A Novel SAR Image Ship Small Targets Detection Method

Yu Song[1,2]*, Min Li[1], Xiaohua Qiu[1], Weidong Du[1], Yujie He[1], Xiaoxiang Qi[1]

[1]Xi'an Institute of High Technology, Xi'an, China
[2]School of Information and Communications, National University of Defense Technology, Xi'an, China
Email: *songyu888@yeah.net

Abstract

To satisfy practical requirements of high real-time accuracy and low computational complexity of synthetic aperture radar (SAR) image ship small target detection, this paper proposes a small ship target detection method based on the improved You Only Look Once Version 3 (YOLOv3). The main contributions of this study are threefold. First, the feature extraction network of the original YOLOV3 algorithm is replaced with the VGG16 network convolution layer. Second, general convolution is transformed into depthwise separable convolution, thereby reducing the computational cost of the algorithm. Third, a residual network structure is introduced into the feature extraction network to reuse the shallow target feature information, which enhances the detailed features of the target and ensures the improvement in accuracy of small target detection performance. To evaluate the performance of the proposed method, many experiments are conducted on public SAR image datasets. For ship targets with complex backgrounds and small ship targets in the SAR image, the effectiveness of the proposed algorithm is verified. Results show that the accuracy and recall rate improved by 5.31% and 2.77%, respectively, compared with the original YOLOV3. Furthermore, the proposed model not only significantly reduces the computational effort, but also improves the detection accuracy of ship small target.

Keywords

1. Introduction

Owing to the development of the aviation and satellite industries, the demand for synthetic aperture radar (SAR) image target detection has become more and more important, particularly SAR image small target detection.

Currently, deep learning techniques are widely used in mainstream target detection algorithms and have achieved desirable results. Target detection algorithms based on deep learning can be categorized into two types: two-stage and one-stage algorithms. Two-stage detection algorithms primarily include R-CNN target detection algorithms [1] [2] [3], whereas one-stage detection algorithms primarily include the YOLO algorithm [4] [5] [6], SSD [7], and Retina-Net [8]. After the extraction of image features, the two-stage detection method first inputs the feature information into a regional proposal network (RPN), which conducts the initial classification of targets. Subsequently, the part of the region with a high score is input to the final detection network for a more accurate detection. Because the calculation of the final detection comprises two stages, the real-time performance is poor. The one-stage detection algorithm does not require the RPN module as it uses the extracted feature information to directly conduct classification and position regression calculations for the target. Therefore, these algorithms are suitable for accomplishing real-time detection.

Commonly, studies regarding target detection using deep learning are increasingly performed in the field of remote sensing images, and the detection effect of some scenes in remote sensing images is remarkable. However, to extract more abundant target features, the network complexity is typically increased, and few studies have focused on small target detection in SAR images. You Only Look Once Version 3 (YOLOv3) are a real-time target detection algorithm that was initially trained and detected invisible images. By contrast, SAR images contain fewer features, such as color information and texture details; therefore, some of the convolutional layers in the feature extraction network will be redundant. In addition, ship targets in SAR images are primarily small targets or near-port targets, and missing detection or false alarm will occur during detection owing to the weak feature extraction capability of the network for SAR image targets.

To solve the problems above, a novel SAR image target detection algorithm based on the improved YOLOv3 is proposed herein. Because SAR images are characterized by less color information and texture details, the main contributions of this paper include the following:

First, the feature extraction network of the original YOLOv3 algorithm is optimized to adapt to the imaging characteristics of the SAR images.

Second, the conventional convolution operation in the feature extraction network is improved to a deep separable convolution to further reduce redundant parameters and computation.

Third, the residual structure is introduced into the feature extraction network, which enhances the training ability of the network, reuses the feature extracted from the previous layer, and retains the key details significantly such that it is more sensitive to small target detection.

Based on the improvements above, the precision of the SAR image target detection algorithm improved by approximately 5.31% compared with that of the original YOLOv3 algorithm, and the recall rate improved by approximately 2.77%.

2. Related Work

With the maturity of deep learning technology, an increasing number of innovative algorithms have been developed in remote sensing image target detection research.

2.1. Remote Sensing Image Target Detection

In Ref. [9], the fusion of both optical data and SAR data sets is performed applying two different approaches. The authors conclude that in refugee camp areas, the results of the independent analyses can be improved significantly by the proposed fusion approaches. In Ref. [10], the presented algorithm works comparatively well with images of the ocean in freezing temperatures and strong wind conditions, common in the Amundsen Sea. In Ref. [11], the continuous learning of a residual convolution neural network that is applicable to middle- and high-resolution optical remote sensing images was proposed; it demonstrated good recognition accuracy for airport targets in optical remote sensing images with complex backgrounds. In Ref. [12], an improved remote sensing image target detection method was proposed based on the faster R-CNN, which can yield better detection results owing to the fusion of multiscale features and features extracted by the convolutional neural network of the rotating regional network. An improved convolutional neural network-based method for SAR image target recognition was proposed in a previous study [13]. After a linear weighted combination of the classification results of the original image and its multi-resolution representation, the classification of the test samples was assessed based on the combined results, and the validation of the proposed method using the MSTAR dataset demonstrated the effectiveness and robustness of the proposed method. The FCD-EMD [14] algorithm combines detailed information in different directions such that the results yielded are more accurate compared with those yielded by individual methods. Furthermore, it can reduce the effect of speckle noise in SAR images via feature selection.

2.2. SAR Image Ship Target Detection

In Ref. [15], an improved convolutional neural network-based SAR image ship target detection algorithm was proposed to detect multiscale ship targets in multiple scenes; the algorithm indicated good adaptability to the detection of ship targets of different sizes in complex scenes. In Ref. [16], a ship detection algorithm based on a depth feature pyramid and a cascade detector was proposed. The feature extraction network of the original target detection algorithm improved, the cascade structure was used to adjust the network, and finally, good detection results were obtained. In Ref. [17], an improved detection method was proposed based on a regional full convolution network, which can suppress the effect of speckle noise, effectively extract the features of ships, and yield a good detection effect. In Ref. [18] a neural network with hybrid algorithm of CNN and multilayer perceptron (CNN–MLP) is suggested for image classification. In

this proposal, the algorithm is trained with real SAR images from Sentinel-1 and RADARSAT-2 satellites, and has a better performance on object classification than state of the art. In Ref. [19], the authors propose a modified topology, utilizing superpixels (SPs) in lieu of rectangular sliding windows to define CFAR guardbands and background. The aim is to achieve better target exclusion from the background band and reduced false detections.

Many advances have been made in previous works, but few studies have focused on the lightweight processing of small target detection algorithms. This paper presents a lightweight neural network for small target detection. The propored method can not only improve the detection accuracy, but also reduce the computational complexity.

3. Methods

Joseph Redmon proposed three versions of YOLO, including YOLOv1, YOLOv2 and YOLOv3. YOLOv3 is the last one proposed by the author, which is also a mainstream target detection algorithm at present. The algorithm firstly extracts the features of different scales through the feature extraction network, then completes classification and regression about location using the extracted features. These features are divided into three scales: large scale, medium scale and small scale on classification and position regression calculation, which can well adapt to the detection characteristics of multi-scale targets.

YOLOv3 was originally used for target detection on visible images, and it needs to make some adaptive improvement in its network when it is used for object detection in SAR images. Visible images have diverse colors, rich textures and clear contours. However, the color of a SAR image is relatively poor, without much color and texture information, and even part of the target contour and background boundary may be unclear. Therefore, SAR image feature information is relatively less, and the feature extraction network should also make corresponding changes. In order to reduce the amount of parameter calculation in feature extraction network, the model proposed in this paper adopts separable convolution operation, which can greatly reduce the amount of computation while ensuring the detection performance of the network. For improving the detection accuracy of small targets, the idea of residual network structure is introduced in this paper. In the feature extraction network, the feature information after each sub-sampling is mapped to the next sub-sampling layer in the way of skip connection. In this way, the feature information of the shallow layer is reused and the feature extraction ability of the network is improved.

3.1. Optimizing Feature Extraction Network

In order to be more suitable for SAR image feature extraction, the feature extraction network structure proposed in this paper is shown in **Figure 1**, in which there are 13 convolutional layers and 5 downsampling layers. The 13 convolutional layers are divided into five blocks by the downsampling layer. Each part

contains several convolutional layers with different scales and different number of convolution kernels. The specific scale and number of convolution kernels are shown in **Figure 1**. The feature extraction network used by YOLOV3 is dar-net53. The network has a total of 52 convolution layers, 5 sampling under the convolution operation and 23 residual network structures, namely the skip connections. For the sake of extracting the image features more rich information, the characteristic of each layer channel more and more, up to 1024 channels, which is result to learn the number of arguments. The network structure proposed in this paper has fewer convolutional layers, and each layer has fewer characteristic channels, with at most 512 channels.

3.2. Reducing Computational Complexity

In the detection of SAR image targets, the computation amount of model parameters can be further reduced on the basis of the original algorithm, which will further improve the real-time performance of the algorithm. In addition, the lightweight processing of parameters is currently an important direction of neural network research. In view of the large networks, the researchers hope to further reduce the computational burden of the model without changing the effect of feature extraction. The current mainstream method is Separable Convolution [20]. This convolution operation is divided into two steps: Depthwise Convolution and Pointwise Convolution. To illustrate the problem, we use a simple convolution operation:

Assuming that A is a matrix of $m \times m$ and B is a matrix of $p \times p$, then B can be represented as Formula (1):

$$B_{p \times p} = M_{p \times 1} \times N_{1 \times p} \tag{1}$$

where, $M_{p \times 1}$ is a matrix of $p \times 1$ and $N_{1 \times p}$ is a matrix of $1 \times p$, then the convolution of the matrices A and B can be represented as Formula (2):

$$A_{m \times m} * B_{p \times p} = \left(A_{m \times m} * M_{p \times 1} \right) * N_{1 \times p} \tag{2}$$

where, $*$ is the convolution operation. The above equation can be generalized to a tensor convolution operation. Assuming that $A_{k \times k \times m}$ is a tensor of $k \times k \times m$ and $B_{p \times p \times n \times s}$ is a tensor of $p \times p \times n \times s$, then the convolution of the tensor A and B can be represented as Formula (3):

$$A_{k \times k \times m} * B_{p \times p \times n \times s} = \left(A_{k \times k \times m} * M_{p \times p \times 1 \times 1} \right) * N_{1 \times 1 \times n \times s} \tag{3}$$

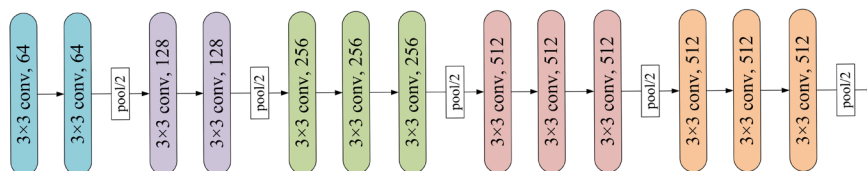Separable convolution operation is shown in **Figure 2**.

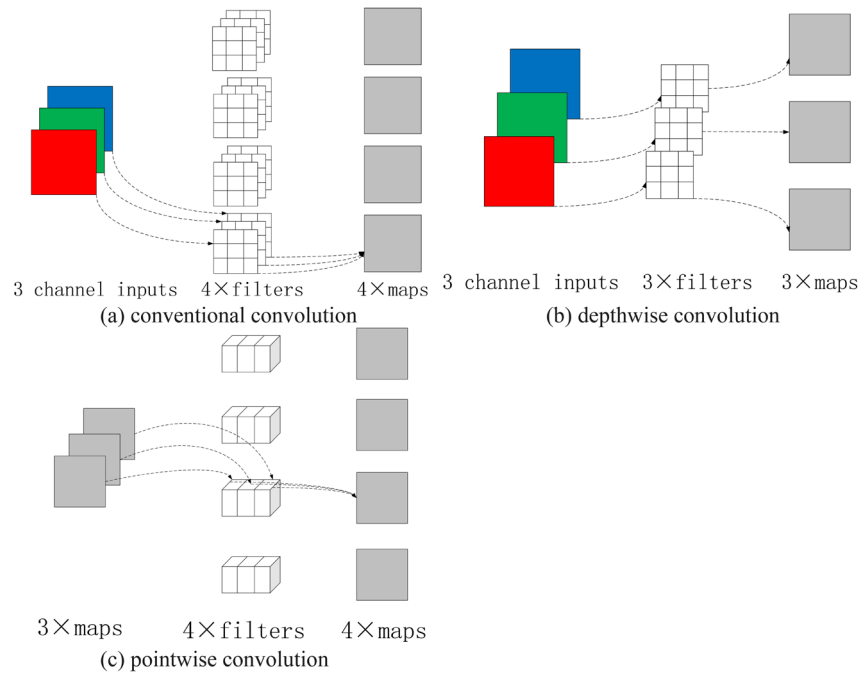**Figure 1.** Convolutional layer part of VGG16.

Figure 2. Schematic of conventional convolution and separable convolution.

As shown in **Figure 2(a)**, it is a conventional convolution operation. For the input of three channels, there are four convolution kernels, and finally four feature maps corresponding to the convolution check will be outputted. Separable operations are divided into two steps: First, Depthwise Convolution (showed in **Figure 2(b)**) and Pointwise convolution (showed in **Figure 2(c)**) are performed. During Depthwise Convolution, Corresponding to three inputs, there will be three convolution kernels perform convolution, then three feature maps will be obtained. Second, perform Pointwise Convolution operations on them. Four convolution kernels (1 × 1 × 3) are used for the feature maps of the three input channels, and four feature maps are generated According to the above, the first step of Separable Convolution is to ensure that the number of channels in the convolution kernel is equal to the number of input channels. The second step is to ensure that the convolution kernel with the size of 1 × 1 is used and the number of channels is equal to the number of pre-set output channels. Although the conventional convolution operation is divided into two steps, the computational amount is greatly reduced and better results are obtained at the same time. It can be seen from **Figure 2** that a general convolution operation is changed to Depthwise convolution and Pointwise Convolution, whose computational complexity is greatly reduced. The time complexity of each convolution layer in the convolutional neural network can be shown as Formula (4) [21].

$$\text{Time} \sim O\left(M^2 \times K^2 \times C_{in} \times C_{out}\right) \tag{4}$$

where, M represents the size of the output feature map, K represents the size of the convolution kernel, C_{in} represents the number of input channels, and C_{out} represents the number of output channels. Taking the first convolution layer of

VGG-16 network as an example, its algorithm complexity is $4162 \times 32 \times 3 \times 64$ = 299,040,768, and the improved algorithm complexity is $4162 \times 32 \times 3 \times 3 + 4162 \times 1 \times 3 \times 64 = 47,244,288$. It can be seen that the computational complexity of convolutional layer is greatly reduced, only 1/8-1/9 of that before. In this way, the real-time performance of the whole network is enhanced, while the performance of feature extraction is not affected.

3.3. Reusing the Feature Information of Shallow Layer

Usually, the shallow layer features of neural network mainly contain the detail information in the image, while the deep layer features mainly include the semantic information. The deep semantic information is easy to lose the feature information of the small target. Moreover, as the number of network layer increases, the training accuracy tends to be saturated, and then falls into Network Degradation. When using a large number of samples to train the deep neural network, the learning mechanism of the network, Chain Rule, is easy to lead to the gradient gradually approaching zero, namely Gradient Vanishment. Assume that the output of each layer of the network is $f(x_i)$, where i is the i_{th} layer, represents the input of the i layer, namely the output of the $(i-1)$ layer. f is the activation function as shown in Formula (5):

$$f_{i+1} = f\left(f_i * w_{i+1} + b_{i+1}\right) \tag{5}$$

where, w_{i+1} denotes the weight of the $(i+1)$ layer and b_{i+1} denotes the bias of the $(i+1)$ layer.

For a depth network with n layers, its final output is shown in Formula (6):

$$\Delta w_i = \frac{\partial Loss}{\partial w_i} = \frac{\partial Loss}{\partial f_n} \frac{\partial f_n}{\partial f_{n-1}} \cdots \frac{\partial f_{i+1}}{\partial f_i} \frac{\partial f_i}{\partial w_i} \tag{6}$$

Taking the derivative of the activation function, if this part is less than 1, then with the increase of layers, the gradient update information obtained will decay in Gradient Vanishment.

In order to enhance the feature information of small targets, this paper uses skip connection to form residual network structure [22], which not only effectively prevents network degradation, but also enhances the detail information of target features. By using identity mapping, the feature information of the shallow layer is directly input to the deeper convolution layer, which preserves more target details and helps to improve the detection accuracy of small targets. The structure of skip connection is shown in **Figure 3**.

Taking **Figure 3** for example, in the forward propagation of neural network, $a^{[l]}$ represents the output of layer l, while in general neural network, it needs to pass through Layer $l + 1$ to reach Layer $l + 2$. In the residual block, it is not only necessary to pass through $l + 1$ layer, but also to skip connection of the output $a^{[l]}$ to $l + 2$ layer, as shown in Formula (7):

$$f^{[l+2]} = g\left(z^{[l+2]} + a^{[l]}\right) \tag{7}$$

The output of the $f^{[l+2]}$ activation unit is related not only to the z of the $l + 2$

layer, but also to the *a* of the *l* layer. In this case, the derivative object has an additional identity mapping term. For example, the derivative of the residual node is shown in Formula (8):

$$\frac{\mathrm{d}h}{\mathrm{d}x} = \frac{\mathrm{d}(f+x)}{\mathrm{d}x} = 1 + \frac{\mathrm{d}f}{\mathrm{d}x} \tag{8}$$

It can be seen that even if the original derivative $\frac{\mathrm{d}f}{\mathrm{d}x}$ approaches 0, it can effectively back propagate, greatly reducing the impact brought by Gradient Vanishment. From the perspective of front-propagation, as the number of network layer increases, the image information contained in the feature graph will be less and less. Skip Connection introduces the features of the lower layer, ensuring that the features of the higher layer will contain more detail information of targets.

Figure 4 shows the network model of the improved method, with the emphasis on the improved part in the feature extraction network. There are four skip connections and five downsampling layers. The residual network structure is adopted to prevent the gradient vanishment of the network as well as to reuse the shallow feature information of the target to enhance the feature of small targets, when the input SAR image first goes through the feature extraction network. Separable Convolution operation is used, which greatly reduces the computation and maintains the detection effect of the network. After these improvements, the experiment in the third part of this paper verifies that compared with the original YOLOv3, the proposed method improves the accuracy of SAR image small target detection.
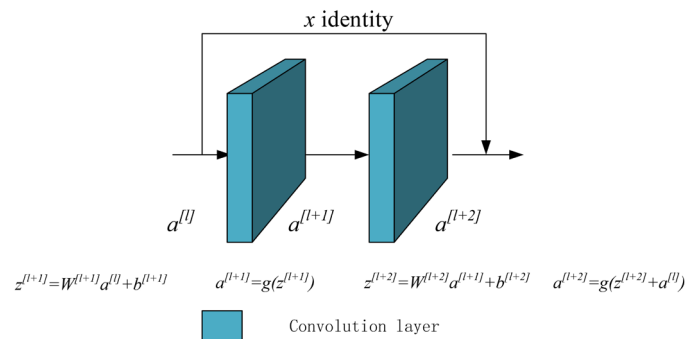

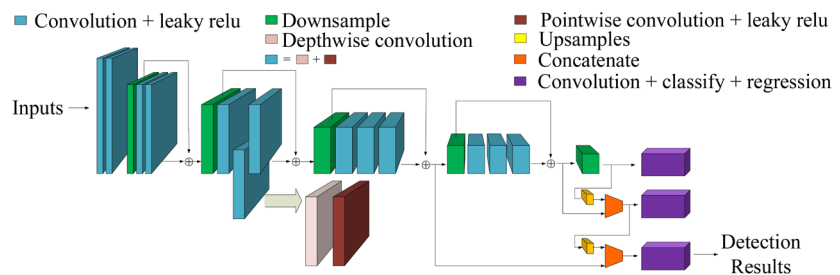**Figure 3.** Schematic diagram of residual block structure.

Figure 4. The proposed model.

4. Experiment and Evaluation

4.1. Experimental Environment

The experiment in this paper runs on the Ubuntu 16.04 operating system, the code runs on python3.6, and the model training runs on the Titan Xp (12G video memory) GPU, CUDA 10.0 and cuDNN 7.0 configuration environment. SSDD [23] is the classic publicly available data set specially used for SAR image ship target detection. It can be used for training and test detection algorithm and has been used by more than 30 universities and research institutes. For each ship target, detection algorithm predicts the boundary of ship target and gives the confidence degree of ship target. The number of iterations was 200, the learning rate was set to 0.001, and the momentum was set to 0.9. SGD(Stochastic Gradient Descent) was used in the optimization algorithm, and the momentum attenuation coefficient was 0.00004.

4.2. Experimental Evaluation Criteria

In this paper, *Precision*, *Recall* and *F*1, a criterion that comprehensively measures the accuracy and recall rate, is used to conduct quantitative analysis on the detection results. The accuracy rate and recall rate are shown in Formula (9) and Formula (10):

$$\text{Precision} = \frac{N_{tp}}{N_{tp} + N_{fp}} \tag{9}$$

$$\text{Recall} = \frac{N_{tp}}{N_{tp} + N_{fn}} \tag{10}$$

where N_{tp} represents the number of positive samples detected as positive samples, N_{fp} represents the number of negative samples detected as positive samples, and N_{fn} represents the number of positive samples detected as negative samples. Precision reflects false alarms in detection, that is, the higher the precision rate, the less false alarms. Recall represents the phenomenon of missing detection in detection, that is, the higher the recall rate is, the fewer missed targets will be. The definition of F1 is shown in Formula (11):

$$\text{F1} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{11}$$

where F1 is an indicator used to comprehensively measure the accuracy rate and recall rate. The higher the indicator is, the better the detection effect will be.

4.3. The Experimental Results and Analysis

SSDD data set was used to train the original YOLOv3 network, and part of the test set was used to test and evaluate the trained model. The experiment is divided into six cases, as shown in Table 1. To be faired, the training and test data set used for the original YOLOv3 algorithm and the algorithm in this paper were consistent.

Table 1. SAR image ship target detection.

SAR image ship target detection	Detection in Complex Background	Nearshore targets detection
		Near islands targets detection
		Docked targets detection
	Small target detection	Target detection in sea clutter background
		Multiple small targets detection
		Dense small targets detection

4.3.1. Detection Results in Complex Background

There are many ships in ports, docks and inlets, etc., so the accuracy of detection in this area is required to be higher. The method in this paper obtains relatively high accuracy and recall rate for detection of ship targets against a complex background. Some detection results are shown in Figure 5.

It can be seen from the figure above that these scenes are SAR images in a typical coastal or port. Column a is the original image, column b is the detection result obtained by using the original YOLOv3 algorithm, and column c is the detection result obtained by the model in this paper. Line (1-2) is the complex scene of the ship target nearshore, line (3) is the scene of the ship target near islands, and line (4) is the scene of the docked ship target. In the coastal complex scenes and near shore scenes from lines (1) to (3), the original YOLOv3 algorithm may misdetect some areas of the coast or port as ships in the detection, while the algorithm in this paper can accurately detect the ship targets. In line (4), the original YOLOv3 algorithm missed detection in the detection. The algorithm in this paper can detect the ship target well. Especially in line (4), the ship target size in the port is obviously inconsistent. Compared with the original YOLOv3 algorithm, this algorithm can detect the ship target with inconsistent size better.

4.3.2. Detection Results of Ships Small Target

In real applications, there are often small ships on the sea surface, or a large number of densely packed small ships targets. In these cases, the detection algorithm needs to have good sensitivity to small-scale targets and be able to detect targets accurately. The method presented in this paper has a good performance in the detection accuracy and recall rate of small ship targets, as shown in Figure 6.

Figure 6 shows the detection of small ship targets and dense small targets. Column *a* is the test image, column b is the detection result obtained by using the original YOLOv3 model, and column c is the detection result obtained by the proposed model. Line (1) are results of small target detection in sea clutter background Line (2) and Line (3) are simple results of multiple small targets detection, and Line (3) is a case of large background noise. Lines (4) and (5) are typical dense small targets. It can be seen that the original YOLOv3 algorithm is prone to missing some small targets in the detection, while the proposed method

shows stronger detection performance in the detection, almost all of which can achieve correct detection of small targets.

First, the complexity of the two algorithms is compared. The Time Complexity of the convolutional neural network is shown in Formula (12) [21]:

$$\text{Time} \sim O\left(\sum_{l=1}^{D} M_l^2 \times K_l^2 \times C_{l-1} \times C_l\right) \tag{12}$$

According to the network structure, Darknet53 in original YOLOv3 contains 52 ordinary convolutional layers with an increasing number of feature channels in each layer, up to 1024 channels at most. The VGG-16 network in the proposed method contains 26 Separable Convolutional layers, with a small number of characteristic channels in each layer, at most 512 channels. The Time Complexity of the two is compared as shown in Formula (13) [21]:

$$\text{Time}\left(\text{darknet53}\right) \sim O\left(\sum_{l=1}^{52} M_l^2 \times K_l^2 \times C_{l-1} \times C_l\right)$$

$$> \text{Time}\left(\text{VGG-16}\right) \sim O\left(\sum_{l=1}^{26} M_l^2 \times K_l^2 \times C_{l-1} \times C_l\right) \tag{13}$$

a(1)The original image	b(1)YOLOv3 detection results	c(1)Our detection results
a(2)The original image	b(2)YOLOv3 detection results	c(2)Our detection results
a(3)The original image	b(3)YOLOv3 detection results	c(3)Our detection results
a(4)The original image	b(4)YOLOv3 detection results	c(4)Our detection results

Figure 5. Comparison of ship target detection results under complex background.

a(1) The original image   b(1) YOLOv3 detection results   c(1) Our detection results

a(2) The original image   b(2) YOLOv3 detection results   c(2) Our detection results

a(3) The original image   b(3) YOLOv3 detection results   c(3) Our detection results

a(4) The original image   b(4) YOLOv3 detection results   c(4) Our detection results

a(5) The original image   b(5) YOLOv3 detection results   c(5) Our detection results
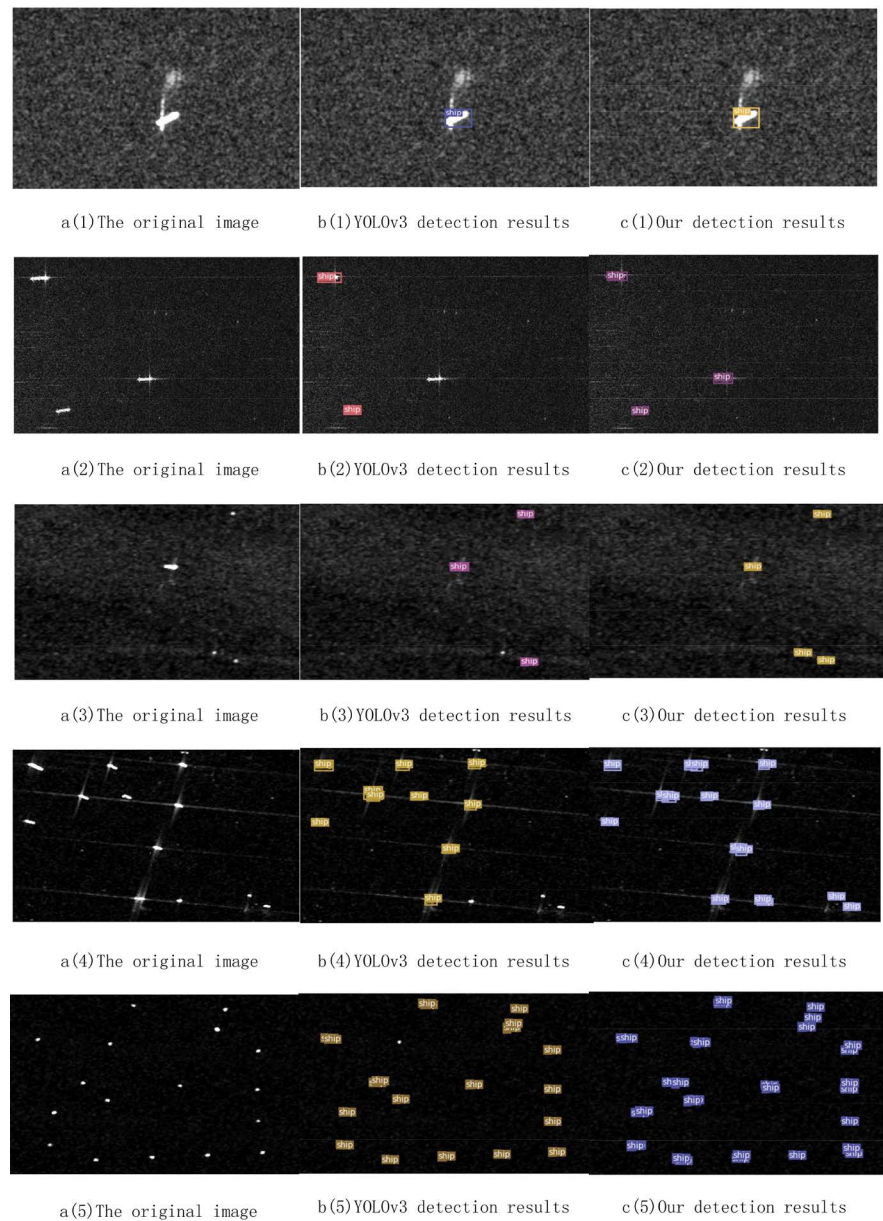
Figure 6. Comparison of small target detection results.

From the above analysis, the Time Complexity of the proposed method is far less than that of the original YOLOv3, which enhances the real-time performance of detection.

For the sake of fairness, the original YOLOv3 algorithm and the proposed method are trained with the same training set. The resulting model is verified with the same test set. The test results obtained on the test set are shown in Table 2.

It can be seen from Table 2 that the method proposed in this paper is higher than the other two algorithms. Table 2 shows Precision is 5.31% higher than the original algorithm, Recall rate is 2.77% higher than that of the original algorithm, and F1 is about 4.24% higher than that of the original algorithm. Compared with

Table 2. Comparison of results about the proposed algorithm between the other algorithms.

algorithm	criterion		
	Precision	Recall	F1
YOLOv3	72.53%	85.36%	78.42%
Literature [24] algorithm	76.70%	66.00%	64.60%
The proposed algorithm	**77.84%**	**88.13%**	**82.66%**

the other algorithms, the proposed algorithm is more suitable for detecting ship targets in SAR images. In this paper, six groups of experiments are carried out for different scenarios, including nearshore, near island, dock, single small target, multiple small targets and dense small target in sea clutter background.

5. Conclusion

This paper proposes an improved YOLOV3 algorithm for SAR image target detection, which not only reduces the algorithm complexity, but also improves the accuracy and recall rate of SAR image target detection. Our key idea is to use the convolutional layer part of VGG16 network as the feature extraction network, and convert conventional convolution operations to Separable Convolution. Then we introduce the skip connection in network. After the above improvement of feature extraction network, SSDD data set was used to train the neural network of the algorithm in this paper, and test set was used to verify the trained model. The detection effect and experimental results obtained were better than the original YOLOv3 algorithm. In the future work, further research will be made on training strategy and network structure optimization, which will be one of the emphases in the following work.

Funding

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/CVPR.2014.81

[2] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision*, Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[3] Ren, S.Q., He, K.M., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[4] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[5] Redmon, J. and Farhadi, A. (2017) YOLO9000: Better, Faster, Stronger. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 21-26 July 2017, 6517-6525. https://doi.org/10.1109/CVPR.2017.690

[6] Redmon, J. and Farhadi, A (2018) You Only Look Once: An Incremental Improvement. arXiv e-prints.
https://arxiv.org/pdf/1804.02767.pdf

[7] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., *et al.* (2016) SSD: Single Shot Multi Box Detector. 2016 *European Conference on Computer Vision*, Amsterdam, 8-16 October 2016, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2

[8] Lin, T.-Y., Goyal, P., Girshick, R., He, K.M. and Dollar, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2999-3007. https://doi.org/10.1109/ICCV.2017.324

[9] Sprohnle, K., Fuchs, E.-M. and Aravena Pelizari, P. (2017) Object-Based Analysis and Fusion of Optical and SAR Satellite Data for Dwelling Detection in Refugee Camps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **10**, 1780-1791. https://doi.org/10.1109/JSTARS.2017.2664982

[10] Mazur, A.K., Wåhlin, A.K. and Krężel, A. (2017) An Object-Based SAR Image Iceberg Detection Algorithm Applied to the Amundsen Sea. *Remote Sensing of Environment*, **189**, 67-83. https://doi.org/10.1016/j.rse.2016.11.013

[11] Li, Z.Q., Zhu, D.F., Ma, J.Y., Meng, X.Y., Wang, D. and Liu, S.Y. (2020) Airport Detection Method Combined with Continuous Learning of Residual-Based Network on Remote Sensing Image. *Acta Optica Sinica*, **40**, Article ID: 1628005.

[12] Dai, Y., Yi, B.S., Xiao, J.S., Lei, J.F., Tong, L. and Cheng, Z.Q. (2020) Object Detection of Remote Sensing Image Based on Improved Rotation Region Proposal Network. *Acta Optica Sinica*, **40**, Article ID: 0111020.

[13] Qiao, L.C. (2020) SAR Image Target Recognition Method Combining Multi-Resolution Representation and Complex Domain CNN. *Laser & Optoelectronics Progress*, **57**, Article ID: 241007.

[14] Huang, S.Q., Liu, Z.G., Liu, Z. and Wang, L. (2017) SAR Image Change Detection Algorithm Based on Different Empirical Mode Decomposition. *Journal of Computer and Communication*, **5**, 9-20. https://doi.org/10.4236/jcc.2017.58002

[15] Yang, L., Su, J., Huan, H. and Li, X. (2020) SAR Ship Detection Based on Convolutional Neural Network with Deep Multiscale Feature Fusion. *Acta Optica Sinica*, **40**, Article ID: 0215002. https://doi.org/10.3788/AOS202040.0215002

[16] Zhao, Y.F., Zhang, B.H., Zhang, Y.Y., Gu, Y., Wang, Y.M., Li, J.J., *et al.* (2020) Ship Detection Based on SAR Images Using Deep Feature Pyramid and Cascade Detector. *Laser & Optoelectronics Progress*, **57**, Article ID: 121019.

[17] Wang, J.L., Lv, X.Q., Zhang, M. and Li, J. (2019) Remote Sensing Image Ship Detection Based on Improved R-FCN. *Laser & Optoelectronics Progress*, **56**, Article ID: 162803. https://doi.org/10.3788/LOP56.162803

[18] Sharifzadeh, F., Akbarizadeh, G. and Seifi Kavian, Y. (2019) Ship Classification in SAR Images Using a New Hybrid CNN-MLP Classifier. *Journal of the Indian Society of Remote Sensing*, **47**, 551-562. https://doi.org/10.1007/s12524-018-0891-y

[19] Pappas, O., Achim, A. and Bull, D. (2018) Superpixel-Level CFAR Detectors for Ship Detection in SAR Imagery. *IEEE Geoscience and Remote Sensing Letters*, **15**, 1397-1401. https://doi.org/10.1109/LGRS.2018.2838263

[20] Wang, C.-F. (2018) A Basic Introduction to Separable Convolutions. https://towardsdatascience.com/a-basic-introduction-to-separable-convolutions-b99ec3102728

[21] He, K.M. and Sun, J. (2014) Convolutional Neural Networks at Constrained Time. 2015 *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 5353-5360. https://doi.org/10.1109/CVPR.2015.7299173

[22] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/CVPR.2016.90

[23] Li, J.W., Qu, C.W., Peng, S.J. and Deng, B. (2018) Ship Detection in SAR Images Based on Convolutional Neural Network. *Systems Engineering and Electronics*, **40**, 1953-1959.

[24] Hu, C.H., Chen, C., He, C., Pei, H. and Zhang, J.X. (2019) SAR Detection for Small Target Ship Based on Deep Convolutional Neural Network. *Journal of Chinese Inertial Technology*, **27**, 397-405.