Small Ship Detection of SAR Images Based on Optimized Feature Pyramid and Sample Augmentation

Yicheng Gong, Zhuo Zhang, Jiabao Wen, *Member, IEEE*, Guipeng Lan, and Shuai Xiao, *Member, IEEE*

***Abstract*—Synthetic aperture radar images have become the latest high-resolution imaging equipment, which can monitor the Earth 24 h a day. More and more deep-learning technologies are applied to ship target detection; however, in complex environments, due to the small target of the ship, problems, such as false detection and miss detection, often occur. For this reason, SSPNet is proposed with several small-target-augmentation strategies to complete the detection of small ships on the sea. This network is an improvement of FPN. The model uses a context attention module (CAM), scale enhancement module (SEM), and scale selection module (SSM). CAM introduces the attention heat map, SEM uses the residual module to make the network pay more attention to specific scale targets, and SSM introduces deep semantic features into shallow features. A weighted negative sampling strategy is proposed to enable the network to select more representative samples. These modules make the network more suitable for small-target detection. The results on the SSDD dataset show that the model is superior to the existing object detection network, and the average precision (AP_{50}) reaches 91.57%.**

***Index Terms*—Deep learning, object detection, sample enhancement, ship detection, small-target detection, synthetic aperture radar (SAR) images.**

I. Introduction

In TERMS of marine resource management, field safety, shipboard operations, and maritime rescue, marine ship detection is of great significance. However, under the uncontrollable natural factors, it is difficult to achieve success by assigning marine police ships or ship target monitoring based on visible light. Ship detection is a traditional and important task for coastal countries and has become a research hotspot in the world in recent years. The research on ship target detection methods is of great significance for sea area management, marine resources development, and national security. In terms of civilian use, relevant technologies can help monitor maritime traffic and fishery management in specific sea areas of the country, monitor and crack down on illegal fishing, illegal oil dumping, illegal smuggling, and other illegal acts, and can timely detect problems in specific sea areas and issue early warnings so as to timely deal with relevant marine problems and effectively rectify and rescue them. In the military field, synthetic aperture radar (SAR) can be used to monitor the ship dynamics in key ports and sea areas, mine, and analyze important data intelligence, such as ship types and positions, so as to evaluate and analyze the adversarial strength of both sides, which is of great significance to ensure the correctness of military operations. At present, infrared images, optical remote sensing images, and SAR images are often used for ship target detection.

SAR is an advanced active microwave Earth observation system [1], [2], [3]. Nowadays, SAR systems are increasingly used in marine traffic control, illegal fishing detection, and maritime emergency rescue [4], [5]. If the resolution of the SAR images is low or the actual size of the ship's target is small, the ship may appear in the SAR image only as a bright dot [6]. Common target detection algorithms are easy to lose some important feature information when extracting features of these small targets, thus leading to false detection and missing detection. Not only that but also lightweight and real-time factors need to be considered. SAR, as a special detection and imaging tool, is different from other remote sensing means and has been widely concerned and used by countries all over the world [7]. SAR is an active remote sensing equipment. It adopts a coherent imaging mechanism and forms images by actively transmitting and receiving microwaves at specified frequency bands. It has strong detection capability for metal targets and geographical textures and can maintain stable and continuous observation imaging [8]. SAR has a certain surface penetration capability. With the use of the synthetic aperture principle and pulse compression technology, SAR can easily obtain large area, high-resolution remote sensing images. It is precisely because SAR has the advantages of all-weather, all-day, multiangle, and certain penetration capability that many spaceborne SAR imaging technologies and airborne SAR imaging technologies have emerged at home and abroad. SAR image resources are growing at a tremendous speed. In order to effectively detect ships in SAR images in real time, it is urgent to carry out deep research on SAR image ship detection algorithms.

With the fast development of deep-learning technology, deep neural networks have shown excellent performance in various fields [9], [10], [11], [12], [13]. The theoretical system of deep

Manuscript received 17 March 2023; revised 27 May 2023 and 4 July 2023; accepted 28 July 2023. Date of publication 10 August 2023; date of current version 14 August 2023. This work was supported by China Postdoctoral Science Foundation under Grant 2021M702441. *(Corresponding author: Zhuo Zhang.)*

The authors are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: gongyicheng@tju.edu.cn; z_zhuo@tju.edu.cn; wen_jiabao@tju.edu.cn; lgp@tju.edu.cn; xs611@tju.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3302575

learning belongs to the artificial neural network system, which simulates the learning mechanism of the human brain. Through the superposition of multilayer networks, the network results at different levels can be learned from shallow to deep, and the characteristics of the original data can be learned [14]. After Hubel and Wiesel put forward the convolutional neural network (CNN), the CNN developed rapidly. CNN was first applied to handwritten font recognition. Later, deep learning began to be used in object classification [15], [16], [17], object detection [18], [19], object tracking [20], [21], image assessment [22], and other fields, and achieved very good results. There are also research articles on deep-learning-based active learning [23], [24], [25] and few-shot learning [26] to make up for the defects of data distribution.

Although the research on ship detection algorithms around SAR images has yielded fruitful results due to the unique imaging mechanism and special image characteristics of SAR images, the ship detection algorithms of SAR images still need further development. In addition, the feature of ship targets in different scales in the images also brings great challenges to the detection algorithms, and the research on small targets and end-to-end detection is also in urgent need of development.

In this article, the network SSPNet, which is based on the faster R-CNN and feature pyramid, is used to solve the SAR small ship detection on the sea surface. The performance of small-target detection is improved through the context attention module (CAM), scale enhancement module (SEM), and scale selection module (SSM). SSPNet can study the relationship between adjacent layers to ensure proper functional separation between depth and bottom layers so as to avoid gradient mismatch between different layers. The contributions are listed as follows.

1) We propose a novel network named SSPNet for SAR small ship detection on the sea surface.
2) We propose CAM, SEM, and SSM modules to enhance the performance of small-target detection.
3) Tested on the SSDD dataset, the SSPNet performs better than normal object detection models, and all modules can improve the small ship detection.

The rest of this article is organized as follows. In Section II, we will introduce the related work of SAR ship detection. In Section III, we will introduce the specific network structure. In Section IV, we will introduce the performance comparison between the method used in this article and the other methods on the dataset. Section V discusses the difficulties analysis and future work. Finally, Section VI concludes this article.

II. Related Work

Before the outbreak of deep-learning technology, the mainstream SAR ship detection methods are mainly divided into methods based on constant false alarm rate [27] (CFAR) and methods based on machine learning. The traditional methods of detection of SAR ships include mainly methods of contrast information, geometric and structural properties, and statistical analysis. The CFAR method is one of the most commonly used methods. The CFAR calculates the adaptive threshold based on the speed of the false alarm and the statistical distribution of the background confusion and compares the detected point to the threshold to distinguish the object from the background. The modeling of complex sea clutter and the selection of model parameters are the main factors affecting the method. Generally, when the scene of ship detection is relatively simple, CFAR method will achieve good performance, but it is difficult to deal with small ships and complex near-shore scenes.

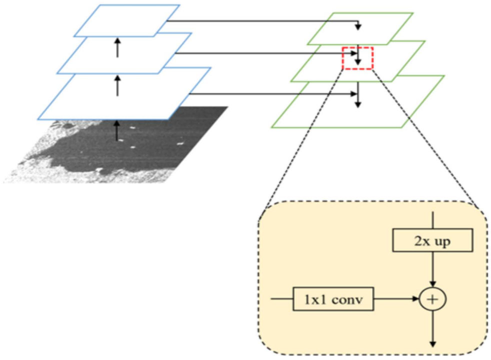


Fig. 1. Structure of FPN.

In recent years, with the vigorous development and wide application of AI technology, the deep-learning object detection method has demonstrated excellent detection accuracy and efficiency across the ages [28]. At present, the research on ship detection algorithms of SAR images mainly uses CNN to perform classification tasks or uses computer vision-related network algorithms to carry out the corresponding ship detection research. Kang et al. [29] used faster R-CNN to detect ships in SAR images and obtained the location information of ship targets and the corresponding confidence score. For areas with low confidence score, they used CFAR again for secondary detection. Wang et al. [30] used SSD to make relevant network changes to detect targets in SAR images. Li et al. [31] improved faster R-CNN network, combined transfer learning, feature fusion, and other techniques, so that the detection accuracy of the network was further improved.

III. Methods

A. Backbone

The SSPNet we use is an improved network based on FPN [32] and faster R-CNN [33]. FPN is used to extract multiscale features. Faster R-CNN is a very classic two-stage method in the field of object detection.

FPN draws on the idea of an image pyramid to extract the feature pyramid of the input image. The process is shown in Fig. 1. Specifically, FPN includes bottom-up, top-down, horizontal connection, and convolution fusion. Bottom-up refers to sending the preprocessed images to the pretrained network (such as ResNet) to obtain feature maps C_2, C_3, C_4, and C_5 of different sizes. From top to bottom means that the top C_5 doubles up sampling to get T_5, then conducts upper sampling to obtain T_4, T_3, and T_2. Each of the top-down and bottom-up layers corresponds to each other. The feature maps obtained from the

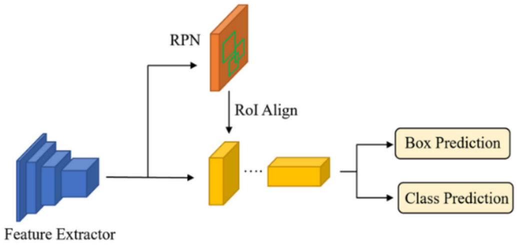


Fig. 2. Structure of faster R-CNN.

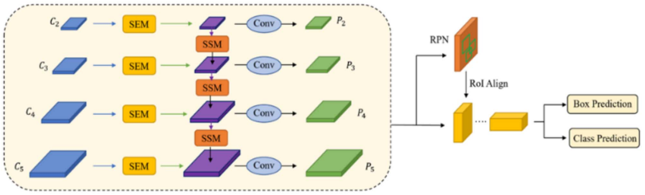


Fig. 3. Framework of SSPNet.

two samples are fused together through a horizontal connection, and finally a 3 × 3 convolution to reduce the aliasing effect of upsampling.

Faster R-CNN consists of three parts: feature extractor, regional proposal network (RPN), and RoI header, see Fig. 2 for the specific network structure. Faster R-CNN concentrates these three parts on the same network, so the detection speed is faster than the previous two versions. The RPN consists of two parts. The first is to determine whether all preinstalled anchors are direct or negative. The second is to return to the boundary and correct the anchor points to obtain more accurate recommendations, so the RPN network is a part of the initial inspection to determine whether there is a target (in this case, no specific category is determined), and to change the anchor points to make the boxes more accurate. After we re-extract the deep features of the image, RPN uses the softmax function to determine whether the anchor in the feature map is the foreground or the background. Then, anchor points with foreground are used as area suggestions, and bounding box regression is performed for these area suggestions. The RoI pooling layer collects the original features and suggestions given by RPN, extracts the suggested feature map, and performs final classification and location regression.

B. Framework

The SSPNet network structure we use is shown in Fig. 3. Faster R-CNN and FPN are used as the backbone of the network, and CAM, SEM, SSM, and WNS modules are added to improve the detection performance of small objects. Among them, the role of CAM is to generate thermal maps of different scales and increase the context information of the network. The SEM module connects the thermal diagram and characteristic diagram of the same level. The SSM module connects the attention heatmap and feature map of different levels. The WNS module filters out more representative samples to the detector.

The loss function is similar to faster R-CNNs loss function, but the difference is the loss function of CAM, which is added, as follows:

$$L_{\text{RPN}} = \frac{1}{N_{\text{cls}}} L_{\text{BCE}} + \frac{1}{N_{\text{reg}}} L_{\text{reg}} \tag{1}$$

$$L_{\text{head}} = \frac{1}{N_{\text{cls}}} L_{\text{CE}} + \frac{1}{N_{\text{reg}}} L_{\text{reg}} \tag{2}$$

where L_{RPN} and L_{Head} are the loss functions used in faster R-CNN. L_{RPN} and L_{Head} use the smooth $L1$ loss to do the regression of the target box, L_{RPN} uses the binary cross-entropy (BCE) as the loss function of classification, while L_{Head} uses the cross-entropy (CE). N_{cls} represents the size of minibatch, and N_{reg} represents the number of target boxes. L_{CAM} is the loss function of CAM. The specific formula is

$$L_A = \alpha L_A^d + \beta L_A^d \tag{3}$$

where L_A^b represents the BCE, L_A^d represents the dice loss, and α and β represent the superparameters of the two loss functions. The purpose of using dice loss is to distinguish the priority of prospects. The purpose of using BCE is to deal with the disappearance of the gradient when the attention heatmap and the s supervised attention heatmap do not intersect. The total loss function can be described as follows:

$$L \ = L_{\text{RPN}} \ + L_{\text{head}} + L_A. \tag{4}$$

C. Context Attention Module

CAM module is designed to extract different scales of attention heatmaps and embedded into SEM and SSM modules. The input of the CAM module is the feature maps C_2, C_3, C_4, and C_5 of different layers obtained by FPN. After upsampling, the feature sizes of all layers are unified, and then the concatenation operation is performed. The features obtained after splicing are highlighted in the attention heatmap through the atmosphere spatial pyramid pooling (ASPP) [34]. Although the traditional downsampling can increase the receptive field, it will reduce the spatial resolution, and the use of ASPP can ensure the resolution while expanding the receptive field. This is very suitable for detection and segmentation tasks. The increase of receptive field can detect and segment small targets, and the high resolution can accurately locate targets. Not only that, ASPP can capture multiscale context information, and the division rate indicates filling according to this value. Setting different division rates brings different receptive fields to the network, that is, multiscale information is obtained. After the attention heatmaps are obtained, the final outputs A_2, A_3, A_4, and A_5 are obtained through several activation functions. The formula is given as follows:

$$a_k = \delta \left(\emptyset_k \left(f_C, w, s \right) \right) \tag{5}$$

where f_C is the context-aware feature extracted by ASPP, w is the convolution layer parameter, s is the convolution step in ASPP, $\emptyset_k$ is a layer of 3 × 3 convolution, and δ refers to the sigmoid activation function. The specific process is shown in Fig. 4.

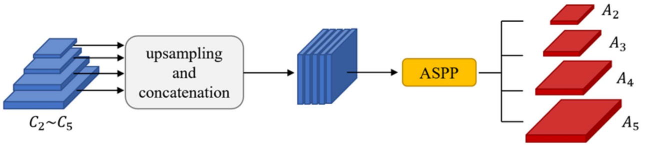


Fig. 4. Overview of CAM module.

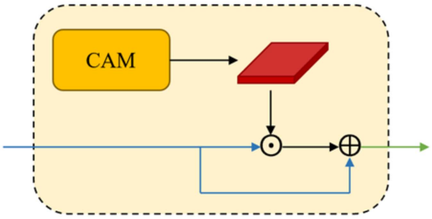


Fig. 5. Overview of SEM module.

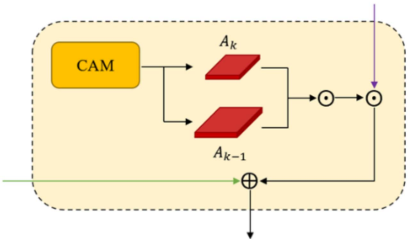


Fig. 6. Overview of SSM module.

D. Scale Enhancement Module

The SEM module includes the CAM module, which uses operations similar to shortcut to generate scale-aware features and is used as feature augmentation. The specific process is shown in Fig. 6, and the formula is given as follows:

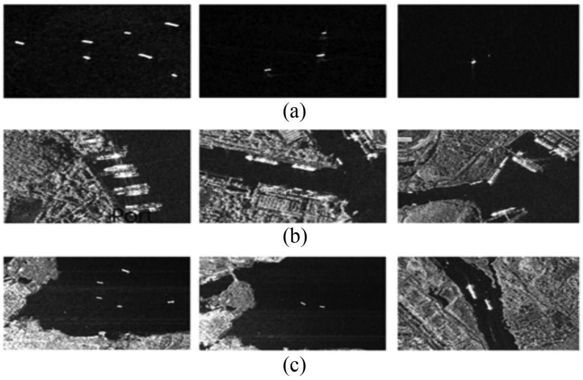
$$f_k^o = f_k^i + A_k \odot f_k^i \tag{6}$$

where f_k^i is the input feature map, and the input of this module is $C_2 \sim C_5$ layer from FPN (shown by the blue arrow in Fig. 5). f_k^o is the output feature map, and A_k is the attention heatmap of layer k obtained by CAM module. Note that the attention heatmap and the input feature map are multiplied elementwise, and then added to the input. This step is similar to the residual structure of ResNet [35] in order to make full use of the context information and avoid the disappearance of the gradient. This operation is added to each layer of the FPN.

After the SEM module, the attention heatmap generated by CAM module is successfully introduced into FPN, and the connection of residual structure can increase network parameters without gradient disappearance or gradient explosion. An attention heatmap is added to enable features to notice small-target information at a specific scale rather than broad target information.

E. Scale Selection Module

In order to make full use of shallow features to locate small targets, feature selection needs to be completed. The SSM module transfers more semantic information from deep features to shallow features. Here, we choose features between adjacent layers. The targets detected between adjacent features will be more similar, but the deep features have richer semantic features. The selection method is shown in the following formula:


(a)

(b)

(c)

Fig. 7. This is a sample of SSDD dataset. (a) SSDD dataset simple background image. (b) SSDD dataset port background image. (c) SSDD dataset complex background image.

$$P'_{k-1} = (A_{k-1} \odot f_u(A_k)) \odot f_u(P'_k) + C_{k-1} \tag{7}$$

where P'_{k-1} is the output feature map, f_u refers to the upsampling operation, A_{k-1} and A_k are the attention heatmaps between two adjacent layers output by CAM, P'_k refers to the feature map after layer k fusion (as shown by the purple arrow at the top of Fig. 6), and C_{k-1} refers to the output of residual blocks in the layer $k-1$ SEM structure (as shown by the green arrow in Fig. 7). The specific process is shown in Fig. 7.

F. Weighted Negative Sampling (WNS)

Because SAR images will have a lot of noise, and ship targets will be obscured, blurred, and difficult to detect, SPPNet uses the WNS module to select more representative samples to enhance the generalization ability of the detector. Confidence is the probability distribution of a model's judgment on a sample and is the most direct factor to judge the representativeness of the sample. In addition, SSPNet also considers the degree of incompleteness of the target and uses the criterion of intersection over foreground (IoF) [36] to measure it. Next, a weighting function is used to fuse confidence and IoF

$$s = \frac{e^{\lambda c_i + (1-\lambda) I_i}}{\sum_{i=1}^{N} e^{\lambda c_i + (1-\lambda) I_i}} \tag{8}$$

where c_i represents the confidence level of the ith sample test, I_i represents the IoF value, and λ is a weighting parameter. The calculated score can represent the representativeness of a sample and further guide the work of the detector.

IV. Results

A. Implementation Details

In order to ensure the reliability and scientificity of the experimental results, the experiments are conducted on the basis of reliable physical and software resources. Two NVIDIA RTX 3080Ti GPUs are used in the server. As for the experimental conditions of the software, the Ubuntu 18.04 system is used, equipped with CUDA, GCC 7.3, Python, PyTorch, and other packages.

In the experiment, the network uses SGD as the optimizer, and the learning rate is 0.005, 0.1 times lower than that of ten epochs. In the programming and evaluation training phase, the number of the proposed frameworks was set at 2000, 1000 in the testing phase, and a total of 20 epochs were prepared. In the aspect of superparameters, set α, β, and λ as 0.01, 1, and 0.5, respectively.

In order to prove the effectiveness of SSPNet, we have done two experiments. First, ablation experiments are implemented to prove the effectiveness of SEM, SSM, and WNS modules in SSPNet. Faster R-CNN-FPN is used to be the baseline. The second is to compare the average precision (AP) on the SSDD dataset with many existing object detection models.

SSDD dataset [37] is used for small-target detection of all ships by downloading public SAR images on the Internet. This dataset is produced and released by the Naval Aviation University. The dataset follows the process similar to PASCAL VOC [38] to build the dataset, and the ratio of train set, verification set, and test set is 7:2:1. The target area of the dataset is trimmed to a size of 500 × 500 pixels and is obtained by manually marking the position of the ship target. The SAR image data in the SSDD dataset comes from many different satellites, such as RadarSat-2 and so on. The resolution of these SAR images is from 1 to 15 m. There are four polarization modes totally (HH, HV, VV, and VH). There are a total of 1160 images in the dataset, with an average of 2.12 ships in each image, totally 2456 ships.

Fig. 7 shows SAR image with simple background, SAR image of port, and SAR image with complex background. The length and width of ship targets in the SSDD dataset cover 0.04–0.24 of the length and width of the whole image. Most ship targets belong to small targets, of which the size of small targets is less than 100 pixels and that of large targets is about 50 000 pixels. It can effectively test the performance of deep-learning algorithm on small targets and multiscale targets.

B. Ablation Experiments

SSPNet adds CAM, SEM, SSM, and WNS modules on the basis of faster R-CNN-FPN. These modules help to enhance the detection of small targets by the model. In this experiment, we compare the difference in model performance by adding these modules one by one to baseline. In the inspection task, we usually use AP to judge whether a model is good or not. The higher the AP value, the better the model performance. In this experiment, we selected AP50 as the evaluation index. The “50” means that the threshold value of IoU is 0.5. In the experiment, except for different network models, the experimental configuration and parameters are the same. The specific experimental results are shown in Table I.

TABLE I
RESULTS ON SSDD DATASET AFTER ADDING DIFFERENT MODULES

SEM	SSM	WNS	AP_{50}
			88.72
√			89.87
	√		90.01
		√	89.50
√	√		90.19
	√	√	90.07
√		√	90.09
√	√	√	**91.57**

The bold value denotes the best result.

It can be seen from the data in the table that when SEM, SSM, and WNS modules are added, respectively, the performance on the SSDD dataset is improved by 1.15%, 1.29%, and 0.78%, respectively. When two modules are added, respectively, the performance of the model will be improved more (1.47%, 1.35%, and 1.37%). After all modules are added, the model performance is 2.12% higher than the baseline. It can be seen that the SEM module can effectively focus on the target with a specific scale, the SSM module can effectively provide deep semantic information for shallow features, and the WNS module can effectively guide the detector to identify small ship target images that are difficult to judge. The introduction of these modules can enable the original FPN to focus more on small-target detection and improve the detection accuracy of baseline.

Among the three modules, SSM plays the most important role. It can be seen that in the multilevel pyramid, the texture information extracted from the bottom feature is mostly the target's texture information, and the semantic information extracted from the top feature is mostly the target's semantic information. The addition of semantic information makes the network more effective in identifying the location of small targets. When the three modules are joined at the same time, the gain of their interaction is greater.

C. Comparison Experiment

At present, there are many classical object detection models. In this experiment, we selected the classic two-stage model to compare with SSPNet and proved that SSPNet can effectively detect small targets through various indicators on the SSDD dataset. In this experiment, the SSPNet model is adapted from faster R-CNN, and faster R-CNN is a classic two-stage model. Therefore, we choose other classic two-stage models to compare with SSPNet and prove that SSPNet can effectively detect small targets through various indicators on the SSDD dataset. In this experiment, precision, recall, average precision (AP50), mean average precision (mAP), and other indicators are used to measure the model performance. The specific results are shown in Table II.

In the table, faster R-CNN-FPN indicates that FPN is used for feature extraction in faster R-CNN, which is the baseline of our experiment. In FPN, shallow feature map can detect small targets more easily. While faster R-CNN uses depth feature map,

TABLE II
PERFORMANCE OF DIFFERENT MODELS ON SSDD DATASET

Network	Precision	Recall	AP_{50}	mAP
Faster R-CNN	73.29	76.10	70.25	43.67
RetinaNet	97.62	93.62	89.83	57.72
RetinaNet+SSPNet	98.71	94.23	91.11	58.10
Cascade RCNN-FPN	98.70	94.32	90.24	57.91
Cascade RCNN+SSPNet	98.87	93.39	91.51	58.07
Faster R-CNN-FPN	97.88	93.54	88.72	57.50
SSPNet	98.87	94.46	91.57	58.29

The bold values denote the best result.

TABLE III
PERFORMANCE OF THE MODEL AT DIFFERENT VALUES OF λ

λ	0	0.1	0.2	0.3	0.4	0.5
AP_{50}	89.68	89.91	90.53	90.12	91.20	91.57

λ	0.6	0.7	0.8	0.9	1
AP_{50}	91.12	90.80	89.88	89.93	89.40

therefore, the detection performance on the SSDD dataset is poorer than that of the faster R-CNN-FPN. RetinaNet [39] is another excellent work in object detection. Its core network is FPN, which uses two subnets to classify and regress the output of each feature map. The purpose of the loss function (focal loss) is to solve the extreme imbalance problem of positive and negative boxes in the object detection. Therefore, the retinal network is more accurate than the high-speed R-CNN-FPN. The SSPNet we used uses an improved FPN; therefore, CAM, SEM, and other modules can be added to the basic network as an external module. We selected cascade RCNN-FPN, faster R-CNN-FPN, and RetinaNet as the basic networks to compare the model performance before and after adding modules. From the results in Table II, we can see that when we use the cascade RCNN-FPN model to replace our improved model with FPN, AP_{50} increases by 1.27%. When FPN in RetinaNet is replaced, AP_{50} increases by 1.28%. The SEM module is used to connect concerns to take the advantage of context information. It introduces the deep semantic features into the shallow features, successfully making the network more suitable for small-target detection.

D. Comparison of Different λ Values

In the WNS module, the λ coefficient is used to fuse the confidence and IoF. This experiment is to explore what value λ takes will make the best fusion score s. The λ value range is [0,1], so we can obtain the experimental results every 0.1 increase from 0. The model used in this experiment is a complete SSPNet, and the parameters of all experiments are the same except λ. The result curve is shown in Table III.

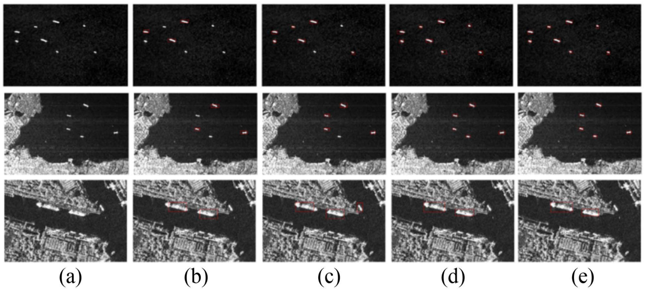


Fig. 8. Visualization of different network detection results. (a) Original image. (b) Faster R-CNN. (c) Faster R-CNN-FPN. (d) RetinaNet. (e) SSPNet.

It can be seen from the results in the figure that the performance is not very good when the λ value is 0 or 1. The detection performance of the model fluctuates when the value is 0.1–0.9, and the model performance is the highest when the λ value is 0.5. It can be seen that by only relying on these two factors at the same time, the model can better identify small-target ships.

E. Visualization of Results

To more intuitively show the effectiveness of SSPNet, we show the detection of faster R-CNN, faster R-CNN-FPN, RetinaNet, and SSPNet on the SSDD dataset, as shown in Fig. 8.

Fig. 8(a) represents the original figure, and Fig. 8(b), (c), (d), and (e) represent the detection results using faster R-CNN, faster R-CNN-FPN, RetinaNet, and SSPNet, respectively. It can be seen intuitively from the results in the figure that when using faster R-CNN, there will be a large number of missed inspections because it is not suitable for small-size targets. When faster R-CNN-FPN and RetinaNet are used, due to the introduction of the FPN structure, the detection accuracy of small targets is improved a lot, but there is still missing and false detection, and the detection frame cannot fit well with the foreground target. When SSPNet is used, it can be seen from the figure that there are no errors or omissions, and the boundary of the detection frame is more accurate than RetinaNet. It can be seen that SSPNet can better locate small-target ships.

V. DISCUSSION

A. Difficulties Analysis

Object detection is currently an important research direction in deep learning, and many highly successful models have now emerged. However, small-target detection remains a difficult problem in object detection. In the real scene, due to the large number of small targets, small-target detection has a broad application prospect and plays an important role in many fields, such as automatic driving, intelligent medical treatment, defect detection, and aerial image analysis.

The difficulties of small-target detection mainly include the following.

1) *Less features available:* Compared with large/medium targets, small targets have lower resolution and less information, and it is difficult to extract distinguishing features.

2) *High positioning accuracy requirements:* Small targets are located too small in the image and vulnerable to environmental interference, and a pixel offset in network prediction has a huge impact on small targets.
3) *The proportion of small targets in the existing dataset is small:* The existing dataset pays less attention to the special type of small targets. At the same time, small targets are not easy to label, the human cost is huge, and they are more sensitive to errors.
4) *Sample imbalance:* During training, set a threshold value to judge whether the anchor frame belongs to a positive sample, which will lead to the sample imbalance problem of different size targets. Therefore, when there is a large difference between the manually set anchor frame and the real frame, the model will ignore the detection of small targets.
5) *Small-target aggregation:* Small targets are more likely to gather. At this time, the prediction frame of the network model may filter out a large number of correct frames due to nonmaximum suppression, resulting in missing small targets, or the frame distance is too close, which makes the model difficult to converge.
6) *Network structure:* At present, there are not many optimization designs for small-target characteristics of existing algorithms, combined with the difficulty caused by the characteristics of small target itself, leading to the poor performance of the existing algorithms in small-target detection.

B. Future Work

In SAR images, small-target ship recognition becomes more difficult due to background interference, and false detection often occurs. The network SSPNet used in our work makes the network more suitable for small-target scenarios due to the attention heatmap, residual connection, and incorporating deep-seated semantics into shallow-seated features. Although the model used in this article can achieve good performance, there are still many areas to be improved in the future. For example, the introduction of several modules adds network parameters. In order to facilitate model application, pruning can be considered.

VI. Conclusion

In this work, we focus on small-target ship positioning in SAR images. We use SSPNet model, which improves the structure of faster R-CNN and FPN and introduces small-target sample augmentation modules (CAM, SEM, SSM, and WNS). Attention heatmaps are added to CAM module to enrich the feature information of samples. SEM alleviates the problem of gradient disappearance by residual structure, connecting sample features with attention heatmaps. SSM makes each layer more suitable for small-target detection by introducing the upper layer's semantics' features into the current layer's features. We use SSDD dataset on which the AP of SSPNet model detection can reach 91.57%, higher than the other classical object detection networks.

References

[1] G.-C. Sun, M. Xing, X.-G. Xia, J. Yang, Y. Wu, and Z. Bao, "A unified focusing algorithm for several modes of SAR based on FrFT," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 3139–3155, May 2013.

[2] A. Reigber and A. Moreira, "First demonstration of airborne SAR tomography using multibaseline L-band data," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2142–2152, Sep. 2000.

[3] F. Lombardini and A. Reigber, "Adaptive spectral estimation for multibaseline SAR tomography with airborne L-band data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2003, pp. 2014–2016.

[4] X. X. Zhu and R. Bamler, "Superresolving SAR tomography for multidimensional imaging of urban areas: Compressive sensing-based TomoSAR inversion," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 51–58, Jul. 2014.

[5] X. X. Zhu and R. Bamler, "Very high resolution spaceborne SAR tomography in urban environment," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 12, pp. 4296–4308, Dec. 2010.

[6] P. Wan, J. Wang, Z. Zhao, and S. Huang, "Edge extraction of small reflection target in SAR image," *Proc. SPIE*, vol. 4029, pp. 141–146, 2000.

[7] E. Sansosti, P. Berardino, M. Manunta, F. Serafino, and G. Fornaro, "Geometrical SAR image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2861–2870, Oct. 2006.

[8] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, Dec. 2017.

[9] S. Xiao, G. Lan, J. Yang, W. Lu, Q. Meng, and X. Gao, "MCS-GAN: A different understanding for generalization of deep forgery detection," *IEEE Trans. Multimedia*, to be published, doi: 10.1109/TMM.2023.3279993.

[10] Y. Guo et al., "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016.

[11] A. C. Mater and M. L. Coote, "Deep learning in chemistry," *J. Chem. Inf. Model.*, vol. 59, no. 6, pp. 2545–2559, 2019.

[12] D. Ravì et al., "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[13] G. Lan, S. Xiao, J. Wen, D. Chen, and Y. Zhu, "Data-driven deepfake forensics model based on large-scale frequency and noise features," *IEEE Intell. Syst.*, to be published, doi: 10.1109/MIS.2022.3217391.

[14] J. Ahmad, H. Farman, and Z. Jan, "Deep learning methods and applications," in *Deep Learning: Convergence to Big Data Analytics*. Berlin, Germany: Springer, 2019, pp. 31–42.

[15] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[16] S. Xiao, G. Lan, J. Yang, Y. Li, and J. Wen, "Securing the socio-cyber world: Multiorder attribute node association classification for manipulated media," *IEEE Trans. Comput. Social Syst.*, to be published, doi: 10.1109/TCSS.2022.3213832.

[17] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[18] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019.

[19] M. Zhang et al., "Deep-learning detection of cancer metastases to the brain on MRI," *J. Magn. Reson. Imag.*, vol. 52, no. 4, pp. 1227–1236, 2020.

[20] S. M. Marvasti-Zadeh, L. Cheng, H. Ghanei-Yakhdan, and S. Kasaei, "Deep learning for visual tracking: A comprehensive survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 3943–3968, May 2022.

[21] A. Brunetti et al., "Computer vision and deep learning techniques for pedestrian detection and tracking: A survey," *Neurocomputing*, vol. 300, pp. 17–33, 2018.

[22] Y. Zhao, S. Xiao, J. Yang, W. Lu, and X. Gao, "No-reference quality index of tone-mapped images based on authenticity, preservation, and scene expressiveness," *Signal Process.*, vol. 203, 2023, Art. no. 108782.

[23] G. Attardi, "DeepNL: A deep learning NLP pipeline," in *Proc. 1st Workshop Vector Space Model. Natural Lang. Process.*, 2015, pp. 109–115.

[24] S. Wu et al., "Deep learning in clinical natural language processing: A methodical review," *J. Amer. Med. Inform. Assoc.*, vol. 27, no. 3, pp. 457–470, 2020.

[25] N. Zhu et al., "Deep learning for smart agriculture: Concepts, tools, applications, and opportunities," *Int. J. Agricultural Biol. Eng.*, vol. 11, no. 4, pp. 32–44, 2018.

[26] T. R. Andersson et al., "Seasonal arctic sea ice forecasting with probabilistic deep learning," *Nature Commun.*, vol. 12, no. 1, 2021, Art. no. 5124.

[27] P. P. Gandhi and S. A. Kassam, “Analysis of CFAR processors in nonhomogeneous background,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 427–445, Jul. 1988.

[28] M. Yasir et al., “Ship detection based on deep learning using SAR imagery: A systematic literature review,” *Soft Comput.*, vol. 27, pp. 63–84, 2023.

[29] M. Kang, X. Leng, Z. Lin, and K. Ji, “A modified faster R-CNN based on CFAR algorithm for SAR ship detection,” in *Proc. Int. Workshop Remote Sens. Intell. Process.*, 2017, pp. 1–4.

[30] Y. Wang, C. Wang, and H. Zhang, “Combining a single shot multibox detector with transfer learning for ship detection using sentinel-1 SAR images,” *Remote Sens. Lett.*, vol. 9, no. 8, pp. 780–788, 2018.

[31] J. W. Li et al., “Ship detection in SAR images based on convolutional neural network,” *Syst. Eng. Electron.*, vol. 40, no. 9, pp. 1953–1959, 2018.

[32] 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.

[33] S. Ren et al., “Faster R-CNN: Towards real-time object detection with region proposal networks,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, vol. 28.

[34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018.

[35] 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.

[36] X. Yu, Y. Gong, N. Jiang, Q. Ye, and Z. Han, “Scale match for tiny person detection,” in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2020, pp. 1246–1254.

[37] M. Everingham et al., “The pascal visual object classes challenge: A retrospective,” *Int. J. Comput. Vis.*, vol. 111, no. 1, pp. 98–136, 2015.

[38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2999–3007.

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

Yicheng Gong received the B.S. degree in communication and information engineering in 2021 from Tianjin University, Tianjin, China, where he is currently working toward the M.S. degree in information and communication engineering with the School of Electrical and Information Engineering.

His research interests include pattern recognition and deep learning.

Zhuo Zhang received the M.S. degree in electronic information engineering in 2021 from Tianjin University, Tianjin, China, where he is currently working toward the Ph.D. degree in information and communication engineering with the School of Electrical and Information Engineering.

His research interests include data information assessment and deep learning.

Jiabao Wen (Member, IEEE) received the Ph.D. degree in information and communication engineering from Tianjin University, Tianjin, China, in 2021.

He is currently a Postdoctor with the School of Electrical and Information Engineering, Tianjin University. His research interests include ocean information processing, pattern recognition, and cloud computing.

Guipeng Lan received the B.S. degree in communication and information engineering in 2020 from Tianjin University, Tianjin, China, where he is currently working toward the M.S. degree in information and communication engineering with the School of Electrical and Information Engineering.

His research interests include face analysis, computer vision, and pattern recognition.

Shuai Xiao (Member, IEEE) received the Ph.D. degree from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2022.

He is currently a Postdoctor with the School of Electrical and Information Engineering, Tianjin University. His research interests include image processing, computer vision, and image forensics.