

Intelligent Pavement Diseases Detection Model based on YOLOv5s

Weihua Pan, Mengying Nie

North China Electric Power University Department of computer, Baoding, Hebei 071003, China

Abstract: There are many kinds of pavement diseases, and the detection background in the actual scene is complex, so the accuracy of the detection model is not high. Based on YOLOv5s, a pavement diseases detection model YOLOv5s-CBP is proposed, and the CBAM (Convolutional Block Attention Module) attention module was introduced into the backbone network to enhance the model's attention to the characteristics of pavement diseases. The BiFPN network was constructed in the neck, which realizes the cross-layer exchange and integration of information and retains more high-level semantic information. A small target detection layer P2 was added to the detection head to improve the detection ability of the model for small road diseases. The experimental results on data sets show that the precision and recall of the model are improved by 6.1% and 4.4% respectively, and the mAP is improved by 5.7%, which effectively improves the performance of the detection model.

Keywords: Target Detection; Pavement Diseases; Neural Network; Attention Mechanism.

1. Introduction

Pavement diseases are the result of many factors. Long-term road load, sun exposure and snow immersion have led to frequent road diseases. The increasingly number of pavement diseases has greatly reduced the service life of asphalt pavement, increased the pressure of road maintenance, and had a negative impact on road traffic. Thus, timely and effective detection and identification of road diseases can eliminate these potential safety hazards as soon as possible and prolong the service life of pavement, which is very important for road maintenance and management.

At the early stage, pavement diseases detection was mainly based on manual detection, which was time-consuming and laborious, and the detection efficiency was low, so it could not meet the large-scale road detection task. With the rapid development of computer software and hardware, the automatic detection methods of pavement diseases have been enriched. The methods of disease detection based on digital images can be summarized as: threshold segmentation method[2-3], edge detection method[4-5] and region growth method[6-7], but these methods have low accuracy and recall rate and complicated steps. The detection method based on neural network has two-stage algorithm [8] and one-stage algorithm[9]. In reference [10], the faster RCNN model was used to detect pavement diseases, and the regional expert network was introduced to improve the network performance. The two-stage algorithm has high accuracy, but the detection speed is slow. Reference [11] used YOLOv3 model to identify pavement cracks. This model can meet the real-time requirements with high accuracy, but there has a problem of missing detection. Yu Lu[12] introduced SENet attention and ASPP module on the basis of YOLOv5 model, which effectively improved the accuracy of pavement diseases detection model. YOLOv5 model can balance precision and speed well, and it is a mature detection model with good performance at present. However, the ability to extract the characteristics of small road diseases under complex background needs to be improved, and it is necessary to further analyze the characteristics of pavement diseases and put forward targeted improvement measures.

Therefore, based of YOLOv5s model, this paper introduced CBAM attention mechanism to enhance the feature extraction ability of the model, used BiFPN structure to realize cross-layer fusion of characteristic information, and added a small target detection layer to detect small target road diseases, enhancing the application ability of the model in complex background.

2. YOLOv5s Model

Unlike other target detection objects, pavement diseases have their own characteristics. Firstly, the background of pavement diseases images in the actual scene is complex and noisy, which usually contains interference information such as pedestrians, vehicles, lane lines, road garbage, etc. Moreover, due to factors such as light intensity and weather, road images often appear problems such as exposure and blurring, which greatly affects the quality of acquired images. Second, the types of pavement diseases are complex and diverse, the area gap of pavement diseases is large, and the proportion of road diseases in the image is uneven, which makes the width-height ratio of the anchor frame different from other data sets, so it is necessary to retrain the anchor frame according to the road datasets; In the end, the existing open source datasets of pavement diseases contains limited types of pavement diseases, and the labeling of pavement diseases is not ideal, and there are some problems such as labeling errors, omissions and the position deviation of labeling boxes. As a result, although YOLOv5s model has been widely used in pavement diseases detection, there are still some problems, such as false detection, missed detection and low accuracy of small target detection. Therefore, the road disease detection model in actual scenes needs further adjustment and improvement.

2.1. YOLOv5s Model

YOLOv5 is a representative model of one-stage detection algorithm, which is widely used in various fields and has good detection performance. YOLOv5 has four different versions, namely YOLOv5s, YOLOv5m, YOLOv5l and Yolov5x. The difference between these four versions is that the depth and width of the network increase in turn. As shown in Figure 1,

YOLOv5s model is the lightest network in YOLOv5 model, which includes four parts: input, backbone, neck and prediction head. At the input, the image data was adaptively scaled to a uniform size: 608*608, and the mosaic data enhancement was used to expand the datasets. Backbone network was responsible for feature extraction of image data to obtain high-level semantic and shallow features of image data; The neck network realizes the exchange and fusion of different levels of characteristic information through up-sampling; The prediction head includes three prediction layers for feature map prediction.

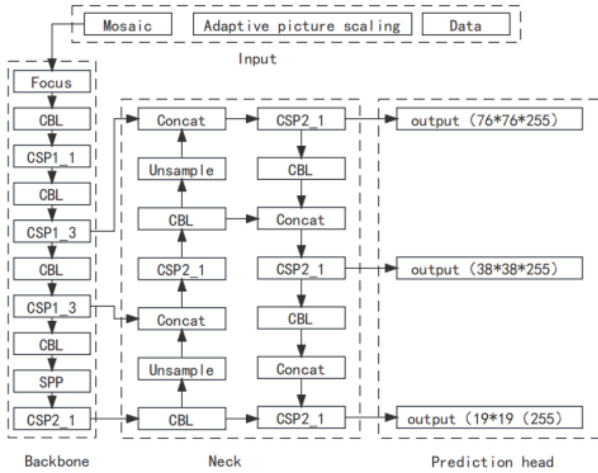


Fig 1. YOLOv5s Network Structure

2.2. Small Target Detection Layer

In the original YOLOv5s model, the detector includes three detection layers: P3, P4 and P5, which detect small, medium and large targets by 8, 16 and 32 times down sampling respectively, and the detection effect of large targets is better than that of small targets. In the datasets of this paper, there are not only large cracks and net cracks, but also small pits and cracks that are difficult to detect. There is a big difference in the proportion of road diseases, and the original three detection layers can not meet the detection requirements, which leads to the problem of missed detection of small target diseases in the detection process. In this paper, on the basis of the original three detection layers, the P2 detection layer with the scale of 152*152 is added, which mainly detects small pavement diseases. Detection layer P2 has a smaller receptive field, which can retain more position information and contour features of small targets. It can significantly improve the network's ability to detect small target diseases and effectively solve the negative impact of target scale differences on the model.

In addition, the anchor frame mechanism of the original YOLOv5s model was trained based on COCO data set, which has different detection target area with the pavement diseases detection in this study, which is not suitable for the actual needs of pavement diseases detection in this study. Therefore, it is necessary to design the target anchor frame size by re-clustering. On the data set constructed in this paper, the K-means++ method was used to recombine the classes, and the anchor frame size that can better match the pavement diseases target is obtained.

2.3. CBAM Attention Mechanism

The attention mechanism is an imitation of human visual system, which can focus on pavement diseases information in the process of feature extraction and improve the feature

extraction ability of the model. On the road disease datasets, the background is complex and the colors are diverse, and the image data containing a lot of unimportant semantic information accounts for a large proportion, which increases the difficulty of feature extraction of the detection model. The attention mechanism can help the model pay more attention to useful information in the training process and ignore the influence of redundant background information to some extent.

CBAM (Convolutional Block Attention Module)[14] is a widely used convolution attention module, which is plug and play, simple and efficient, different from channel attention mechanism or spatial attention mechanism. CBAM module realizes the fusion of cross-channel features and spatial information by convolution operation. As shown in Fig.2, the input feature map $F(H*W*C)$ is subjected to global maximum pooling and global average pooling operations, respectively, to obtain one-dimensional channel features $(1*1*C)$, and the feature maps of these two branches are respectively connected to the two-layer shared neural network (MLP) for training, and then the obtained output results are added by elements. Finally, the channel attention diagram M_c can be obtained through Sigmoid activation function:

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

The spatial attention module (SAM) compresses the input features into one dimension by maximum pooling and average pooling, and obtains two feature maps $(H*W*1)$, then obtains the single-channel features by concat and convolution operations, and finally obtains the spatial attention map M_s by Sigmoid activation function:

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) \quad (2)$$

Multiply and weight the channel attention diagram M_c and the spatial attention diagram M_s to get the final feature map. By adding CBAM module, the flow of feature information is promoted, the ability of model feature extraction is enhanced, and the detection performance and accuracy of the network are improved.

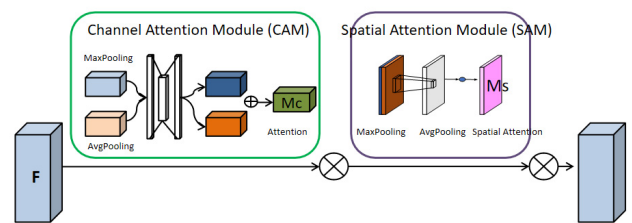


Fig 2. CBAM Attention Mechanism Module

2.4. BiFPN Cross-scale Feature Fusion

YOLOv5s model uses FPN-PAN structure to fuse features. FPN uses up-sampling to construct a one-way path, which fuses high-level semantics and low-level features. However, shallow features are seriously lost after many up-sampling operations, which directly affects the prediction effect of the detection head. As shown in Fig.3, PAN module adds a bottom-up path to FPN, constructs a bidirectional fusion path, and realizes the fusion of different levels of features, but frequent up-sampling and down-sampling operations also increase the parameters and calculation of the model. Therefore, this paper uses the BiFPN structure as the neck network, and BiFPN[15] keeps the bidirectional feature fusion path in PANet, and makes the following improvements: deleting a single input node, and the node with only one input

framework, is used to build a neural network. The experiment is conducted on Windows10 operating system, with Intel(R) Core (TM)i7-13700H as the CPU version and NVIDIA GeForce RTX3050 as the GPU version. The initial parameters of the experiment are shown in Table 1.

Table 1. Parameter Configuration

Experimental hyperparameter	Parameter setting
Input picture	608*608*3
Number of categories	five
Batch size	eight
Training times	100
Learning rate	0.01
optimizer	Adam

3.3. Evaluation Indicators

In order to verify the performance of the improved detection model, accuracy (Pr), recall (Re) and average accuracy (mAP) are used as evaluation indexes. Accuracy is the ratio of the number of correctly detected diseases to the number of positive samples judged by the model, recall is the ratio of correctly identified samples to the number of positive samples, and mAP is a comprehensive index to evaluate the

overall performance of the detection model, which can measure the average accuracy of detection of all disease categories in the model.

$$Pr = \frac{TP}{TP+FP} \quad (3)$$

$$Re = \frac{TP}{TP+FN} \quad (4)$$

$$AP = \frac{p_1+p_2+L+p_n}{n} \quad (5)$$

$$mAP = \frac{1}{n} \sum_{i=0}^{i=n} AP_i \quad (6)$$

Among them, TP represents the number of positive samples correctly identified by the model, FN represents the number of negative samples wrongly detected by the model, and FP is the number of positive samples wrongly detected by the model.

3.4. Ablation Experiment

The influence of the improved method on the model detection performance is verified by ablation experiments. Based on YOLOv5s network, model YOLOv5s-p is obtained by adding small target detection layer, model YOLOv5s-c is obtained by adding CBAM attention mechanism, and model YOLOv5s-b is obtained by adding BiFPN module. The experimental results are shown in Table 2.

Table 2. Results of ablation experiment

model	P2 detection layer	CBAM	BiFPN	Pr	Re	mAP@0.5
YOLOv5s	×	×	×	82.1%	76.3%	79.8%
YOLOv5s-p	√	×	×	82.1%(+2.1%)	76.3%(+5.2%)	79.8%(+1.5%)
YOLOv5s-c	×	√	×	82.1%(+3.5%)	76.3%(+2.8%)	79.8%(+2.2%)
YOLOv5s-b	×	×	√	82.1%(+2.9%)	76.3%(+1.6%)	79.8%(+1.7%)
YOLOv5s-CBP	√	√	√	82.1%(+6.1%)	76.3%(+4.4%)	79.8%(+5.7%)

The results of ablation experiments show that the accuracy and recall of the original model are improved by 3.5% and 2.8% respectively by adding CBAM attention mechanism to the model detector, which shows that the introduction of CBAM attention mechanism pays more attention to pavement diseases information, effectively improves the feature extraction ability of the model and further improves the performance of the model.

Using BiFPN instead of PANet in the neck network improves the accuracy and recall of the original model to 85% and 77.9%, which shows that BiFPN, as a feature fusion network, can better realize cross-layer communication and weighted fusion of feature information and effectively solve the problem of losing high-level semantic information.

After adding P2 detection layer to the network prediction head, the accuracy and recall of the model increased by 2.1 and 5.2 percentage points respectively, which is a targeted improvement measure. The newly added small target detection layer has a smaller receptive field and is responsible for detecting small target pavement diseases, which improves the detection efficiency and recall of small targets such as small cracks and pits. It can be seen that the three model improvement approaches proposed according to the characteristics of pavement diseases are effective.

3.5. Comparative Experiment

In order to verify the detection performance of the improved model, the YOLOv5s-CBP model is compared with the mainstream target detection algorithms Faster R-CNN, YOLOv3, YOLOv4 and YOLOv5s, and the comparison test results are shown in Table 3.

Table 3. Performance comparison with other detection models

model	Pr	Re	mAP@0.5
Faster RCNN	79.3%	73.1%	75.4%
YOLOv3	76.1%	72.3%	74.7%
YOLOv4	79.5%	74.0%	76.5%
YOLOv5s	82.1%	76.3%	79.8%
YOLOv5s-CBP	88.2%	80.7%	85.5%

As can be seen from Table 3, the improved YOLOv5s-CBP model has the highest detection accuracy (88.2%) and recall rate (80.7%). Compared with the original YOLOv5s, the accuracy has increased by 6.1 percentage points, and the recall rate has increased by 4.4%. The overall performance is superior to other models, which shows that the improved model method used in this paper is a practical and effective measure to improve the model performance. The detection

results of each model are shown in Figure 6. After careful observation, it can be found that there is little difference in the recall rate among the models except YOLOv3 model, and they all have high recall rate, but the confidence of the improved v5 model is as high as 94%, which can realize the accurate detection of the disease. For minor diseases (potholes), both v3 and v4 versions can't identify them, but the improved YOLOv5s model can accurately identify them,

which shows that the improved measures can strengthen the model's attention to minor road diseases and optimize the model's ability to detect small targets. Judging from the detection effect, YOLOv5s-CBP model not only can accurately identify road diseases, but also has high confidence and better small target detection ability, which can better cope with complex background problems in actual scenes.



Fig 6. Comparison of test results

4. Conclusion

In order to improve the detection ability of the road disease detection model in the actual scene, this paper obtains a neural network model with higher detection accuracy from two aspects: data set construction and model improvement. Through the analysis of the characteristics of road diseases, the open source road disease data set is sorted out and marked twice, and the local road image data is obtained by field shooting. At the same time, the data set is expanded by data enhancement method to obtain high-quality road disease data set. Based on YOLOv5s model, small target detection layer P2 is added to detect smaller road diseases, which effectively improves the small target detection performance of the model. Add CBAM attention to the backbone network, optimize feature extraction, focus on important feature information, suppress unimportant features, and promote information flow in both channel and space dimensions; The BiFPN bidirectional connection feature fusion network is constructed in the neck network, which simplifies the bidirectional fusion path, adds jumping connection, retains the semantic features

of higher layers, and effectively prevents information loss caused by up-sampling. Compared with other algorithms, YOLO-PCB model can better detect small target road diseases in complex background and has stronger practical application ability. At the same time, the actual detection results also show that for some uncommon diseases, such as potholes, the detection effect of the model needs to be further improved. The next step will mainly focus on optimizing the data set construction, increasing the number of uncommon road diseases through measures such as GAN, ensuring that the number of various road diseases in the data set is basically the same, reducing the difference of road disease categories, and further improving the detection performance of the model in practical application.

References

- [1] Ning Tingting. Research on pavement crack detection algorithm based on digital image processing [D]. Liaoning Normal University, 2023.

- [2] Chu Xiumin, Yan Xinping, Chen Xianqiao. Research on binarization method of pavement damage image [J]. Computer Engineering and Application, 2008(28):161-165.
- [3] Jiang Wei. Research and implementation of adaptive threshold segmentation algorithm for pavement crack diseases [D]. Nanjing University of Information Science and Technology, 2020.
- [4] Wang G, Peter W T, Yuan M. Automatic internal crack detection from a sequence of infrared images with a triple-threshold Canny edge detector[J]. Measurement Science and Technology, 2018, 29(2): 025403.
- [5] Rolling, Meng Hua, Liu Tangzhi. Research on edge detection algorithm of asphalt pavement crack image [J]. Northern Transportation, 2015(03):56-59.
- [6] Wang Wei, Meng Lina, Yan Zhaojin. Pavement damage detection based on regional growth and outermost edge tracking [J]. Journal of civil aviation university of china, 2016, 34(06):51-54.
- [7] Xiao Xiaoming, Ma Zhi, Cai Zixing et al. An Adaptive Region Growing Algorithm for Road Segmentation [J]. Control Engineering, 2011,18(03):364-368.
- [8] Chen Yimin, Li Wanyi, Weng Hanrui, et al. Overview of two-stage target detection algorithms based on deep learning [J]. Information and Computer (Theoretical Edition), 2023,35 (14): 112-114.
- [9] Zhu Hao, Zhou Shunyong, Liu Xue, et al. Overview of single-stage target detection algorithm based on deep learning [J]. Industrial Control Computer, 2023,36(04):101-103.
- [10] Kortmann F, Talits K, Fassmeyer P, et al. Detecting various road damage types in global countries utilizing faster r-cnn[C]//2020 IEEE International Conference on Big Data (Big Data). IEEE, 2020: 5563-5571.
- [11] Nie M, Wang C. Pavement Crack Detection based on yolo v3[C]//2019 2nd international conference on safety produce informatization (IICSPI). IEEE, 2019: 327-330.
- [12] Yu Lu, Dai Tianjie, Yu Lihua. Intelligent detection of road diseases based on improved YOLOv5 [J]. journal of fujian university of technology, 2023,21(04):332-337.
- [13] Wang Linyi, Bai Jing, Li Wenjing et al. Research progress of YOLO series target detection algorithms [J]. Computer Engineering and Application, 2023,59(14):15-29.
- [14] Qi xuanhao, Zhi Min. a summary of attention mechanism in image processing [J/OL]. computer science and exploration: 1-20 [2023-10-16]. <http://kns.cnki.net/kcms/detail/11.5602.tp.20629.10606060606>
- [15] Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.