A Fast and Accurate Small Target Detection Algorithm Based on Feature Fusion and Cross-Layer Connection Network for the SAR Images

Ming Sun, Yanyan Li, Xiaoxuan Chen, Yan Zhou, Jinping Niu, and Jianpeng Zhu

***Abstract*—Target detection technology has been greatly improved for synthetic aperture radar (SAR) images recently, due to the advancement in the deep learning domain. However, because of the existence of clutter in the SAR images, it is still a challenge to detect small targets with high accuracy and low computational complexity. To solve this problem, a detection algorithm based on a feature fusion and cross-layer connection network is proposed in this article. First, attention feature fusion is applied to improve the feature fusion ability for the small targets by allocating weights to various feature maps adaptively. Meanwhile, the depthwise separable convolution (DW-Conv) is used to reduce the computational complexity caused by the increasement of network layers. Then, a cross-layer connection (Cross-Connect) submodule is proposed to fuse shallow features with deep features further. Finally, a multiscale target detection (Multi-Detect) submodule is designed to improve the detection ability for small targets. We compare the proposed algorithm with the other representative methods on the SAR-Ship-Dataset and SSDD, quantitative evaluations show that our proposed algorithm can reach the highest computational efficiency. Therefore, because of the superior performance in terms of accuracy and efficiency, the algorithm proposed in this article is more suitable to detect small targets for the SAR images.**

***Index Terms*—Attentional feature fusion, cross-layer connection, deep learning (DL), small target detection, synthetic aperture radar (SAR) images.**

I. Introduction

SYNTHETIC aperture radar (SAR) [1] is a powerful technology to obtain high-resolution images under various conditions, and noncooperative targets can be detected through detection algorithms, which is important for the military and civilian applications. Since accuracy and efficiency are two metrics that are commonly utilized in target detection algorithms, current target detection studies for SAR images have focused on these two factors.

The traditional target detection algorithms for SAR images include the constant false alarm rate (CFAR) [2], multiple image feature fusion [3], support vector machine [4], Bayesian classifier [5], and enhanced adaptive window CFAR method [6]. The CFAR method determines a threshold through processing the input noise; if the input signal's energy is larger than the threshold, a target is judged to be present. However, in addition to the accuracy being reduced due to the interference caused by the ground, rain, snow, and thermal noise, the false alarm rate will also deteriorate along with the increment of data amount. Wang et al. [7] proposed a new ship detector based on the rich target scattering information available from polarized SAR, which combines two-parameter CFAR for detection that can suppress false alarms caused by SAR image cross partials in addition to further reducing the number of missed targets. A moving target identification and tracking approach based on the Gaussian mixed probability hypothesis density filter was suggested [8] to increase the detection accuracy in cluttered environments; this method can detect targets by removing stationary strong ground clutter. Considering the scattered noise, Li et al. [9] proposed a two-domain sparse reconstruction method, in which the intensity information of the image domain and the structure information of the feature domain are both used to improve the robustness of detection performance. Although these above detection algorithms have solved some problems, their computational complexity is high and efficiency is low; thus, they are only suitable for simple SAR images.

Since the deep learning (DL) technology [10], [11], [12], [13], [14], [15] can complete target detection without requiring feature creation by humans, it has been promoted to be applied to target detection recently. Because of the principle of SAR imaging, some small targets that have little feature information usually exist in the images, and it is difficult to separate these small targets from the background correctly. To solve this problem, the algorithm is proposed combining the features of SAR image and time series [16] and the algorithm constructing global attention modules in the spatial and channel domains [17]. Lin et al. [18] proposed a new network architecture based on the faster R-CNN to further improve the detection performance by using squeeze and excitation mechanism to improve the detection performance. Sun et al. [19] proposed an anchor-free method

Manuscript received 23 February 2023; revised 7 May 2023 and 17 July 2023; accepted 2 September 2023. Date of publication 18 September 2023; date of current version 4 October 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61901372, Grant 61901371, and 62072373, in part by the Natural Science Research Program of Shaanxi Province under Grant 2020JQ-599, in part by the China Postdoctoral Science Foundation under Grant 2020M683541, and in part by the Key Research and Development Program of Shaanxi Province of China under Grant 2021KW-05. *(Corresponding author: Yanyan Li.)*

The authors are with the School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: s_m1234@163.com; liyanyan_xd@163.com; chenxx@nwu.edu.cn; yanzhou@nwu.edu.cn; jpingniu@nwu.edu.cn; 202133517@stumail.nwu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3316309

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

for ship target detection in HR SAR images to address the complex surroundings, targets defocusing, and diversity of the scales, and obtain encouraging detection performance compared other networks. Sun et al. [20] proposed a novel YOLO-based arbitrary-oriented SAR ship detector using bidirectional feature fusion and angular classification (BiFA-YOLO) to address the multiscale, arbitrary directions and dense arrangement issues; this method shows strong robustness and adaptability in HR SAR images. Kuang et al. [21] proposed an elaborately designed deep hierarchical network based convolutional neural network with multilayer fusion to improve the detection performance for small-sized ships. Song et al. [22] proposed an attention-guided end-to-end change detection network (AGCDetNet) based on the fully convolutional network and attention mechanism to the detection performance of high-resolution remote sensing images. The algorithm introduces spatial attention and channel attention mechanisms during the feature-extracting stage [23]. The algorithm introducing a rotating bounding box-based target detection algorithm that can effectively reduce the interference of background pixels and avoid overlapping detection boxes of dense targets [24] and the algorithm embedding an enhanced attention module into the RCNN [25] are proposed. Detection accuracy of the aforementioned algorithms has been improved for small targets; however, their detection efficiency is affected by the complex computational processes. Considering the detection efficiency, the algorithm embedding a convolutional attention module into the feature extraction network to extract useful features [26], the algorithm reducing the number of deep convolutions [27], [28] through modified lightweight RetinaNet, and the method segmenting the coastline by CNN [29] are proposed. Although these methods have improved detection efficiency, some small targets are still dismissed, which causes the detection accuracy to be unsatisfactory. According to the above analysis, it can be seen that achieving a good balance between detection accuracy and efficiency is a challenging task for small target detection in SAR images.

To solve the aforementioned issues, in this article, for the SAR images, we present a small target detection algorithm based on feature fusion and cross-layer connection (FFCLC) network. DL-based target detection algorithms are usually grouped into two classes: one-stage and two-stage, in which the two-stage class pursues accuracy and the one-stage class pursues speed. After years of development, the one-stage methods are no less accurate than the two-stage ones and show a great lead in speed. Therefore, in this article, we choose the one-stage network YOLOv5 with high performance as the baseline. In the proposed FFCLC-based detection algorithm, our main work is as follows.

1) To raise the small target detection accuracy for SAR images, an attention feature fusion (AFF) which allocates weights to various feature maps adaptively, a cross-layer connection (Cross-Connect), and a multiscale small target detection submodule (Multi-Detect) are designed.
2) To improve the computational efficiency and reduce the number of redundant parameters, the depthwise separable convolution (DW-Conv) [30] is used.

The SAR-Ship-Dataset and SSDD have been used to validate the effectiveness of the proposed algorithm. When compared to other representative detection techniques, the suggested algorithm's mAP is the greatest and it has the highest computing efficiency of the methods mentioned.

The rest of this article is organized as follows. Section II introduces the detailed principle of the proposed algorithm. Section III analyses the outcomes of the ablation experiments and compares the results with other representative algorithms. Section IV discusses the performance of submodules. Section V summarizes the entire study and outlines potential future study directions.

II. Methods

The network structure of the proposed FFCLC-based detection algorithm for SAR images is given in Fig. 1. As shown, the CBS, SPPF [31], and CSPDarknet53 [32] compose the backbone of FFCLC, which can extract most features. The CBS structure consists of the convolutional layers, batch normalization (BN), and SiLU [33] activation function, and the SPPF structure transmits the input information through several MaxPool layers having sizes 5×5 sequentially.

A feature pyramid structure based on PANet [34] is used in the neck subnetwork, which is located between the detection subnetwork and the backbone subnetwork. In this layer, strong semantic characteristics are conveyed from the top down while strong location features are transmitted from the bottom up. Semantic features and localization features are combined to aggregate parameters. To further improve the diversity and robustness of the features, this article proposed two structures: AFF and Cross-Connect, which can be fused into PANet to raise the accuracy of small target detection for SAR images; meanwhile the DW-Conv is utilized to reduce the redundant parameters in the network.

At the Multi-Detect stage, four different scales of feature maps F_1, F_2, F_3, and F_4 are obtained, and the corresponding detection box for each scale is generated. Then, the nonmaximum suppression [35] is used to filter out the lower confidence detection box. Finally, the loss function, which consists of classification loss, target box regression loss, and confidence loss, is used to further optimize the detection box to complete the target detection. In this article, the binary cross entropy loss function [36] is used for classification loss and confidence loss, whereas the complete intersection over union loss function [37] is used for target box regression loss.

In the following, the key points of AFF, Cross-Connect, and Multi-Detect are described in detail.

A. AFF Submodule

By continually convolving the images through the backbone network, some basic features, such as the edges and position of the images, may be recovered in the shallow network. As the network goes deeper, more sophisticated features can be extracted. YOLOv5 fuses shallow and deep features through the Concat operation, in which a simple stitch of the feature maps is completed, and some default fixed weights are allocated to the extracted feature maps. Since the resolution of feature maps extracted through the deep network is low, if some fixed weights

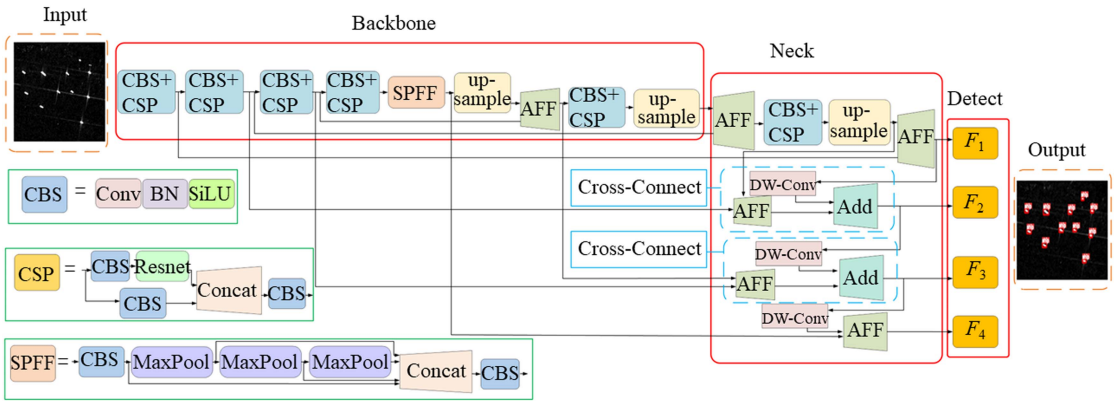


Fig. 1. Structure of the proposed network.

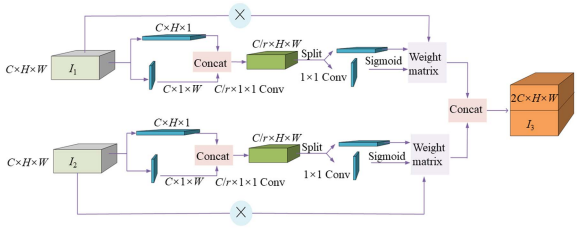


Fig. 2. Structure of AFF.

are utilized for these feature maps, the small targets usually are judged to be noisy by the network due to their little information, which will affect the detection performance for the YOLOv5.

The feature fusion network is enhanced by the addition of coordinate attention [38] in this research to address the aforementioned problem, and the modified network is known as AFF. In the AFF submodule, since the model can be instructed to concern more useful information by the coordinate attention, the feature information for various feature maps can be learned, which can make the weights to be allocated adaptively to the various feature maps, and the detection performance can be improved. The proposed AFF submodule is shown in Fig. 2, where *C*, *H,* and *W* denote the number of channels, height, and width of the image, respectively. I_1 is the high-resolution feature map extracted by the shallow network, I_2 is the feature map with more semantic information but lower resolution extracted by the deep network, *C*/*r* denotes the dimensionality is reduced to *r* times of the original, and Split denotes the feature map is divided into two directions along the height and width.

To obtain feature maps I_1^{h}, I_1^{w}, I_2^{h}, and I_2^{w} in the horizontal and vertical directions, as shown in Fig. 2, feature maps I_1 and I_2 with the sizes of $C\times H\times W$ first undergo a global average pooling operation, using pooling kernels of (H, 1) and (1, W) to encode the horizontal and vertical directions, respectively, which can be expressed as

$$I_{\mathrm{c}}^{\mathrm{h}} = \frac{1}{W} \sum_{0\le i\le W} I_{\mathrm{c}}(h,i) \tag{1}$$

$$I_{\mathrm{c}}^{\mathrm{w}} = \frac{1}{H} \sum_{0\le i\le H} I_{\mathrm{c}}(w,i) \tag{2}$$

where $I_{\mathrm{c}}^{\mathrm{h}}$ and $I_{\mathrm{c}}^{\mathrm{w}}$ stand for the feature maps of the *c*th dimensional channel in row *h* and the *c*th dimensional channel in column *w*, separately, whereas I_{c} denotes the *c*th dimensional feature of I_1 or I_2. From (1) and (2), we can see that the result is a feature map with two directions, which indicates that the attention mechanism will concentrate on not only the channel direction but also the features of the spatial direction. Through this operation, it is possible to guarantee that when information is searched in a different direction, the spatial information from one direction will be kept.

After that, to create intermediate feature maps having features in both horizontal and vertical directions, Concat splicing, feature transformation by convolution, BN, and ReLU nonlinear activation function [39] are applied to $I_{\mathrm{c}}^{\mathrm{h}}$ and $I_{\mathrm{c}}^{\mathrm{w}}$, which is expressed as

$$f = \mathrm{Relu}\left(\mathrm{Conv}\left(\left[I^{\mathrm{h}}, I^{\mathrm{w}}\right]\right)\right) \tag{3}$$

where Conv denotes 1×1 convolution.

Then, two convolutions of size 1×1 and Sigmoid [40] operations are used for feature transformation to obtain the attention weights in vertical and horizontal directions, which are calculated as shown in the following:

$$\eta^{\mathrm{h}} = \mathrm{sigmoid}\left(\mathrm{Conv}\left(f^{\mathrm{h}}\right)\right) \tag{4}$$

$$\eta^{\mathrm{w}} = \mathrm{sigmoid}\left(\mathrm{Conv}\left(f^{\mathrm{w}}\right)\right) \tag{5}$$

where f^{h} and f^{w} denote the horizontal and vertical feature maps divided by the feature map *f* produced by (3), respectively. The output of coordinate attention is produced by multiplying I_1 and I_2 with the relevant attention weight matrices.

Finally, I_1 and I_2, which have been assigned distinct weights, are fused by the Concat operation, and the feature fusion process of this network is represented as

$$I_3 = \mathrm{Concat}\left(CA\left(I_1\right), CA\left(I_2\right)\right) \tag{6}$$

where I_3 represents the result obtained after coordinate attention and Concat operations, and *CA* represents the operation procedure from (1) to (5).

The addition of coordinate attention can complete the weights' allocation adaptively, and the model's capacity to learn the features of small targets can be increased; nevertheless,

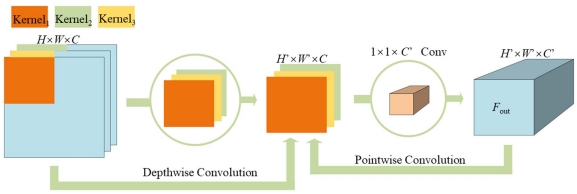


Fig. 3. Structure of DW-Conv.

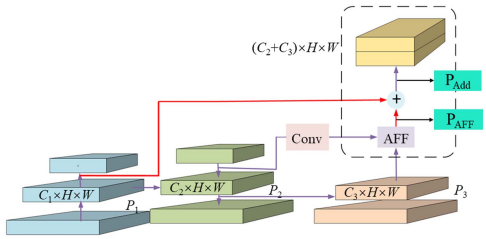


Fig. 4. Structure of Cross-Connect.

the number of network layers grows, and the computational cost rises. To raise the model's computational efficiency, the DW-Conv as given in Fig. 3 is used in the neck subnetwork.

As can be seen from Fig. 3, the DW-Conv consists of depthwise convolution and pointwise convolution. For the input with the size of $H \times W \times C$, where H, W, and C denote the number of height, channels, and width of the feature map, respectively. DW-Conv uses convolution kernels with different parameters such as Kernel_1, Kernel_2, and Kernel_3 for different input channels. In the depthwise convolution, a one-to-one relationship between the channel and convolution kernel exists; therefore, the channels' number of the feature map obtained after the depthwise convolution is the same as the input. For the depthwise convolution, its parameters' number can be calculated through the following equation:

$$P_{\text{depth}} = K_{\text{W}} \cdot K_{\text{H}} \cdot C_{\text{input}} \tag{7}$$

where K_{w} and K_{H} stand for the width and height of the convolution kernel, respectively, and C_{input} denotes the total number of input channels.

Because the number of feature maps is the same as the number of channels in the input layer, the feature maps cannot be enlarged; also, the feature information of the spatial locations of distinct channels has not been efficiently exploited. To achieve a new feature map, the pointwise convolution is necessary. In the depthwise convolution, the size of $1 \times 1 \times C'$ convolution kernel is used to weigh these feature maps obtained through the depthwise convolution in the depth direction, where the C' stands for the number of convolution kernels, and then a new feature map F_{out} can get. For the pointwise convolution, its parameters' number can be calculated through the following equation:

$$P_{\text{point}} = 1 \cdot 1 \cdot C_{\text{input}} + C_{\text{input}} \cdot C_{\text{output}} \tag{8}$$

where C_{output} stands for the number of output channels.

According to the above description, after adding (7) and (8), we can get the parameters' total number of the DW-Conv as follows:

$$P_{\text{DW-Conv}} = K_{\text{H}} \cdot K_{\text{W}} \cdot C_{\text{input}} + C_{\text{input}} \cdot C_{\text{output}}. \tag{9}$$

In the conventional convolution, the parameters' number is indicated as

$$P_{\text{Conv}} = K_{\text{W}} \cdot K_{\text{H}} \cdot C_{\text{input}} \cdot C_{\text{output}}. \tag{10}$$

Therefore, the ratio between P_{conv} and $P_{\text{DW-Conv}}$ can be given as follows:

$$\frac{P_{\text{Conv}}}{P_{\text{DW-Conv}}} = \frac{K_{\text{H}} \cdot K_{\text{W}} \cdot C_{\text{output}}}{K_{\text{H}} \cdot K_{\text{W}} + C_{\text{output}}}. \tag{11}$$

Since the parameters of the convolution kernel are all greater than 1, the value of (11) is much greater than 1, and the number of parameters for the DW-Conv is much fewer than the conventional convolution.

It is clear from the above description that using the suggested AFF improves the ability to learn the feature of small targets, and that using DW-Conv reduces the amount of network parameters.

B. Cross-Connection Submodule

An improved extraction capability called Cross-Connect is introduced to increase the small target feature's detection accuracy. The structure of the proposed Cross-Connect submodule based on the feature pyramid network is shown in Fig. 4. The ability to fuse shallow features like details, edges, and contours with deeper features can be improved while maintaining the same computational cost by adding Cross-Connect structures to input and output nodes of the same size.

In this submodule, the AFF structure proposed in the above subsection is used to fuse features. For the Cross-Connect, because each input feature has a different resolution, the contribution of these input features to the output features is unequal. To learn the importance of each feature, here, the fast normalized fusion [41] algorithm is utilized to add additional weights to each input, which can be expressed as

$$\text{out} = \sum_i \frac{u_i \cdot x_i}{\varepsilon + \sum_i u_i} \tag{12}$$

where x_{i} denotes the ith feature map, u_{j} denotes the weight coefficient of the ith feature map, and ε denotes a very small value to prevent instability during the fusion process. This method is similar to softmax [42], which extends the range to [0, 1], and simplifies the softmax operation by removing its exponential operators to improve the computational efficiency.

In Fig. 4, "+" denotes the operation to fuse two feature maps, $P_1 \in \mathbb{R}^{c_1 \times h \times w}$ denotes the feature maps generated through the backbone network, $P_2 \in \mathbb{R}^{c_2 \times h \times w}$ denotes the feature maps generated through the top-down paths in the feature pyramid network, $P_3 \in \mathbb{R}^{c_3 \times h \times w}$ denotes the feature maps generated through the bottom-up paths, C_1, C_2, and C_3 refer to the channel numbers for the feature maps P_1, P_2, and P_3, respectively, and W

and H refer to the width and height. After introducing the learned weight coefficient of the branches that are colored with red in Fig. 4 to the "+" operation, the size of the final feature map becomes $(C_2 + C_3) \times H \times W$. To describe more intuitively, in Fig. 4, the output after the AFF and the "+" operation are expressed as P_{AFF} and P_{Add}, respectively. P_{Add} is calculated as

$$P_{\mathrm{Add}} = \frac{u_1 \cdot P_{\mathrm{AFF}} + u_2 \cdot P_1}{u_1 + u_2 + \varepsilon}. \tag{13}$$

Two weights u_1 and u_2 of the branches colored with red in Fig. 4 are substituted into the "+" operation to produce the feature map P_{Add}. This equation can further speed up the localization and classification of small targets by performing the "+" operation.

C. Multi-Detect Submodule

The input images are subsampled 8 times, 16 times, and 32 times in the YOLOv5 backbone subnetwork to create feature maps with sizes of large, medium, and small. These feature maps have different target feature information. The feature maps obtained through the shallow network have better resolution and more detailed target location data, however, they lack some complex features, such as texture. The feature maps derived from the deep network have the advantage of the semantic information but lack underlying features. As a result, the feature fusion network can fuse the feature maps that were obtained through the shallow and deep networks. At the input side of the network, the input image is resized to 640 × 640, after subsampled with different sizes and feature fusion, three different scales of 80 × 80, 40 × 40, and 20 × 20 feature maps are obtained at the detection layer, respectively.

SAR images are different from optical images in that the targets' pixels are smaller and some targets are even only visible as a bright spot in some SAR images. The receptive field [43] of each feature map can be determined based on the three scale feature maps created by the original feature extraction network, and the size of the receptive field can reflect the range of target boxes recognized by the network, it can be calculated as follows:

$$\mathrm{RF}(i) = (\mathrm{RF}(i+1) - 1) \cdot \mathrm{Stride} + \mathrm{Ksize} \tag{14}$$

where Ksize is the size of the convolution kernel or pooling kernel, RF(i) is the receptive field of the ith layer, and Stride is the ith layer's step.

Among the three scaled feature maps extracted, the feature map with a size of 80 × 80 is the feature map for the smallest pixel target in the image, and its receptive field mapped to the input image can be calculated as 8 × 8 according to (14); thus, if the target's pixels are smaller than 8 × 8, this target cannot be detected. Since some image details will be lost if the image is continuously subsampled, the original feature extraction submodule is improved in this article.

Fig. 5 shows that the size of each feature map is expressed as Channel × Height × Width. The line colored with red in Fig. 6 represents the feature map created by extracting the input four times subsampled; in the feature fusion network, this feature map is fused with the features extracted by the deeper network,

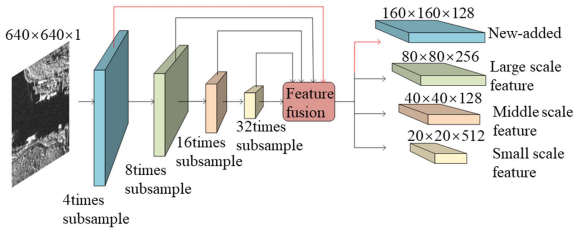


Fig. 5. Structure of Multi-Detect.

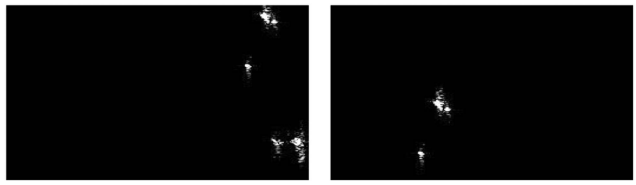


Fig. 6. Visualization of partial datasets.

TABLE I
DETAILED PARAMETERS OF SSDD

Parameter	Value
sensors	RadarSat-2, TerraSAR-X
resolution	1m-15m
scale	1:1, 1:2 and 2:1
total images	1160
total ships	2456

a feature map with a size of 160 × 160 can be got at the detection layer, and the receptive field is 4 × 4 according to (14). It can be seen that the receptive field of this feature map obtained through the proposed structure is smaller; therefore, small targets can be detected.

In this section, the proposed FFCLC network in this article is introduced in detail in aspects of AFF, Cross-Connect, and Multi-Detect. The FFCLC network enhances the fusion of the shallow features and deep features, and the detection accuracy and detection efficiency can be improved according to the theory analyses. In the next section, we will use some experiments to demonstrate this conclusion.

III. EXPERIMENTS AND ANALYSES

A. Dataset and Preprocessing

The current SAR image datasets include SSDD (SAR Ship Detection Dataset, SSDD) [44], LS-SSDD-v1.0 (Large-Scale SAR Ship Detection Dataset-v1.0, LS-SSDD-v1.0) [45], and SAR-Ship-Dataset [46]. In this experiment, the SSDD and SAR-Ship-Dataset are used. SSDD dataset has many different scales of large, medium, and small images. We have added experiments on the SSDD dataset in the comparison test and further verified the multiscale target detection characteristics of the algorithm proposed in this article; more information about the SSDD dataset is shown in Table I. In the SAR-Ship-Dataset, 102 Chinese Gaofen-3 and 108 Sentinel-1 images are the main data source, which were processed as 43 819 images with size of 256 pixels in both range and azimuth exist; we chose to conduct

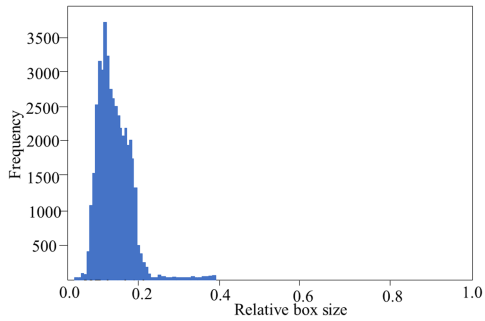


Fig. 7. Distribution of ship size.

ablation experiments on this dataset because of its large data volume. Some of the dataset visualizations are shown in Fig. 6.

To illustrate the distribution of ship size in this dataset, the bounding box size of the ship target relative to the original image is calculated as

$$\text{rate} = \frac{\sqrt{W_{\text{bbox}} \times H_{\text{bbox}}}}{\sqrt{W_{\text{img}} \times H_{\text{img}}}} \tag{15}$$

where the rate indicates the ratio of the target box to the image, and W_{bbox}, H_{bbox}, W_{img}, and H_{img} indicate the width and height of the target box and the image, respectively. The statistics result of the ship size distribution is given in Fig. 7, in which Frequency and Relative box size indicate the number of times the image appears and the proportion of the target box occupying the image, respectively. It can be seen the relative sizes of most of the targets are smaller than 0.2; therefore, the dataset belongs to small target data, which is beneficial to verify the small target algorithm proposed in this article.

The dataset for the input network is preprocessed through the Mosaic [47] data enhancement method; in this method, four images are cropped randomly and scaled, then, these four images are stitched into a single image in a random arrangement. This proprocessing can enrich the dataset, increase the number of small targets, and raise the training speed of the network. Additionally, the images also have been scaled adaptively, which reduces information redundancy by adding just the right amounts of black borders to the original image.

B. Experimental Environment and Parameter Setting

This experiment is conducted on the PyCharm 2022 platform using the Ubuntu 18.04 operating system, based on the Pytorch 1.7.0 framework, and model training is accelerated by the GPU under the NVIDIA RTX 2080 Ti (11GB of memory) GPU in the CUDA 10.2 environment. The dataset was randomly divided into a training set, a validation set, and a test set in the ratio of 7:2:1 for the experiments. The initial learning rate is 0.01, the momentum parameter is 0.937, and the warm-up method with epoch of 3 and momentum parameter of 0.8 is used to warm up the learning rate. In the warm-up phase, the learning rate is updated using 1-D linear interpolation up to 0.1. The number of training epochs is 300, and the learning rate is finally updated by the cosine annealing method. Some hyperparameters used in the experiments are listed in Table II.

TABLE II
NETWORK HYPERPARAMETERS

Parameter	Value
initial learning_rate	0.01
final learning_rate	0.001
momentum	0.937
warm-up_epochs	3
Warm-up_momentum	0.8
warm-up bias lr	0.1
batchSize	16
epoch	300

TABLE III
RESULT OF YOLOV5 IN DIFFERENT MODELS [48]

Model	mAP	Params(M)	FLOPS(B)
YOLOv5n	45.7	1.9	4.5
YOLOv5s	56.8	7.2	16.5
YOLOv5m	64.1	21.2	49.0
YOLOv5l	67.3	46.5	109.1
YOLOv5x	68.9	86.7	205.7

C. Evaluation Indicators

In this experiment, the Precision, Recall and mean average precision (mAP) are used as evaluation metrics. The Precision is an index to evaluate the performance of the model, and its calculation is shown as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{16}$$

where TP indicates the number of ships that can be correctly detected in the image, and FP indicates the number of targets in the image that are misclassified as ships.

Recall, also known as the check-all rate, is the number of samples that are positive that were predicted to be positive, and its calculation is shown as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{17}$$

where FN indicates the number of undetected ship targets in the image.

The mean average precision is derived as stated in (17), and it is a measure of the algorithm's capacity to carry out classification and bounding box regression on various targets.

$$\text{mAP} = \int_0^1 \text{Precision}\,(\text{Recall})\text{d}\,(\text{Recall})\,. \tag{18}$$

D. Ablation Experiments

The following list outlines the different types of graphics published in IEEE journals. They are categorized based on their construction, and use of color/shades of gray.

According to the width and depth of the network, YOLOv5 algorithms are classified into YOLOv5s, YOLOv5m, YOLOv5l, YOLOv5x, and YOLOv5n. Table III shows that we can obtain the detection results of five models of YOLOv5 on the COCO dataset according to the authors of YOLOv5, and we find that

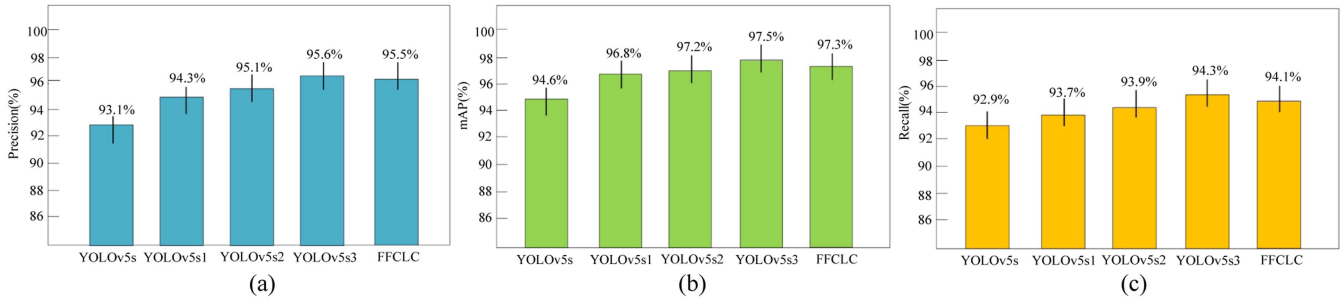


Fig. 8. Results of ablation experiments. (a) Precision. (b) mAP. (c) Recall.

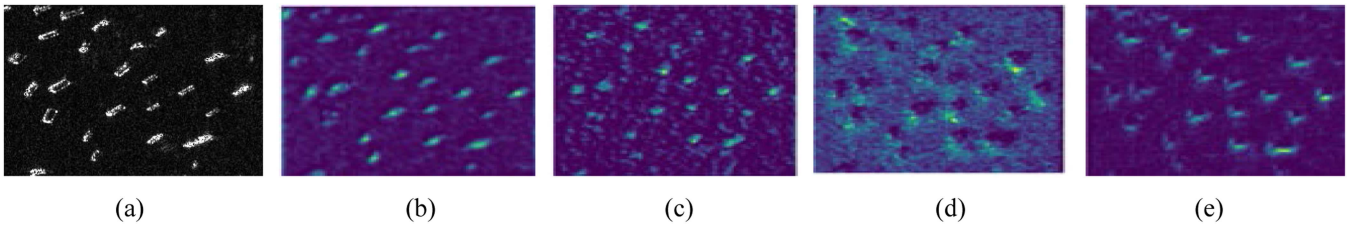


Fig. 9. Visualization of feature maps. (a) Input image. (b) Output of Cross-Connect. (c) Shallow feature map. (d) and (e) Output of PANet.

TABLE IV
COMPARISON NETWORKS

Model	AFF	Cross-Connect	Multi-Detect	DW-Conv
YOLOv5s				
YOLOv5s1	√			
YOLOv5s2	√	√		
YOLOv5s3	√	√	√	
FFCLC	√	√	√	√

TABLE V
DETECTION TIME OF DIFFERENT NETWORKS

Network	Time/ms
YOLOv5s	14.4
YOLOv5s1	14.9
YOLOv5s2	15.3
YOLOv5s3	15.8
FFCLC	11.5

YOLOv5s has a low number of parameters and has high detection accuracy. So we choose YOLOv5s as the baseline in this article, and the YOLOv5s.pth trained on the COCO dataset is used as the pretrained model to speed up the convergence of the network. To reflect the performance improvement of the proposed network for SAR image target detection more intuitively, the proposed AFF, Cross-Connect, and Multi-Detect submodules, as well as DW-Conv are added separately for ablation experiments on SAR-Ship-Dataset. Table IV shows that YOLOv5s1, YOLOv5s2, and YOLOv5s3 refer to the additions of the AFF module, Cross-Connect module, and Multi-Detect module in YOLOv5s, respectively. On replacing the regular convolution with the DW-Conv in YOLOv5s3, the FFCLC is obtained.

The experimental results shown in Fig. 8 (a)–(c) represent the experimental results of Precision, mAP, and Recall for different networks, respectively. From the experimental results, we can see that after adding the AFF submodule, the Precision, mAP, and Recall of YOLOv5s1, YOLOv5s2, YOLOv5s3, and FFCLC are improved compared with YOLOv5s since the original network lacks the ability to extract the features of small targets, and the AFF makes the network focus on the small targets' features to raise the detection accuracy for the small target. Fig. 9 shows the feature maps fused in the Cross-Connect. Fig. 9(a) represents the input image. Fig. 9(b) represents the heat maps obtained after the Cross-Connect. Fig. 9(c) represents the heat maps extracted after the shallow network. It can be seen that the resolution of Fig. 9(c) is higher and the location information of the target is more obvious. Fig. 9(d) and (e) represent the heat maps obtained after the PANet. It can be found that Fig. 9(d) and (e) has more texture features of the target, but after deep convolution, the resolution is lower and many small targets have less information, so the shallow feature map is added to the Cross-Connect network and the feature map obtained in Fig. 9(b) has more target information. So after adding the Cross-Connect and Multi-Detect submodules, the model's ability to extract small target features can be enhanced, and the performance of localization and classification for the small targets is improved; therefore, the Precision, mAP, and Recall of YOLOv5s3 are superior to those of YOLOv5s1 and YOLOv5s2.

For the FFCLC, because the DW-Conv used in the neck subnetwork reduces the number of parameters in the network, a little loss in the Precision, mAP, and Recall of the FFCLC present compared with YOLOv5s3. However, the DW-Conv will raise the computational efficiency greatly. The computational time cost of the YOLOv5, YOLOv5s1, YOLOv5s2, YOLOv5s3, and FFCLC are listed in Table V.

According to the above analyses, it can be known that the proposed FFCLC has better detection accuracy and detection

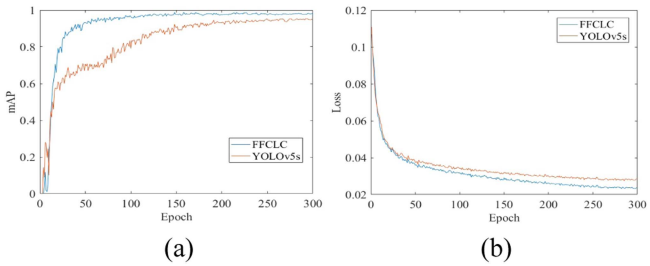


Fig. 10. Variances of mAP and loss versus with epoch. (a) mAP. (b) Loss.

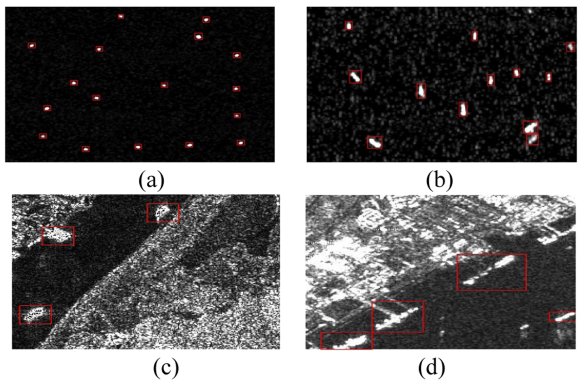


Fig. 11. FFCLC detection results. (a) and (b) Deep-sea small target scenario. (c) and (d) Near-shore complex small target scenario.

efficiency than the YOLOv5, and the results shown in Fig. 10 also demonstrate this conclusion. According to the results in Fig.10, we can obtain that the mAP can reach larger than 97%, and the loss is smaller than in YOLOv5s. Therefore, the proposed FFCLC is much more proper than the YOLOv5s to detect small targets for the SAR images.

To demonstrate the effectiveness of the proposed FFCLC, Fig. 11 gives some detection results through the FFCLC under different scenarios. As shown in Fig. 11, the small targets in the deep-sea scenarios can be detected accurately. In the nearshore complex scenarios, the FFCLC can distinguish the targets from the complex backdrop, and the targets can be located and categorized reliably.

E. Comparative Experiment

In the previous section, the effectiveness of the proposed FFCLC has been verified. Now, we choose some representative algorithms to compare with the FFCLC in different datasets of SAR-Ship-Dataset and SSDD. The detection results of Faster R-CNN, SSD, YOLOv3, TPH-YOLOv5 [49], QueryDet [50], YOLOv8, YOLOv5, and FFCLC in four different scenarios of SAR-Ship-Dataset are given in Figs. 12 and 13 which show the detection results of Faster R-CNN, SSD, YOLOv3, TPH-YOLOv5, QueryDet, YOLOv8, YOLOv5, and FFCLC in four different scenarios of SSDD, in which green boxes indicate the real targets in the images, red boxes indicate the targets detected by the network, yellow boxes indicate the missed targets, and blue boxes indicate the mis-detected targets.

From the detection results of the two deep-sea scenes in the first column and the second column of Figs. 12 and 13, it can be seen that TPH-YOLOv5, YOLOv8, YOLOv5, QueryDet, Faster R-CNN, SSD, and YOLOv3 have missed and mis-detected targets, and from the detection results of two complex backgrounds in the first column and the second column of Figs. 12 and 13, it can be seen that TPH-YOLOv5, YOLOv8, YOLOv5, QueryDet, Faster R-CNN, SSD, and YOLOv3 also have missed and mis-detected targets, whereas the results detected by the FFCLC network proposed in this article in these two scenes can match with the real ones perfectly. The proposed algorithm can accurately detect the target in both different datasets, and it is obvious from Figs. 12 and 13 that it has higher detection accuracy compared with other algorithms. In addition, the detection accuracy is higher in the deep sea and nearshore scenes, therefore, the algorithm proposed in this article will not only improve the detection performance of targets in complex scenes but also enhance the localization and detection performances for the small targets.

Tables VI and VII show the experimental results of the Faster-RCNN, SSD, YOLOv3, TPH-YOLOv5, QueryDet, YOLOv8, YOLOv5, and FFCLC based on the SAR-Ship-Dataset and SSDD, respectively. SAR-Ship-Dataset has more small ships, and the comparison results with other algorithms in this dataset can further verify the efficiency of the proposed algorithm for small target detection, as we can see in the table, the FFCLC network can acquire better performances in Precision, Recall, and mAP than the other algorithms. For the FFCLC, its mAP can reach 97.3%, which is 0.7 percentage points higher than YOLOv8, and the detection time is 0.8 ms faster than the YOLOv8. SSDD has different scale images, and the comparison results with other algorithms under this dataset can further verify the multiscale property of the proposed algorithm; as we can see in Table VII, FFCLC has the highest mAP and can reach 97.7%. It is 2.4 percentage points higher than YOLOv5 and also reaches the highest speed of 1.3 ms faster than the fastest YOLOv8. Therefore, we can see that the proposed algorithm in this article has higher detection performance and is suitable for more detection scenarios.

IV. Discussion

From the results shown in Fig. 8, it can be seen that the AFF module improves the mAP by 2.2% compared with YOLOv5s. Since the AFF module can adaptively assign weights to the feature maps, it enables the model to focus more on small ship features, which can effectively distinguish the ships from the background and improve the learning ability of small ships. As the network gets deeper, part of the ship's edge and contour information will be lost in the features recovered by the deep network, leading to missed ship detection. Incorporating Cross-Connect into the neck network can improve the ability to fuse deep features and underlying features and increase the sensing area of the model. As a result, the network can improve ship localization and lessen incorrect and missing ship identification. For the problem of small ships with small areas and sparse features, the detection accuracy of small ships is enhanced by mapping four different

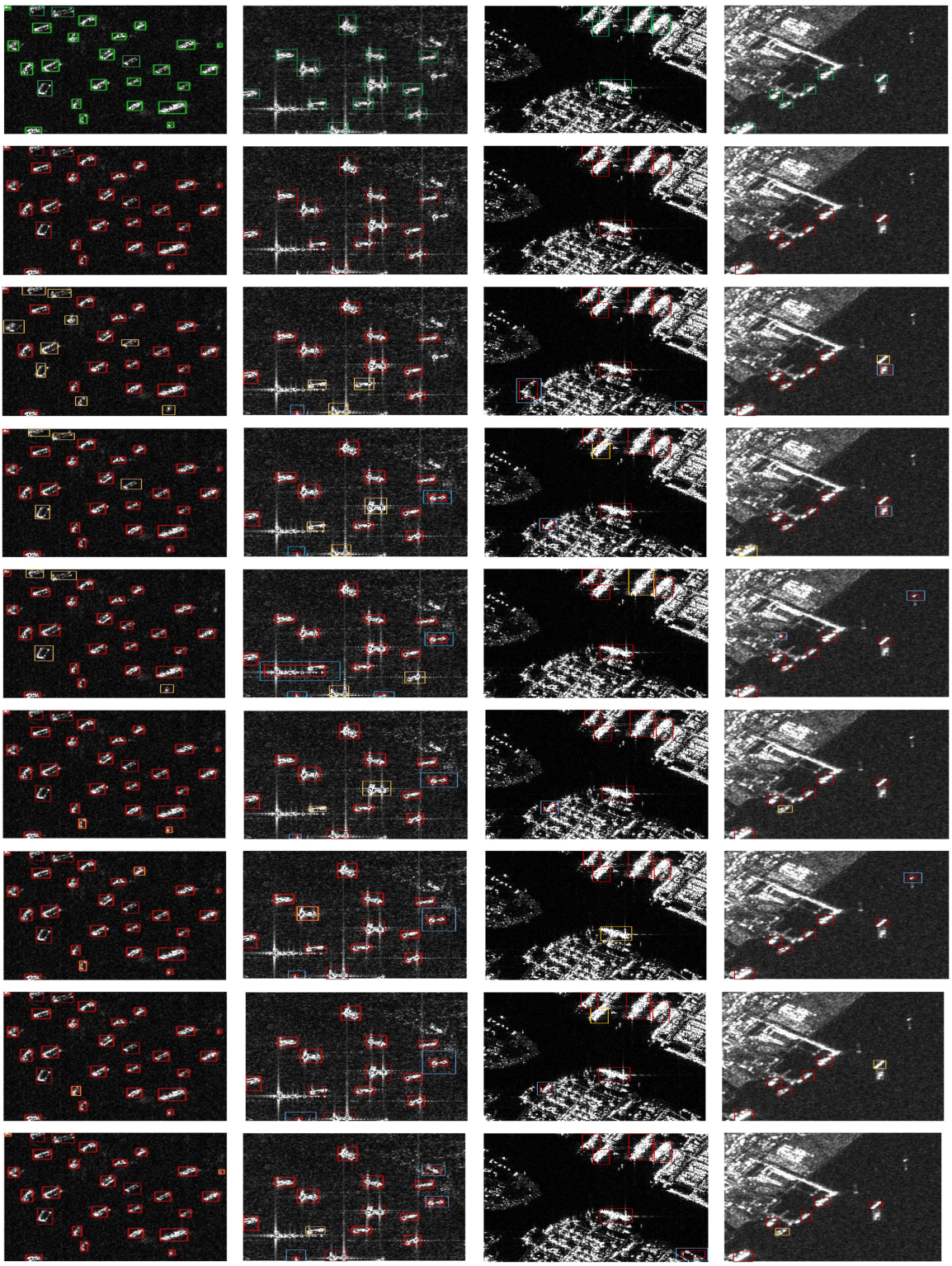


Fig. 12. Comparison of experimental results of different networks on SAR-Ship-Dataset. The first row indicates the ground truth, and the second row to the ninth row indicates the detection result of FFCLC, TPH-YOLOv5, YOLOv5, QueryDet, YOLOv8, Faster R-CNN, SSD, and YOLOv3, respectively.

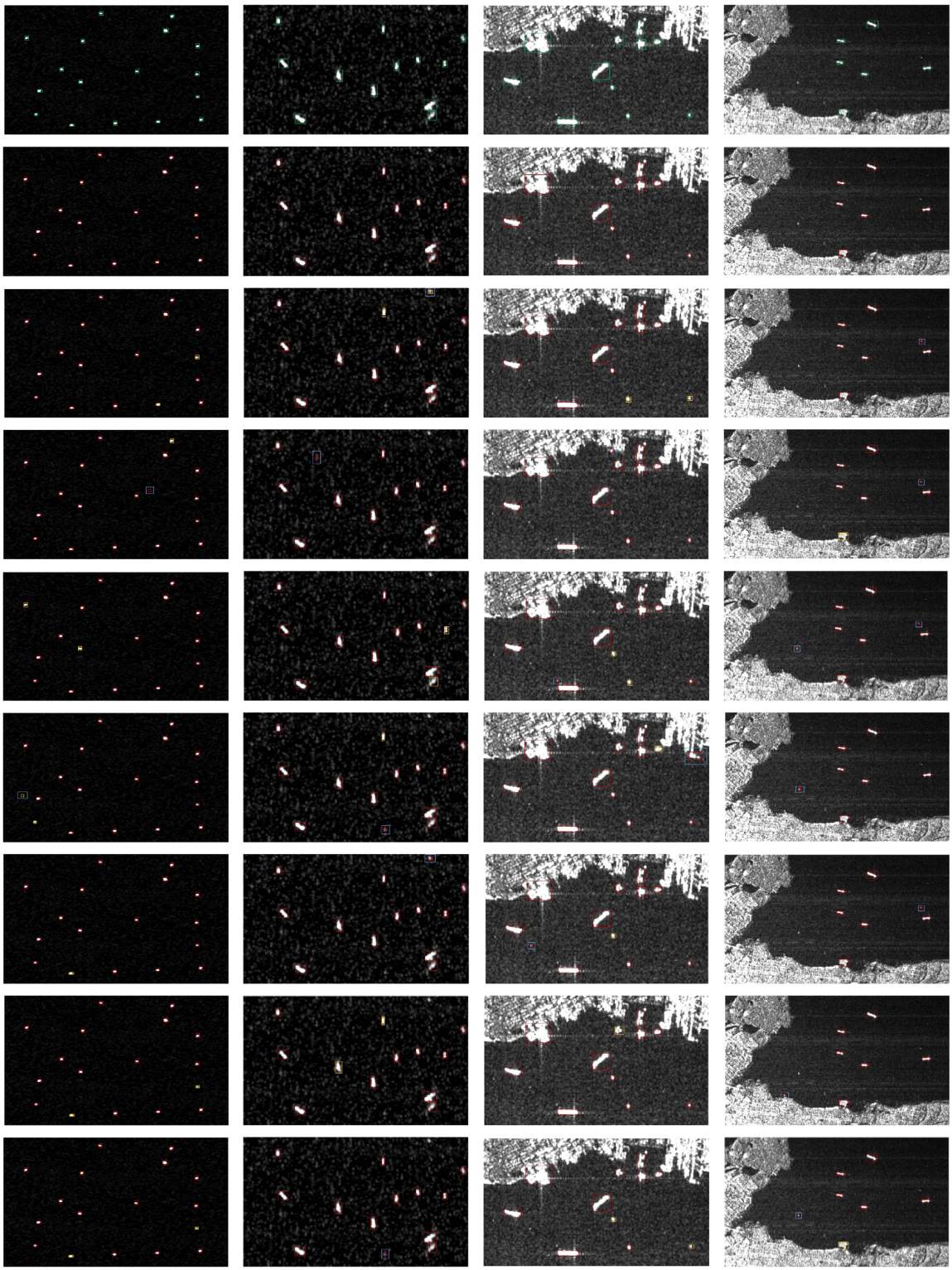


Fig. 13. Comparison of experimental results of different networks on SSDD. The first row indicates the ground truth, and the second row to the ninth row indicates the detection result of FFCLC, TPH-YOLOv5, YOLOv5, QueryDet, YOLOv8, Faster R-CNN, SSD, and YOLOv3, respectively.

TABLE VI
COMPARISON OF EXPERIMENTAL RESULTS ON SAR-SHIP-DATASET

Network	Precision	Recall	mAP	Speed/ms
Faster R-CNN	81.5%	85.6%	82.6%	98.5
SSD	85.1%	94.5%	92.6%	31.7
YOLOv3	90.6%	92.8%	91.7%	19.4
TPH-YOLOv5	91.7%	92.5%	94.6%	15.8
QueryDet	94.6%	93.5%	95.5%	14.8
YOLOv8	94.3%	93.8%	96.6%	12.3
YOLOv5	93.1%	92.9%	94.6%	14.4
FFCLC	95.5%	94.1%	97.3%	11.5

TABLE VII
COMPARISON OF EXPERIMENTAL RESULTS ON SSDD

Network	Precision	Recall	mAP	Speed/ms
Faster R-CNN	82.5%	85.7%	83.4%	105.6
SSD	85.3%	91.6%	89.3%	28.7
YOLOv3	91.2%	92.6%	92.5%	20.5
TPH-YOLOv5	93.6%	92.8%	93.5%	17.2
QueryDet	93.5%	93.0%	94.7%	15.2
YOLOv8	94.8%	94.1%	95.8%	13.9
YOLOv5	94.5%	93.8%	95.3%	14.7
FFCLC	95.3%	95.6%	97.7%	12.6

scales of feature maps to the input image, which can enable the network to sense the target information present in the input image. By adding the DW-Conv to the neck network, the number of parameters in the network is reduced compared to YOLOv5s3, making the accuracy slightly lower than YOLOv5s3, but the detection speed can be increased by 4.3 ms. According to the above discussion, taking into consideration the detection performance and detection efficiency, the proposed FFCLC is more suitable in application compared with YOLOv5s1, YOLOv5s2, and YOLOv5s3.

According to the detection results obtained by the algorithm of this article with TPH-YOLOv5, YOLOv5, QueryDet, YOLOv8, Faster R-CNN, SSD, and YOLOv3 in four different datasets shown in Figs. 12 and 13, it can be seen that in the scenarios of sparse distribution of small ships in deep sea and dense distribution of small ships in deep sea, the algorithms of TPH-YOLOv5, YOLOv5, QueryDet, YOLOv8, Faster R-CNN, SSD, and YOLOv3 have missed and mis-detected targets for small ships, and since the Multi-Detect structure proposed in this article enhances the multiscale detection of targets, the number of ships detected by the algorithm of this article is basically consistent with the actual number of ships which exist; in the two scenes of complex scenes alongside the shore and complex background, the algorithms of TPH-YOLOv5, YOLOv5, QueryDet, YOLOv8, Faster R-CNN, SSD, and YOLOv3 have both wrong and missed detection of ships. Since the three structures, i.e., AFF, Cross-Connect, and Multi-Detect, proposed in this article enhance the learning ability of small ships and can distinguish small ships from the complex background near the shore, which improves the detection accuracy of ships; therefore the detected ships are basically the same as the actual number and location of ships. At the same time, according to the detection results obtained in four different scenarios, it can be reflected that the algorithm in this article has strong robustness and is more suitable for practical detection needs.

V. CONCLUSION

To resolve the problems present in the DL methods for small target detection in SAR images, a small target detection network based on the FFCLC network is proposed in this article. Its effectiveness is demonstrated on the SAR-Ship-Dataset and SSDD, and the proposed method is compared with the Faster-RCNN, SSD, YOLOv3, TPH-YOLOv5, YOLOv8, QueryDet, and YOLOv5 networks. Theory analyses and experimental results show that this proposed method has better detection accuracy and detection efficiency; it is more suitable for the small target detection in SAR images. In future research, the focus will continue to be on improving the lightweight of the model, and the next research directions are as follows.

1) To reduce the consumption of resources and the number of parameters of the model and remain the same detection accuracy, we will improve the existing backbone network by modifying the regular convolution.
2) Attempt to port the model to a hardware platform that enables hardware acceleration of the neural network to realize faster detection of targets.

REFERENCES

[1] J. Wu et al., "Texture and intensity fusion based SAR image change detection," in *Proc. SAR Big Data Era*, Oct. 2021, pp. 1–4.
[2] C. Xu, F. Wang, Y. Zhang, L. Xu, M. Ai, and G. Yan, "Two-level CFAR algorithm for target detection in mmWave radar," in *Proc. Int. Conf. Comput. Eng. Appl.*, 2021, pp. 240–243.
[3] S. Parisotto, L. Calatroni, A. Bugeau, N. Papadakis, and C.-B. Schönlieb, "Variational osmosis for non-linear image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 5507–5516, Apr. 2020.
[4] B. Krishnapuram, J. Sichina, and L. Carin, "Physics-based detection of targets in SAR imagery using support vector machines," *IEEE Sensors J.*, vol. 3, no. 2, pp. 147–157, Apr. 2003.
[5] Y. S. Sumanto, A. Supriyatna, I. Carolina, R. Amin, and A. Yani, "Model naïve Bayes classifiers for detection apple diseases," in *Proc. IEEE 9th Int. Conf. Cyber IT Serv. Manage.*, 2021, pp. 1–4.
[6] W. Li, B. Zou, Y. Xin, L. Zhang, and Z. Wu, "An improved CFAR scheme for man-made target detection in high resolution SAR images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 2829–2832.

[7] G. Wang, X. Zhang, and J. Meng, "A small ship target detection method based on polarimetric SAR," *Remote Sens.*, vol. 11, no. 24, Nov. 2019, Art. no. 2938.

[8] Y. Zhang, H. Mu, Y. Jiang, C. Ding, and Y. Wang, "Moving target tracking based on improved GMPHD filter in circular SAR system," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 559–563, Dec. 2019.

[9] L. Li, L. Du, and Z. Wang, "Target detection based on dual-domain sparse reconstruction saliency in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4230–4243, Nov. 2018.

[10] V.-T. Hoang, V.-D. Hoang, and K.-H. Jo, "Realtime multi-person pose estimation with RCNN and depthwise separable convolution," in *Proc. Int. Conf. Comput. Commun. Technol.*, 2020, pp. 1–5.

[11] K. S. Htet and M. M. Sein, "Event analysis for vehicle classification using fast RCNN," in *Proc. IEEE 9th Glob. Conf. Consum. Electron.*, 2020, pp. 403–404.

[12] X. Xiao and X. Tian, "Research on reference target detection of deep learning framework faster-RCNN," in *Proc. IEEE 5th Annu. Int. Conf. Data Sci. Bus. Anal.*, 2021, pp. 41–44.

[13] S. Zhai, D. Shang, S. Wang, and S. Dong, "DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion," *IEEE Access*, vol. 8, pp. 24344–24357, 2020.

[14] Y. Song, Z. Xie, X. Wang, and Y. Zou, "MS-YOLO: Object detection based on YOLOv5 optimized fusion millimeter-wave radar and machine vision," *IEEE Sensors J.*, vol. 22, no. 15, pp. 15435–15447, Aug. 2022.

[15] Q. Guo, J. Liu, and M. Kaliuzhnyi, "YOLOX-SAR: High-precision object detection system based on visible and infrared sensors for SAR remote sensing," *IEEE Sensors J.*, vol. 22, no. 17, pp. 17243–17253, Sep. 2022.

[16] C. Xu, Z. He, and H. Liu, "An effective method for small targets detection in synthetic aperture radar images under complex background," *IEEE Access*, vol. 10, pp. 44224–44230, 2022.

[17] C. Zhu, D. Zhao, Z. Liu, and Y. Mao, "Hierarchical attention for ship detection in SAR images," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2020, pp. 2145–2148.

[18] Z. Lin, K. Ji, X. Leng, and G. Kuang, "Squeeze and excitation rank faster R-CNN for ship detection in SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 751–755, May 2019.

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

[20] Z. Sun, X. Leng, Y. Lei, B. Xiong, K. Ji, and G. Kuang, "BiFA-YOLO: A novel YOLO-based method for arbitrary-oriented ship detection in high-resolution SAR images," *Remote Sens.*, vol. 13, no. 21, 2021, Art. no. 4209.

[21] 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, no. 8, 2017, Art. no. 860.

[22] K. Song and J. Jiang, "AGCDetNet: An attention-guided network for building change detection in high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4816–4831, 2021.

[23] G. Jianxin, W. Zhen, and Z. Shanwen, "Multi-scale ship detection in SAR images based on multiple attention cascade convolutional neural networks," in *Proc. Int. Conf. Virtual Reality Intell. Syst.*, 2020, pp. 438–441.

[24] Q. An, Z. Pan, L. Liu, and H. You, "DRBox-v2: An improved detector with rotatable boxes for target detection in SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8333–8349, Nov. 2019.

[25] M. Li, S. Lin, and X. Huang, "SAR ship detection based on enhanced attention mechanism," in *Proc. IEEE 2nd Int. Conf. Artif. Intell. Comput.Eng.*, 2021, pp. 759–762.

[26] B. Chai, L. Chen, H. Shi, and C. He, "Marine ship detection method for SAR image based on improved faster RCNN," in *Proc. IEEE SAR Big Data Era (BIGSARDATA)*, pp. 1–4, Oct. 2021.

[27] T. Miao et al., "An improved lightweight RetinaNet for ship detection in SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4667–4679, 2022.

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

[29] Y. Li, W. Zhu, and B. Zhu, "SAR image nearshore ship target detection in complex environment," in *Proc. IEEE 5th Adv. Inf. Technol., Electron. Autom. Control Conf.*, 2021, pp. 1964–1968.

[30] Y. Gong, J. Peng, S. Jin, X. Li, Y. Tan, and Z. Jia, "Research on YOLOv4 traffic sign detection algorithm based on deep separable convolution," in *Proc. IEEE Int. Conf. Emerg. Sci. Inf. Technol.*, 2021, pp. 333–336.

[31] K. He, X. Zhang, J. Sun, and S. Ren, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Jan. 2015.

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

[33] N. Ma et al., "Activate or not: Learning customized activation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Virtual Conf.*, 2021, pp. 8032–8042.

[34] W. Wang et al., "Efficient and accurate arbitrary-shaped text detection with pixel aggregation network," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 8439–8448.

[35] E. R. Capia, A. M. Sousa, and A. X. Falcão, "Improving lung nodule detection with learnable non-maximum suppression," in *Proc. IEEE 17th Int. Symp. Biomed. Imag.*, 2020, pp. 1861–1865.

[36] S. Janthakal and G. Hosalli, "A binary cross entropy U-net based Lesion segmentation of granular parakeratosis," in *Proc. Int. Conf. Adv. Elect., Elect., Commun., Comput. Autom.*, 2021, pp. 1–7.

[37] Z. Zheng et al., "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, Aug. 2022.

[38] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13713–13722.

[39] D. Dũng and V. K. Nguyen, "Deep ReLU neural network in high-dimensional approximation," *Neural Netw.*, vol. 142, pp. 619–635, Oct. 2021.

[40] A. Antipov and S. Krasnova, "Using of sigmoid functions in the control system of the overhead crane," in *Proc. IEEE 16th Pyatnitskiy's Conf.*, 2022, pp. 1–4.

[41] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10778–10787.

[42] W. Liu et al., "Large-margin softmax loss for convolutional neural networks," in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 507–516.

[43] Y. Li, Y. Chen, N. Wang, and Z.-X. Zhang, "Scale-aware Trident networks for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6054–6063.

[44] 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.

[45] 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, 2020, Art. no. 2997.

[46] 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, 2019, Art. no. 765.

[47] D. Kumar and V. Kukreja, "Image-based wheat mosaic virus detection with Mask-RCNN model," in *Proc. IEEE Int. Conf. Decis. Aid Sci. Appl.*, 2022, pp. 178–182.

[48] G. Jocher, 2022. [Online]. Available: https://github.com//ultralytics/yolov5

[49] 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. Workshops*, 2021, pp. 2778–2788.

[50] C. Yang, Z. Huang, and N. Wang, "QueryDet: Cascaded sparse query for accelerating high-resolution small object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 13658–13667.

Ming Sun received the B.S. degree in information science and technology in 2021 from Northwest University, Xi'an, China, where he is currently working toward the M.S. degree in information and communication engineering.

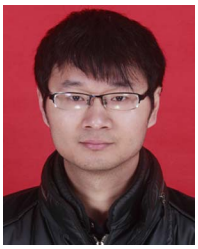
His research interests include radar signal processing and synthetic aperture radar image target detection.

Yanyan Li received the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 2011 and 2016, respectively.

She is currently an Associate Professor with the School of Information Science and Technology, Northwest University, Xi'an, China. Her research interests include ISAR imaging, target detection, and time-frequency analysis.

Xiaoxuan Chen received the Ph.D. degree in information and communication engineering from Xi'an Jiaotong University, Xi'an, China, in 2014.

She is currently an Associate Professor with the School of Information Science and Technology, Northwest University, Xi'an, China. Her research interests include image processing and deep learning.



Yan Zhou received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2015.

He is currently an Associate Professor with the School of Information Science and Technology, Northwest University, Xi'an, China. His research interests include space–time adaptive processing and array signal processing.

Jinping Niu received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2014.

She is currently an Associate Professor with the School of Information Science and Technology, Northwest University, Xi'an, China. Her research interests include signal processing and resource allocation for wireless communication systems.

Jianpeng Zhu received the B.S. degree in information science and technology in 2021 from Northwest University, Xi'an, China, where he is currently working the M.S. degree in electronics and communication engineering.

His research interests include video superresolution.