ARBITRARY-ORIENTED SHIP DETECTION METHOD BASED ON IMPROVED REGRESSION MODEL FOR TARGET DIRECTION DETECTION NETWORK

Bohao Ran, *Yanan You*, *Zezhong Li*, *Fang Liu*

School of Information and Communication Engineering, Beijing University of Posts and Telecommunication

bohaor@163.com; youyanan@bupt.edu.cn; lizezhong@bupt.edu.cn; lindaliu@bupt.edu.cn

IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323265

ABSTRACT

Arbitrary-oriented ship detection is one of the main applications of high-resolution remote sensing images. The current target direction detection methods, based on deep convolution neural network (DCNN), can estimate most of the ship directions (i.e. represented by angles). However, the performance of these networks is limited by the boundary discontinuity problem due to the angle-based regression model. In this paper, we propose an improved regression model based on coordinates in the complex plane to solve the boundary discontinuity problem. Experiments on real remote sensing dataset verify the effectiveness and robustness of our method.

Index Terms— ship detection, regression model, remote sensing, deep learning

1. INTRODUCTION

High-resolution remote sensing images are capable of accurately exposure and locate surface entities, which is widely used in the field of wide-area object recognition. Ship detection is a typical application of high-resolution remote sensing images and plays an important role in ship navigation, ship communication, and maritime rescue. In recent years, the object detection method based on Deep Convolutional Neural Network (DCNN) has become a research hotspot. By extracting and learning the feature of the ship in remote sensing images, the network could achieve the classification and location of the targets.

Although the conventional DCNN-based methods [1-3] have achieved good detection performance in natural images, due to the lack of information on the target direction, these methods are not suitable for multi-directional ship detection in remote sensing images. Currently, many target direction detection networks can be used to predict ship direction, such as R2CNN [4], R2CNN++ [5], R2PN [6], and R-DFPN [7]. However, these methods have the same shortcoming, that is the boundary discontinuity problem. It means that the predicted direction is inaccurate when the ship's orientation is near the angular boundary. This problem can be attributed

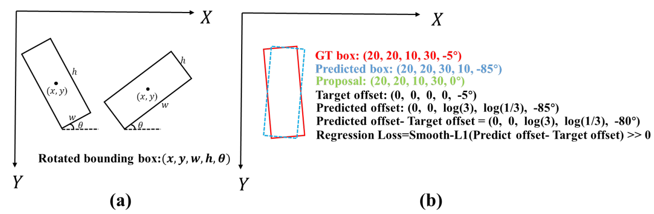

Fig. 1. Representation of rotated bounding box and calculation of regression loss in target direction detection networks in R-DFPN and R^2CNN++. (a) representation of rotated bounding boxes. (b) calculation of regression loss. (No proposals are drawn for easy viewing).

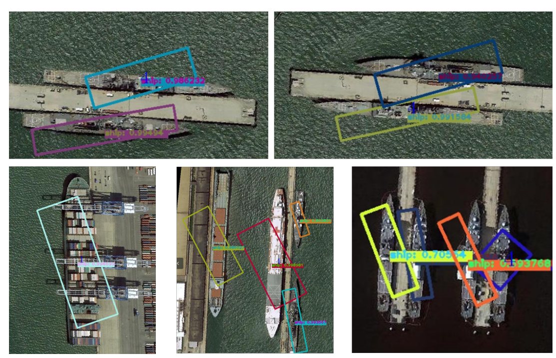

Fig. 2. Inaccurate detection results of R-DFPN at angular boundary.

to the fact that these methods adopt the angle-based regression model. Since the angle is periodic and the angular boundary is discontinuous, the regression loss cannot converge near the boundary, which deteriorated the results of detection. Limited to the boundary discontinuity problem, these methods could not truly realize ship detection in arbitrary directions.

In this paper, an arbitrary-oriented ship detection method based on an improved regression model is proposed to solve this problem. This new regression model can be easily integrated into the existing target direction detection networks. Experiments on real remote sensing dataset prove that networks with the proposed regression model achieve more competitive detection results.

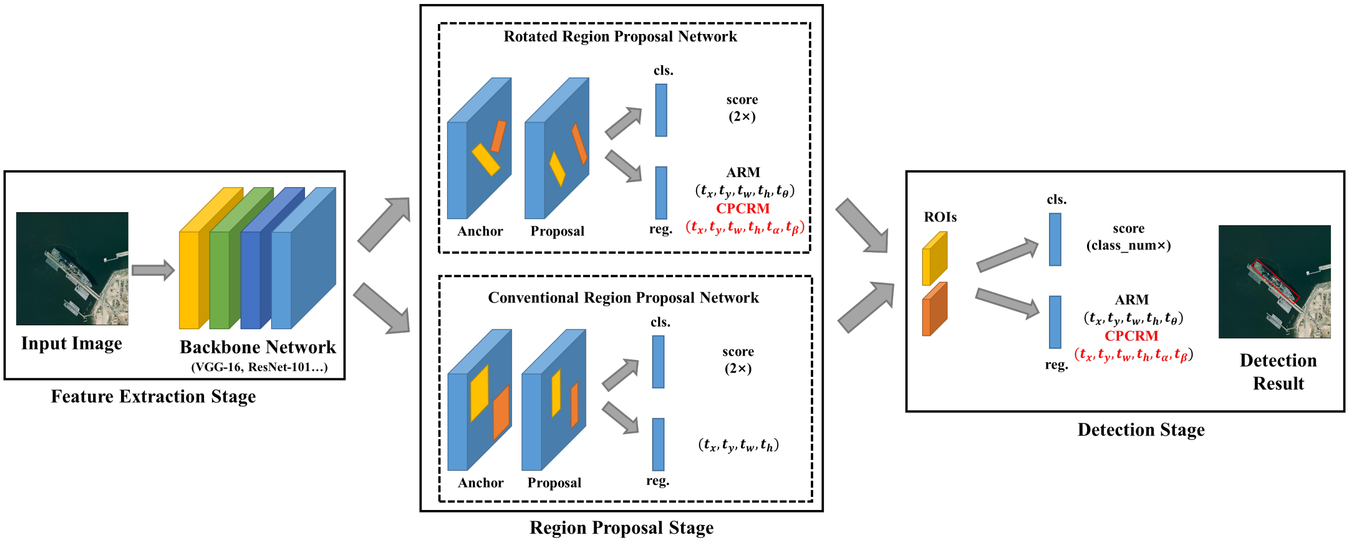

Fig. 3. The framework of target direction detection networks.

The rest of this paper is organized as follows. In Section 2, we analyze the boundary discontinuity problem. The proposed regression model is detailed in Section 3. Experimental results on real remote sensing images are presented in Section 4, and Section 5 gives the conclusion.

2. Boundary discontinuity problem

In the target direction detection networks, a five-element vector (x, y, w, h, θ) is usually used to define the rotated bounding box [8] in regression. The parameterized coordinate regression model is as follows:

$$
\begin{aligned}
& t_x = (x_p - x_a)/w_a, t_y = (y_p - y_a)/h_a \\
& t_w = \log(w_p/w_a), t_h = \log(h_p/h_a) \\
& t_\theta = \theta_p - \theta_a \\
& t_x^* = (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a \\
& t_w^* = \log(w^*/w_a), t_h^* = \log(h^*/h_a) \\
& t_\theta^* = \theta^* - \theta_a
\end{aligned} \tag{1}
$$

where x, y, w and h represents the box's center coordinates and its width and height. θ represents the ship direction. Variables x_p, x_a, x^* are for the predicted box, anchor box, and ground-truth box, respectively (likewise for y, w, h and θ). The regression loss function is defined as:

$$L_{reg}(t, t^*) = smooth_{L1}(t - t^*) \tag{2}$$

where t is the predicted offset and t^* is the target offset. $smooth_{L1}$ is the smooth L1 loss defined in [9]. However, this definition of target direction, i.e. using the angle to represent direction, will cause the boundary discontinuity problem, that is, the calculation of the regression loss is inaccurate near the angular boundary.

For instance, in R-DFPN and R^{2}CNN++, ranging in $[-90°, 0°)$, θ is defined as the acute angle to the x-axis, and w represents the other side, as shown in Fig. 1 (a). When the ship direction approaches horizontal or vertical, w and h are reversed and the θ will change abruptly at $-90°$ and $0°$. This causes the large difference between the target offset and the predicted offset, so the regression loss turns to be large. As is illustrated with Fig. 1 (b), even though the prediction box and the ground-truth box have the same shape and their directions are close, their w, h and θ is much different, resulting in a large loss of regression. The boundary discontinuity problem will deteriorate the detection results near the angular boundary in these angle-based ship detection methods, as shown in Fig. 2.

3. METHODOLOGY

3.1. Target direction detection framework

The target direction detection networks mentioned above have a similar framework which can be summarized as three stages: feature extraction stage, region proposal stage, and detection stage, as shown in Fig. 3. In the first stage, the feature map is extracted from input images by the backbone network such as VGG-16 and ResNet-101. Additionally, R-DFPN builds a feature pyramid of dense connections to fuse multi-level information. In the next stage, there are two types of region proposal network (RPN): conventional RPN and rotated RPN. Both R^{2}CNN and R^{2}CNN++ adopt conventional RPN which uses axis-aligned anchors to generate axis-aligned proposals. To better cover ship targets at different directions, in R-DFPN and R^{2}PN, rotated RPN generates the rotated proposals by regressing the rotated anchors. During the detection stage, the features of proposals are fed into two branches with fully connected layers for classification and regression. In particular, the angle-based regression model (ARM) as Eq. (1) is used both

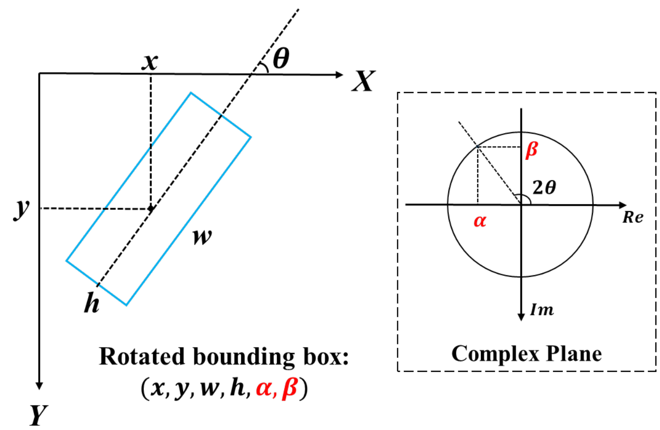

Fig. 4. The new representation of rotated bounding box in our method.

in the rotated RPN and detection stage, which will cause the boundary discontinuity problem. We thereby propose a complex plane coordinates regression model (CPCRM) as a substitution. This will be detailed in the next part.

3.2. Complex plane coordinates regression model

According to the analysis in section 2, we believe that the fundamental cause of the boundary discontinuity problem is the angle-based regression model. In this model, ship direction is represented by an angle in a rotated bounding box, while the angle is periodic and thus the angular boundary is not continuous. Therefore, it is necessary to find a new regression model, in which the definition of ship direction is unique and the boundary is continuous in any direction. For the above reasons, we first design a new representation of rotated bounding boxes that uses the coordinates on the unit circle in the complex plane to represent the ship direction, as illustrated in Fig. 4. A six-element tuple $(x, y, w, h, \alpha, \beta)$ represents the arbitrary-oriented rotated bounding box, where (x, y) is the coordinates of the center of the bounding box, w is the length of the long side, and h is the short side. The (α, β) is the coordinates of ship direction on the unit circle in the complex plane, which can be calculated as follow:

$$\alpha = \cos 2\theta, \beta = \sin 2\theta \quad (3)$$

where θ is the angle between the long side and the x-axis. Starting from the positive x-axis, it increases clockwise and decreases counterclockwise, and its range is $[0°, 180°)$. It is worth noting that the coefficient of θ is 2 because the period of ship direction is 180 degrees in the current target direction detection networks. Although θ is still discontinuous at the boundary (there is still a 180° mutation from 180° to 0°), the coordinates (α, β) are continuous in any direction that avoids the boundary discontinuity problem. When predicting the ship direction, we first calculate the predicted coordinates (α_p, β_p) on the complex plane:

$$(\alpha_p, \beta_p) = (\alpha_a, \beta_a) + (t_\alpha, t_\beta) \quad (4)$$

where (t_α, t_β) represents the regression vector of the ship direction, and (α_a, β_a) represents the coordinates of the anchor direction on the complex plane.

The complex plane coordinates regression model (CPCRM) of the rotate bounding box in our method is represented as:

$$\begin{aligned} & t_x = (x_p - x_a)/w_a, t_y = (y_p - y_a)/h_a \\ & t_w = \log(w_p/w_a), t_h = \log(h_p/h_a) \\ & t_\alpha = \alpha_p - \alpha_a, t_\beta = \beta_p - \beta_a \\ & t_x^* = (x^* - x_a)/w_a, t_y^* = (y^* - y_a)/h_a \\ & t_w^* = \log(w^*/w_a), t_h = \log(h^*/h_a) \\ & t_\alpha^* = \alpha^* - \alpha_a, t_\beta^* = \beta^* - \beta_a \end{aligned} \quad (5)$$

where x, y, w, h represent the center coordinates, width and height, respectively. (α, β) represent the coordinates of ship direction on complex plane. Variables x_p, x_a, x^* represent the predicted box, anchor box, ground-truth box respectively (likewise for y, w, h, α, β). The regression loss is defined same as Eq. (2).

Compared with the angle-based regression model, CPCRM does not introduce additional calculations and can be easily embedded in existing networks.

4. EXPERIMENTS

4.1. Dataset

The dataset was collected from Google Earth with 95 optical remote sensing images sized from 4000×4000 pixels to 10000×10000 pixels. All ship targets in the images are labeled by rotated rectangles. Then these images are cut into 800×800 sub-images by taking the center point of the targets as the center of the images. Rotation augmentation with angles $[90°, 180°, 270°]$ is adopted for the purpose of increasing the diversity of ship direction. Finally, 5492 pictures were obtained and divided into a training set and a test set by 3:1.

4.2. Implementation and results

All experiments were executed based on a server with an NVIDIA GeForce GTX 1080Ti GPU and 64GB of memory, and were performed on the deep learning framework Tensorflow. We use the pre-trained model ResNet-101 [10] from the ImageNet classification to initialize the network. The total iterations of training are 50k, with a learning rate of 0.0003 for the first 30k iterations, 0.00003 for the next 10k iterations, 0.000003 for the last 10k iterations. Besides, Momentum Optimizer is used as an optimizer, setting the weight decay and momentum to 0.0001 and 0.9, respectively.

In order to verify the effectiveness of the proposed method, we conducted comparative experiments based on three target direction detection networks: R^2CNN, R-DFPN, and R^2PN. Each network uses two different regression

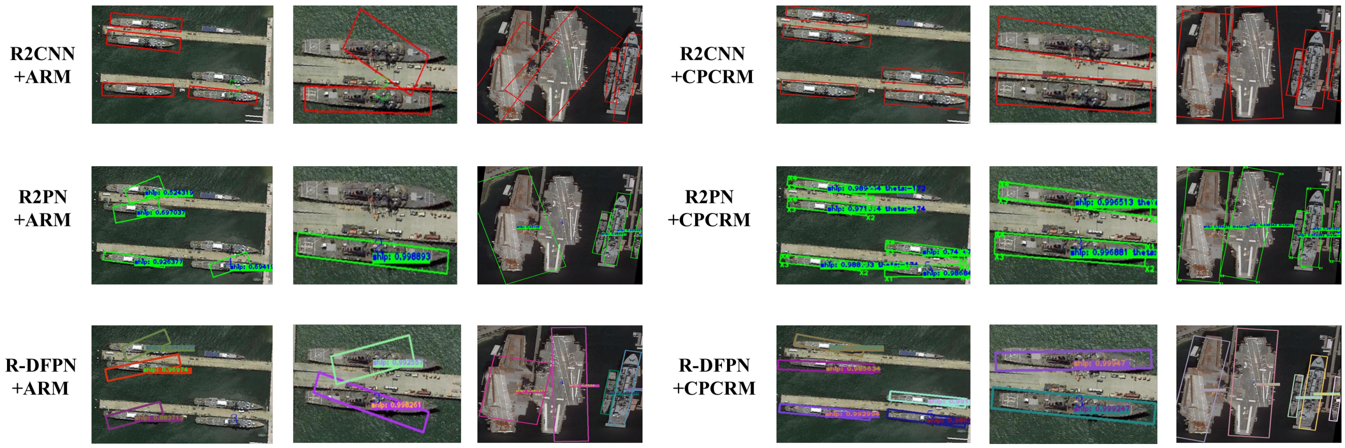

Fig. 5. Some detection results of three target direction detection networks with ARM and CPCRM.

models, i.e. ARM and CPCRM. Some detection results are shown in Fig. 5. It can be seen that, when ships are closed to horizontal or vertical, the predicted ship directions of networks with ARM are inaccurate and even some of the targets are lost. In contrast, the rotated bounding boxes of networks with CPCRM still effectively match the target in this situation. Furthermore, to quantitatively evaluate the performance of networks with different regression models in ship detection, we adopt the average precision (AP) under two IoU thresholds (0.5 and 0.7) as evaluation indicators. The threshold 0.5 is the common setting in object detection, and 0.7 is used to further evaluate the fitting ability of the rotated bounding box on the ship targets. Table 1 shows the quantitative comparison results. For each network in our experiments, CPCRM achieves a higher AP under both IoU thresholds than ARM. The results indicate that the CPCRM can promote the performance of the networks in ship detection.

Table 1 Quantitative comparison of the performance of three networks with ARM and CPCRM.

Methods	AP (IoU=0.5)	AP (IoU=0.7)
R^2CNN + ARM	81.95	53.33
R^2CNN + CPCRM	**87.53**	**65.51**
R^2PN + ARM	84.85	45.58
R^2PN + CPCRM	**91.56**	**61.45**
R-DFPN + ARM	83.01	53.52
R-DFPN + CPCRM	**91.12**	**63.97**

5. CONCLUSION

In this paper, we first analyze the boundary discontinuity problem in current target direction detection networks. Aiming at this problem, we proposed an arbitrary-oriented ship detection method based on complex plane coordinates regression model. With this regression model, networks can not only solve the boundary discontinuity problem, but also improve performance in ship detection. The results of comparison experiments on real remote sensing dataset validates the effectiveness and feasibility of our method.

6. REFERENCES

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A.C. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision*, pp. 21-37, 2016.
[2] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779-788, 2016.
[3] S. Ren, K. He, R. Girshick and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, pp. 91-99, 2015.
[4] Y. Jiang, X. Zhu, X. Wang, S. Yang, W. Li, H. Wang, P. Fu and Z.J.a.p.a. Luo, "R2cnn: Rotational region cnn for orientation robust scene text detection.", 2017.
[5] X. Yang, K. Fu, H. Sun, J. Yang, Z. Guo, M. Yan, T. Zhan and S.J.a.p.a. Xian, "R2cnn++: Multi-dimensional attention based rotation invariant detector with robust anchor strategy,", 2018.
[6] Z. Zhang, W. Guo, S. Zhu, W.J.I.G. Yu and R.S. Letters, "Toward arbitrary-oriented ship detection with rotated region proposal and discrimination networks," pp. 1745-1749, 2018.
[7] X. Yang, H. Sun, K. Fu, J. Yang, X. Sun, M. Yan and Z.J.R.S. Guo, "Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks," pp. 132, 2018.
[8] Z. Liu, H. Wang, L. Weng, Y.J.I.G. Yang and R.S. Letters, "Ship rotated bounding box space for ship extraction from high-resolution optical satellite images with complex backgrounds," pp. 1074-1078, 2016.
[9] R. Girshick, "Fast r-cnn," *Proceedings of the IEEE international conference on computer vision*, pp. 1440-1448, 2015.
[10] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778, 2016.