Detection of Artificial Objects in Remote Sensing Image Based on Deep Learning

Yuan Dai [1] , Jinsheng Xiao [1, *], Benshun Yi[1], Junfeng Lei[1], Zhiyi Du[1]
Electronic Information School, Wuhan University, Wuhan, Hubei, China - (dai_yuan, xiaojs,yibs)@whu.edu.cn

Commission VI, WG VI/4

KEY WORDS: Object detection, Remote sensing image, Deep learning, Rotation Region Proposal Network

ABSTRACT:

Aiming at multi-class artificial object detection in remote sensing images, the detection framework based on deep learning is used to extract and localize the numerous targets existing in very high resolution remote sensing images. In order to realize rapid and efficient detection of the typical artificial targets on the remote sensing image, this paper proposes an end-to-end multi-category object detection method in remote sensing image based on the convolutional neural network to solve several challenges, including dense objects and objects with arbitrary direction and large aspect ratios. Specifically, in this paper, the feature extraction process is improved by utilizing a more advanced backbone network with deeper layers and combining multiple feature maps including the high-resolution features maps with more location details and low-resolution feature maps with highly-abstracted information. And a Rotating Regional Proposal Network is adopted into the Faster R-CNN network to generate candidate object-like regions with different orientations and to improve the sensitivity to dense and cluttered objects. The rotation factor is added into the regional proposal network to control the generation of anchor box's angle and to cover enough directions of typical man-made objects. Meanwhile, the misalignment caused by the two quantifications operations in the pooling process is eliminated and a convolution layer is appended before the fully connected layer of the final classification network to reduce the feature parameters and avoid over-fitting. Compared with current generic object detection method, the proposed algorithm focus on the arbitrary oriented and dense artificial targets in remote sensing images. After comprehensive evaluation with several state-of-the-art object detection algorithms, our method is proved to be effective to detect multi-class artificial object in remote sensing image. Experiments demonstrate that the proposed method combines the powerful features extracted by the improved convolutional neural networks with multi-scale features and rotating region network is more accurate in the public DOTA dataset.

1. INTRODUCTION

Object detection aims to recognize and localize object, including pre-processing, feature extraction and classification. It is divided into two categories, one based on combination of traditional image processing and machine learning algorithm and the other based on deep convolution neural network. The machine learning based algorithms mainly extract hand-crafted or shallow features with limited representation power, such as the histograms of oriented gradients (Dalal and Triggs, 2005) and then input these features into the classifier such as support vector machine(Cortes and Vapnik, 1995). Meanwhile, the diversity and complexity of the background as well as different perspectives change interference detection performance.

So far, many object detection(Liu et al., 2016) research based on deep learning (LeCun et al., 2015) have been proposed due to the large success of deep learning in the natural scene object detection. However, the detection in very high resolution remote sensing images performs poor when achieved by directly applying the existing algorithms because of the scale diversity, the object orientation and complex background(Xiao et al., 2018). Since the AlexNet won in the ImageNet(Deng et al., 2009; Krizhevsky et al., 2012) competition, some kinds of object detection algorithms were proposed successively. The object detection algorithms are mainly divided into One-Stage and Two-Stage. The YOLO(Redmon et al., 2016), proposed in the 2016, was a typical One-Stage method and improved to be faster. While RCNN(Girshick et al., 2014), Fast RCNN(Girshick, 2015) and Faster RCNN (Ren et al., 2017) are the representative of the Two-Stage algorithms. The Mask RCNN(He et al., 2017) is based on the instance object segmentation by appending a mask branch and proposing a more accurate matching strategy.

One of the key challenge in the remote sensing image target detection is the arbitrary orientation and scale variability. The remote sensing image was captured through aerial photography equipment or artificial satellite. There are some similarities between the object detection of remote sensing and the text detection in regards of the orientation. To solve this problem, some methods have been proposed (He et al., 2016b; Yao et al., 2016; Zhang et al., 2016)[14-16]. For example, the (Jiang et al., 2017)[17] proposed a text detection algorithm based on the region proposal network which uses a region proposal network to produce the horizon candidate bounding box and then predict the rotated text box. And the (Ma et al., 2018)[18] improves the RPN in the Faster RCNN and adds rotation information to predict the rotated text boxes.

Above all, this article investigates several public remote sensing image datasets and collects high resolution remote sensing images from the Google Earth and CRESDA [19]. We proposed an end-to-end multi-category object detection algorithm based on the Faster RCNN and the deep neural network theory. As showed in the figure 1, we adopt the rotated region proposal network to generate rotated proposal. This network can propose candidate bounding boxes with angles, which are much more suitable for dense object in the remote sensing image, and improved the detection accuracy. Meanwhile, the RoIAlign is applied to eliminate the unnecessary misalignment in the pooling stage. And a modification is made in the classification layer by appending a convolution layer to reduce the parameters in the feature map, which can avoid the over-fitting and improve the classifier performance. After comprehensive

[*] Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W10-321-2020 | © Authors 2020. CC BY 4.0 License.

evaluation with several state-of-the-art algorithms such as YOLO v2, YOLO V3 and Faster RCNN, the experimental results showed that the method we proposed can achieve better performance in the remote sensing image object detection which demonstrated the effectiveness of our method

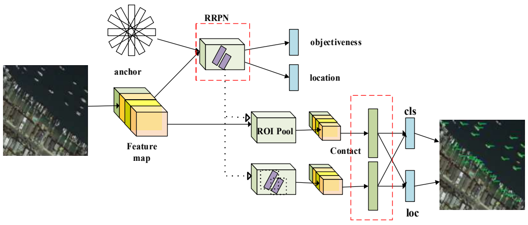


Figure.1 The proposed network

2. FASTER R-CNN

Faster RCNN was proposed as one of the state-of-the-art detection framework. The framework is shown in Figure 2. It takes VGG 16 (Simonyan and Zisserman, 2015)to act as the backbone network .

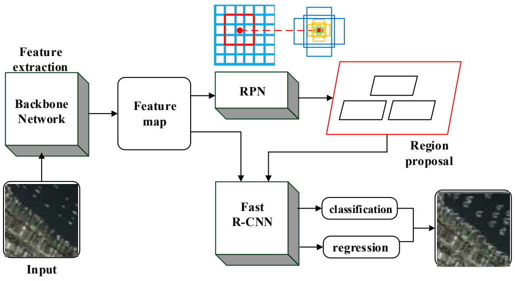


Figure.2 The framework of Faster RCNN

The first stage of Faster RCNN is a backbone feature extraction network. Each layer of the convolutional network utilizes the feature map obtained from the former layer to extract more abstract feature.

The RPN is critical in the Faster RCNN to propose candidate region box. The architect of the RPN is showed in the figure 2. It can share the feature extraction network with the subsequent classification network. The RPN utilizes a 3*3 window to slide on the feature map generated after the shared convolution layer and each window corresponds to 9 bounding box with different scale and ratio mapping to the input image.

The last stage of Faster RCNN is the RCNN, a classification network. The input is the object proposal region produced by the RPN network. Then the feature of the proposal region is extracted and the network accomplishes classification by using these features. Fully connected layer is used to output a confidence score for each probable object.

The original Faster RCNN use part of VGG Net to function as the feature extractor and ignore some features which are vital to the small object detection. Due to the orientation of the objects in the satellite images, some objects, for example, the ship and the car, are densely peaked in the images. And the recognition and detection would be influenced by the direction of the target. In order to address these challenges, three improvements will be described in the next section including the backbone network, the region proposal network and the classification network.

3. THE DETSILS OF OUR METHODS

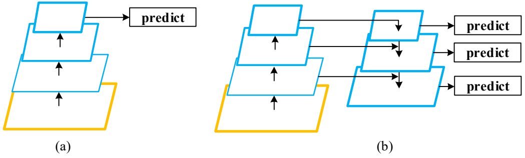
In this paper, an improved algorithm is proposed to achieve the multi-class object detection and object orientation, including: first, apply ResNet and FPN in the backbone network to construct feature pyramid network and achieve multi-scale object detection; and second, redesign the RRPN to replace the RPN in the Faster RCNN to propose rotated proposal region; third, use RoIAlign to accomplish more accurate matching and add a convolution layer in the classification network. The overall framework of the proposed method is shown in the Fig.1.

3.1 Multi-scale feature extraction

Due to the high-resolution of the remote sensing images, the ResNet-101 is used as the backbone feature network. At the same time, multi-scale feature is extracted to make the network more robust to the small object detection.

The Faster RCNN apply the feature maps pooled by the topmost layer of VGG Net with low resolution. However, this method only concentrates on the high-level semantic feature while neglecting the low-level feature that brings much more location details, which may lose information of small objects. At present, ResNet(He et al., 2016a) is one of the state-of-the-art backbone networks. Innovated by the grate performance of ResNet, we improved the detection accuracy by replacing the VGG Net with the ResNet.

When the high-level and low-level features are combined, multi-scale information will be used. According to the(Lin et al., 2017) [22], feature pyramid is used in the feature extraction stage to improve the final detection performance, as shown in the Fig.3.



(a) Faster RCNN feature extraction; (b) Improved feature extraction

Figure.3 multi-feature extraction

Therefore, a bottom-to-up, top-to-down path and a horizontal connection pathway are built. As the equation shows, the ROI with width w and height h are allocated to the feature pyramid:

$$k = [k_0 + \log_2(\sqrt{wh} / 224)] \qquad (1)$$

The top-to-down pathway combines the low-level and high-resolution function to up sample in the high-level semantic information. And then these features are connected to the former features horizontally to strength the high-level features.

So the small object can be processed with the following reasons: First is that this structure can utilize much more high-level semantic information compared to the method only using the last convolution features; Second is that operating in the larger feature map can increase the resolution and then acquire more useful information of the small target.

3.2 Rotated Region proposal network

Faster RCNN uses 9 anchor boxes to produce candidate bounding boxes with 3 different scales and 3 different ratios. However, for the very high resolution satellite image, the shape and scales have much difference. Thus, the original parameters cannot detect all of the objects perfectly.

Firstly, an angle parameter is used to control the direction of the proposal anchor boxes, and we tried different angles in the experiments to cover as much directions as possible. And to ensure the balance between the computation complexity and detection performance, we selected six different angles in the following tests. Secondly, the aspect ratios are set to 1, 1:2, 1:3, 1:4, 1:5, 1:6, 1:7, 1:8 due to the special ratios of the

This contribution has been peer-reviewed.

targets in the remote sensing images. Moreover, we designed the scales as 4, 8, 16 and 32 in that some object are relatively small.

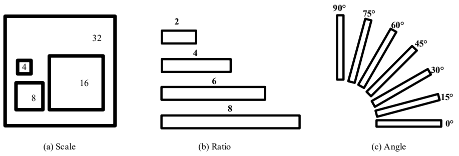
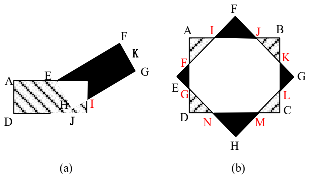


Fig.4 the anchor strategy in our method

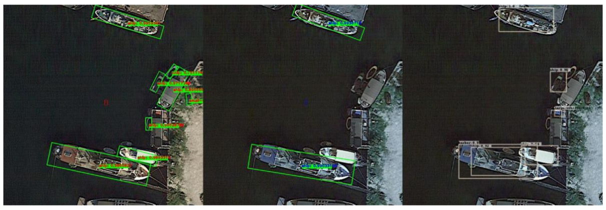
The proposal generated by our network is rotated with angles. The traditional IoU calculation used in the axisymmetric boxes is not suitable in the RRPN network, which may cause inaccurate learning and lead to poor detection performance. The IoU value between two horizontal bounding boxes are very simple. However, it becomes troublesome to calculate the intersecting location of two rectangles with rotated angles. As shown in the Fig.5, the intersecting shape of two rotated boxes is uncertain. In our method, we firstly calculate the intersecting point of the two rectangles and the vertex of one of the rectangles in the other rectangles. And then we can calculate the area of the polygon encircled by these points to compute the IoU value.



(a) Regular intersection; (b) Irregular intersection
Fig.5 the IoU calculation in the rotated box

The conventional NMS merely consider the IoU factor but it' s not suitable in the rotated box processing. For example, the IoU of the box with extreme aspect ratio, such as 1： 8 and relatively low angel, such as π /12, is 0.31 (less than the threshold) but it can be treated as a positive sample. In our method, the IoU and the angle are considered simultaneously. It consists of two phase: （i） the maximal IoU is reserved of the box with higher IoU than 0.7; （ii） the minimal angle is reserved if all IoU is between 0.3 and 0.7 (which should be less than π /12） .

The Fig.6 shows three different detection results for different NMS and boxes. The rotated NMS means to calculate IoU between two rotated boxes. Compared to (a), (b) missed several objects which located near each other. Figure (c) shows that some horizontal boxes are over-lapping. For the dense rotated targets, the traditional NMS would miss some object because the IoU may be high between the axisymmetric boxes while low in the rotated NMS.

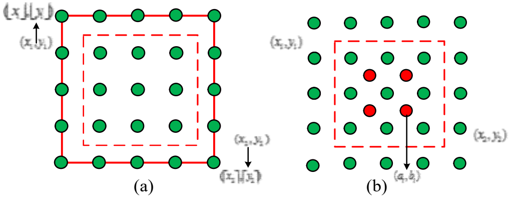


(a) Rotated NMS for rotated boxes (b)Traditional NMS for rotated boxes (c)Traditional NMS for horizontal boxes
Fig.6 Traditional NMS vs. Rotated NMS

3.3 Pooling layer and classification layer

There are two quantification steps in the RoIPooling. Firstly, the input image is fed into the feature map through the convolution layer and the candidate frame location changes from float to the nearest whole number. Secondly, there are rounding operation when the RoIPooling was executed to localize each anchor boxes.

RoIAlign was proposed in Mask R-CNN [13] and it can eliminate the misalignment produced in the quantification. As shown in Fig.7, the key idea of the RoIAlign is to cancel the quantification steps and use bilinear interpolation method to convert the pixels of the image to float(Jiang et al., 2018). Thus, the whole procedure was transformed to continuous operation.



(a) RoI Pooling; (b) RoI Align
Fig.7 RoI Pooling and RoI Align

The fully connected layer of the CNN contains too much parameters. According to the research of Min Lin (Lin et al., 2014)[23], the fully connected layer easily leads to over-fitting and weaken the generalization ability of the network. On the basis of the theory, some modification is conducted by adding one convolution layer before the fully connected layer to reduce the parameters to strength the classifier. 3*3 convolution kernel is used to accelerate computing at the same time. Besides, the operation can avoid the over-fitting which is caused by the large dimension of fusion feature and reduce half size of the feature to speed up the subsequent calculation.

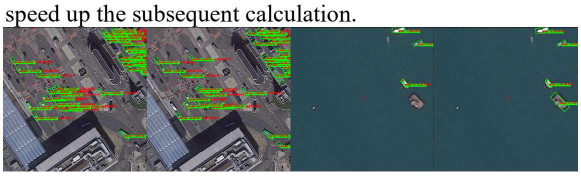


Fig.8 RRPN vs our method

As the Fig.8 shows, we compared our method with the original RRPN and the first and third figures represent the original RRPN and the second and fourth one stand for our method results. It shows that our methods can improve the detection performances.

Our improvements to the primary Faster R-CNN algorithm are three aspects: 1) the ResNet is used to function as the feature extraction network and the FPN network is introduced to fusion the multi-feature; 2) the RRPN is used to propose rotated region; 3) substitute the RoIPooling with the RoIAlign and then add a convolution layer in the classification network.

4. EXPERIMENTS

4.1 Dataset and Implementation Details

In this paper, we investigated and analysed several public dataset of satellite images for the task of object detection in remote sensing images. The DOTA(Xia et al., 2018) is selected as base dataset and we request some data from the websites of Google Earth and CRESDA at the same time. We labelled the collected data and applied them in the experiments. The dataset contains 2806 images for 15 categories, including 1411 images for training, 458 images for validation and 937 images for testing. And in this paper, we mainly focus on 10 categories object, including bridge, small vehicle, harbor, storage tank, large vehicle, plane, tennis court, helicopter, ship and swimming pool.

An end-to-end training strategy is used, which proved to be better than the alternative training. In the whole training period, there are 4 loss function, 2 in the RPN phase while 2 in the RCNN phase. We utilize the pre-trained ResNet101 model that trained on the Image Net [8] dataset in our training process.

In the training and the testing process, the parameters setting are as follows:

(1) Pre-processing: Input images size is set to 800×800 and the learning rate is 0.0003;

(2) Parameters of the RRPN: location loss weight is 1/7, classification loss weight is 2, IoU positive sample threshold is 0.7, negative sample threshold is 0.3, the threshold of the NMS is 0.7, anchor scale is [4, 8, 16, 32], anchor aspect ratios are [1, 1 / 2, 2., 1 / 3, 3, 5, 1 / 4, 4, 1 / 5, 6, 1 / 6, 7, 1 / 7], anchor angle is [-90, -75, -60, -45, -30, -15];

(3) Parameters of the Fast R-CNN: location loss weight is 4, classification loss weight is 2, the ROI pooling size is 14, the threshold of the NMS is 0.2, IoU positive sample threshold is 0.6.

4.2 Experiments and Analysis

Different Feature Extraction Network.

In order to evaluate whether the backbone network improvement can make a difference to the final detection or not, we use the VGG 16 and ResNet-101+FPN as our backbone and experiments are carried out on the DOTA dataset.

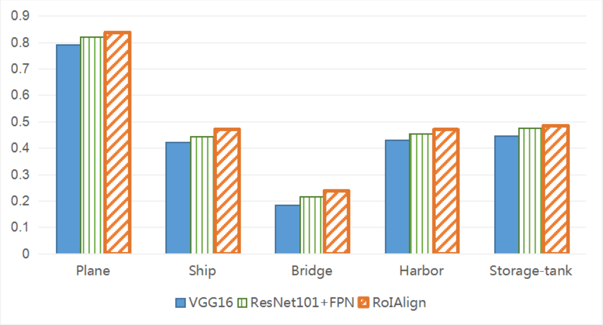


Fig.9 AP of different feature extraction network. (blue: with VGG backbone; green: with ResNet101+FPN; orange: RoIAlign)

We adopted one of the widely used criteria to quantitatively evaluate the detection performance, namely, average precision (AP). The higher AP value means the better the performance. And the Figure 9 shows the AP value we calculated and the green line denotes the improvement with backbone network.As shown in Figure 9, accuracy was improved apparently when the Res-Net101+FPN was used to act as the backbone network. The inference time cost of the VGG 16 is 1.87s while the Res-Net101+FPN cost 2.14s. This computation increase may be caused by the latter's deeper layers and much more feature parameters.

Different Pooling methods.

The RoIPooling method and the RoIAlign method are used separately for evaluation on the DOTA dataset. The Figure 9 shows the AP value of 5 categories object and the orange line denotes the improvement with RoIAlign method. In Figure 9, we evaluate the comparison of the two approaches of the five categories of object. In Table 1, the test accuracy and recall rate of the 5 categories of typical ocean targets are statistically analysed.

Categories	Precision	Recall	AP
Bridge	55.9	32.5	23.9
Harbor	70.1	54.0	47.0
Storage-tank	82.6	50.2	48.5
Plane	92.5	55.8	83.8
Ship	74.1	52.8	47.4

Table 1. Precision, Recall and AP of the RoIAlign.

For the target is relatively dense, RoIAlign reduces the pixel deviation during the pooling process, thus the detection accuracy is improved compared with the original RoIPooling method. The costing time is 2.14s for the RoIPooling method while 2.18s for the RoIAlign method, which means that the modified network will not add too much calculation complexity.

Different classification network.

The original classification network and a classification network with a newly appended convolution layer are evaluated. And the corresponding statistic values of the two methods are shown in Figure 10.

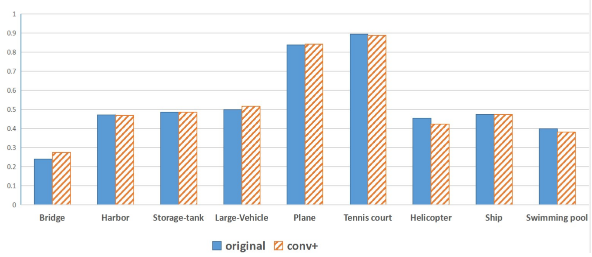


Fig.10 AP of different Classification network. (blue: original; orange: with convolution layer appended)

The improved classification network method increased the detection accuracy of some categories, such as bridges, vehicles, track and field fields, soccer fields, airplanes and etc. At the same time, the average time for the original classification network is 2.14s, while 1.98s for the improved classification n network. This proved that the improvement remains the computation complexity

Overall algorithm comparison.

As shown in Table 2, the detection performance of airplanes and tennis courts is better (the precision value is over 80%), while the bridge detection is of poor performance (the precision value is less than 30%). This may be caused by the fact that airplanes and tennis courts have apparent shape, colour and texture features, and the environment is relatively simple and easier to recognize. However, ships and ports are generally densely packed, with varying shapes and scales, and bridges tend to be occluded by the complex background, making the accurate recognition difficult.

Categories	Precision	Recall	AP
Bridge	59.15	32.92	26.38
Small -Vehicle	68.65	44.12	34.15
Harbor	77.95	61.41	56.18
Storage-tank	81.40	52.82	51.55

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLII-3-W10-321-2020 | © Authors 2020. CC BY 4.0 License.

Large-Vehicle	62.68	76.20	56.91
Plane	94.10	87.36	86.52
Tennis court	97.04	91.29	91.15
Helicopter	82.05	65.31	61.91
Ship	74.01	55.35	50.15
Swimming pool	71.60	53.45	47.55

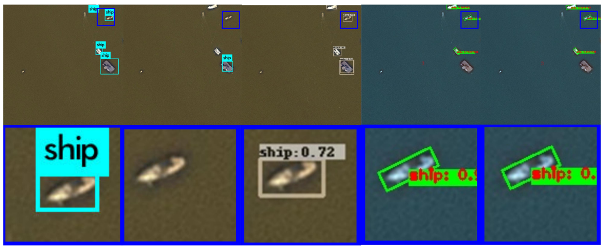
Table 2. Statistical results of the experimental. %

In order to demonstrate the effectiveness of our proposed algorithm, we also conduct comprehensive comparison with several state-of-the-art deep learning object detectors, including YOLO v2, YOLO v3, Faster R-CNN and the original RRPN algorithm. Here, we adopted three widely used criteria to quantitatively evaluate the performance of detection, namely, the precision, recall and average precision (AP). These quantization values of detection accuracy are shown in Table 3.

Categories	YOLO v2	YOLO v3	Faster RCNN	RRPN	Proposed
Bridge	14.18	10.03	41.82	23.88	**26.38**
Small -Vehicle	13.08	14.79	3.85	34.65	34.15
Harbor	51.99	17.07	59.04	47.3	56.18
Storage-tank	40.21	24.59	5.31	48.77	**51.55**
Large-Vehicle	22.02	9.09	38.94	49.74	**56.91**
Plane	80.91	49.44	38.74	83.89	**86.52**
Tennis court	72.52	15.18	89.75	89.4	**91.15**
Helicopter	21.22	0.02	40.64	45.44	**61.91**
Ship	46.73	30.31	3.99	47.19	**50.15**
Swimming pool	34.31	7.54	22.71	39.78	**47.55**
mAP	39.72	17.81	34.48	51.00	**56.25**

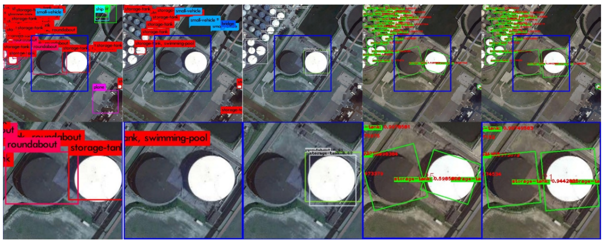
Table 3. Different algorithm comparison results. %

As the Table 3 shows, the optimal values in each category are emphasized by the bold numbers. It's noteworthy that for most class of objects detection, including bridge, storage-tank, large-vehicle, plane, tennis court, helicopter, ship and swimming pool, our proposed method can achieve the best performances. It's proved that the rotated candidate region boxes predicted from the improved network by adopting RRPN and earlier mentioned improvements can truly improve the detection performance of the remote sensing image objects with different directions, shapes and scales apparently. However, although our method has achieved the best mAP in the 10 class object detection, the detection performances of some categories, such as bridge and small vehicle are still relatively lower than the other categories. To further improve the detection accuracy of these objects is our future task.



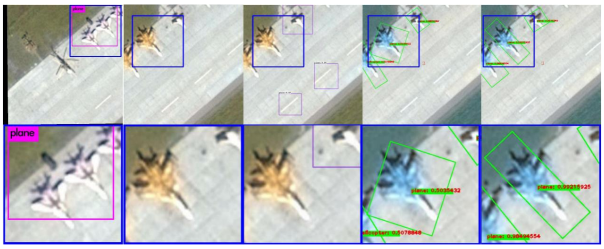
(a)YOLO v2 (b) YOLO (c) Faster RCNN (d) RRPN (e) Ours

Fig.11 Small-vehicle.



(a)YOLO v2 (b) YOLO (c) Faster RCNN (d) RRPN (e) Ours

Fig. 12 Storage-tank.



(a)YOLO v2 (b) YOLO (c) Faster RCNN (d) RRPN (e) Ours

Fig. 13 Helicopter.

As shown in Figure 11-13, the target of our test contains 15 categories. It can be seen that the method is useful for oriented and dense object detection, such as ships, storages-tanks, and helicopters. The detection results in targets such as helicopters and ships are better than the rest of the algorithms. In summary, the experimental results show that our algorithm can effectively deal with multi-class object detection problem of high resolution remote sensing images.

5. CONCLUSION

With the increasing spatial resolution of remote sensing images, convolutional neural networks have been widely used in remote sensing image scene classification, target recognition, segmentation and other fields. Aiming at the typical targets in satellite images, this paper focuses on the CNN based object detection. Considering the distinctiveness of remote sensing images, the network structure of Faster R-CNN is improved, and the accuracy of typical target detection of remote sensing images is achieved. For most of the targets in remote sensing images, they are characterized by orientation and denseness. A rotation factor is added to the regional proposal network so that it can generate candidate regions with angles. At the same time, a convolution layer is appended in front of the fully connected layer of the final classification network to reduce the feature parameters and avoid over-fitting. The experimental evaluation shows that the improved algorithm can get better detection results.

6. ACKNOWLEDGEMENTS

This work is supported by the National Key Research and Development Program of China (No. 2016YFB0502602), and National Natural Science Foundation of China (Grant No. 61471272).

REFERENCES

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273-297.

Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 886-893 vol. 881.

Deng, J., Dong, W., Socher, R., Li, L., Kai, L., Li, F.-F., 2009. ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248-255.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation, 2014 IEEE Conference on Computer Vision and Pattern Recognition, pp. 580-587.

Girshick, R.B., 2015. Fast R-CNN, 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440-1448.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN, 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980-2988.

He, K., Zhang, X., Ren, S., Sun, J., 2016a. Deep Residual Learning for Image Recognition, pp. 770-778.

He, T., Huang, W., Qiao, Y., Yao, J., 2016b. Accurate Text Localization in Natural Image with Cascaded Convolutional Text Network. CoRR abs/1603.09423.

Jiang, B., Luo, R., Mao, J., Xiao, T., Jiang, Y., 2018. Acquisition of Localization Confidence for Accurate Object Detection, in: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018. Springer International Publishing, Cham, pp. 816-832.

Jiang, Y., Zhu, X., Wang, X., Yang, S., Li, W., Wang, H., Fu, P., Luo, Z., 2017. R2CNN: Rotational Region CNN for Orientation Robust Scene Text Detection. CoRR abs/1706.09579.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks. Commun. ACM 60, 84-90.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436.

Lin, M., Chn, Q., Yan, S., 2014. Network In Network. ICLR abs/1312.4400.

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature Pyramid Networks for Object Detection, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936-944.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C., 2016. SSD: Single Shot MultiBox Detector, in: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp. 21-37.

Ma, J., Shao, W., Ye, H., Wang, L., Wang, H., Zheng, Y., Xue, X., 2018. Arbitrary-Oriented Scene Text Detection via Rotation Proposals. IEEE Transactions on Multimedia 20, 3111-3122.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You Only Look Once: Unified, Real-Time Object Detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788.

Ren, S., He, K., Girshick, R., Sun, J., 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence 39, 1137-1149.

Simonyan, K., Zisserman, A., 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition.

Xia, G., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., Zhang, L., 2018. DOTA: A Large-Scale Dataset for Object Detection in Aerial Images, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3974-3983.

Xiao, J., Tian, H., Zhang, Y., Zhou, Y., Lei, J., 2018. Blind video denoising via texture-aware noise estimation. Computer Vision and Image Understanding 169, 1-13.

Yao, C., Bai, X., Sang, N., Zhou, X., Zhou, S., Cao, Z., 2016. Scene Text Detection via Holistic, Multi-Channel Prediction. CoRR abs/1606.09002.

Zhang, Z., Zhang, C., Shen, W., Yao, C., Liu, W., Bai, X., 2016. Multi-oriented Text Detection with Fully Convolutional Networks, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4159-4167.