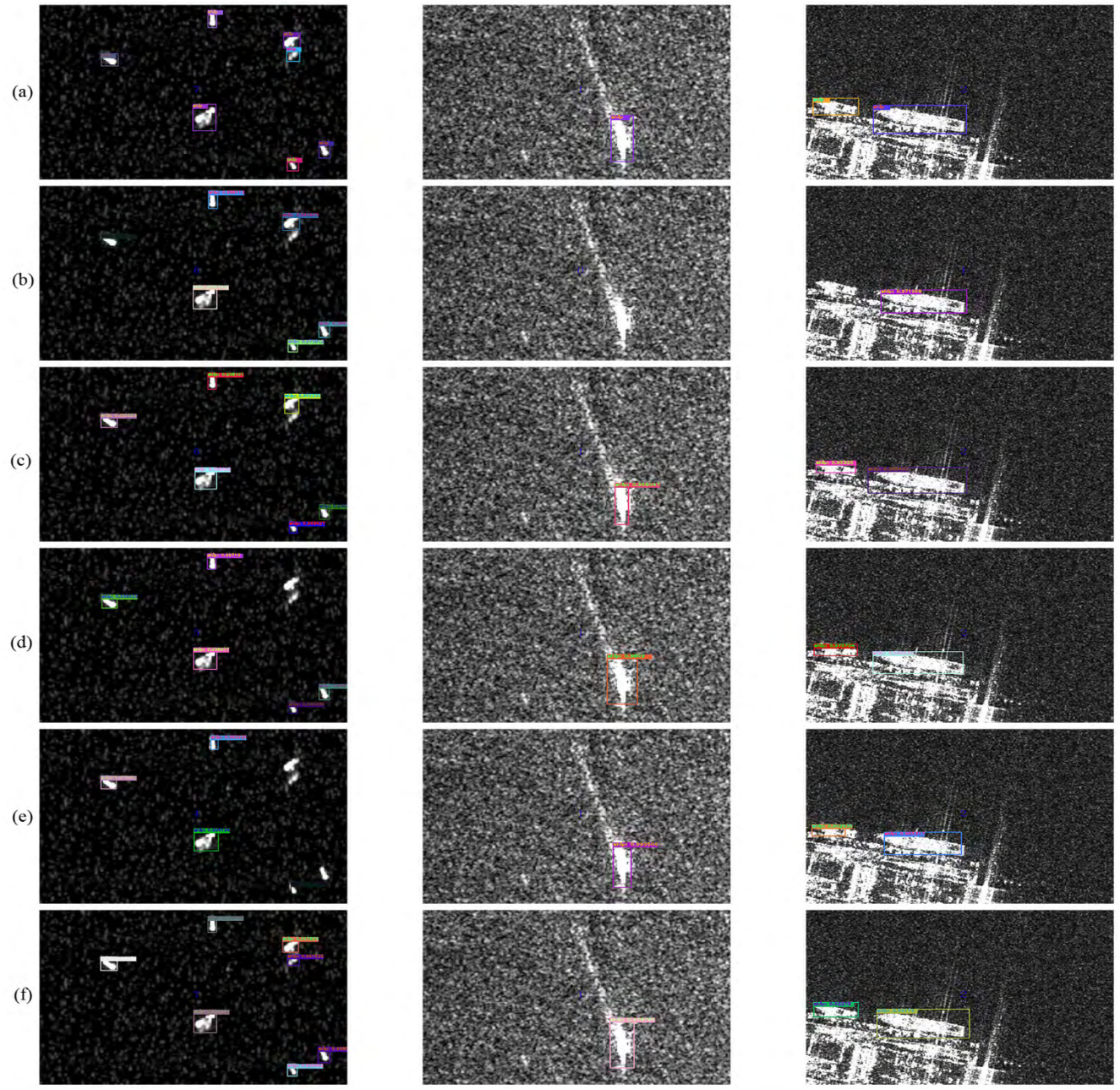


FIGURE 5. The test result of RPN is attached to different fused feature maps in SSDD. (a) is the ground truth. (b) to (e) is the RPN is attached to P_2, P_3, P_4, P_5 respectively. (f) is the PRN is attached to the multi fused feature maps.

detection recall and F1 scores of mapping RoIs to different fused feature maps. Mapping RoIs to P_2 has the highest value either in accuracy, recall and F1 score. In contrast, mapping RoIs to P_5 has the lowest value either in accuracy, recall or F1 score. Mapping RoIs to P_3 and mapping RoIs to P_4 are similar in accuracy, mapping RoIs to P_4 has a 2.5 higher than mapping RoIs to P_3 in recall.

As shown in Fig. 6, (a) is the ground truth, (b) to (e) is the RoIs are mapped to the P_2, P_3, P_4, P_5. Mapping RoIs to P_2 has the best performance in both large-scale and small-scale object detection. In summary, mapping RoIs to the feature maps with higher resolution fused more semantic information can improve the performance of network.

4) THE INFLUENCE OF THE ADJUSTABLE PARAMETER IN MODIFIED LOSS FUNCTION

To avoid easy examples dominating the training process, we multiply a modulation factor $(1 - p_t)^\gamma$ to cross-entropy loss function. To explore the influence of the modulation factor γ, in this section we change the from 0 to 5, when $\gamma = 0$, the modified loss function degenerated into cross-entropy. The models use dense connection, the RPN

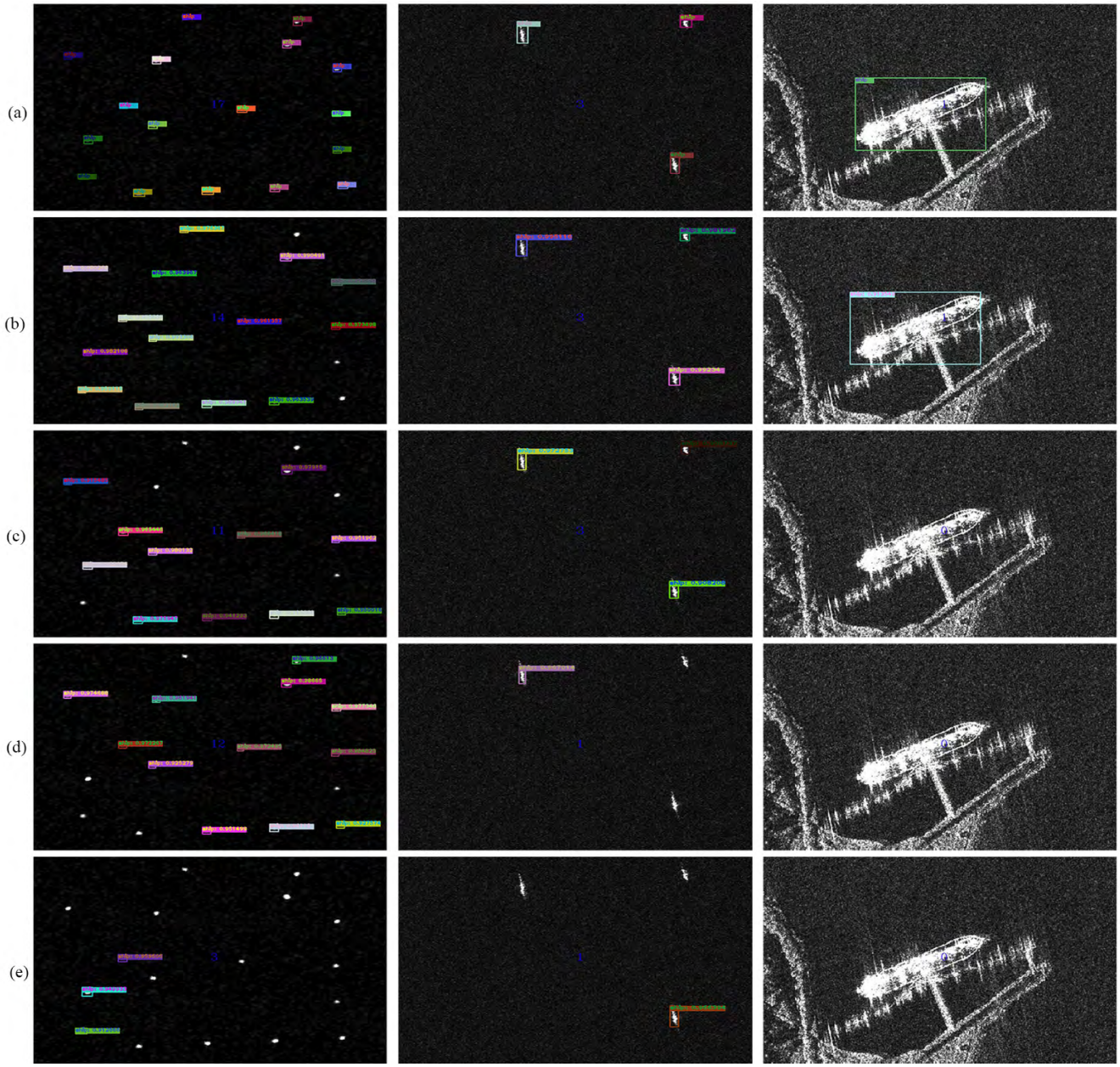


FIGURE 6. The test result of RoIs are mapping to different fused feature maps in SSDD (a) is the ground truth, (b) to (e) is the RoIs are mapped to the P_2, P_3, P_4, P_5.

subnetwork is attached to each fused feature maps, RoIs are mapped to the lowest fused feature maps, the same detection subnetwork is used, all models have the same experiment settings. Table 5 shows the performance of models with different γ. When $\gamma = 3$, the model has the same recall as $\gamma = 0$ but it has higher accuracy than $\gamma = 0$ and reduce the false alarm effectively.

5) THE INFLUENCE OF OTHER FACTORS

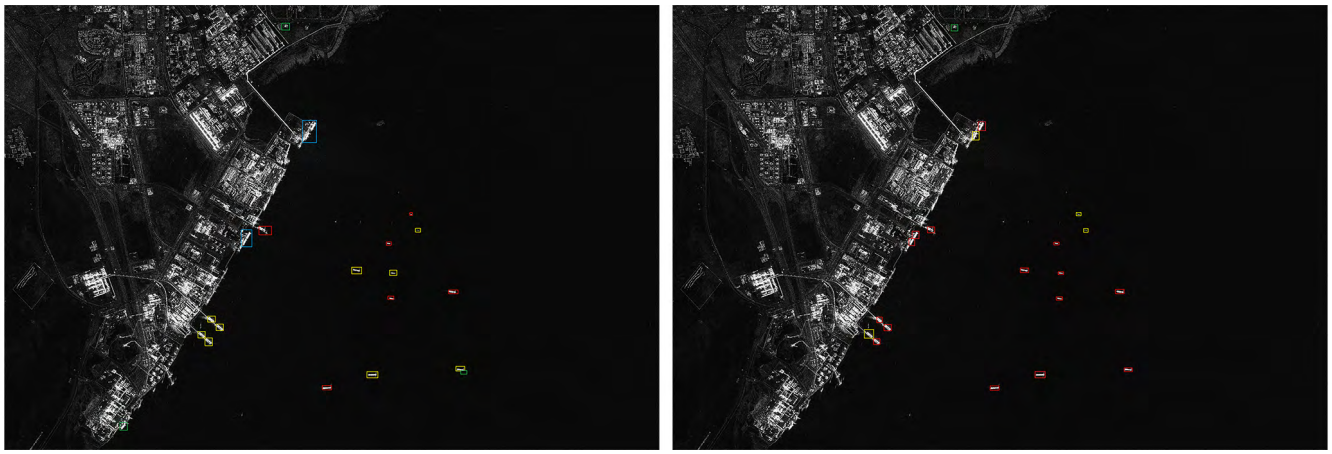
As mentioned earlier, when train the neural network, for an image we sample 512 anchors and 256 proposals where the sampled positive and negative anchors have a ratio of 1:1. It can ensure the balance of positive and negative examples during the training. If cancel the sample heuristic, the performance of model will be reduced a lot as shown in Table 6. In addition, as can be seen in Table 6, when cancel detection

TABLE 5. Detection performance of model with different modulation factor.

Method	Accuracy	Recall	F1
$\gamma = 0$	92.8%	**83.4%**	87.9%
$\gamma = 1$	90.5%	81.9%	85.9%
$\gamma = 2$	91.8%	83.0%	86.3%
$\gamma = 3$	**96.7%**	**83.4%**	**89.6%**
$\gamma = 4$	91.3%	77.9%	84.1%
$\gamma = 5$	94.1%	76.3%	84.3%

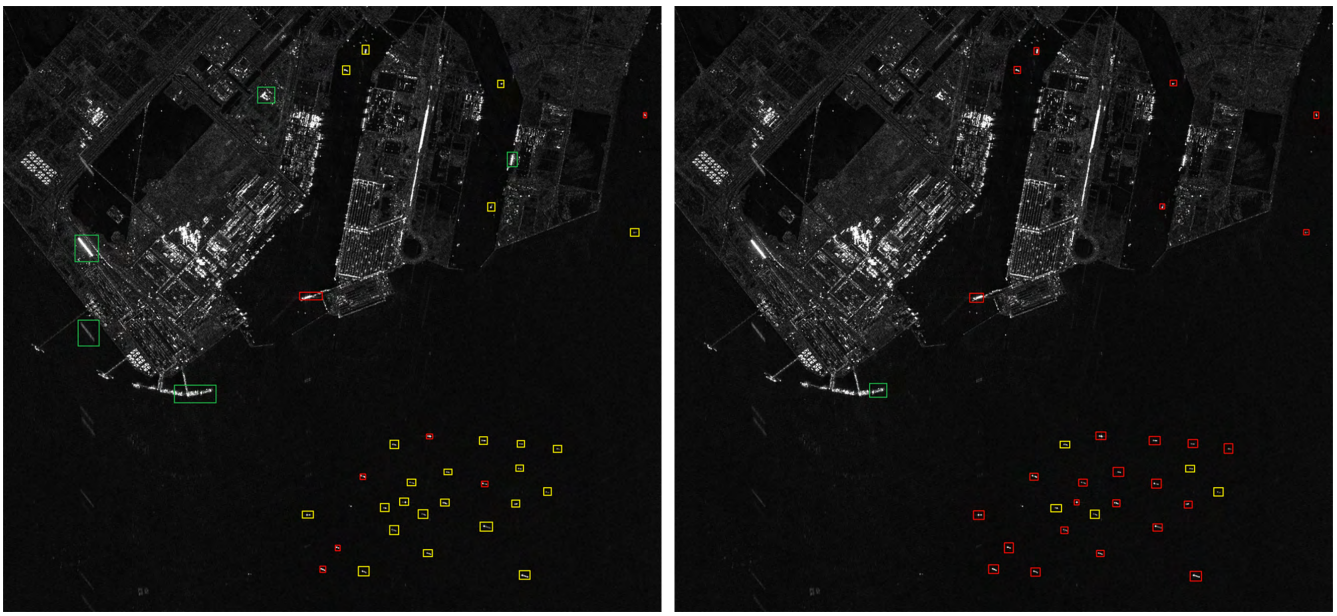
TABLE 6. Detection performance of different models.

Method	Modified Loss	Sample	Detection subnetwork	Accuracy	Recall	F1
Base model	√	√	√	**96.7%**	83.4%	**89.6%**
		√	√	92.8%	83.4%	87.9%
	√		√	57.1%	36.4%	44.4%
	√	√		63.7%	**85.3%**	72.9%



(a) (b)

FIGURE 7. The detection result of Gaofen3 SAR images, (a) uses Faster-RCNN, (b) uses DCMSNN. The red, yellow, and green rectangles represent the correct detection ships, the missing ships, and the false alarms respectively. The blue rectangles represent multiple vertical ships next to each other are detected as one ship.



(a) (b)

FIGURE 8. The detection result of Sentinel-1 SAR images, (a) uses Faster-RCNN, (b) uses DCMSNN. The red, yellow, and green rectangles represent the correct detection ships, the missing ships, and the false alarms respectively. The blue rectangles represent multiple vertical ships next to each other are detected as one ship.

subnetwork will get the highest recall but the accuracy is low, there will be a lot of false alarms in the test image. So, using the detection subnetwork to refine the proposals will improve the performance of model. While using the modified loss, the γ is 3, otherwise the γ is 0. The modified loss can decrease false alarm by distinguishing the objects are highly

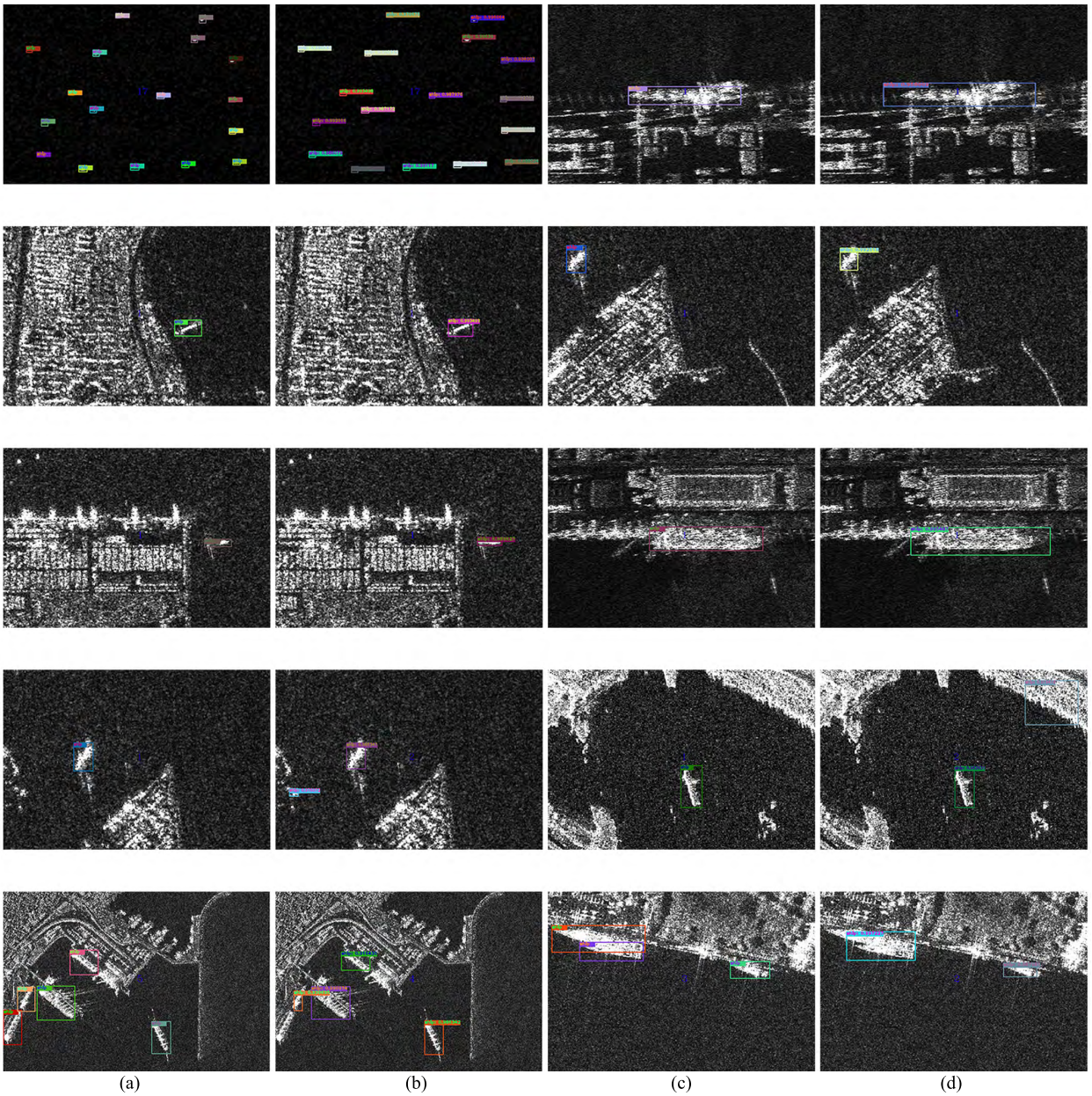


FIGURE 9. The detection result of test SAR images in SSDD with our proposed method, (a) and (c) are ground truth, (b) and (d) are the corresponding results.

like ships well. In this section, the base model is the same, using dense connection, attaching RPN to each fused feature maps, mapping RoIs to the lowest fused feature map, using the same detection subnetwork and experiment settings.

6) COMPARISON WITH OTHER METHODS

To validate the effectiveness of the proposed method, Faster RCNN and SSD is applied to SSDD, Table 7 displays the performance of the three methods. With densely connecting the feature maps from top to down, attaching RPN to each fused feature maps, mapping RoIs to the lowest fused feature map and training with the modified loss function, the proposed method increases detection accuracy by 4% and increases the detection recall from 70.8% to 83.4% compared with Faster-RCNN. Compared to the SSD, the F1 score of our proposed method has been increased from 75.6% to 89.6%.

TABLE 7. Detection performance of proposed method and faster-RCNN.

Method	Accuracy	Recall	F1
Proposed method	**96.7%**	**83.4%**	**89.6%**
Faster RCNN	92.9%	70.8%	80.4%
SSD	80.7%	71.1%	75.6%

By changing the confidence score threshold of detection results on one testing SAR images can get different values of accuracy and recall.

Fig. 7 and Fig. 8 show the detection result of Gaofen3 and Sentinel-1 SAR images. (a) shows the detection result with Faster-RCNN. (b) shows the detection result with our proposed method DCMSNN. The red, yellow, and green rectangles represent the correct detection ships, the missing ships, and the false alarms respectively. The blue rectangles represent multiple vertical ships next to each other are detected as one ship. It loses many small ships and appears many false alarms on the shore while using Faster-RCNN to detect. In addition, Faster-RCNN detects the multiple vertical ships next to each other as one ship. While using the DSMSNN, it effectively alleviates the above problems.

V. DISCUSSION

Experiments on expanded pubic SSDD verify the effectiveness of our proposed method DCMSNN in multiscale and multiscene SAR ship detection. Using dense connection can merge the feature maps have more semantic information with the high-resolution feature maps. The lower-level feature maps are suitable to detect small-scale ships while the higher-level feature maps are suitable to detect large-scale ships. Attaching RPN to each fused feature maps makes the model more suitable for multiscale SAR ship detection. According to the (1) and (2), the lowest fused feature map not only has higher resolution but also incorporates more semantic information, mapping the RoIs to the lowest fused feature maps can achieve the best performance. Using the modified loss function can solve the problem of easy examples dominate the training process which alleviate the inference of inshore complex background to get higher detection accuracy. When $\gamma = 3$, the model has the best performance. Fig. 9 shows the detection results of test SAR images with our proposed method (a) and (c) are ground truth, (b) and (d) are the corresponding results. Multiscale SAR ship detection both inshore and offshore have a better performance, but false alarms and missing ships also exist as shown in fourth and fifth lines in the Fig. 9. This is because there are still cases where objects like ships are misclassified. Due to accuracy and recall are mutual restraint, we can adjust the confidence score threshold to get the best performance. In addition, ships side by side will be detected one ship with our proposed method, this may be caused by the ships are relatively close, only leaving one proposal at NMS. It can be solved by modifying the method of NMS, but it is beyond the scope of this paper, it will be carried out in the future. In summary, this paper proposes an end-to-end method of multiscale and multiscene SAR ship detection which does not need land-sea segmentation, in addition, the features extracted by neural network are better than the features selected by hand. However, using deep learning to detect ships of SAR images needs plenty of annotated SAR data. The quality of the data is important for network, it is hoped that many high-quality datasets for SAR ship detection will be provided by research scholars in the future.

REFERENCES

[1] X. Yang *et al.*, "Automatic ship detection in remote sensing images from Google Earth of complex scenes based on multiscale rotation dense feature pyramid networks," *Remote Sens.*, vol. 10, no. 1, p. 132, 2018.
[2] M. F. Fingas and C. E. Brown, "Review of ship detection from airborne platforms," *Can. J. Remote Sens.*, vol. 27, no. 4, pp. 379–385, 2001.
[3] Y. D. Yu, X. B. Yang, S. J. Xiao, and J. L. Lin, "Automated ship detection from optical remote sensing images," *Key Eng. Mater.*, vol. 500, pp. 785–791, Mar. 2012.
[4] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3446–3456, Sep. 2010.
[5] D. J. Crisp, "A ship detection system for RADARSAT-2 dual-pol multi-look imagery implemented in the ADSS," in *Proc. IEEE Int. Conf. Radar*, Adelaide, SA, Australia, Sep. 2013, pp. 318–323.
[6] X. Leng, K. Ji, S. Zhou, X. Xing, and H. Zou, "An adaptive ship detection scheme for spaceborne SAR imagery," *Sensors*, vol. 16, no. 9, p. 1345, 2016.
[7] D. J. Crisp, "The state-of-the-art in ship detection in synthetic aperture radar imagery," *Org. Lett.*, vol. 35, no. 42, pp. 2165–2168, 2004.
[8] C. Wang, F. Bi, W. Zhang, and L. Chen, "An intensity-space domain CFAR method for ship detection in HR SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 529–533, Apr. 2017.
[9] S.-I. Hwang and K. Ouchi, "On a novel approach using MLCC and CFAR for the improvement of ship detection by synthetic aperture radar," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 391–395, Apr. 2010.
[10] H. Zhao, Q. Wang, J. Huang, W. Wu, and N. Yuan, "Method for inshore ship detection based on feature recognition and adaptive background window," *J. Appl. Remote Sens.*, vol. 8, no. 1, p. 083608, Jan. 2014.
[11] L. Zhai, Y. Li, and Y. Su, "Inshore ship detection via saliency and context information in high-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1870–1874, Dec. 2016.
[12] Q. Wang, H. Zhu, W. Wu, H. Zhao, and N. Yuan, "Inshore ship detection using high-resolution synthetic aperture radar images based on maximally stable extremal region," *J. Appl. Remote Sens.*, vol. 9, no. 1, p. 095094, 2015.
[13] Y. Liu, M.-H. Zhang, P. Xu, and Z.-W. Guo, "SAR ship detection using sea-land segmentation-based convolutional neural network," in *Proc. IEEE Int. Workshop Remote Sens. Intell. Process.*, Shanghai, China, May 2017, pp. 1–4.
[14] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.
[16] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Lille, France, Jul. 2015, pp. 448–456.
[17] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 5353–5360.
[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Amsterdam, The Netherlands, Oct. 2016, pp. 770–778.
[19] G. Huang, Z. Liu, K. Weinberger, and L. van der Maaten. (2016). "Densely connected convolutional networks." [Online]. Available: https://arxiv.org/abs/1608.06993
[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 3431–3440.
[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Columbus, OH, USA, Jun. 2014, pp. 580–587.
[22] R. Girshick, "Fast R-CNN," in *Proc. IEEE Conf. Comput. Vis. (ICCV)*, Boston, MA, USA, Jun. 2015, pp. 1440–1448.
[23] 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.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2016, pp. 779–788.
[26] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.
[27] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie. (2016). "Feature pyramid networks for object detection." [Online]. Available: https://arxiv.org/abs/1612.03144
[28] J. Dai, Y. Li, K. He, and J. Sun. (2016). "R-FCN: Object detection via region-based fully convolutional networks." [Online]. Available: https://arxiv.org/abs/1605.06409
[29] C. Zhaowei, F. Quanfu, R. S. Feris, and N. Vasconcelos, "A unified multi-scale deep convolutional neural network for fast object detection," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 354–370.
[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. (2017). "Focal loss for dense object detection." [Online]. Available: https://arxiv.org/abs/1708.02002
[31] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," in *Proc. Int. Conf. Learn. Represent.*, Banff, Canada, Apr. 2014, pp. 1–16.
[32] M. Kang, X. Leng, Z. Lin, and K, Ji, "A modified faster R-CNN based on CFAR algorithm for SAR ship detection," in *Proc. Int. Workshop Remote Sens. Intell. Process.*, Shanghai, China, May 2017, pp. 1–4.
[33] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sens.*, vol. 9, n. 8, p. 860. 2017.
[34] S. Y. Wang, X. Gao, H. Sun, X. Zheng, and X. Sun, "An aircraft detection method based on convolutional neural networks in high-resolution SAR images," *J. Radars*, vol. 6, no. 2, pp. 195–203, 2017.
[35] J. Li, C. Qu, and J. Shao, "Ship detection in SAR images based on an improved faster R-CNN," in *Proc. BIGSARDATA*, Beijing, China, Nov. 2017, pp. 1–6.
[36] K.-K. Sung and T. Poggio, "Learning and example selectionfor object and pattern detection," Artif. Intell. Lab., MIT, Cambridge, MA, USA, Tech. Memo. 1521, 1994.
[37] H. Rowley, S. Baluja, and T. Kanade, "Human face detection in visual scenes," Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-95-158R, 1995.
[38] A. Shrivastava, A. Gupta, and R. Girshick, "Training region-based object detectors with online hard example mining," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.

JIAO JIAO (S'18) received the B.Sc. degree from Jilin University, Changchun, China, in 2016. She is currently pursuing the M.S. degree with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China.

Her research interests include computer vision, pattern recognition, and SAR image processing, especially on object detection.

YUE ZHANG (M'18) received the B.E. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in 2012, and the Ph.D. degree from the University of Chinese Academy of Sciences, Beijing, China, in 2017. He is currently an Assistant Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing. His research interests include the analysis of optical and synthetic aperture radar remote sensing images.

HAO SUN received the B.Sc. degree from the Beijing Institute of Technology, Beijing, China, in 2007, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2012.

He is currently an Associate Professor with the Institute of Electronics, Chinese Academy of Sciences. His research interests include computer vision, pattern recognition, and remote sensing image processing.

XUE YANG (S'18) received the B.Sc. degree from Central South University, Changsha, China, in 2016. He is currently pursuing the M.S. degree with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China.

His research interests include computer vision, pattern recognition, and remote sensing image processing, especially on object detection and classification.

XUN GAO (S'18) received the B.Sc. degree from Jilin University, Changchun, China, in 2016. He is currently pursuing the M.S. degree with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China.

His research interests include computer vision, pattern recognition, and remote sensing image processing, especially on semantic segmentation.

WEN HONG (M'03–SM'17) received the M.S. degree in electronic engineering from Northwestern Polytechnical University, Xi'an, China, in1993, and the Ph.D. degree from Beihang University, Beijing, China, in 1997.

Her main research interests include polarimetric interferometric, synthetic aperture radar (SAR) data processing and application, 3-D SAR signal processing, circular SAR signal processing, SAR polarimetry application, and sparse microwave imaging with compressed sensing.

KUN FU received the B.Sc., M.Sc., and Ph.D. degrees from the National University of Defense Technology, Changsha, China, in 1995, 1999, and 2002, respectively.

He is currently a Professor with the Institute of Electronics, Chinese Academy of Sciences, Beijing, China. His research interests include computer vision, remote sensing image understanding, geospatial data mining, and visualization.

XIAN SUN received the B.Sc. degree from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2004, and the M.Sc. and Ph.D. degrees from the Institute of Electronics, Chinese Academy of Sciences, Beijing, in 2009.

He is currently an Associate Professor with the Institute of Electronics, Chinese Academy of Sciences. His research interests include computer vision and remote sensing image understanding.