

A Faster RCNN Airport Pavement Crack Detection Method Based on Attention Mechanism

Shougang Hao^{1,2}, Liming Shao^{1,2}, Sibow Wang^{1,2}

¹The Second Research Institute of CAAC, Chengdu 610000, China

²Civil Aviation Electronic Technology Co., Ltd., Chengdu 610000, China

Abstract: Airport pavement inspection is an important link to ensure the safe takeoff and landing of aircraft. At present, the airport pavement safety inspection is still dominated by manual inspection. This method has some problems such as low detection efficiency, strong subjectivity, and unable to fully cover. In this context, Faster RCNN network was used for tunnel crack detection, the backbone network of Faster RCNN was modified, the lightweight network MobileNetV2 was used to improve the network training speed, and CBAM attention mechanism was added to improve the feature extraction ability of the model.

Keywords: Faster RCNN, MobileNetV2, CBAM.

1. Introduction

At present, our country is accelerating the implementation of the strategic goal of building civil aviation power, gradually realizing the transformation from civil aviation power to power. Airport runway is one of the key points of airport construction. The runway surface of airport carries aircraft takeoff and landing, and the quality of the runway surface directly affects the safety of aircraft takeoff and landing. Cracks belong to one of the apparent structural diseases in the airport pavement. If not treated in time, it will shorten the life of the pavement with the passage of time, rain and snow and other weather influences, which will have adverse effects on the takeoff and landing of aircraft and threaten the operation safety of aircraft. If cracks are detected and repaired before other cracks develop, more serious damage to the pavement can be prevented. At present, the detection of pavement cracks in the airport is still based on the traditional manual inspection. After the manual inspection, the problem of pavement cracks is found and the information is recorded by taking photos manually. This type of inspection often requires a large amount of manpower and material resources to complete, and is subject to subjective factors. With the development of science and technology, the use of computer can efficiently and accurately detect the cracks of the surface, so as to prevent the surface diseases becoming more and more serious.

The essence of pavement crack detection is to mark the location and category of the disease in the image containing crack disease. According to the nature of the problem, it can be classified as the problem of target detection, that is, to accurately locate the object with the characteristics of a certain target type from the given image, and assign corresponding category labels to the object. With the rapid development of deep learning in recent years, the feature extraction ability of convolutional neural networks for target images has also been greatly improved, so the performance of object detection algorithms based on convolutional neural networks has been greatly improved compared with traditional object detection algorithms. Using convolutional neural network technology to realize the target detection task of tunnel surface crack image will play a substantial role in solving the problem of tunnel surface crack detection. The

development of object detection in the past two decades can be roughly divided into two historical periods: the "traditional object detection period" before 2014 and the "deep learning-based object detection period" after 2014. In the period of deep learning, object detection based on deep learning can be divided into two categories: "two-stage detection" and "single-stage detection". The two-stage detection algorithm divides the detection task into suggestion generation stage and region classification stage, which is a process from rough to fine. Two-stage detection is characterized by high precision, but relatively slow detection speed. Typical two-stage target detection methods include RCNN, SPPNet, Fast RCNN, Faster RCNN, etc. The target detection method of RCNN was proposed by Girshick in 2014. By combining the classification method of SVM classifier and linear regression, this method proposed the location extraction algorithm of the target object of R-CNN network model. The results show that RCNN can effectively improve the accuracy of target detection, but it also has some defects such as tedious steps, long model training time and slow detection speed. Girshick also proposed Fast RCNN algorithm based on ROI pooling layer in 2015, which carried out target detection and recognition by changing the number of output layers. Fast RCNN algorithm is simple to operate, but it still has the defects of long time and low efficiency. In the same year that Fast RCNN was proposed, Ren proposed Faster RCNN. By introducing RPN, this algorithm solves the problems existing in RCNN and Fast RCNN, and extracts candidate areas within the whole network, realizing end-to-end calculation in a real sense, and greatly improving the detection accuracy and speed.

In view of the good detection effect of Faster RCNN, this paper based on Faster RCNN, added an attention mechanism to the algorithm to further improve the detection accuracy of the algorithm, and modified the backbone network of Faster RCNN to improve the training speed and detection speed of the model.

2. Technical Overview

2.1. Faster RCNN

Faster RCNN is proposed to use regional suggestion network RPN to generate a series of candidate boxes to avoid

the problem of information redundancy during the generation of candidate boxes, which perfectly replaces the selective search method and improves the training and testing speed of

network modules. The structure of Faster RCNN is shown in Figure 1.

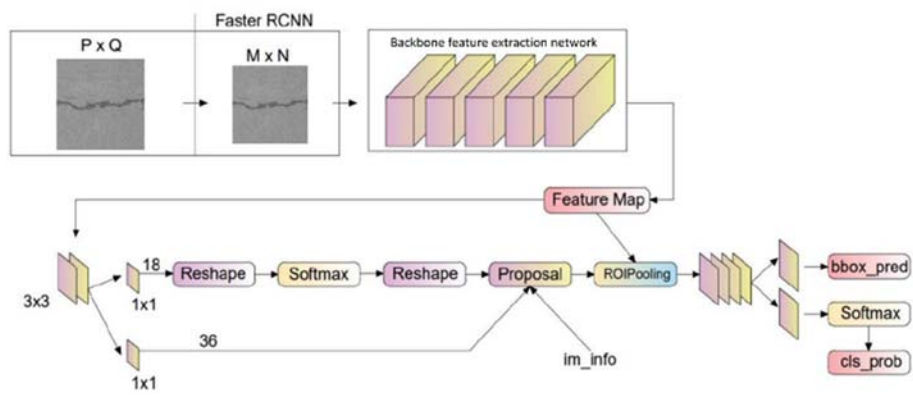


Figure 1. The structure of Faster RCNN

As can be seen from Figure 1, Faster RCNN network can be divided into four parts:

(1) Input images that need to extract features through the backbone feature extraction network. Convolutional neural networks are usually used to construct feature extraction networks. Feature images extracted by the image feature extraction layer are usually shared by regional proposal network and the region of interest pooling layer.

(2) Feature map through the regional suggestion network RPN. Regional suggestion network RPN is used to generate target candidate regions, and the full convolutional network structure is used to generate a specified number of candidate regions for each image, and the position regression is carried out and the existence of the target is preliminarily determined.

(3) Input to the area of interest pool layer. Input the feature map in the first step and the target candidate region in the second step into the region of interest pooling layer at the same time, and extract the features suggested by the corresponding region.

(4) Finally to the classification regression layer. The obtained target candidate region is further classified to output the classification of the region and accurate regression.

2.2. Improved Backbone Network

In the object detection network, the backbone feature extraction network can largely determine the prediction effect of the model. The traditional convolutional neural network has a large memory requirement for hardware equipment and a large amount of computation. In Faster RCNN, the complex network structure leads to slower training and testing speed. Considering the actual service scenario, the training quality of the whole network needs to be greatly improved to improve test efficiency, so the backbone network of Faster RCNN is replaced with lightweight network MobileNetV2.

MobileNetV2 introduces a backward residual module and a linear bottleneck layer. Firstly, the reciprocal residual model is expanded by 1*1 convolution computation channels, so as to obtain more features, then features are extracted by 3*3 convolution computation channels, and finally the channels are compressed back by 1*1 convolution computation compression. The overall process is as follows: expansion - convolution operation - compression. Finally, the basic convolution block of the MobileNetV2 network is the linear bottleneck layer, and the structure is shown in Figure 2.

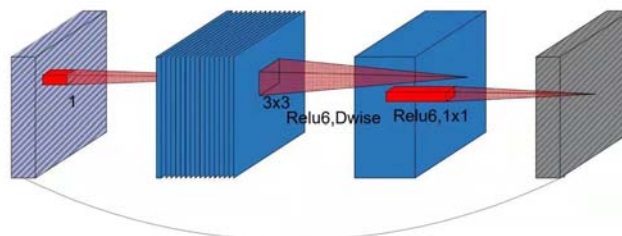


Figure 2. Inverted residual block

As can be seen from Figure 2, when the information goes through the process of expansion -- convolution -- compression in the inverted residual model, the damage characteristic of ReLU6 activation function may occur, so a linear bottleneck layer is introduced. Since the output of ReLU6 function is zero in the interval from zero to minus infinity, when the signal is mapped from low dimension to high dimension, some features will be lost anyway when the

signal is remapped to low dimension after ReLU6. If the dimension reflected by the final information is relatively high, the loss is small.

2.3. Network Convolutional Block Attention Module

The network of the Attention mechanism CBAM (Convolutional Block Attention Module) is shown in Figure

3, which includes the channel attention module and the spatial attention module. The channel attention module is computed as:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (1)$$

$$= \sigma(W_1(W_0(F_{avg}^c))) + \sigma(W_1(W_0(F_{max}^c)))$$

Where, the input is a feature F, and they are respectively sent to a two-layer neural network. The number of neurons at the first layer is C/r, the activation function is ReLU, and the number of neurons at the second layer is C. After adding the

two features, a Sigmoid activation function was used to get the weight coefficient. Multiply by the input feature F to get the new compressed feature.

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F), MaxPool(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s, F_{max}^s])) \quad (2)$$

Similar to channel attention, after a 7*7 convolution layer, the activation function is Sigmoid, and the weight coefficient M_s is obtained. Multiply by the input feature F' to get the scaled new feature.

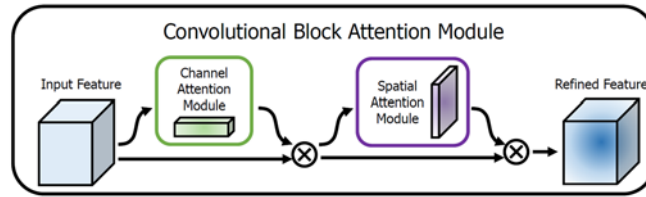


Figure 3. Convolutional Block Attention Module

The backbone network of Faster RCNN is improved. By replacing the original backbone network with lightweight network MobileNetV2, the training and testing speed can be improved. Finally, CBAM attention mechanism is added to MobileNetV2 network, so that the network can further improve the extraction ability of effective information and the extraction and resolution ability of features under high receptive field. The overall structure is shown in Figure 4.

The host video card is Tesla V100 and the video memory is 16GB. In the experimental detection process, the time required for an image to be detected by the original Faster RCNN is 85.645ms. By replacing the backbone network with a lightweight network, the detection speed of the improved Faster RCNN detection model is 62.175ms, indicating a great improvement in the detection speed. The test results of the original Faster RCNN and improved Faster RCNN are shown in Table 1.

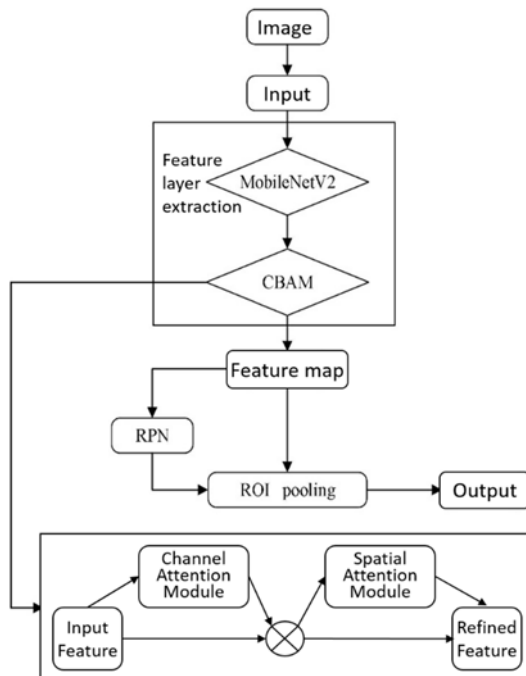


Figure 4. Improved Faster RCNN network structure

3. Experimental Results and Analysis

The pixel size of the training sample used in the experiment is 512*512. Among them, 80% of the crack samples were used to train the improved Faster RCNN network based on attention mechanism, and the remaining 20% samples were tested. The network designed in this paper is implemented with PyTorch deep learning framework in win10 environment.

Model	Precision /%	Recall/%	mAP/%
Faster RCNN	89.3	85.6	87.2
Faster RCNN +MobileV2+CBAM	95.7	88.4	89.5

As shown in Table 1, compared with the original network model, the detection effect of the improved network model has been greatly improved, with the accuracy increased by 6.4% to 95.7%. Recall increased by 2.8% and mAP increased by 2.3%. The detection effect has been significantly improved. Part of the detection effect diagram is shown in Figure 5. The detection confidence of cracks is 94% and 95% respectively, and the detection results can meet the requirements of airport runway crack detection.

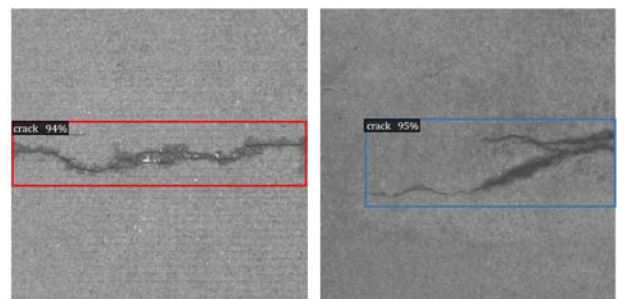


Figure 5. Test results

4. Conclusion

In this paper, for Faster RCNN network, the complex

network structure is modified, lightweight MobileNetV2 network is used to replace the original backbone network, and CBAM attention mechanism is used to improve the feature extraction capability of the network. Through experiments, the improved Faster RCNN network can be used for crack detection of airport pavement, and the detection accuracy of crack diseases meets the application requirements. In the follow-up research work, we can continue to consider improving the detection speed of network models.

References

- [1] Zou Z, Shi Z, Guo Y, et al. 2019. Object Detection in 20 Years: A Survey[J]. arXiv:1905.05055
- [2] Girshick R, Donahue J, Darrell T, et al. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Columbus, OH, USA, 580-587.
- [3] Girshick R. 2015. Fast R-CNN[C]// 2015 IEEE International Conference on Computer Vision(ICCV), Santiago, Chile, 1440-1448.
- [4] Ren S, He K, Girshick R, et al. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks[C] // Advances in neural information processing systems 2015, 91-99.
- [5] Woo S, Park J, Lee J Y, et al. 2018. CBAM: Convolutional Block Attention Module[J]. 2018 European Conference on Computer Vision, 3-19.
- [6] He K, Zhang X, Ren S, et al. 2016. Deep residual learning for image recognition[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, 770-778.
- [7] He K, Zhang X, Ren S, et al. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1904-1916.
- [8] Tong Z, Gao J, Han Z Q, et al. 2017a. Recognition of asphalt pavement crack length using deep convolutional neural networks[J]. Road Material and Pavement Design, 19(6): 1334-1349.