Object Detection of Remote Sensing Airport Image Based on Improved Faster R-CNN

View the article online for updates and enhancements.

Object Detection of Remote Sensing Airport Image Based on Improved Faster R-CNN

HAN Yongsai[1], MA Shiping[2], ZHANG Fei[3], LI Chenghao[4]

[1,2,3,4] Air Force Engineering University, No.1 Baling Road, Baqiao District, Xi'an, Shaanxi China.

Corresponding e-mail: 1013765061@qq.com

Abstract. In order to effectively improve the detection accuracy of remote sensing images in airport areas, basing on the representative deep network Faster R-CNN as the object detection method, a deeper basic network ResNet and feature fusion component FPN are used to extract more robust deep distinguishing features, and add a new fully connected layer to the end detection network and combine the softmax classifier and 4 logistic regression classifiers for object detection according to the inter-class correlation of the object. Experiments show that the improvement of the original network brings a 7.7% mAP improvement to 76.6% of the mAP. Compared with other mainstream networks, it also has a better accuracy rate. At the same time, by appropriately reducing the input amount of the proposals, the speed can be increased 3 times to 0.169s under the premise of reducing the accuracy by 2.2%. According to the specific task, the accuracy and detection speed can be reasonably weighed, which reflects the effectiveness and practicability of the network.

1. Introduction

With the great strides of deep learning, deep learning technology has become a research hotspot in various fields in recent years. Object detection based on deep learning has also shined. Deep learning has become a "standard" in the field of object detection. Zhang [1] et al. used Convolutional Neural Network (CNN) to extract image features and combined with airport linear distribution to make more effective area suggestions to detect airports. Zhu [2] et al. used Residual Net (Residual Net, ResNet) for feature extraction and airport saliency maps for region suggestions to detect airports. Chen [3] et al. combined Deep Belief Network (DBN) for aircraft detection. Wu[4] and others adopted the BING+CNN method to detect the object of the aircraft. The above method applies the powerful feature extraction capability of deep neural networks, which is more robust than artificially designed features, which greatly improves the detection accuracy, but there are still problems that too many proposals make the detection efficiency slow. In response to the above problems, many domestic and foreign improved algorithms have been proposed. Among them, Faster R-CNN [5] is one of the representative detection frameworks. The proposal of its Regional Proposal Net (RPN) has greatly accelerated the detection efficiency, including improvements in various aspects, has also made the detection efficiency reach an unprecedented new height. Its powerful classification, positioning ability and good detection efficiency for multiple types of objects have made people pay more and more attention to it, and in this framework based on their own data sets, corresponding improvements were made and good results were achieved.

Use airport area objects as detection objects, including airports, civil aviation aircraft, fighter jets, transport planes, helicopters, tanks, and bridge objects; use satellite remote sensing images as the

detection source, and use deep learning Faster R-CNN as the detection framework. It is debugged and improved to complete the object detection of the airport area. The main improvements are as follows:

a. Add a deeper basic network and feature fusion detection components to extract deeper discriminative robust features;

b. A new type of end detector with better combination of inter-class correlation characteristics of the detected objects.

2. Principle and improvements of airport area object detection

2.1. Principle

Taking as far as possible to improve the accuracy and completeness of the objects in the remote sensing monitoring tasks, Faster R-CNN is used as the basic framework for detection. The specific detection principle is as followed. When the image to be trained or detected enters the neural network, it will first enter the feature extraction network to extract its features. The feature extraction network is composed of many hidden layers, mainly convolved with many 3*3*n convolution to extract the features of the input picture, and use the pooling operation to control the size of the feature map, which is convenient for subsequent calculations, and finally obtain a complete global feature map. Next, the feature map enters the RPN and maps to the original image. With the help of the softmax classifier and the bounding box regression and the corresponding label information in the data set, the regional proposal is reduced by the NMS algorithm. The number of proposals can be selected and mapped back to the feature map. The next step is to enter the ROI Pooling layer to pool the suggested areas of interest and change it to a fixed size for subsequent input. Finally enter the full connection layer combined with softmax classifier and bounding box regression to achieve accurate classification and positioning of the object.

2.2. Data set preparation

At present, there are few satellite remote sensing image data sets in public airport areas, which is also a major difficulty in the detection of specific objects, because a more complete data set is a necessary condition for training an efficient detection network. Existing literature [6] shows that the pictures obtained from high-precision satellite detection software have high adaptability and generalization to object detection, so Google Earth is used as the source for autonomously obtaining pictures containing object information, and VOC2007 [7] is used as Create the data set in a standard format.

The dataset produced contains more than 200 different airport areas, with a total of 7264 original pictures, including 1982 airport pictures, 1838 civil aviation aircraft pictures, 565 fighter aircraft pictures, 715 helicopter pictures, 813 transport aircraft pictures, 583 pictures Bridge pictures and 768 oil tank pictures. The data set is shown in Figure 1.
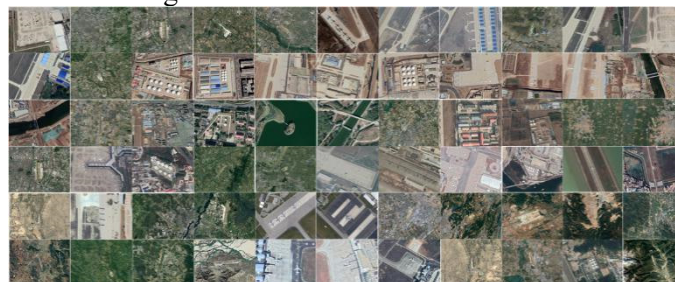

Figure 1. Data set example.

2.3. Improvements

2.3.1. Choose a better basic network and add feature fusion detection components to improve the performance of the neural network detector from the network level. Here, Use ResNet [8] as the basic network of Faster R-CNN, and FPN as the newly added detection component. ResNet puts forward the concept of "layer-jumping connection", that is, the input is directly superimposed on the output. During

back propagation, this gradient is passed back intact, retaining a strong correlation of image information. As shown in Figure 2. FPN can obtain better deep-level distinguishing characteristics by fusing different convolutional features, and send all three features (Predict-1~3) to the feature detection network. As shown in Figure 2.
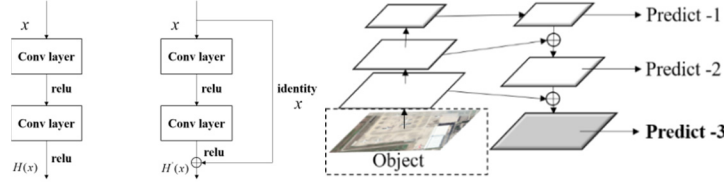

Figure 2. The sample graph of ResNet and FPN.

2.3.2. Construct an end detector that combines softmax classifier and logistic regression based on the correlation characteristics between object categories. In order to further improve the average detection accuracy of civil aviation aircraft, helicopters, fighter aircraft and transport aircraft in the aircraft category with strong correlation, reduce the false detection rate between categories. It is proposed to construct a new type of detection module at the end detection area of the Faster R-CNN detection framework, as follows:
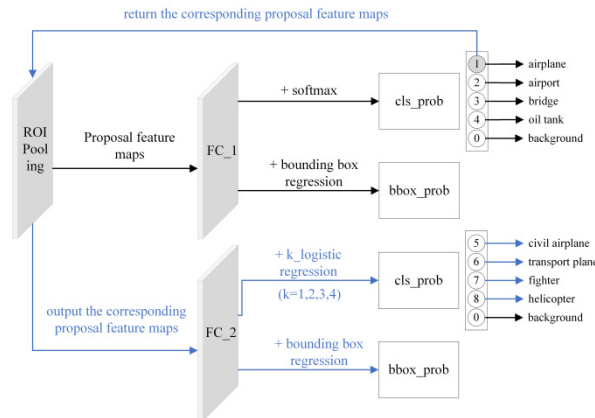

Figure 3. Schematic diagram of the new end detector.

Among the level characteristics of various objects, the features of airports, aircrafts, bridges, and oil tanks have a large degree of distinction, such as color, shape, texture, etc., while the features of aircraft objects have a relatively small degree of distinction, such as similar shapes and sizes, so theoretically there is a mutually exclusive relationship between airports, airplanes, bridges, and tanks, while aircraft objects have a strong correlation.

Softmax is a multi-class object classifier. When all categories are obviously mutually exclusive, the softmax classifier has good classification performance. The classification mechanism is as follows:

As shown in Figure 3, after the candidate feature area enters FC_1, the input information is converted into multiple K*1 dimensional variables and output to the softmax layer, and then the final confidence of each category is output by it, as shown below：

$$a^{[L]} = \frac{e^{z^{[L]}}}{\sum_{i=1}^{K} e_i^{z^{[L]}}} \qquad (1)$$

Where K=5, which is the same as the number of predicted categories in the design; $Z^{[L]}$ is the output variable of FC_1, $a^{[L]}$ is the final output variable of softmax layer.

Logistic regression is the Special cases when the K=2 of softmax, Existing studies [9] have shown that multiple logistic regression has better classification performance than a softmax classifier in a strongly correlated data set, so when identifying aircraft-type objects such as civil aviation aircraft,

transport aircraft, fighter aircraft and helicopters , Choose to use 4 logistic regression instead of softmax classifier.

3. Experimental results

3.1. Environment and parameter setting
CPU: Core i7-7700 @3.60GHz; RAM: 16.0GB; Graphics Card: GTX 1080Ti @6GB; Operating System: Ubuntu; Frame: Caffe. Use the model transfer learning method to pre-train the network with imagenet_models. The initial learning rate is set to 0.001, the momentum is set to 0.9, the weight decay is set to 0.0005; the initial threshold is set to 0.7; the training iteration round is set to 80,000 times; the training set is verified by the random number method according to the training set: validation set: test set = 7: 2 :1.

3.2. Results analysis
In a highly consistent experimental environment, we conducted a comparative experimental test on different basic networks and before and after adding detection components using the independently calibrated data set. The results are shown in Table 1.

Table 1. Basic network performance comparison.

Model	mAP (%)	Average IOU
ZFNet	54.4	0.390
VGG_CNN_M_1024	59.3	0.419
VGG-16	68.9	0.563
VGG-19	70.1	0.569
ResNet-50	70.3	0.570
ResNet-101	71.9	0.571
ResNet-50 + FPN	74.8	0.642
ResNet-101 +FPN	**76.3**	**0.643**

Experimental results show that as the network depth increases, the model's mAP and Average IOU will increase more and more, because as the network depth increases, the feature extraction ability will become stronger and stronger, and the feature expression ability will also become better. ResNet-101 has a 3.0% mAP improvement compared to VGG-16. After adding FPN detection components, the effective feature fusion operation has improved 4.4% mAP and 7.2% Average IOU based on ResNet-101, which proves the effectiveness of adding FPN feature fusion.

Based on the above experiments, the network after adding the new end detector is defined as L1, and the original network is L2. The detection results are shown below.

Table 2. End detector effect comparison.

Model	AP(%)							mAP
	airport	bridge	oil tank	Civil airplane	Transport plane	fighter	helicopter	
L1	84.9	72.4	73.3	84.9	74.9	68.6	74.9	76.3
L2	85.2	72.5	73.4	85.5	75.4	69.1	75.4	76.6

Obtained from the test results, after the new detection framework constructed is applied to the aforementioned Faster R-CNN framework, the ap of aircraft objects in the airport area is increased by an average of about 0.5%, and the mAP of various objects in the airport area is improved by about 0.3%. The reason is that by adding a fully connected layer and subsequent multiple logistic regression with the softmax classifier of the original network, the model detection can be combined with the characteristics of the specific detected objects to combine a better classifier. The experimental results also prove that the method has a beneficial effect on the model detection.

In a highly consistent experimental environment, compared with other mainstream algorithms, the effect is as follows.

Table 3. Comparison of different network performance.

	R-CNN	SPP-Net	Faster R-CNN	HyperNet	R-FCN	**proposed**	**proposed-50_proposal**
mAP（%）	54.2	54.9	68.9	72.0	74.8	**76.6**	**74.4**
detect time（s）	＞10.000	0.401	0.215	0.160	0.167	**0.508**	**0.169**

Through experimental testing, it can be found that the network of this paper has a higher level than the current mainstream object detection methods R-CNN [10], SPP-Net [11], Faster R-CNN, HyperNet [12] and R-FCN [13]. The mAP is 7.7% higher than Faster R-CNN. However, due to the application of deeper feature extraction networks, FPN detection components, and the addition of new fully connected layers at the end detector, the time cost has more than doubled compared to Faster R-CNN, which is also a defect of the proposed network. Experiments have found that when the proposal is reduced to 50 (originally set to 300), the network reduces the mAP by 2.2%, but the detection time is shortened to about 1/3 (0.169s) of the original, so it can be used in specific applications. Reset according to the weight of detection accuracy and detection speed, and find the best method for the task.

4. Conclusion
The deeper ResNet-101 network and the addition of the feature fusion detection component FPN have enabled the model to extract more robust features, which has greatly improved the precise classification and accurate positioning of the model; the new end detector allows the model to allocate more suitable detectors according to the correlation between objects which has a higher mAP, and by reducing the proposals to 50 to reduce the accuracy by a certain rate, the speed is increased by 3 times, and at this time mAP is still good compared with other mainstream detection networks. In specific tasks, the accuracy and speed can be weighed to conveniently select a more suitable method. Experiments verify the effectiveness and significance of the proposed method, which has strong practical value for this type of task.

References
[1] ZHANG P, NIU X, DOU Y, et al. Airport Detection on Optical Satellite Images Using Deep Convolutional Neural Networks[J]. IEEE Geoscience & Remote Sensing Letters, 2017, 14(8):1183-1187.
[2] ZHU T H, LI Y H, YE Q K, et al. Integrating saliency and ResNet for airport detection in large-size remote sensing images[C] //2017 2nd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2017.
[3] CHEN X Y, XIANG S M, LIU C L, et al. Aircraft Detection by Deep Belief Nets[M]. 2013:54-58.
[4] WU H, ZHANG H, ZHANG J, et al. [IEEE 2015 IEEE International Conference on Image Processing (ICIP) - Quebec City, QC, Canada (2015.9.27-2015.9.30)] 2015 IEEE International Conference on Image Processing (ICIP) - Fast aircraft detection in satellite images based on convolutional neural networks[J]. 2015:4210-4214.
[5] REN S, HE K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6).
[6] CHEN F, REN R, De Voorde T V, et al. Fast Automatic Airport Detection in Remote Sensing Images Using Convolutional Neural Networks[J].Remote Sensing,2018, 10(3):P.443.
[7] Everingham M, Gool L V, Williams C K I, et al. The Pascal Visual Object Classes (VOC) Challenge[J]. International Journal of Computer Vision,2010, 88(2):p.303-338.

[8]   HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]. computer vision and pattern recognition, 2016: 770-778.

[9]   Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[J]. 2018.

[10]   Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. computer vision and pattern recognition, 2014: 580-587.

[11]   HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9):1904-16.

[12]   KONG T, YAO A, CHEN Y, et al. HyperNet: Towards Accurate Region Proposal Generation and Joint Object, Detection[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).IEEE,2016.

[13]   DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks[C]. neural information processing systems, 2016: 379-387.