Research on Ship Target Detection in SAR Image Based on Improved YOLO v3 Algorithm

Yang Chen(✉), Shaojie Zhu, Xiuwen Xu, Hui Ye, Yang Liu, and Yuchuan Xu

Shanghai Institute of Satellite Engineering, Shanghai, China

Abstract. Synthetic aperture radar (SAR) has the characteristics of all-weather, all day and multi-application observation. In recent years, ship target detection based on SAR image has been widely concerned by relevant researchers. In this paper, based on the object detection method of deep learning algorithm, the detection performance of ship target in SAR image is studied by using YOLOv3 algorithm. In order to solve the problem of increasing error rate of ship target detection in complex background, YOLOv3 algorithm is improved. By adding a preprocessing layer in the front of the input layer, the accuracy of the ship detection is improved from 92.17% to 95.80%. The algorithm can be applied to other target detection in SAR image.

Keywords: Deep learning · YOLOv3 · Ship detection · SAR image

1 Introduction

Synthetic aperture radar (SAR) has the characteristics of all-weather, all-day and multi-purpose observation. It can make high-resolution imaging of the observation area, and has a wide range of applications in environmental protection, disaster monitoring, marine observation, resource exploration, precision agriculture, geological mapping and other fields [1]. Therefore, for SAR images, it is particularly important to achieve fast and accurate target detection.

With the rapid development of global economy and society, the trade between countries is increasing, the demand for shipping is growing rapidly, and ships are getting bigger and faster. Safety has become an important research content in the field of modern ship research [2]. With the increase of the number of ships, the increase of ship size and tonnage, the risk of ship navigation is also increasing. In March 2021, "Ever Given", a super large freighter of Taiwan Evergreen Marine Corp, ran aground in the Suez Canal. It led to the whole river blocked for several days, resulting in huge economic losses. Therefore, it is of great significance to realize the ship target detection in SAR image for ship monitoring (Fig. 1).

With the continuous progress of SAR imaging technology, the ability to obtain high-resolution and massive data has been greatly improved. Ship target detection based on SAR image has been paid more and more attention by relevant researchers and engineers, and has become one of the important research contents in the field of marine remote sensing technology [3].

X. Wang et al. (Eds.): AICON 2021, LNICST 396, pp. 165–176, 2021.
https://doi.org/10.1007/978-3-030-90196-7_15

Fig. 1. The "Ever Given" stranded in Suez Canal

2 Current Situation of Ship Target Detection

Object detection is one of the basic researches in the field of computer vision. In recent years, with the continuous improvement of computer computing ability, the research of object detection based on deep learning has developed rapidly. The target detection algorithm has also changed from the traditional algorithm based on artificial features to the deep neural network detection technology. In just a few years, many excellent algorithm technologies have emerged in the top-level conference of machine vision. From R-CNN and OverFeat proposed in 2013 to the following Fast R-CNN, SSD and YOlO series, the network structure has changed from two levels to one level, and the algorithm performance has changed from PC oriented to mobile oriented (Fig. 2). The object detection algorithm based on deep learning shows strong detection effect and performance [4].

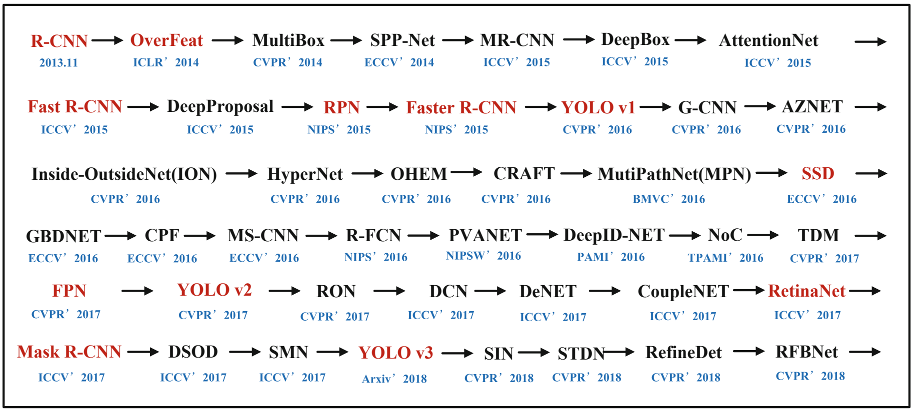

Fig. 2. The history of object detection

The core of ship target detection research is image technology. Through the acquisition of the ship target image, the object detection technology is used to recognize the ship target in the complex environment. The existing methods are mainly divided into two categories: traditional image processing technology or artificial intelligence deep learning technology. Different methods are different in the specific processing process [3], but these methods have general characteristics. The details are shown in the figure below (Fig. 3).

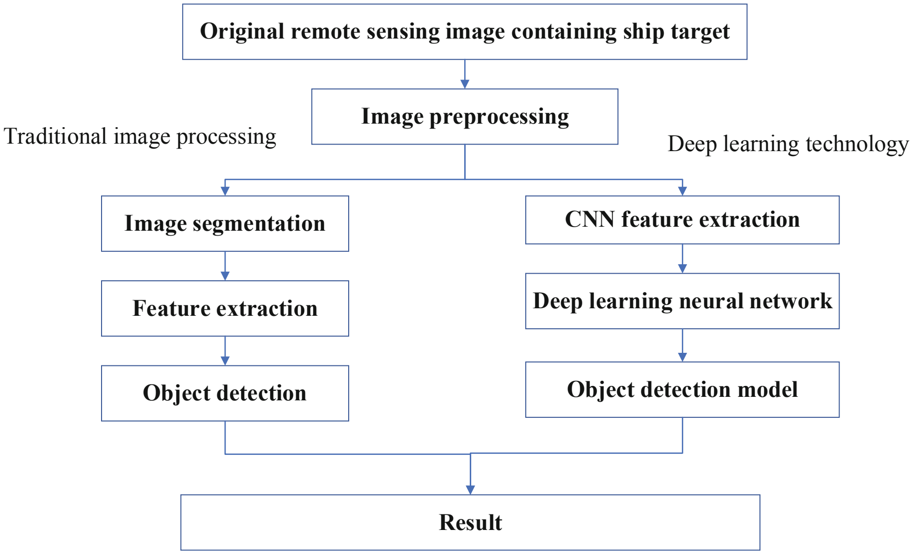

Fig. 3. Flow chart of ship detection

The method of target detection based on deep learning uses convolution neural network or other deep learning models to extract the deep features of the image. Then the network structure is used to transfer the feature map layer by layer to complete the accurate prediction of the target location. In this paper, YOLOv3 algorithm based on deep learning is used to detect ship targets in SAR images. More than 1000 SAR images are trained and tested in SSDD dataset, and the detection effect was evaluated by calculating mAP(mean Average Precision). At the same time, through the improvement of yolov3 algorithm, the detection performance is improved.

3 YOLO v3 Algorithm

The target detection methods based on convolution neural network can be divided into two categories, one is two-stage detection method, the other is single-stage detection method. The two-stage target detection usually starts with region proposal correction and background elimination, and then carries out region proposal classification and bounding box regression; The single-stage target detection algorithm integrates the two processes, and achieves the framework by anchor points and classification refinement [4].

Each of the two methods have their own advantages and disadvantages. The two-stage detection has some advantages in detection accuracy and performance, but its speed and real-time performance are still far behind the single-stage detection. The single-stage detection method only needs one feed-forward network calculation, which greatly improves the detection speed and it is more conducive to the on orbit application in the case of limited satellite resources.

YOLO is the first single-stage target detection method and it is also the first method to realize real-time target detection. The detection speed can reach 45 frames per second, and the mAP is more than twice that of other real-time detection systems.

YOLO algorithm regards the detection problem as an end-to-end regression problem, so the process of image processing is very simple and direct. In 2018, YOLO v3 was officially proposed. On the basis of YOLO v2, it adopts a deeper network structure and extends darknet-19 to darknet-53. The model has 106 layers of network. The network structure of Yolo V3 is shown in the following figure (Fig. 4):

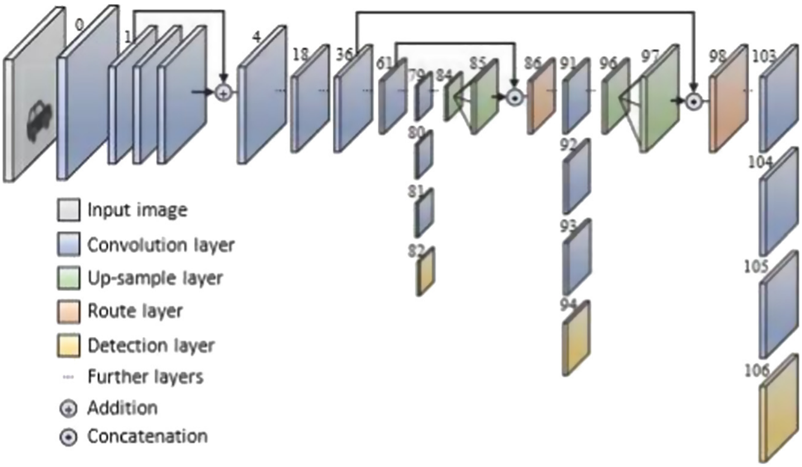

Fig. 4. The network structure of YOLO v3

As shown in the figure above, the number represents the number of layers, and its network structure includes Input layer, Convolution layer (including residual bloack), Up-sample layer, Route layer and Detection layer. As can be seen from the figure, YOLO v3 model realizes three kinds of different scale detection in the detection layer, which are located in the 82nd, 94th and 106th layers. The characteristic maps are 13×13, 26×26 and 52×52, which can detect targets of different sizes.In addition, YOLO v3 uses a total of nine anchor boxes, and each anchor box has three sizes, which can predict more frames [4]. In YOLO v3, the loss functions of predicted and true values are calculated as follows:

$$
\begin{aligned}
loss = & \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}\left[\left(\sigma\left(t_x\right)_i^j - \sigma\left(\hat{t}_x\right)_i^j\right)^2 + \left(\sigma\left(t_y\right)_i^j - \sigma\left(\hat{t}_y\right)_i^j\right)^2\right] \\
& + \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}\left[\left(t_{w_i^j} - \hat{t}_{w_i^j}\right)^2 + \left(t_{h_i^j} - \hat{t}_{h_i^j}\right)^2\right] \\
& + \sum_{i=0}^{S^2}\sum_{j=0}^{B} I_{ij}^{obj}\left(C_i^j - \hat{C}_i^j\right)^2 \\
& + \sum_{i=0}^{S^2}\sum_{j=0}^{B}\sum_{C\in classes} I_{ij}^{obj}\left(p_i^j(c) - \hat{p}_i^j(c)\right)^2
\end{aligned} \tag{1}
$$

Where S2 is the number of grids, and B is the number of bounding boxes in each grid. The means that when the jth bounding box of the ith grid is responsible for predicting the target. If the target is detected, the, otherwise it is 0. The first term in the formula represents the loss of the center point of the bounding box, the second term represents the loss of the width and height of the bounding box. The parameter C is the confidence score. The p is the probability of ship class.

4 Ship Detection in SAR Image Based on YOLOv3

4.1 Experimental Environment

In order to verify the ability of YOLOv3 algorithm in ship target detection of SAR image. In this paper, the framework of deep learning is Tensorflow2.4.0. Tensorflow is a widely used framework developed and maintained by Google for deep learning. It has a training visualization component Tensorboard, which can visualize the network structure and training process, and it is convenient for long-term and large-scale training of the network. Our framework was developed on Windows 10 operating system and Intel (R) core (TM) processor i5-8400cpu@2.80 Ghz. The GPU version is NVIDIA GTX 1060 which contains 5G RAM. The memory capacity is 8G and the programming language is python.

4.2 Experimental Dataset

The common dataset called SSDD used in this paper contains 1160 SAR images and 2456 ship targets. The dataset contains ship targets under various conditions, which is commonly used in this field (Fig. 5).

The dataset in this article uses the same format as the PASCOL VOC dataset, which store the data content in three folders called Annotation、JPEGImages and ImageSets.

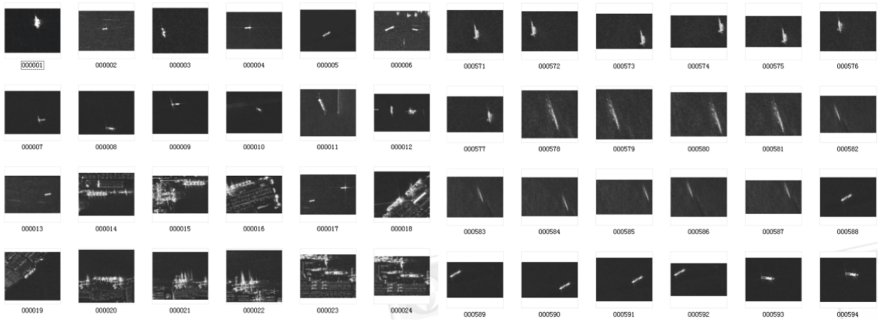

Fig. 5. The SSDD dataset


```
- <annotation verified="no">
    <folder>JPEGImages</folder>
    <filename>000001</filename>
    <path>/home/ljw/FRCN_ROOT/data/VOCdevkit2007/VOC2007/JPEGImages/000001.jpg</path>
  - <source>
      <database>Unknown</database>
    </source>
  - <size>
      <width>416</width>
      <height>323</height>
      <depth>1</depth>
    </size>
    <segmented>0</segmented>
  - <object>
      <name>ship</name>
      <pose>Unspecified</pose>
      <truncated>0</truncated>
      <difficult>0</difficult>
    - <bndbox>
        <xmin>208</xmin>
        <ymin>50</ymin>
        <xmax>273</xmax>
        <ymax>151</ymax>
      </bndbox>
    </object>
  </annotation>
```


Fig. 6. The Label data format

The JPEGImages folder is used to store images, and the Annotation folder is used to store label files corresponding to each image. The XML file format is shown in the following figure (Fig. 6):

The label content includes the category (name) of the target in the bounding box, as well as the width, height and position (x_{min}, x_{max}, y_{min}, y_{max}) of the bounding box. The dataset is divided into training set, validation set and test set according to the ratio of 8:2:1.6.

4.3 Model Training

The weight adjustment method of model training is gradient descent method. The training batchsize value is 4, epoch value is 100, the initial learning rate value is 0.01, and the attenuation weight of learning rate value is 0.0001. The training generates CKPT model file and node file, which saves the weights in the network structure. The weight parameter file occupies about 240 MB of storage capacity. The figures below show the process of the loss value changing with the number of iterations in the training process. With the increase of training times, the training error gradually decreases, and the network fitting effect is getting better and better (Fig. 7).

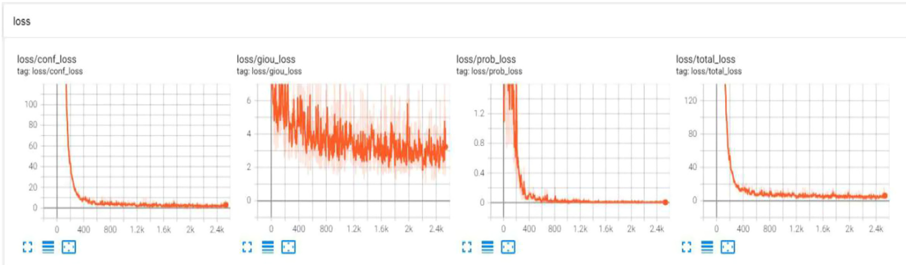

Fig. 7. The process of loss reduction in training

4.4 Experimental Results Analysis

In this paper, mAP (average precision) is used as the measurement index. Since the data only contains lable like ship, the AP value of the ship is mAP value of the experimental results, and the calculation formula is as follows:

$$AP = \int_0^1 P(R)dR \tag{2}$$

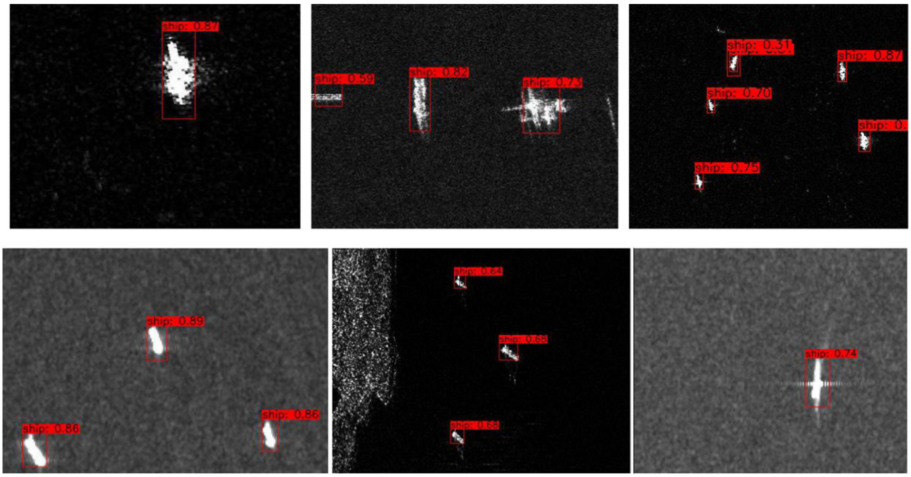

Fig. 8. Research on ship detection effect of SAR image based on YOLOv3 (fishing boats)

The parameter P represents the accuracy rate and the parameter R represents the recall rate.

The ship target detection algorithm based on YOLO v3 has achieved good results on SSDD dataset. The test set contains 188 images, including 348 ship targets, with an average accuracy of 92.17%. The specific test results are shown in the figure below (Fig. 8 and Fig. 9).

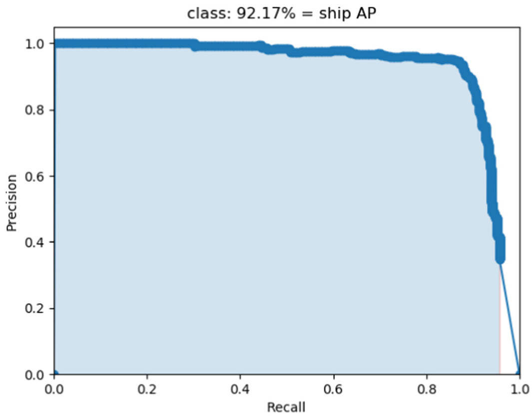

Fig. 9. P-R curve of test results (accuracy 92.17%)

However, the effect of this algorithm is not good in the detection of complex background, and it is easy to mistakenly detect the background as a ship, as shown in the figure below (Fig. 10).

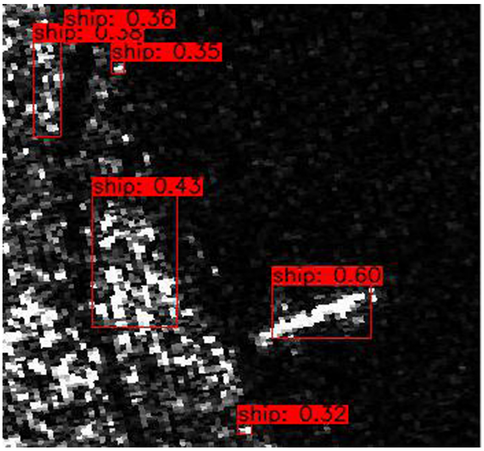

Fig. 10. Weak detection in complex background

5 Ship Detection in SAR Image Based on Improved Algorithm

5.1 Algorithm Improvement

Because there are a lot of coherent noises in SAR images, especially in complex background, it is easy to cause interference to the network, which leads to the false detection of the background as the target and reduces the accuracy. To solve this problem, this paper adds a preprocessing layer before the input layer of the network structure (Fig. 11).

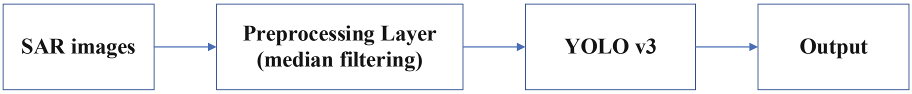

Fig. 11. Improved network structure

Before network training, median filtering is performed on the preprocessing layer. Median filtering is a non-linear smoothing technique, which sets the gray value of each pixel to the median value of all pixels in the neighborhood window. If the filter window length value L is equal to 2N + 1, the calculation formula of window filter output is as follows:

$$y(i) = Med[x(i - N), \cdots, x(i), \cdots, x(i + N)] \tag{3}$$

Median filter is very effective for removing salt and pepper noise, while preserving the edge details of the image. The figures below show a comparison of SAR image before and after median filtering (Fig. 12).

It can be seen from the above figure that the filtered SAR image effectively suppresses the background noise, and the target is more prominent. After preprocessing all training images, the network model is retrained according to the settings in Sect. 4.3. The experimental results have been effectively improved.

5.2 Experimental Results Analysis and Comparison

Meanwhile, 188 images in the testset of SSDD are used to test the improved algorithm model in this paper. The average accuracy value was 95.80%. Compared with single YOLO network, the accuracy is improved by 3.63%. The specific test results are shown in the figure below (Figs. 13 and 14).

It can be seen from the above figures that the proportion of correct prediction is significantly improved after median filtering, which indicates that the network can better identify the image target after preprocessing and the proportion of the correct predicted quantity value has been greatly increased. Especially in the complex background, the accuracy of the method has been greatly improved (Fig. 15).

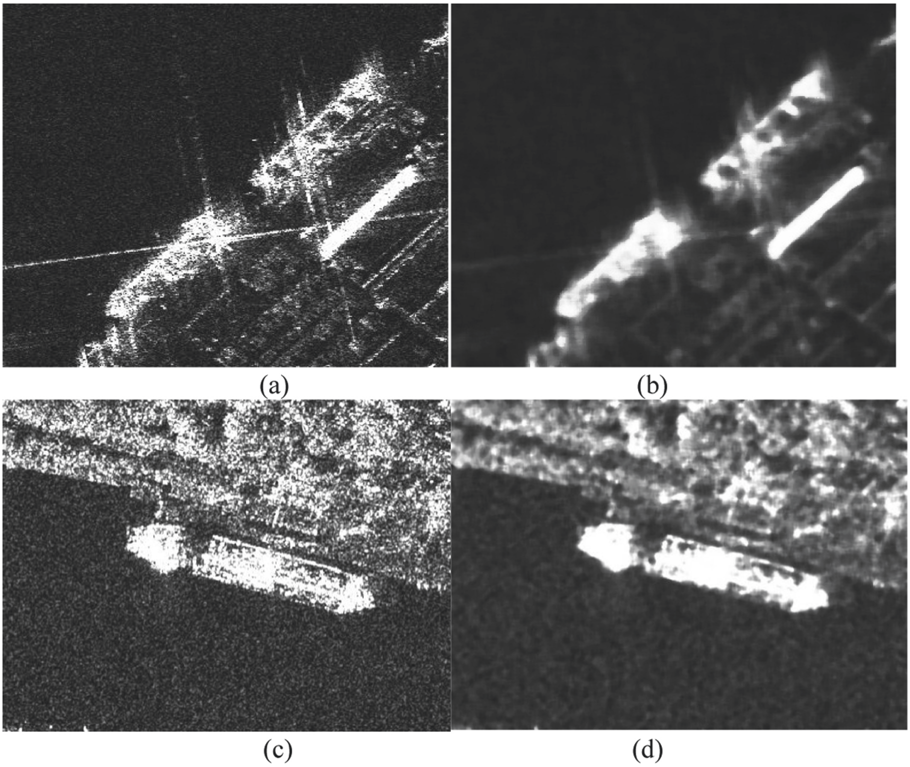

Fig. 12. Comparison of SAR image (Cargo Ship) before and after median filtering: (a) and (c) original image, (b) and (d) after median filtering image

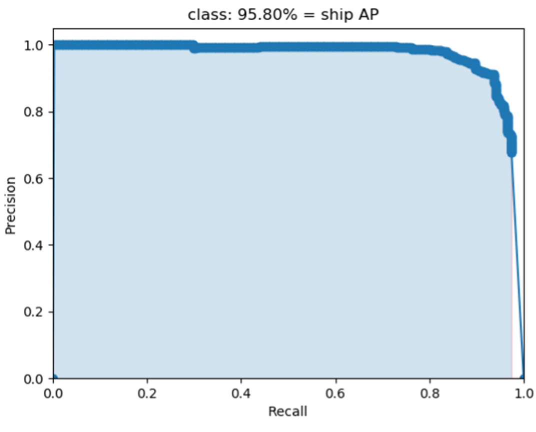

Fig. 13. The P-R curve of the improved algorithm (accuracy 95.80%)

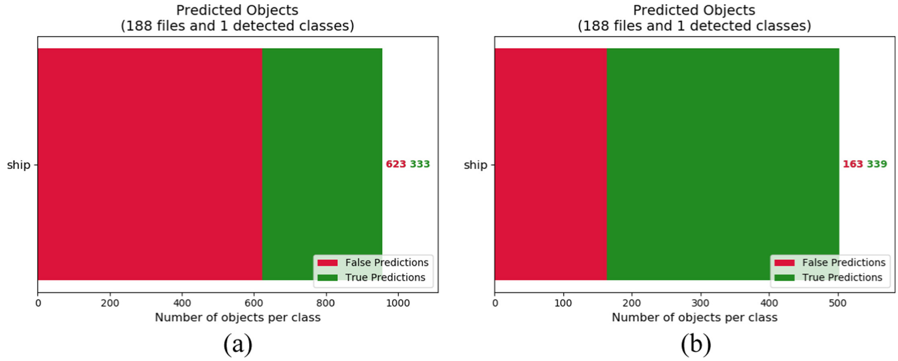

Fig. 14. The proportion change of correct prediction number before and after the improvement of the algorithm: (a) The correct proportion number of the original algorithm; (b) The correct proportion number of the improved algorithm

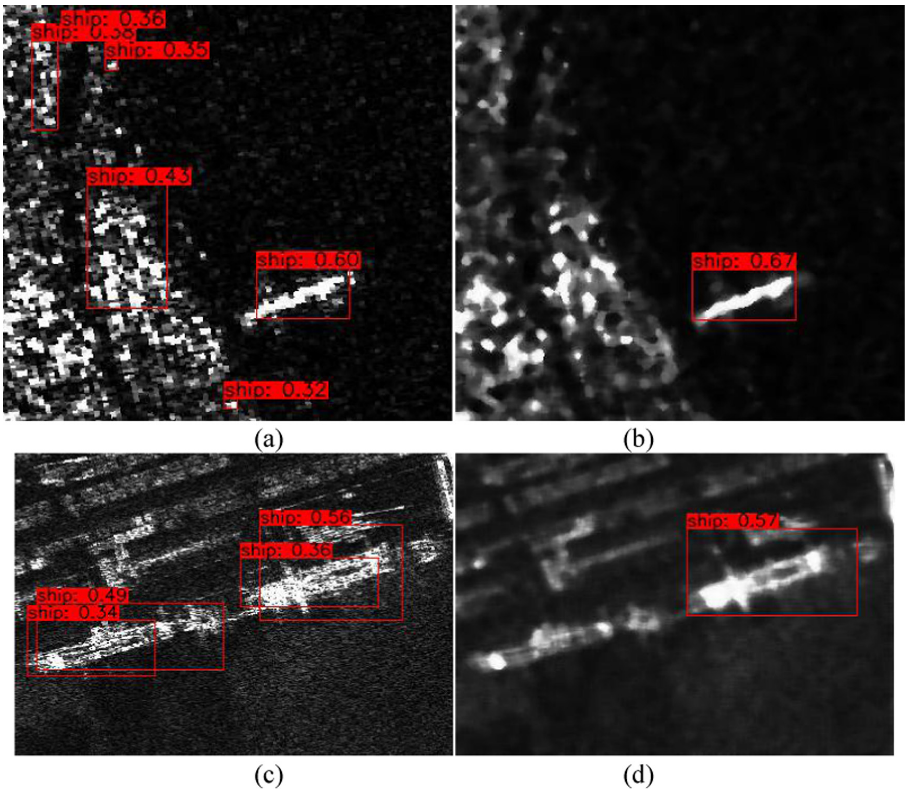

Fig. 15. Comparison of detection results before and after improved algorithm model: (a) and (c) The detection results of the original algorithm; (b) and (d) The detection results of the improved algorithm

6 Conclusion

To solve the problem of ship detection in SAR image, the YOLOv3 algorithm is applied to detect ship targets in SSDD dataset. Through the training and optimization of its network parameters, the accuracy of the test results only reached 92.17%. According to the specificity analysis, the main reason for the decrease of detection accuracy is the interference of complex sea background. Therefore, in this paper the network structure of YOLO v3 is improved. Specifically, we add a preprocessing layer before the input layer of the network structure. In this new layer, the image data is processed by median filtering, so that the data can be enhanced by suppressing the background clutter, then the network parameter training is optimized. Compared with YOLOv3 algorithm, the ship detection accuracy is improved to 95.80% based on the improved algorithm in this paper. The algorithm can be applied to other target detection in SAR images.

References

1. Qingjun, Z., Xiaolei, H., Jie, L.: Technology progress and development trend of spaceborne synthetic aperture radar remote sensing. Spacecraft Eng. **26**(6), 1–7 (2017). (in Chinese)
2. Wei, S.: Overview of ship detection technology based on remote sensing image. Telecommun. Eng. **60**(9),1126–1132 (2020) (in Chinese)
3. Li, J.W., Qu, C.W., Peng, S.J., Deng, B.: Ship detection in SAR images based on convolutional neural network. Syst. Eng. Electr. **40**(9), 1953–1959 (2018). (in Chinese)
4. Peng, D., Ming, C., Hua, S.T.: Deep Learning and Object Detection. Publishing of Electronics Industry, Beijing, March 2020
5. Redmon, J., Divvala, S., Girshick, R.: et al.: You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
6. Redmon, J., Farhadi, A.:Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
7. An, Q., Pan, Z., You, H.: Ship detection in Gaofen-3 SAR images based on sea clutter distribution analysis and deep convolutional neural network. Sensors **18**(2), 1345–1357 (2018)
8. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)