Remote sensing image dense target detection based on rotating frame

Zhixin Wang¹, Shaojun Wan² and Xiaoying Ma*¹

¹ Radio Monitoring Division, The State Radio Monitoring Center, Beijing, 100037, China
² College of Electronics and Information Engineering, Shenzhen University, Shenzhen, Guangdong, 518060, China

*Corresponding author’s e-mail: maxiaoying@srrc.org.cn

Abstract. Detecting densely arranged and arbitrarily oriented targets on optical remote sensing images is a challenge, and there is much room for improvement in existing algorithms. In this paper, an end-to-end two-stage rotating target detection model based on rotating frames is proposed. This model adds FPN+PAN feature fusion structure after the backbone network to obtain enhanced features that fuse the feature information of each layer, and secondly, fine-grained rotation detection is accomplished by introducing Oriented Region Proposal Network and Oriented Region of Interest Pooling layer. By comparing the performance of this paper’s algorithm with the current mainstream rotation detection algorithm on DOTA remote sensing dataset, this paper's algorithm can solve the difficult problem of detecting remote sensing targets with tight arrangement and arbitrary orientation to a certain extent.

1. Introduction

Remote sensing technology is one of the effective means for human to explore the earth. In recent years, with the number of optical remote sensing satellites launched by various countries increasing year by year, the sources of optical remote sensing images have become more and more abundant, and the optical remote sensing image target detection algorithms based on deep learning are also developing rapidly.

Yao et al. [1] first applied the Faster R-CNN algorithm to remote sensing image detection, and its detection accuracy was significantly improved compared with the traditional remote sensing detection algorithm. Li et al. [2] added multi-angle anchor boxes to the RPN module of Faster R-CNN for the problem of changing target orientation in remote sensing images, which effectively solved the problem of changing target orientation in remote sensing images. However, these algorithms are not effective for the detection of densely arranged and arbitrarily oriented targets, and the false detection rate and missed detection rate are much higher compared with other targets; therefore, there is still much room for improvement of optical remote sensing image target detection based on deep learning for densely arranged and arbitrarily oriented scenes.

In this paper, an end-to-end two-stage rotating target detection model based on rotating frames, FFPO+, is proposed to address this problem.

2. Model architecture design

The main architecture of this model is similar to that of Faster R-CNN, which is divided into five
modules: backbone network, feature fusion module, Oriented Region Proposal Network (ORPN), Oriented Region of Interest pooling (OROI pooling) and R-CNN sub-network.

The overall flow of the algorithm is as follows: the backbone network selects ResNet101 [3] for extracting the input image features, and the C2-C5 feature layers extracted by the backbone network are used to build the feature fusion structure. The feature fusion module adopts FPN+PAN (Feature Pyramid Network [4] + Path Aggregation Network [5]) for feature fusion enhancement at each layer, and the fused features will generate several rotated candidate frames with angles after passing through the ORPN rotated region generation network, and then pass through the OROI pooling layer for feature extraction of candidate regions, and finally the extracted features are sent to the sub-network for prediction. The overall structure of the network is shown in Figure 1.

Figure 1. Architecture of remote sensing image target detection algorithm based on rotating frame.

2.1. Feature fusion module
Figure 2 shows the schematic diagram of the FPN+PAN feature fusion module. The C2-C6 feature layers with dimensions of 256×256, 128×128, 64×64, 32×32, and 16×16 extracted by the ResNet backbone network are used to build the FPN+PAN structure, and the input image will generate P2-P6 features with five scales after the backbone network and the FPN+PAN structure. of predicted features, and the subsequent network will predict these enhanced features.

Figure 2. FPN+PAN feature fusion structure.

2.2. Oriented Region Proposal Network
The ORPN rotational candidate region generation network generates rotated anchor frames with angles by introducing rotation angle parameters to the preset anchor frames, which can better cover the multi-directional targets in dense scenes. In this algorithm, the target bounding box is represented by
(x, y, w, h, θ) five variables, where (x, y) is the center coordinate of the rotating box, and (w, h) are the longer and shorter edges of the rotating box, respectively, and θ is the rotation angle, and the rotation direction is counter clockwise. Figure 3 shows the schematic diagram of the rotating anchor box.

For the ORPN network, the regression parameters are defined in Equation (1)

\[
\begin{align*}
    t_x &= (x - x_a) / w_a; \\
    t_y &= (y - y_a) / h_a; \\
    t_w &= \log (w / w_a); \\
    t_h &= \log (h / h_a); \\
    t_\theta &= \theta - \theta_a \\
    t_x^* &= (x^* - x_a) / w_a; \\
    t_y^* &= (y^* - y_a) / h_a; \\
    t_w^* &= \log (w^* / w_a); \\
    t_h^* &= \log (h^* / h_a); \\
    t_\theta^* &= \theta^* - \theta_a
\end{align*}
\]

where (x, y, w, h, θ) is the rotation candidate frame coordinate parameter, (x_a, y_a, w_a, h_a, \theta_a) is the rotation anchor frame coordinate parameter, and (x^*, y^*, w^*, h^*, \theta^*) is the real frame coordinate parameter.

2.3. Oriented Region of Interest Pooling layer

The process of OROI Pooling layer is shown in Figure 4. OROI pooling layer outputs feature dimensions of N × N, and then these feature dimensions are fixed to 7 × 7, by averaging pooling operation to facilitate their classification and regression prediction by subsequent R-CNN sub-networks.

2.4. R-CNN sub-network

This algorithm adjusts the R-CNN network structure: the classification prediction branch remains unchanged and still uses two fully connected layers, while the regression prediction branch uses two
convolutional residual modules (ResBlock), and the adjusted R-CNN network structure is shown in Figure 5.

![Figure 5. Adjusted R-CNN network structure.](image)

### 3. Experimental steps and results

We choose the DOTA [6] remote sensing dataset, which contains a total of 15 target types such as aircraft, baseball fields, ports and vehicles. To ensure experimental fairness, all models in this section use the ResNet101 network trained on the ImageNet dataset [7] as the backbone network, while the original images are cropped into multiple sub-images according to the 200-pixel overlap, and the sub-images are resized to 1024 × 1024 size. The network uses the SGDM gradient optimization algorithm with the momentum set to 0.9, the decay coefficient set to 0.0001, and the initial learning rate set to 5*10^-4, with the learning rate reduced to one-tenth of the original after every 50 iterations, while a random flip in the horizontal-vertical direction is used for data enhancement to avoid overfitting.

We selected four rotating target detection algorithms to compare with this paper's algorithm FFPO+. The four compared algorithms are FR-O [6], R2CNN [8], RRPN [9] and RoI Transformer [10], and the backbone network of all algorithms is ResNet101. The performance of different algorithms on the DOTA remote sensing dataset is shown in Table 1.

| Target types          | FR-O  | R2CNN | RRPN | Transformer | FFPO+ |
|-----------------------|-------|-------|------|-------------|-------|
| plane                 | 79.42 | 80.94 | 88.52| 88.64       | 89.31 |
| baseball diamond      | 77.13 | 65.75 | 71.20| 78.52       | 76.89 |
| bridge                | 17.70 | 35.34 | 31.66| 43.44       | 45.36 |
| ground track field    | 64.05 | 67.44 | 59.30| 75.92       | 72.19 |
| small vehicle         | 35.30 | 59.92 | 51.85| 68.81       | 70.13 |
| large vehicle         | 38.02 | 50.91 | 56.19| 73.68       | 71.87 |
| ship                  | 37.16 | 55.81 | 57.25| 83.59       | 79.46 |
| tennis court          | 89.41 | 90.67 | 90.81| 90.74       | 86.92 |
| basketball court      | 69.64 | 66.92 | 72.84| 77.27       | 79.20 |
| storage tank          | 59.28 | 72.39 | 67.38| 81.46       | 81.03 |
| soccer ball field     | 50.30 | 55.06 | 56.69| 58.39       | 63.71 |
| roundabout            | 52.91 | 52.23 | 52.84| 53.54       | 49.05 |
| harbor                | 47.89 | 55.14 | 53.08| 62.83       | 65.79 |
| swimming pool         | 47.40 | 53.35 | 51.94| 58.93       | 56.37 |
| helicopter            | 46.30 | 48.22 | 53.58| 47.67       | 49.27 |
| mAP50                 | 54.13 | 60.67 | 61.01| 69.56       | 69.10 |

From the table, we can see that the algorithm of this paper and RoI Transformer algorithm achieve the best detection accuracy among the algorithms, among which the accuracy of RoI Transformer algorithm is slightly higher than that of this paper by 0.46%, however, RoI Transformer algorithm uses the engineering method of multiscale testing, and this paper does not use the engineering method of multi-scale training and multi-scale testing to improve the model detection accuracy, and among the
15 types of targets, the algorithm of this paper achieves the best accuracy in 7 types of targets, among which the detection accuracy of aircraft targets is close to 90%, which achieves more accurate target detection.

Figure 6 shows some of the detection results of this paper's algorithm on the DOTA remote sensing dataset, and it can be seen that this paper's algorithm has good detection performance for densely arranged targets with arbitrary orientations.

4. Conclusion
In this paper, an optical remote sensing image target detection algorithm based on rotation region is proposed to solve the detection difficulties of closely spaced and arbitrary orientation of remote sensing image targets. By comparing the performance with the current mainstream rotation detection algorithm on the DOTA remote sensing dataset, it is shown that the algorithm in this paper can solve the difficult problem of detecting closely arranged and arbitrary orientation of remote sensing targets to a certain extent.

Acknowledgments
This work was supported in part by the National Key Research and Development Program of China under Grant (2018YFF0301202).
References
[1] Y. Yao, Z. Jiang, H. Zhang, D. Zhao, and B. Cai, "Ship detection in optical remote sensing images based on deep convolutional neural networks," J. Appl. Remote Sens. vol. 11, no. 04, p. 1, Sep. 2017.
[2] K. Li, G. Cheng, S. Bu, and X. You, "Rotation-Insensitive and Context Augmented Object Detection in Remote Sensing Images," IEEE Trans. geosci. remote Sens. vol. 56, no. 4, pp. 2337-2348, Apr. 2018.
[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.
[4] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 936-944.
[5] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path Aggregation Network for Instance Segmentation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8759-8768.
[6] G.-S. Xia et al., "DOTA: A Large-scale Dataset for Object Detection in Aerial Images," Nov. 2017.
[7] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248-255.
[8] Y. Jiang et al., "R2CNN: Rotational Region CNN for Orientation Robust Scene Text Detection," Jun. 2017.
[9] J. Ma et al., "Arbitrary-Oriented Scene Text Detection via Rotation Proposals," IEEE Trans. on Multimed. vol. 20, no. 11, pp. 3111-3122, Nov. 2018.
[10] J. Ding, N. Xue, Y. Long, G.-S. Xia, and Q. Lu, "Learning Roi Transformer for Oriented Object Detection in Aerial Images," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 2844-2853.