Oriented Ship Target Detection Based on Improved CenterNet in Synthetic Aperture Radar Images

Xinqiao Jiang1, Hongtu Xie1,\*, Liang Kang1, Kaipeng Chen1, Guoqian Wang2

*1School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510006, China*

*2Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510120, China*

*\*Hongtu Xie is the corresponding author. xiehongtu@mail.sysu.edu.cn*

*Abstract*—Deep learning has a wide application prospects in the field of the ship target detection in synthetic aperture radar (SAR) images. The existing researches mainly use anchor-based target detection method, but this kind of methods is not suitable for the ship SAR image with the sparse ship targets. It requires additional computational resources to filter out a large number of overlapped candidate prediction boxes, which tends to result in the inaccurate target position and low detection efficiency. At the same time, most of existing methods use the horizontal box to detect the ship targets, which are not suitable for detecting the large aspect ratio ship targets and densely arranged ship targets. Aiming at the problem of the insufficient data, a data augmentation strategy for the ship SAR images is proposed, which effectively alleviates the problem of over-fitting during training, improves the detection performance and stability of the model. Then, an oriented ship target detection method is presented for the SAR images, which extends the CenterNet to the field of the rotating target detection. Compared with the existing ship detection method of the SAR images based on the deep learning, this method is very easy to implement, has the high detection precision and positioning accuracy, which has been conducive to the ship detection in practice.

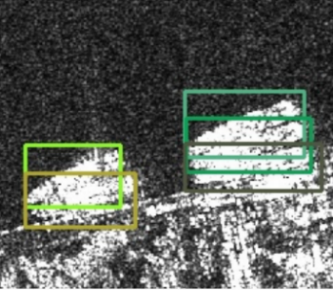
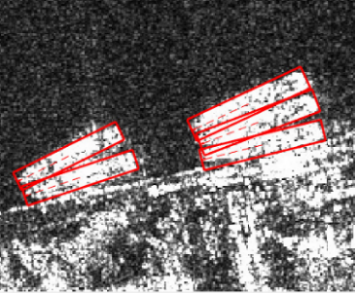
Keywords—Data augmentation, deep learning, improved CenterNet, oriented ship target detection, synthetic aperture radar images

# Introduction

In recent years, the research and technology development of the ship target detection in the synthetic aperture radar (SAR) images have attracted the great attention in the field of the marine remote sensing. However, the features required by the traditional ship detection methods in SAR images depend on the artificial design, which are susceptible to the complex background interference. Besides, these methods have some drawbacks such as the low detection accuracy, low detection efficiency and weak generalization capability [1]. Compared with the traditional methods, the methods based on the deep learning have the advantages of the strong anti-interference ability and high detection accuracy.

Deep learning technology has made the great progress in the field of the ship target detection in the SAR images. The Faster R-CNN is improved to detect the ship target in the SAR images by using strategies such as the feature fusion and migration training [2]. A ship detection model of SAR images based on the improved YOLOv3 is proposed by redesigning the network structure of the bottom residual unit and feature pyramid [3]. In [4], a ship detection model based on the RetinaNet has been presented. In [5], a DRBox-v2 method is presented to detect the ship target in any direction by using the feature pyramid and hard negative mining technique.

However, the existing ship target detection methods based on the deep learning in SAR images usually use the anchor- based strategy with the horizontal bounding boxes to detect the ship targets, which still has some problems. (1) The ship target usually has a large aspect ratio. For the ship target with a certain inclination angle, the horizontal bounding box will surround a large background area. Thus, compared with the horizontal bounding box, the rotating bounding box fits the ship target better, as show in Fig.1.(2) If multiple ship targets are densely arranged, the overlap between adjacent horizontal detection boxes is large (in Fig.1(a)), and the post-processing method prone to result in removing the correct detection boxes [6]. (3) The method based on dense anchor boxes is not suitable for the SAR images with sparse ship targets, which leads to extremely unbalanced positive and negative samples. It needs a lot of computing resources for the post-processing to remove the overlapping candidate prediction boxes.

(a) (b)

1. (a) Dense ship target detection result using horizontal bounding boxes; (b) Dense ship target detection result using rotating bounding boxes.

Based on the existing researches, an oriented SAR ship target detection method based on the improved CenterNet [7] in the SAR images is proposed in this paper. This method uses the rotating bounding box to locate the SAR ship targets, without needing to generate dense anchor boxes. As a result, this method can greatly reduce the computational complexity of the model and enable the more accurate determination of the position of the ship targets.

This work was co-supported by the National Natural Science Foundation of China under Grant 61801510, by Guangdong Basic and Applied Basic Research Foundation under Grants 2021A1515010768 and 2214050002344, and by the National Science Foundation of Hubei Province under Grant 2019CFB263.

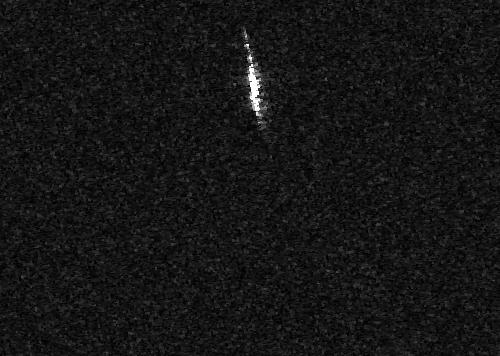
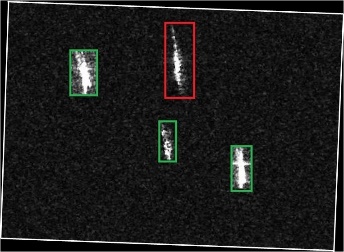
# Methodology

## Data augmentation strategy

Ship target detection method based on the deep learning requires a lot of data, but acquiring the SAR image is difficult. The insufficient data is one of the main shortcomings of the ship target detection in the SAR images. Therefore, this paper presents a data augmentation strategy for the SAR images to expand the data volume of the SAR Ship Detection Dataset Plus (SSDD+) [2].

We take these images as the simple scene images that have a small number of the ship targets and do not contain the terrestrial backgrounds. Then, the simple scene images are selected and divided into three subsets (large, medium and small), according to the size of the targets. The reason for dividing dataset by target size is that the ship target sizes in a SAR image do not differ much, making the augmented images very similar to the real images. Inspired by CutMix data augmentation method [8], in each subset, we cut out the ship targets in the SAR images to generate the ship target slice set. Then, for each picture in each subset, *n* target slices are randomly selected from the slice set and then embedded into the image with a randomly selected position. For the images in the large, middle, small target image subset, *n* is equal to 1, 2 and 3, respectively.

In addition, the random small angle rotation augmentation method is applied to the train set images, that is, each image rotates at any angle within 5 degrees clockwise. This method can increase the diversity of the ship target angles without excessively changing the size of the image to prevent the image distortion. The image after the data augmentation is shown in Fig.2, the target in the red box is originally in this image, while the target in the green box is cut from other images and mixed into this image. If the slice background is similar to the image background, it is difficult to find that the target is mixed from other images. This paper applies these two data augmentation methods above to the SSDD+ dataset and generates the local image files and label files. The amount of the data in the training set is tripled.

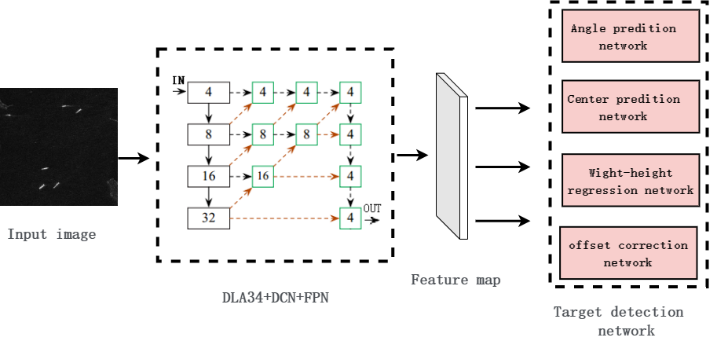
 

(a) (b)

1. (a) The original SAR image; (b) Image after the data augmentation

## Model architecture

In this paper, the CenterNet is extended to the field of the rotating target detection. The model structure is shown in Fig. 3, which includes two parts: the backbone network and target detection network. The Deep Layer Aggregation network (DLA34) [9] is used as the basic backbone network of the model, fusing the deformable convolution and feature pyramid strategy into it, which is same as the backbone of the CenterNet [7]. We first converts the input image size into 512×384, then the backbone network extracts the features in the image and outputs a feature map. At the top of the model is the target detection network, including four sub networks: the center prediction network, offset correction network, width-height regression network and angle prediction network. The structure of the 4 sub networks is the same, which is composed of two convolution layers and one ReLU activation layer.



1. Model structure.

The center prediction network outputs a two-dimensional matrix which called the heatmap to locate the targets. In the ground truth of the heatmap, the position where the value is one indicates that there is a target at the point. At the same time, the values of the points within the radius *r* around this position on the two-dimensional matrix are distributed in the form of the Gaussian function, which is given by

 (1)

where, x and y are the horizontal and vertical coordinates, and  is the standard deviation. The values of the remaining position in the matrix are zeros.

In the inference stage, if the value of a point in the output heatmap is greater than that of the eight surrounding points, then the point is a candidate center point, and the value at the point is taken as the confidence of the target. The candidate center point whose confidence reaches the threshold is taken as the center point of the prediction target.

The target location in the heatmap predicted by the center prediction network should be mapped to the original image coordinates. Considering the discreteness of the coordinates, mapping the center points of the feature map to the original image will bring errors, because the down sampling rate of the model is 1/4. Thus, an additional offset correction network is needed to predict a deviation for each center point to correct the coordinates of the target center point. The offset correction network also outputs a two-dimensional matrix, where we select the value of the target center coordinate determined by center prediction network as the offset for the corresponding target. Similarly, the value of the target center coordinate in the two-dimensional matrix output by width-height regression network and angle prediction network are taken as the width, height and angle of the corresponding target, respectively. By training, these parameters will approach the ground truth.

At present, most of the relevant SAR ship target detection model use the anchor-based strategy, which needs to generate the dense anchors, and the subsequent cumbersome post-process toward the overlapped prediction boxes. However, the computational cost of these models is large. Besides, the anchors with the different sizes and angles are needed in the advance. In addition, for the anchor-based models, the angle regression of the anchors suffers the discontinuous boundaries problem [10].The structure and processing flow of the model proposed in this paper are simple, the amount of parameters is small, and there is no need to generate the anchors and implement the subsequent post-processing.

## Rotating rectangle definitions

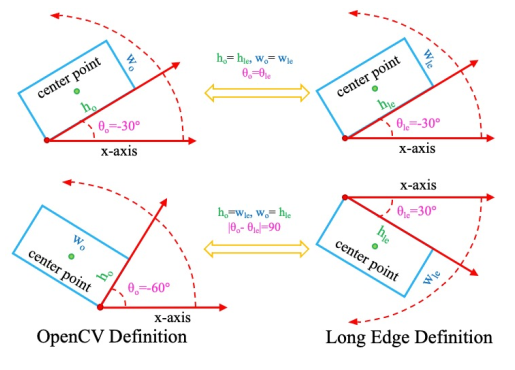
Compared with the ship target detection method using the horizontal bounding box, the key of the rotating bounding box target detection is the target angle prediction. Above all, the angle of the ship target in the two-dimensional SAR image must be clearly defined.

Angle of a rotating rectangle defined in OpenCV4.4.0.46 is the angle between the height and the positive direction of the x-axis. And, the height defined in the OpenCV4.4.0.46 is the line connecting the lowest and rightmost of the four vertices of the rotating rectangle. Long edge definition defines the angle between the long edge of the rotating rectangle and the positive direction of the x-axis as the angle of the rotating rectangle [10]. The two rotating rectangle definitions method are shown in Fig.4. The conversion method between the two rotating rectangle definitions is shown in (2) and (3) [10]. This paper uses the rotating rectangle definition method in the OpenCV 4.4.0.46, which is given by

 (2)

 (3)

where, , **,  and  represent the rotating rectangle, height, width and angle, respectively



1. Two angle definition method.

## Loss function

The model includes four sub networks, which play the roles of the classification, offset correction, width-height regression and angle prediction, respectively. Therefore, the different loss is designed for each task. Suppose the SAR image contains *N* targets, and the corresponding center points are , ,…. The classification loss () of the target center point adopts the focal loss, which is given by

 (4)

where,  is the value of coordinates (x, y) in the predicted heatmap,  is the value of coordinates (x, y) in the ground truth,  and  are the constants, whose values are 2 and 4, respectively, in the experiments. Offset loss () is added to correct the position of the target center points, which adopts the Smooth L1 loss, as shown in (5). The loss function of the wide-height regression network () and angle prediction network () also adopts the Smooth L1 loss.

 (5)

The loss function of the whole model is equal to the weighted sum of the loss functions of the above tasks, which is

 (6)

where, , ,  and  are equal to 1, 0.1, 0.3 and 0.5 , respectively, in the experiments of this paper.

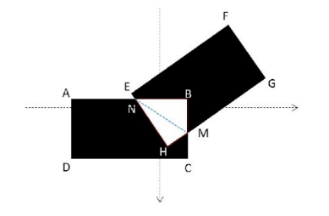
# Experiments Results

## Experimental settings

This paper uses SSDD+ dataset to implement experiments. This dataset contains a total of 1160 SAR images and 2456 ship targets. Given the small amount of data, no validation set is set in the experiments. The training sets and test sets account for 70 percent and 30 percent of the whole dataset, respectively. The images of the SSDD+ dataset come from the open SAR images on the Internet with the resolution from 1m to 15 m. The SAR image data is mainly collected by the RadarSat-2, TerraSAR-X, and Sentinel-1 sensors, which contains four polarization modes: HH, HV, VV and VH. The scene of the image includes complex Nearshore Area and vast sea area. The experimental hardware is a computer with the Intel i7-9700K CPU, NVIDIA RTX 2060s Ti GPU and Ubuntu 18.04 operating system. The software framework used in the experiment is pytorch1.7.1, and cuda10.2 is used to accelerate all models in the training and testing stages.

## Experimental results

Currently, the intersection over union (IoU) is commonly used to assess the accuracy of a target location result, which refers to the ratio of the intersection and union between the prediction box and the ground truth. Rotating target detection can also use this evaluation criterion, as shown in Fig.5, but the angle has a great impact on the IoU between the prediction rotating box and ground truth. Therefore, the accurate angle prediction is required for the rotating target detection.



1. IoU of the rotating box.

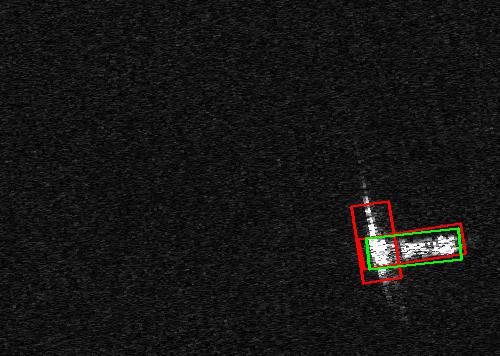
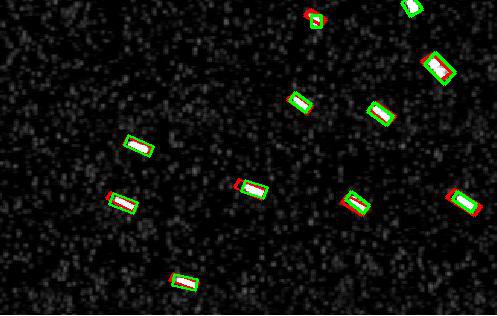
We implemented two sets of these experiments, which selected the stochastic gradient descent (SGD) as the training optimizer. For the experiments implemented in the original dataset, the learning rate, momentum, weight decay is set to 0.00025, 0.9, and 0.0005, respectively. A total of 70 epochs have been trained. For the experiments implemented in the dataset after data augmentation processing, the learning rate, momentum, weight decay is set to 0.0003, 0.9, and 0.0005, respectively. A total of 120 epochs have been trained. The learning rate decay strategy is implemented in the training of both sets of the experiments. Two sets of the experimental results are shown in Table I.

1. Experiments results

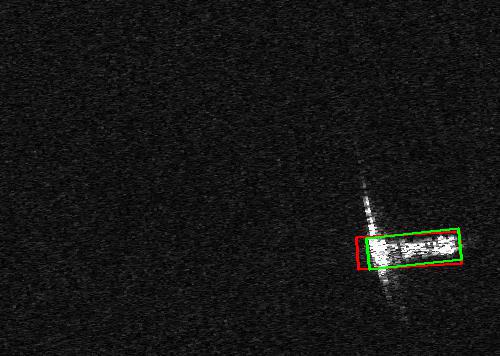
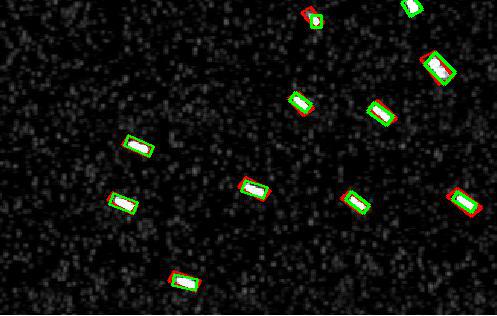
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Indicator**  **Data** |  |  |  |  |  |
| Original data | 0.960 | 0.940 | 0.872 | 0.728 | 0.490 |
| Augmented data | 0.959 | 0.944 | 0.903 | 0.805 | 0.575 |
| **Indicator**  **Data** | **Recall-30** | **Recall-40** | **Recall-50** | **Recall-60** | **Recall-70** |
| Original data | 0.892 | 0.874 | 0.811 | 0.676 | 0.455 |
| Augmented data | 0.880 | 0.866 | 0.828 | 0.739 | 0.527 |

It is considered that the target is successfully detected, if the IoU between the prediction box and the ground truth is greater than *N*%, and  in Table I refers to the proportion of prediction boxes which successfully detected ship target in all prediction boxes, Recall-N indicates the proportion of the successfully detected targets in all targets under this condition.

Based on these experimental results, the proposed ship target detection method achieves a detection precession of 0.872 under the evaluation criterion . After the data augmentation processing, the detection precession reaches 0.903 under the evaluation criterion . However, when the IoU threshold increases, the detection accuracy decreases dramatically, which is caused by the prediction errors of the parameters of the rotating rectangle. Compared with the horizontal rectangle based method, the rotating rectangle based method needs to predict the angle of the rotating rectangle, which has big impact on the detection performance, resulting in a larger cumulative error. Based on the experimental data in Table I, the proposed data augmentation method effectively improves the detection precession and recall rate of the model and then alleviates the problem of the insufficient data.

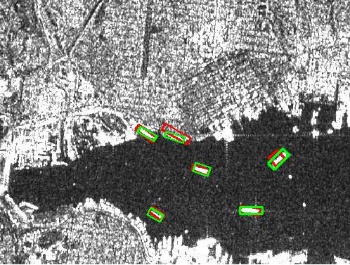
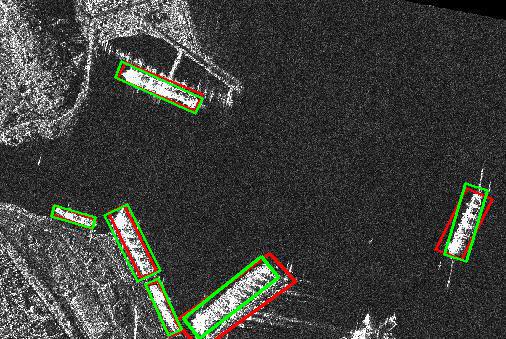


(a) (b)

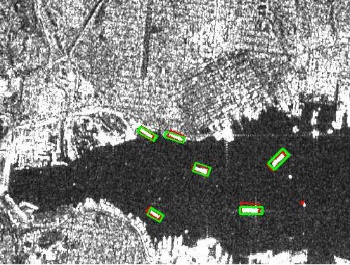
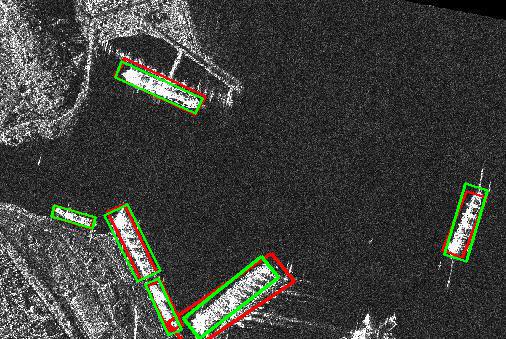


(c) (d)

1. Ship target detection results for the simple scenes. (a) and (b) are the experimental results without the data augmentation processing; (c) and (d) are the experimental results after the data augmentation processing.



(a) (b)



(c) (d)

1. Ship target detection results for the complex scenes. (a) and (b) are the experimental results without the data augmentation processing; (c) and (d) are the experimental results after the data augmentation processing.

Fig.6 and Fig.7 show the ship target detection results for the simple and complex scenes, respectively. The green boxes are ground truth, and the red boxes are the prediction results. In both scenes, the model effectively detects the ship targets in the SAR images. In Fig.6(b), the model takes the bright spot at the bottom of the ship target which maybe the mast of the ship as another object, by mistake. But in Fig.6(d), the model trained under the augmented dataset doesn’t interfered by the bright spot. The reason is that the data augmentation strategy enhances the diversity of the samples and enriches the information learned by the model, therefore it enhances the detection performance and anti-interference ability of the model. It is seen that the prediction boxes of the model trained under the data augmented dataset fit the ship targets better.

# Conclusion

The method proposed in this paper extend CenterNet to the field of the oriented ship detection in SAR images. The experimental results show that the model performs well on the SDDD+ dataset. The data augmentation strategy proposed in this paper effectively alleviates the problem caused by the insufficient data and improves the detection performance and stability of models. Finally, the average detection precession reaches 0.903 under the evaluation criterion . The angle prediction is the key to detecting the ship targets with the rotating rectangle, which is also the focus of the subsequent research work.

##### Acknowledgment

This work has been co-supported by the National Natural Science Foundation of China under Grant 61801510, by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515010768 and 2214050002344, and by the National Science Foundation of Hubei Province under Grant 2019CFB263. Hongtu Xie is the corresponding author.

##### References

1. D. Chen and Y.W. Ju, “SAR ship detection based on improved YOLOv3,” Systems Engineering and Electronic, vol.43, no.4, pp. 937-947, 2021.
2. J. Li, C. Qu and J. Shao, “Ship detection in SAR images based on an improved faster R-CNN,” IEEE SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 2017, pp. 1-6.
3. C.H. Hu, C. Chen, C. He, et al, “SAR detection for small target ship based on deep convolutional neural network,” Journal of Chinese Inertial Technology, vol.27, no.03, pp.397-405+414, 2019.
4. Y. Wang, C. Wang, H. Zhang, et al, “Automatic ship detection based on RetinaNet using multi-resolution Gaofen-3 imagery,” Remote Sensing, vol.11, no.5, pp.1-14, 2019.
5. T.Y. Lin, P. Goyal, R. Girshick, et al, “Focal loss for dense object detection,” Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017, pp.2980-2988.
6. Q. An, Z. Pan, L. Liu, et al, “DRBox-v2: an improved detector with rotatable boxes for target detection in SAR images,” IEEE Transactions on Geoscience and Remote Sensing, vol.57, no.99, pp. 8333-8349, 2019.
7. X. Zhou, D. Wang, and P. Krähenbühl, “Objects as points,” 2019, arXiv preprint arXiv:1904.07850.
8. S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo and J. Choe, “CutMix: regularization strategy to train strong classifiers with localizable features,” IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 6022-6031.
9. F. Yu, D. Wang, E. Shelhamer and T. Darrell, “Deep layer aggregation,” IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 2403-2412.
10. X. Yang and J. Yan, “Arbitrary-oriented object detection with circular smooth label,” [European Conference on Computer Vision](https://link.springer.com/conference/eccv), Glasgow, UK, 2020, pp. 677-694.