Research on Remote Sensing Image Target Detection Methods Based on Convolutional Neural Networks

Haiwen Chen¹, Lu Wang¹*, Lei Zhang², Yanping Li¹, Yuting Wu¹ and Jun Qi¹

¹ Qinghai University, Xining 810016, Qinghai, China
² College of Computer Science, Sichuan University, Chengdu, Sichuan, China
Email: haiwen5020@163.com

Abstract. Remote sensing image is extremely important for the study of natural and social phenomena and environmental analysis, remote sense image with a complex background, the goal is relatively small, arbitrary direction and disordered distribution, difficult access to information. With the continuous development of convolutional neural networks, great achievements have been made in target detection on natural images, which provides a good solution for obtaining remote sensing image information. How to improve the existing detection algorithms to make them more suitable for the characteristics of remote sensing images has become the main research direction in the current exploration of remote sensing image information. This paper reviews the existing research results of target detection on remote sensing images in detail, and finally looks forward to the future research directions of target detection methods in remote sensing images.

Keyword. Remote-sensing image; target detection; neural network.

1. Introduction
Remote sensing satellite image can obtain a wide range of target image, has the characteristics of comprehensive and macro. Aeronautic photos or satellite images obtained from a commanding position have a much larger field of view than those on the ground, and it is not affected by the terrain and objects, provides convenient conditions for people to study various natural and social phenomena on the ground and their distribution laws. With the rapid development of artificial intelligence, the use of the convolutional neural network in target detection on natural images has become more and more mature and has been applied to robot navigation, industrial detection, automatic driving, and other fields. However, compared with natural images, remote sensing images have a more complex background and a relatively smaller percentage of targets. The biggest difference is that targets in natural images are all vertically distributed in direction, while the target direction in remote sensing images is arbitrary due to different shooting angles, in order to adapt to the characteristics of remote sensing images, many scholars proposed to use the anchor with an angle to locate the target more accurately. At present, the rotating target detection networks are all improved on the basis of single-stage or two-stage target detection networks. This paper summarizes several rotating target detection networks that appeared in recent years.

2. Remote Sensing Image Rotation Target Detection
The design of existing general object detectors is often based on the implicit assumption that the bounding box is in a horizontal position. The use of horizontal anchors to locate the target position will
lead to problems such as missed detection and inaccurate positioning of some targets. Therefore, the rotating objects detection has great research value in remote sensing images.

2.1. SCRDet

SCRDet [1] is a rotating target detection network designed and improved based on Faster-RCNN. The article recognizes that the main difficulties of remote sensing target detection mainly include small targets, densely arranged, and arbitrary directions, and improve the above problems. The network structure is shown in figure 1. For small targets, articles believed that feature fusion and effective sampling are the keys to better detection of small targets. The article improves the sampling rate of positive samples by resizing the original image to different sizes and paving anchors of different lengths. The proposed area generated by RPN may introduce a lot of noise information. The article designs a supervised multi-dimensional attention network (MDA-Net). In the pixel-based attention network, a two-channel saliency map is learned. Which shows the foreground and background scores, and multiply it with the feature map to get the enhanced object information and reduce the noise. In the last module (Rotation Branch), the smooth L1 function is improved by adding the IOU constant factor to eliminate the sudden increase in loss, and the loss function is used to enhance the efficiency of the horizontal box regression to the detection box in any direction. The validity of the network was verified in the DOTA [2] data set labeled with the rotating bounding box and the NWPU VHR-10 [3] data set labeled with the horizontal bounding box. The results were superior to the results based on R2CNN [4], and the average accuracy of multiple categories (mAP) reached 72.61% and 91.75%.

![Figure 1. SCRDet model.](image)

2.2. ReDet

Reference [5] proposed that ordinary CNN networks are sensitive to the direction of the target, and do not explicitly model the direction changes. The same target is input to the neural network at different rotation angles to obtain different target features. Therefore, the ordinary CNN network requires a large number of rotational the augmented data to train an accurate target detector. The article added rotation-equivariant network to backbone network, generates the rotation-equivariant features. And propose new the RoI Align (RiRoI Align), complete rotation-invariant features in spatial dimension and directional dimension are extracted from the rotation-equivariant features, the network structure is shown in figure 2. Extended experiments in versions 1.0 and 1.5 of the DOTA [2] dataset and in the HRSC2016 [6] dataset show that the proposed method can achieve the most advanced performance for aerial object detection tasks. Compared to the previous best results, the performance of 1.2%, 3.5%, and 2.6% mAP was improved, respectively, while the number of parameters was reduced by 60%.
2.3. CFC-Net

Reference [7] proposes most target detection algorithms rely on shared features for classification and regression. The article decouples shared features and proposes the concept of key features, which represents the features required for accurate classification or positioning. By experiments and phenomena observation to prove the importance of key features in high precision target detection. The article optimizes the single-stage detector from three aspects: feature representation, anchor refinement, and training sample selection. Specifically, the article proposes Polarization Attention Module (PAM) to avoid feature interference between different tasks, effectively extract key features of specific tasks, and generate rotation-equivariant feature pyramids that are favorable for classification tasks and targets border feature pyramid that is favorable for regression tasks. The generation of the rotating anchor method is different from the increase in the number of anchors in most target detection networks. The article proposes a rotating anchor optimization module (R-ARM), which generates high-quality anchors based on key regression features, the number of anchors can be reduced while the dependence of anchors on prior geometry knowledge can be reduced, after refining the initial anchor, the rotating anchor which is more consistent with the key regression feature is obtained. Finally, different from the traditional classification and regression method of selecting anchor with the ground truth value of IOU greater than 0.5, the paper proposes a dynamic anchor learning method (DAL) to measure the degree of matching capacity to capture the key features of the anchor, thereby screening positive samples with high positioning potential are generated to achieve high-quality detection performance, the network structure is shown in figure 3. The mAP on the UCAS-AOD [8] data set reached 89.49%, and the mAP on the DOTA [2] data set reached 73.5%.

**Figure 2.** ReDet model.

**Figure 3.** CFC-Net model.
2.4. R3Det

R3Det [9] designed a fast, accurate and end-to-end rotating target detector based on the improvement of the RetinaNet single-stage network. In order to combine the advantages of the horizontal anchor can achieve a higher recall rate with less quantity and the adaptability of the rotating anchor to dense scenes, the whole network is divided into two stages. The horizontal anchor is used in the first stage to obtain faster speed and produces more candidate anchor, and then the horizontal anchor is refined and rotated in the refining stage to adapt to the dense scene and reduce the missed detection rate. Many detectors use the same feature mapping for classification and regression, without considering the feature offset caused by changes in the position of the bounding box, which is not good for categories with large aspect ratios or small sample sizes. The article proposes a refinement module to address the problem of feature misalignment. The refinement module uses bilinear interpolation to re-encode the position information of the current refined bounding box into the corresponding feature points, and then rebuild the entire feature map to achieve feature alignment. R3Det with MobilenetV2 [10] as the backbone on the HRSC2016 [6] data set can achieve an accuracy rate of 86.67% and a speed of 20fps. The mAP on the UCAS-AOD [8] dataset was 96.17%.

3. Method Comparison

The second part is a detailed introduction to the existing rotating target detection methods. Now the performance of these methods is compared. The results are shown in table 1. The evaluation index is the average accuracy of mAP in the DOTA data set. It shows that the ReDet network adopts the rotation-equivariant network as the backbone network and achieves the highest average accuracy of 76.2%. Among the three networks with ResNet-101 as the backbone, R3Det has the highest average accuracy.

| Algorithm   | Backbone network | mAP (%) |
|-------------|------------------|---------|
| SCRDet [1]  | ResNet-101       | 72.6    |
| ReDet [5]   | ReR50-ReFPN      | 76.2    |
| CFC-Net [7] | ResNet-101       | 73.5    |
| R3Det [9]   | ResNet101        | 73.7    |

4. Remote Sensing Target Detection Data Set

Target detection algorithms based on convolutional neural networks are driven by huge data sets. However, because the imaging technology of high-resolution remote sensing images is difficult, the source image is difficult to obtain, and the processing steps are more complicated and difficult to label, so there are fewer target detection data sets for remote sensing images, it becomes an important factor restricting the development of remote sensing image target detection. In recent years, with the continuous exploration of the value of remote sensing images and the successful launch of many remote sensing satellites in China, multiple high-resolution remote sensing image target detection datasets have been published one after another.

(1) DOTA [2]: Remote sensing target detection field data set, using oriented bounding box labeled, Version 1.5 contains 400,000 annotated object instances in 16 categories, Including many small object instances of about 10 pixels or less.

(2) VEDAI [11]: Vehicle detection data set in aerial images, labeled with oriented bounding box. There are a total of 1210 images, 3640 vehicle instances, including 9 categories, including boat, car, camping car, plane, pickup, tractor, truck, vans and other categories. Raw images have 4 uncompressed color channels (three visible color channels and one near-infrared channel).

(3) UCAS-AOD [8]: Remote sensing target detection field data set, labeled with a horizontal bounding box, containing two categories of aircraft and vehicles. Specifically, the aircraft data set consists of 600 images and 3,210 aircraft instances, and the vehicle data set consists of 310 images and 2,819 vehicle instances.
5. Conclusions and Prospects

The focus of research on rotating target detection algorithms is how to locate a relatively small target from the complex background, how to get the feature layer with rich target information, and how to efficiently obtain the most accurate detection box in any direction with only adding a small number of parameters. Combining with the existing detection algorithms, this part looks forward to several potential research directions in view of the difficulties in remote sensing image research.

1. Use the attention mechanism to distinguish the foreground and background of the image. Because the remote sensing image has a wide field of view to obtain the target image, it is inevitable to introduce other messy instance noises. The essence of the attention mechanism is to quickly find the area of interest, ignore other unimportant information. In the future, for target detection in remote sensing images, the attention mechanism can be used to effectively select information and perform a more accurate detection of regions of interest.

2. Based on the network structure of the non-anchor detection, there are many small targets on the remote sensing image and the distribution is irregular. To ensure a positive sample rate, increasing the number of preset anchors is the most direct method, but too many anchors will slow down the detection speed. Therefore, an anchorless detection network is designed to ensure detection accuracy and avoid the decrease of detection speed.

3. Utilize the multi-spectral information of remote sensing images. Hyperspectral remote sensing images are different from natural images and only have RGB three-channel information. Hyperspectral remote sensing images have more spectral information. And each material has different spectral response characteristics in different wavelength bands. How to use this spectral information to improve the performance of target detection is also an interesting direction.

Acknowledgments

This paper is supported by the Youth Natural Science Foundation of Qinghai Provincial Department of Science and Technology (No. 2021-ZJ-952Q).

References

[1] Yang X, Yang J, Yan J, et al. 2019 SCRDet: Towards more robust detection for small, cluttered and rotated objects Proceedings of the IEEE/CVF International Conference on Computer Vision pp 8232-8241.

[2] Xia G S, Bai X, Ding J, et al. 2018 DOTA: A large-scale dataset for object detection in aerial images Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 3974-3983.

[3] Cheng G, Han J, Zhou P, et al. 2014 Multi-class geospatial object detection and geographic image classification based on collection of part detectors ISPRS Journal of Photogrammetry and Remote Sensing 98 119-132.

[4] Jiang Y, Zhu X, Wang X, et al. 2017 R2CNN: Rotational region CNN for orientation robust scene text detection arXiv preprint arXiv:1706.09579.

[5] Han J, Ding J, Xue N, et al. 2021 ReDet: A Rotation-equivariant detector for aerial object detection arXiv preprint arXiv:2103.07733.

[6] Liu Z, Yuan L, Weng L, et al. 2017 A high resolution optical satellite image dataset for ship recognition and some new baselines International Conference on Pattern Recognition Applications and Methods pp 324-331.

[7] Ming Q, Miao L, Zhou Z, et al. 2021 CFC-net: A critical feature capturing network for arbitrary-oriented object detection in remote sensing images arXiv preprint arXiv:2101.06849.
[8] Zhu H, Chen X, Dai W, et al. 2015 Orientation robust object detection in aerial images using deep convolutional neural network 2015 IEEE International Conference on Image Processing (ICIP) pp 3735-3739.

[9] Yang X, Liu Q, Yan J, et al. 2019 R3det: Refined single-stage detector with feature refinement for rotating object arXiv preprint arXiv:1908.05612.

[10] Sandler M, Howard A, Zhu M, et al. 2018 Mobilenetv2: Inverted residuals and linear bottlenecks Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition pp 4510-4520.

[11] Gleason J, Nefian A V, Bouyssounousse X, et al. 2011 Vehicle detection from aerial imagery 2011 IEEE International Conference on Robotics and Automation pp 2065-2070.