

# Research on a Road Defect Detection Method based on Improved YOLOv8

Yanyan Wang, Menghao Zhou, Haojie Chai and Ran Xue

School of Artificial Intelligence, Henan Institute of Science and Technology, Xinxiang 453003, China

**Abstract:** Traditional road defect detection methods mainly rely on manual detection, which is inefficient and has many limitations. In order to detect road defects more accurately and quickly, this paper proposes a road defect detection method based on improved YOLOv8. In this paper, a large number of image datasets containing road defects are collected and annotated, and the YOLOv8 algorithm is used to train the model on the dataset, one group uses the original model, one group replaces the convolutional kernel (DCNv2), and one group adds the attention mechanism (CBAM), and then a comparative evaluation is performed based on the training results of the models in each group. The experimental results show that the improved model reflects higher accuracy in road defect detection compared to the untuned model. The model with the replaced convolutional kernel improved about 2% compared to the original model Mean Average Precision (mAP), and the model with the added attention mechanism improved about 3.5% compared to the original model mPA. This study can efficiently identify different types of road defects, which can help road maintenance work to be carried out more scientifically and efficiently, and provide important support for improving road safety and maintenance efficiency.

**Keywords:** Road Defect Detection; YOLOv8; Neural Network; Deep Learning.

## 1. Introduction

With the acceleration of urbanization and the continuous growth of the number of vehicles, the safety and accessibility of roads are of great significance to the normal operation of urban traffic. However, various defects, such as cracks, potholes, and damages, will appear on roads in the process of use, which will bring safety hazards to drivers and pedestrians, and even cause damages to vehicles. Therefore, timely detection and repair of road defects are very important to ensure the safety and accessibility of roads [1]. Current road defect detection methods mainly rely on manual inspection, and there are always differences in monitoring and evaluation standards, and they are also prone to the limitations of weather, light and other natural conditions, resulting in omissions and misdetections. Therefore, finding an efficient and accurate road defect detection method is of great significance to improve the efficiency of road maintenance and ensure the safety of people traveling [2, 3].

At present, the defect detection method based on deep learning has gradually become a hot spot of research [4-6]. Sun et al [4] optimized the aspect ratio of crack candidate frames by fusing VGG16 network and Faster R-CNN, which further improved the detection accuracy despite sacrificing part of the detection speed. Jiang et al [5] improved YOLOv5s by introducing an attention mechanism and an adaptive feature fusion mechanism to optimize the accuracy rate of road detection, but the method still has low detection accuracy for transverse cracks. Yan et al [6] proposed a road crack detection method based on Faster R-CNN, which is able to extract the crack morphology and size, but the misdetection rate is on the high side, which makes it difficult to be applied in practice. Sun et al [7] achieved intelligent identification of road marking defects through deep learning, using data augmentation, modular optimization, and anchor point redesign, to locate the road markings and classify their defects. Wang et al [8] proposed a lightweight SSD-based road crack detection method, which improved the detection speed and

accuracy, but the accuracy was reduced for some types of detection.

Although many research results have been achieved [9-12], there are still shortcomings such as high detection difficulty and poor generalization ability, and there is still room for progress in the accuracy of detection. Therefore, by analyzing the existing methods, this paper proposes a road defect detection method based on improved YOLOv8 using deep learning technology. The original YOLOv8 [5] target detection algorithm model is improved by adding a variable convolutional kernel (DCNv2) with an attention mechanism (Convolutional Block Attention Module (CBAM)), and then these three models are trained on the dataset separately, i.e., one group uses the original model, one group replaces the convolutional kernel (DCNv2) and one group adds the attention mechanism (CBAM), and the training results of different models are recorded separately. Then the best modeling scheme is selected based on the comparative evaluation of the model training results of each group to improve the accuracy of road defect detection.

## 2. Related Theories

### 2.1. YOLOv8 Model

YOLOv8 is an efficient, accurate and easy-to-use target detection algorithm, and its model structure is an important milestone in the field of target detection, which has been improved and optimized in several ways on the basis of the YOLO series of models, and its network structure is shown in Figure 1.

The following is a brief description of the structure and role function of each layer of the model:

#### (1) Input

The input side of YOLOv8 is similar to the traditional approach, accepting images as input and performing the necessary preprocessing such as scaling and normalization.

#### (2) Backbone

The Backbone part is responsible for extracting the basic



set, a test set, and a validation set according to the ratio of 8:1:1.

The four typical road surface defects are shown in Figure 2.

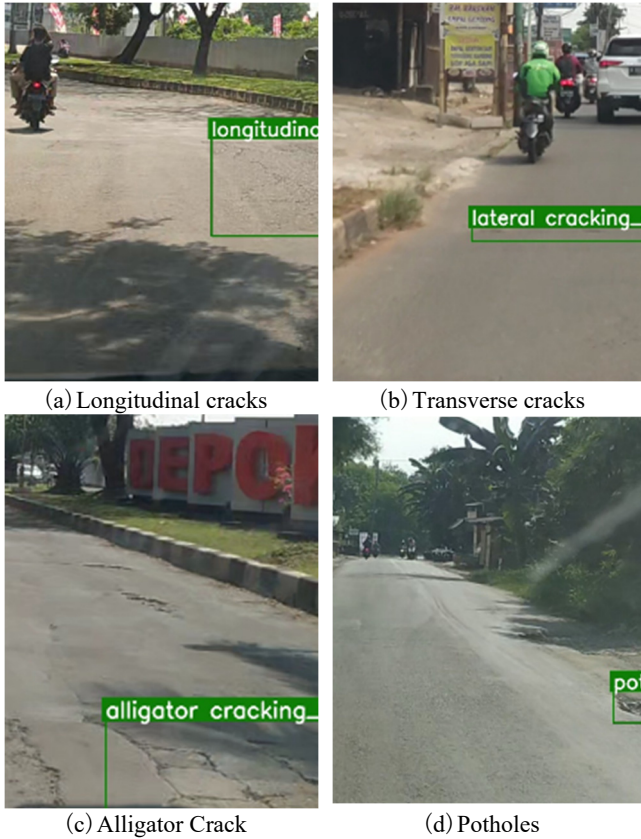


Fig 2. Four typical road surface defects

## 4. Experimentation and Analysis

### 4.1. Environment Configuration

Prior to model training, a series of preparations are required to ensure that the training process runs smoothly. A brief description of the preparations before training the YOLOv8 model:

Environment configuration: Install Python and its related libraries, including PyTorch and YOLOv8. Install CUDA and cudnn as needed, and ensure that the Python environment is configured correctly, including Python version and library dependencies [16].

The main experimental environment configuration is shown in Table 1.

Table 1. Experimental environment

Parameters	Configuration
Operating system	Windows 11
CPU	AMD Ryzen 5 4600H with Radeon Graphics
Memory	16G
Python environment	3.8
Deep learning framework	torch1.9.0+cu111

Dataset preparation: prepare the dataset for road defect detection, ensuring that the dataset contains sufficient samples and diversity. And annotate the dataset to generate annotation files compatible with YOLOv8.

Configuration file preparation: prepare model

configuration files, which define the model architecture, training parameters, etc. Prepare dataset configuration files, which specify the path, category, and other information of the dataset. Adjust the default configuration file (default.yaml) as needed to fit specific training needs.

After all preparations are complete, start training the model.

### 4.2. Model Training

epochs (number of rounds): the complete number of times the entire dataset will be traversed. In this paper, the model will traverse the entire dataset 200 times.

patience (number of rounds to wait for early-stop training): when using the early-stop strategy, this parameter defines how many epochs the model needs to wait before stopping training when its performance on the validation set is no longer improving.

batch (number of images per batch): the number of samples used in a weight update.

lr0 (initial learning rate): the learning rate determines the step size of the weight update during training of the model.

lrf (final learning rate): this parameter defines the minimum value that the learning rate decays to.

The specific model training parameters are shown in Table 2.

Table 2. Model training parameters

Key	Value	Description
epochs	200	Number of training rounds
patience	50	Number of waiting rounds for early stop training
batch	16	Number of images per batch
lr0	0.01	Initial learning rate
lrf	0.01	Final learning rate

### 4.3. Model Training Results

The model training results of this paper are shown in Figure 3.

(1) weights file

This directory holds two weights saved during training, respectively:

last.pt: "last.pt" usually refers to the last weights file saved during model training.

best.pt: "best.pt" usually refers to the weight file of the model that performs best on the validation set or test set.

(2) Confusion Matrix

The confusion matrix is used to evaluate the performance of the model, which can visualize the classification of the model on each category. The rows of the matrix represent the actual categories and the columns represent the predicted categories, and the number of different category combinations is obtained by comparing and counting the real categories of the samples and the categories predicted by the model.

Precision is the proportion of all samples predicted as Positive by the model that are actually Positive, which measures the accuracy of the model in Positive prediction, and is given by the formula

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

Recall is the proportion of all samples that are actually positive cases that are successfully predicted to be positive by the model, and it measures the model's ability to recognize positive cases, which is given by the formula

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

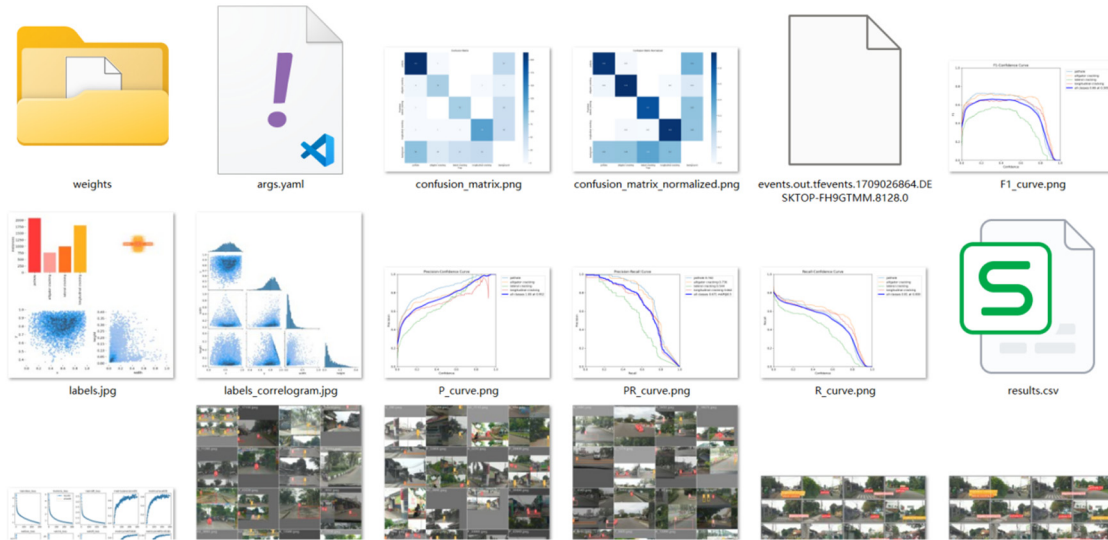


Fig 3. Model Training Results

Based on the confusion matrix, we can make a preliminary evaluation of the performance of the classifier, and the confusion matrix of the model trained in this paper is shown in Fig. 4.

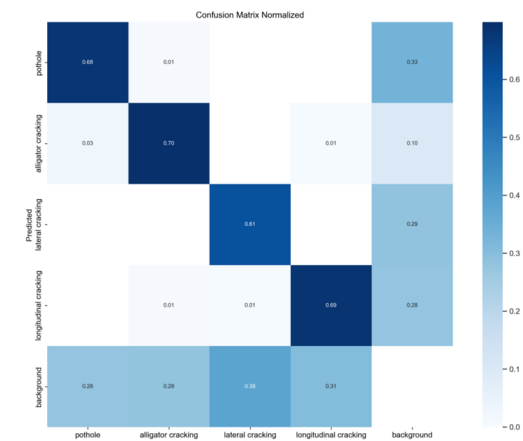


Fig 4. Confusion Matrix

(3) F1 curve

The F1 curve is a performance evaluation tool commonly used in multicategorization problems. It is based on the F1 score, which is the harmonic mean of precision and recall, and takes values between 0 and 1. The F1 curve is defined as the value of F1 that is closer to 1 than the F1 curve. The closer the value of F1 is to 1, the better the model performance is,

and F1 is defined as

$$F1 = 2 * \frac{precision * recall}{precision + recall} \quad (3)$$

The F1 curves for the models trained in this paper are shown below.

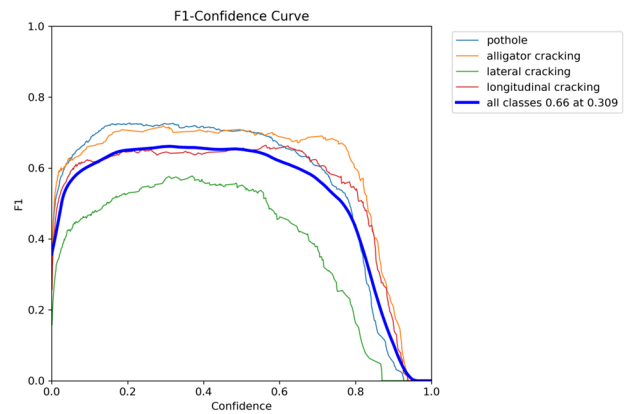


Fig 5. F1\_curve

(4) Labels (labeled statistical charts)

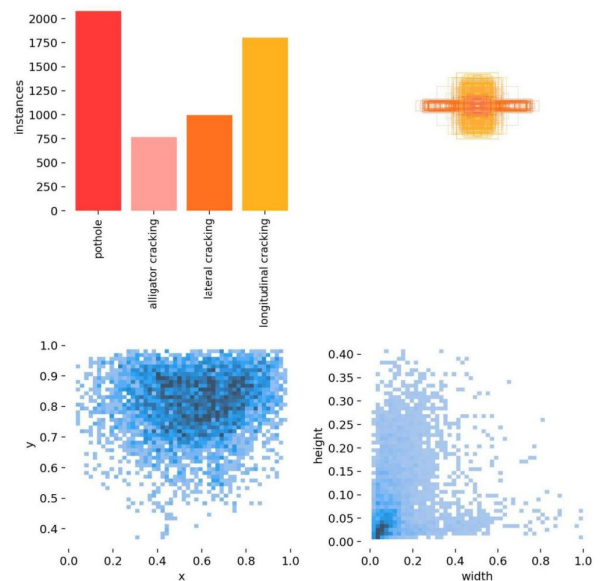


Fig 6. Labels

In Figure 6, they are listed in order from left to right:

Palace 1: The amount of data in the training set, showing the number of samples contained in each category.

Palace 2: Size and number of boxes, showing the size distribution and corresponding number of bounding boxes in the training set.

Palace 3: Position of the center point relative to the whole image, describing the distribution of the position of the bounding box center point in the image.

Grid 4: the height-to-width ratio of the target in the image relative to the whole image, reflecting the distribution of the height-to-width ratio of the target in the training set.

(5) Labels\_correlogram (label correlation graph)

The results of this model training show the distribution of data with different heights and widths, as shown in Figure 7. The scatterplot and histogram on the left focus on the distribution of "height", while the graph on the right focuses on the distribution of "width". By further analyzing these data, we can optimize the model parameters and improve the prediction accuracy and generalization ability of the model.

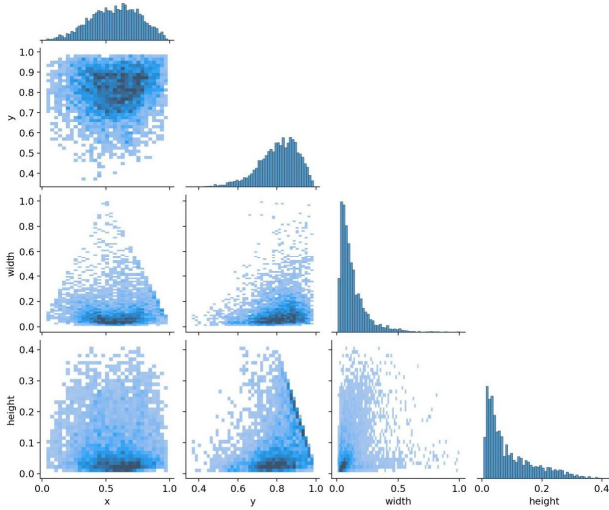


Fig 7. Labels\_correlogram

(6) P\_curve (precision-confidence curve)

The precision-confidence curve (P\_curve) plot of the model trained in this paper is shown in Fig. 8, where the horizontal coordinate represents the confidence level of the detector and the vertical coordinate represents the precision (or recall). The shape and position of the curve reflect the performance of the detector at different confidence levels.

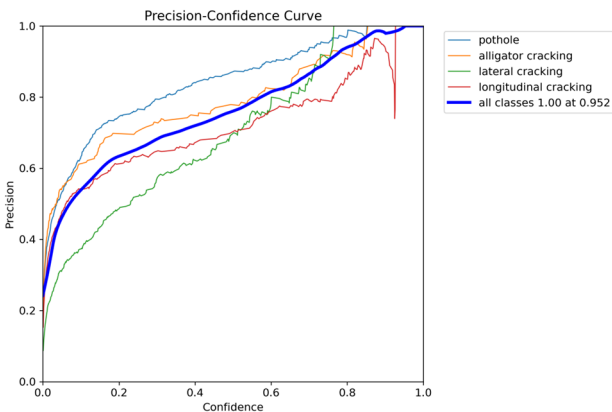


Fig 8. P\_curve

This result reflects the efficiency and accuracy of the model on the crack detection task, especially in the high confidence

interval, the model performs well and provides a strong support for the automatic identification and assessment of road cracks.

(7) PR\_curve (precision-recall curve)

The precision-recall curve plot of this model training result demonstrates the performance of the model in detecting four different types of cracks (pothole, alligator cracking, lateral cracking, and longitudinal cracking), and the results are shown in Fig. 9.

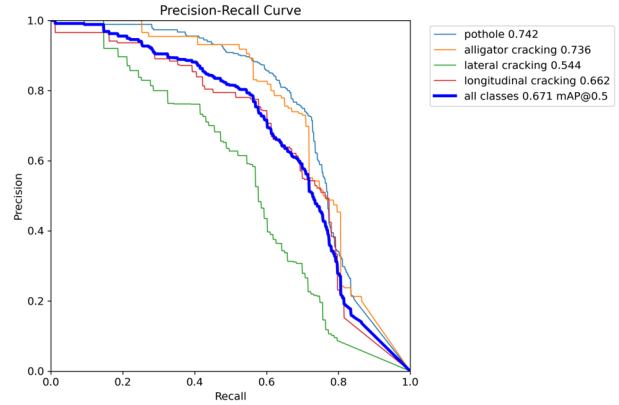


Fig 9. PR\_curve

(8) R\_curve (recall-confidence curve)

Recall indicates the proportion of all positive examples that are correctly predicted as positive, while confidence reflects how certain the model is about its predictions. With Figure 10, we can observe that the recall is decreasing as the model's confidence in its predictions increases.

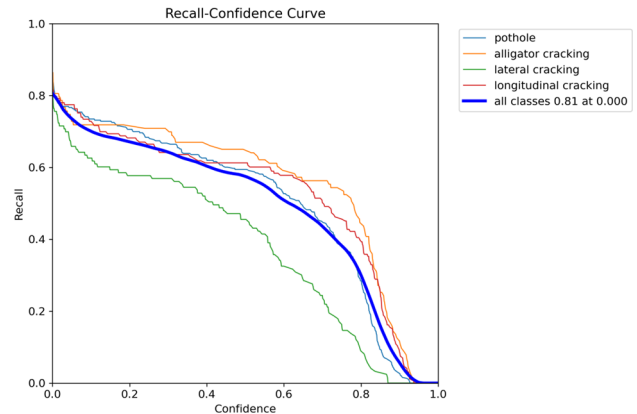


Fig 10. R\_curve

(9) Results (loss function)

As can be seen in Fig. 11, the training process involves losses in several aspects, which gradually decrease with the increase of training data points, indicating that the model is being gradually optimized. At the same time, the precision and recall of the model on the training set are gradually improving, which indicates that the accuracy of the model in recognizing the target is improving. The improvement of these two metrics indicates that the model is performing better and better on the classification task.

#### 4.4. Model Performance Comparison

In this paper, the original YOLOv8 model is compared with the model of YOLOv8 with added CBAM attention mechanism and replaced convolutional kernel, and its training results are shown in Table 3. The results show that the

YOLOv8 model with added attention mechanism is better than the YOLOv8 model with replaced convolutional kernel

and the original YOLOv8 model in terms of real-time performance, accuracy, and training stability.

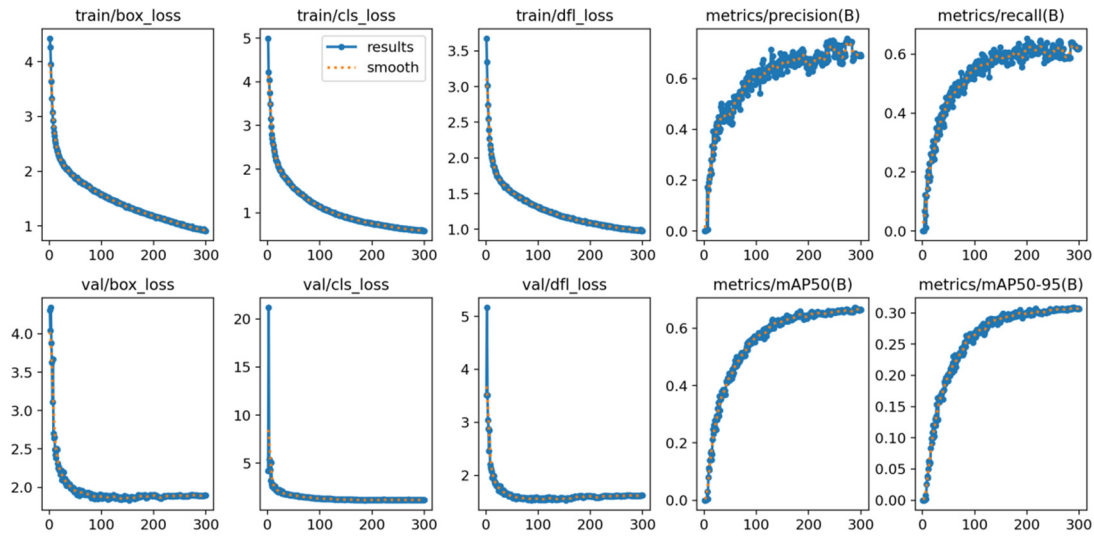


Fig 11. Results (loss function)

Table 3. Model Performance Comparison

Weighting Model	Number of Iterations	Training Batches	P(%)	R(%)	mAP(%)
YOLOv8 model	200	16	0.937	0.80	0.652
Replacement of convolution kernel	200	16	0.945	0.79	0.665
Add CBAM attention mechanism	200	16	0.952	0.81	0.671

## 5. Summary

In order to detect road defects more accurately and efficiently, this paper proposes a road defect detection method based on improved YOLOv8. The original YOLOv8 model is improved by introducing the CBAM attention mechanism and replacing the convolutional kernel, and the two improved models and the original YOLOv8 model are trained on well-labeled road defects dataset respectively. Finally, the training results of each group of models are then compared. The results show that the improved models reflect higher accuracy in road defect detection compared with the original model. Among them, the YOLOv8 model with the added attention mechanism outperforms the other two models in terms of real-time performance, accuracy, and training stability. The method proposed in this paper can identify and detect road defects efficiently, which is of great significance in ensuring public travel safety, optimizing traffic operation, and saving resources.

## Acknowledgments

Science and Technology Research Project of Henan Province, Research on Key Technologies of Agricultural Robot Edge Cloud Computing Unloading based on D2D Collaboration, 242102210050, Yanyan Wang (host);

Henan Province Science and Technology Research Project, Research on key technologies of Abnormal behavior Detection for University security surveillance video, 242102210057, Xue Ran (host).

## References

- [1] Li Wenwen. Design of Road Defect Detection and Recognition System based on Machine vision [D]. Fujian University of Technology, 2023.
- [2] Feng Shujie. Research on Pavement defect Detection Algorithm based on Deep Learning [D]. Shaanxi University of Science and Technology, 2023.
- [3] Munish R, Boris B, Maryam D. Automated Road Defect and Anomaly Detection for Traffic Safety: A Systematic Review. [J]. Sensors (Basel, Switzerland), 2023, 23(12):
- [4] Sun Chaoyun, Pei Lili, Li Wei, et al. Pavement filling crack detection Method based on improved Faster R-CNN [J]. Journal of South China University of Technology (Natural Science Edition), 2024, 48(02): 84-93.
- [5] Jiang Dawei, Wu Zhengping, Jing Siwei. Research on Road Defect Detection and Classification based on improved YOLOv5 [J]. Information Technology and Informatization, 2024 (02); 31-34.
- [6] Yan Banfu, Xu Guanya, Luan Jian, et al. Pavement distress recognition based on Faster R-CNN and Morphological Method [J]. China Journal of Highway and Transportation, 2021, 34(9):181-193.
- [7] Yazhen S, Haixiang T, Huaizhi Z. Automatic Detection of Pavement Marking Defects in Road Inspection Images Using Deep Learning [J]. Journal of Performance of Constructed Facilities, 2024, 38(2).
- [8] Wang Bo, Li Qi, Liu Jiao. A lightweight SSD road crack detection algorithm [J]. Journal of Shangluo University 2022, 36 (4): 83-90.
- [9] WANG Xueqiu, GAO Huan Bing, JIA Zemeng. Improved road defect detection algorithm of YOLOv8 [J/OL]. Computer Engineering and Applications, 1-16[2024-06-28].
- [10] Cheng Z, Gang L, Zekai Z, et al. AAL-Net: A Lightweight Detection Method for Road Surface Defects Based on Attention and Data Augmentation [J]. Applied Sciences, 2023, 13 (3):1435-1435.
- [11] Li Weixiang, Li Wujin, Chen Siyuan. Road defect detection based on image point cloud [J]. Application of Computer Systems, 2019, 33(03):220-225. (in Chinese)

- [12] Xu Tiefeng, Huang He, Zhang Hongmin, et al. Lightweight Road Disease Detection Method Based on improved YOLOv8 [J]. Computer Engineering and Application; 2024, 1-16.
- [13] Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional Block Attention Module. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
- [14] Anghyun, Woo., Jongchan, Park., Joon-Young, Lee. CBAM: Convolutional Block Attention Module[J]. arXiv: 1807.06521
- [15] Jifeng Dai, Haozhi Qi, Yuwen Xiong. (2019). Deformable ConvNets v2: More Deformable, Better Results. arXiv preprint arXiv: 1908.01891.
- [16] Deng Huan. Steel Surface defect detection Method based on regularized YOLO [J]. Science and Technology Innovation and Application, 2019,14(11); 168-172.