Advances in Computer, Signals and Systems (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2026.100107
ISSN 2371-8838 Vol. 10 Num. 1

Research on pavement disease detection algorithm based on YOLOv5s

Hanpeng Wu, Yuxi Guo, Zhengbing Zheng

Shaanxi University of Technology, Hanzhong, Shaanxi, 723001, China

Keywords: Pavement defect detection; data enhancement; CBAM; NWD

Abstract: In response to the problems such as significant background interference, large differences in target scales, and low detection accuracy for small targets in the current pavement defect detection task, this paper proposes an improved YOLOv5s pavement defect detection algorithm. The training dataset is obtained by integrating the open-source RDD2022 dataset and the self-built dataset, and then the data is augmented through Mosaic method before being input into the network for training. In addition, the CBAM attention mechanism is introduced into the backbone network to enhance the model's ability to focus on the pavement defect features through both channel and spatial attention. Finally, the NWD optimization loss function is adopted to improve the model's detection accuracy for small targets and ablation experiments are conducted. The experimental results show that the improved algorithm has an mAP@0.5 of 3.8% higher than that of the original YOLOv5s on the dataset, effectively enhancing the model's detection accuracy.

1. Introduction

According to the statistical bulletin[1] of the development of the transportation industry of the Ministry of Transport of China in 2024, the total length of highways in China has been increasing. At the end of 2024, the length of highways has been 5.4904 million kilometers, and there are 12.1224 million road vehicles in operation in the country. The commercial freight volume has been 41.88 billion tons and the turnover of goods has been 7.6848 billion tonnage kilometers. The total number of road personnel movements was 59.299 billion. In the process of highway operation, due to the influence of vehicle load and environmental factors, different types of diseases will occur on the road surface. Common road surface disease detection methods include physical detection methods, manual measurement methods and machine vision-based detection[2] methods. Among them, the detection methods of machine vision are divided into traditional machine learning methods and deep learning methods. Deep learning algorithms can effectively extract target features in complex environments and overcome the limitations of traditional algorithms.

The object detection algorithms of deep learning include two types: two-stage and one-stage. In 2014, Girshick et[3] al proposed R-CNN (Region-CNN, R-CNN), which introduced two-stage detection algorithm for the first time. In the first stage, candidate boxes are generated by the algorithm, and image features are extracted in the second stage. This kind of method has high detection accuracy, but a large number of redundant operations greatly improve the space and time cost, and the efficiency is low, and it is difficult to promote. Compared with the two-stage algorithm, the one-stage algorithm removes the generation stage of candidate box, and only needs to extract the feature value once to realize the object detection, which greatly reduces the calculation time and meets the needs of object detection. In the single-stage algorithm, YOLO algorithm has been widely used.

Based on the YOLOv5s algorithm, this paper introduces a Convolutional Block Attention Module (CBAM), which includes a channel attention module and a spatial attention module to realize deep learning of two dimensions of channel and space. By optimizing the loss function, The ability of the model to detect small objects is improved, and the data enhancement technology is combined to further strengthen the model's attention to small objects. The improved YOLOv5s algorithm is used for road detection research.

2. Introduction of the Model

2.1. YOLOv5s algorithm

YOLOv1 was proposed in 2015, and researchers have continuously optimized the model based on it. The optimization measures include the data set of model training, the introduction of multi-module architecture, etc., to improve the autonomous learning ability of the network and the calculation accuracy of the model. YOLOv5[4] algorithm was proposed in June 2020. Compared with other versions, YOLOv5 simplifies the model structure, and optimizes the speed and accuracy. In practical applications, YOLOv5 can maintain real-time performance while still having high accuracy. YOLOv5s is the model with the smallest depth and width among YOLOv5, which has good real-time performance and is convenient for practical embedded deployment. The network structure of YOLOv5 is shown in Figure 1.

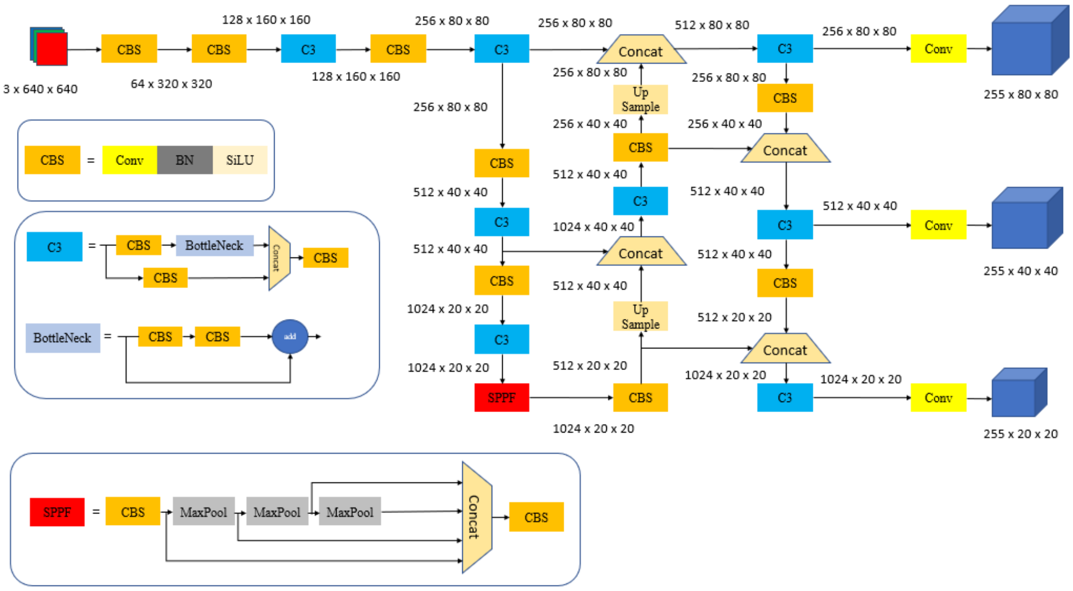


Figure 1. YOLOv5 network structure

2.2. CBAM module

In order to further improve the detection accuracy of YOLOv5s model and improve the feature extraction ability, this paper introduces the CBAM module to construct the CBAM-YOLOv5S algorithm. The CBAM module was proposed by Woo S et[5] al in 2018, which innovatively combines the channel attention module and the spatial attention module to realize the accurate optimization of the feature map in two dimensions of channel and space. The function of this module is to make the model focus on extracting important features, so as to suppress unimportant features, so as to improve the accuracy of multi-scale feature attention. Figure 2 shows the structure of the CBAM module, in which the channel attention sub-module preserves 3D information by using 3D arrangement, and then uses a two-layer multi-layer perceptron (MLP) to amplify the cross-dimensional channel dependence. The spatial attention sub-module focuses on the spatial information, and uses two convolutional layers to fuse the spatial information. Firstly, the number of channels is reduced by convolution with a convolution kernel of 7 to reduce the amount of calculation, and then the number of channels is increased by a convolution operation with a convolution kernel of 7 to maintain the consistency of the number of channels. Finally, the output is output after the Sigmoid function. Figure 3 and Figure 4 show the structure diagram of the channel attention module and spatial attention module, respectively.

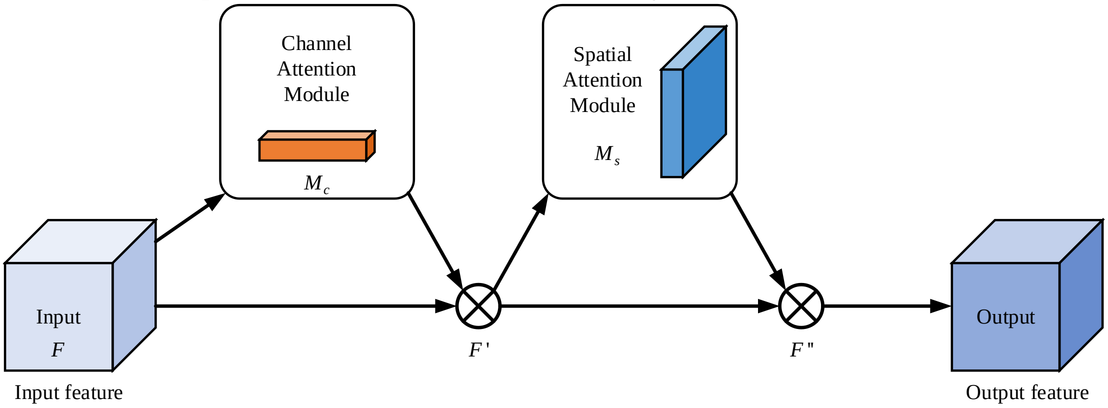


Figure 2. CBAM module structure

$$F' = M_c(F)\ conv\ F \tag{1}$$

$$F'' = M_s(F')\ conv\ F' \tag{2}$$

Where: F is the input feature map; $M_c(F)$ is the convolution operation of channel attention module; $M_s(F')$ is the convolution operation of spatial attention module; F' is the channel feature map generated after the channel attention module; F'' is the spatial feature map generated after the spatial attention module.

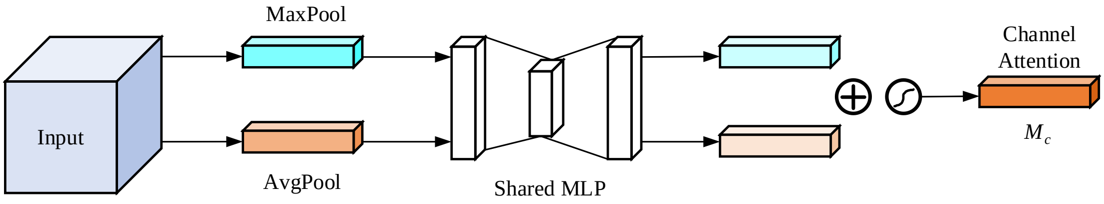


Figure 3. Channel Attention Module

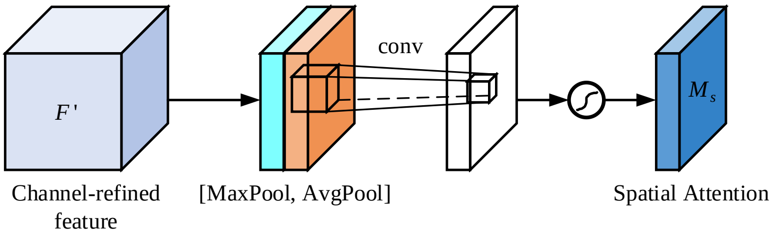


Figure 4. Spatial Attention Module

3. Loss function optimization

The road surface disease detection task has the problems of dense small targets and large difference in target aspect ratio. By introducing Normalized Wasserstein Distance (NWD), and using the characteristics of NWD based on distribution distance, the problem of vanishing gradient of small targets is eliminated, so as to improve the performance[6] of the model for small target detection.

In order to solve the problem that IoU loss cannot provide an effective gradient, Wasserstein Distance (WD) is introduced as an alternative metric, which can give a continuous and derivable distance value even if the predicted box and the real box do not overlap. The original WD is sensitive to image scale, and then optimized to obtain NWD, which is suitable for images of any scale. NDW is defined as follows:

For a horizontal bounding box $R = (cx, cy, w, h)$, Where, (cx, cy), w and h represent the center coordinates, width and height, respectively, modeling the horizontal bounding box as a two-dimensional Gaussian distribution, then, $N(\mu, \Sigma)$, Where:

$$\mu = \begin{bmatrix} cx \\ cy \end{bmatrix} \tag{3}$$

$$\Sigma = \begin{bmatrix} \frac{w^2}{4} & 0 \\ 0 & \frac{h^2}{4} \end{bmatrix} \tag{4}$$

The Wasserstein distance from the optimal transport theory is used to calculate the distribution distance. The secondorder Wasserstein distance between two two-dimensional Gaussian distributions is defined as,

$$W_2^2 = \left|\left|\mu_1 - \mu_2\right|\right|_2^2 + Tr\left(\Sigma_1 + \Sigma_2 - 2\left(\Sigma_2^{\frac{1}{2}}\Sigma_1\Sigma_2^{\frac{1}{2}}\right)^{\frac{1}{2}}\right) \tag{5}$$

After eliminating the effect of scale, the normalized new metric NWD is obtained.

$$NWD(N_1, N_2) = \exp\left(-\frac{W_2}{C}\right) \tag{6}$$

Where C is the normalization constant, which is usually taken as the diagonal length of the image to ensure.$NWD \in (0,1]$.

4. Experiment process and result analysis

4.1. Dataset preprocessing

The dataset used in this paper integrates the self-built dataset and the public road disease dataset RDD2022 [7], The self-built dataset contains 1000 road disease images, and the RDD2022 dataset contains 38385 road disease images from six countries, including China, India, Japan, Czech Republic, Norway and the United States. In total, there are more than 42000 road damage instances.

The dataset mainly contains four types of road surface diseases longitudinal cracks (D00), lateral cracks (D10), cracking (D20) and potholes (D40). The dataset was shuffled and a total of 5000 images were randomly selected from it to form a new dataset. The new dataset was divided into training set, validation set and test set according to the ratio of 8:1:1. Then, the images in the training set were enhanced by Mosaic.

Mosaic is to concatenate multiple images in accordance with a certain proportion to form a new image, and each image is more likely to contain small objects, which is used as the data set, so that the model can recognize objects in a smaller range, and effectively improve the detection accuracy of the model for small objects. Some schematic diagrams of Mosaic data augmentation are shown in Figure 5:

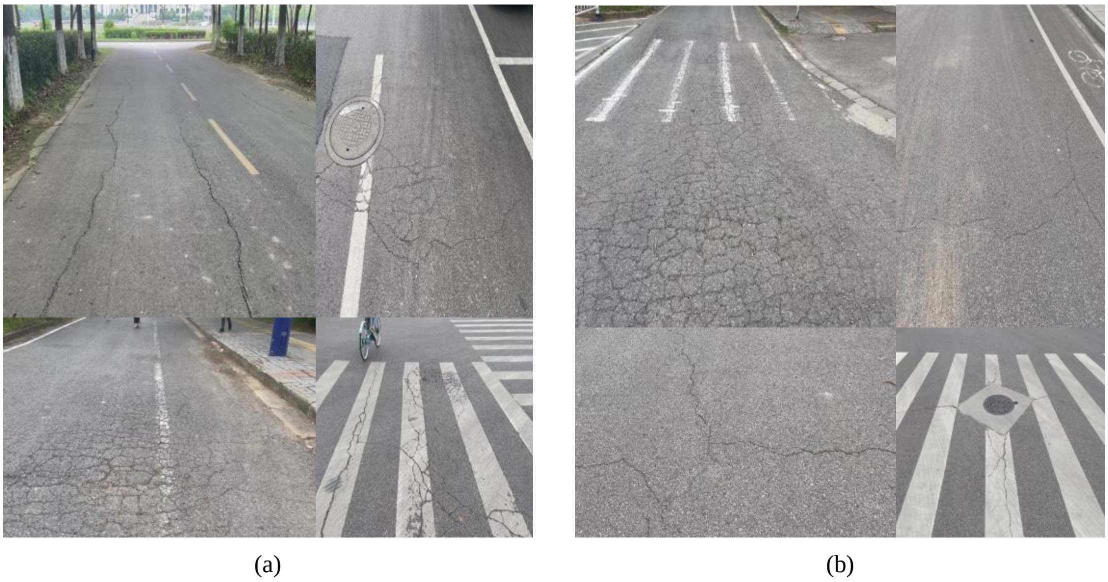

(a) (b)

Figure 5. Mosaic data augmentation

4.2. Evaluation Metrics

In order to understand the model training, it is necessary to evaluate the model after the training is completed. In addition to the intuitive evaluation results through the detection box, the model can also be evaluated through some quantifiable parameters. By comparing the model detection category with the actual category, it can be divided into the following four categories:

(1)True Positive(TP) means the sample is judged to be positive and the result is correct;

(2)True Negative(TN) means that the sample is judged to be negative and the result is correct;

(3)False Positive(FP), indicating that the sample was judged as positive and the result was wrong;

(4)False Negative(FN), which means the sample is judged as negative and the result is wrong.

Model evaluation metrics include precision P, recall R and mean average precision (mAP) $m_{\bar{p}}$.

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$\bar{P} = \int_0^1 P\, dR \tag{9}$$

$$m_{\bar{p}} = \frac{1}{n}\sum_{i=1}^{n} \overline{P_i} \tag{10}$$

Where: n is the number of sample categories; i is the current number; TP indicates the positive example of correct prediction; FP is a positive incorrectly predicted example; FN is a negative case of misprediction; $\bar{P}$ is the integral of precision-recall (P-R) curve within [0,1], which is used to evaluate the detection effect of the model on a certain type of disease. $m_{\bar{p}}$ is the average value of $\bar{P}$ for all diseases, which can evaluate the recognition effect of the model on road surface diseases. mAP@0.5 represents the average accuracy of all classes when the intersection over union (IoU) is set to 0.5. In practical applications, two metrics are usually used: mAP@0.5 and mAP@0.5:0.95, 0.5 and 0.95 represent the confidence, which refers to the overlap between the predicted box and the true box.

4.3. Experimental Results

In this experiment, Pytorch 1.12.1 was used as the deep learning development framework, Python language programming was used, and the experiment was carried out by renting a server. The GPU was NVIDIA GeForce RTX 3080, and the running memory was 8G. The hyperparameter configuration includes: input image normalization to 6640 pixels, training times 100, batch size 32, learning rate 1×10^{-2}, weight decay value 5×10^{-4}, momentum value 0. 937, and stochastic gradient descent (SGD) as the optimizer.

The model detection results of the original YOLOv5s algorithm, trained by the CBAM module, SE module, and ECA module are listed in Table 1. It can be found that the evaluation indicators of CBAM-YOLOv5S are higher than those of the other three cases.

Table 1. Comparative experimental results of each attention mechanism

Module			Evaluation metrics			
CBAM	SE	ECA	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95/%
×	×	×	70.4	78.4	71.1	48.2
√	×	×	81.0	88.8	73.6	50.5
×	√	×	69.3	82.0	71.8	48.8
×	×	√	70.1	79.7	72.0	48.6

By comparing the improved algorithm with some classical object detection networks, it can be found that after the introduction of CBAM module, the accuracy rate, recall rate and mAP@0.5 of the algorithm are significantly improved, and the evaluation indicators are shown in Table 2:

Table 2 Comparison with the classical object detection network

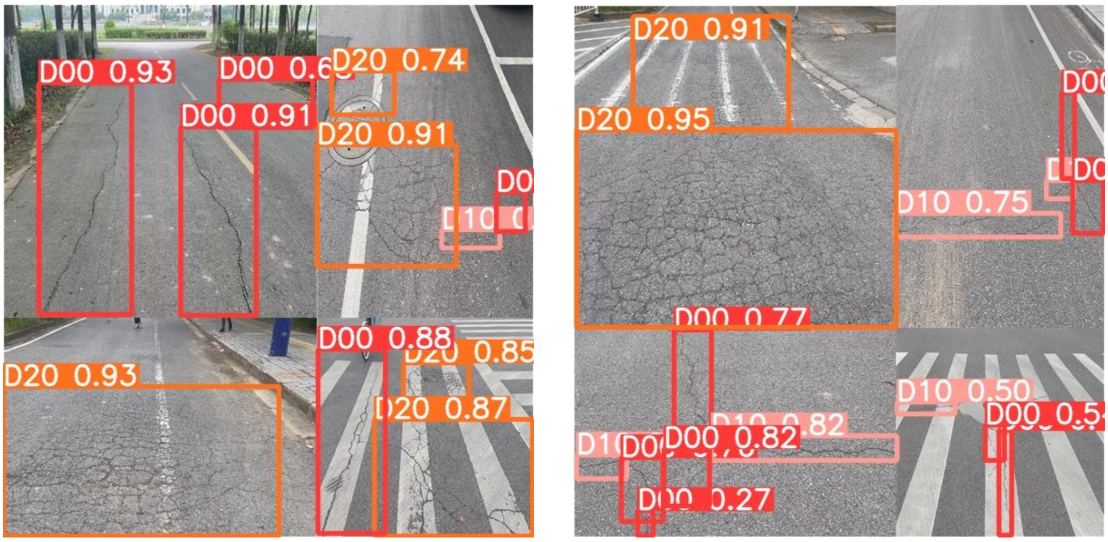
Module	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95/%
RCNN	84.9	76.7	74.6	49.6
YOLOv5s	70.4	78.4	71.1	48.2
CBAM-YOLOv5	81.0	88.8	73.6	50.5

At the same time, the CBAM module and NWD loss function were introduced, and ablation experiments were carried out. The experimental results show that the introduction of CBAM module and NWD loss function can improve the detection performance of the model, and the introduction of both can make the model obtain better detection performance. The ablation experiment results are shown in Table 3:

Table 3 Ablation experiments

CBAM	NWD	*P*/%	*R*/%	mAP@0.5/%	mAP@0.5:0.95/%
×	×	70.4	78.4	71.1	48.2
√	×	81.0	88.8	73.6	50.5
×	√	79.6	84.4	72.2	48.3
√	√	84.2	90.4	74.9	51.7

An example of the improved model checking results is shown in Figure 6:



(a)

(b)

Figure 6. Road surface disease detection effect

5. Conclusions

The CBAM module can significantly improve the degree of optimization, the most important reason is that the spatial attention module is added to the CBAM module, and the model is deeply learned through two dimensions. In the learning process of the model, after several layers of convolution, each position of the model feature contains the information of a local area of the original image. By taking the maximum and average value of multiple channels at each position as the weighting coefficient, the optimal result is obtained. By optimizing the loss function, NWD is used to eliminate the problem of gradient disappearance of IoU loss on small targets. At the same time, the characteristics of NWD based on distribution distance are used to help the algorithm enhance the attention to small cracks and effectively improve the efficiency of the algorithm.

In this paper, based on YOLOv5s, by introducing different modules to improve the algorithm, through experiments, the following conclusions are drawn:

(1) Compared with the YOLOv5s algorithm, the precision, recall rate, mAP@0.5, and mAP@0.5:0.95 of the algorithm in this paper have increased by 13.8%, 12.0%, 3.8%, and 3.5% respectively. Through comparative experiments, the superiority of the combination analysis of the channel attention module and the spatial attention module in two dimensions has been verified.

(2) By introducing the NWD loss function, the algorithm can effectively enhance the attention of the algorithm to small targets, which has a good effect on the slender cracks and other pavement disease detection tasks. Compared with introducing CBAM alone, the model's accuracy, recall rate, mAP@0.5, and mAP@0.5:0.95 have increased by 3.2%, 1.6%, 1.3%, and 1.2% respectively.

(3) Although the improved algorithm performs well on the dataset, it is necessary to balance the accuracy and real-time performance in practical application scenarios. In the future, it is still necessary to strengthen the model detection performance from the aspects of dataset preprocessing, accuracy and real-time balance, so as to provide theoretical basis for actual production.

References

[1] Ministry of Transport of the People's Republic of China. Statistical Bulletin on the Development of the Transportation Industry in 2024 [EB/OL]. The Central People's Government of the People's Republic of China website, 2025. https://xxgk.mot.gov.cn/2020/jigou/zhghs/202506/t20250610_4170228.html.

[2] Chen Yun. Research on the fusion of traditional and intelligent technology of highway pavement disease detection [J]. Intelligent Building and Smart City2025,(04):27-29.

[3] GIRSHICK Ross, DONAHUE Jeff, DARRELL Trevor, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE,Ohio,USA,2014:580-587

[4] Liu P Y, Yuan J, Gao Q, et al. Road surface disease detection method based on improved YOLOv5 [J]. Journal of Beijing University of Technology, 2025, 51 (05): 552-559.

[5] WOO Sanghyun, PARK Jongchan, LEE Joon Young, et al. CBAM: Convolutional Block Attention Module[C]. Spring:Cham,Switzerland,2018:3-19

[6] Jinwang Wang, Chang Xu, Wen Yang, Lei Yu. A Normalized Gaussian Wasserstein Distance for Tiny Object Detection[J]. arXiv preprint, arXiv; 2110.13389.2021.

[7] DEEKSHA Arya, HIROYA Maeda, SANJAY Kumar Ghosh, et al. RDD2022:A multi-national image dataset for automatic Road Damage Detection[J]. IEEEBigData,2022,1-22.