

Research Article

Research on Asphalt Pavement Disease Detection Based on Improved YOLOv5s

Lingxiao Wu,¹ Zhugeng Duan ,¹ and Chenghao Liang²

¹School of Civil Engineering, Central South University of Forestry and Technology, Changsha, Hunan 410004, China

²School of Civil Engineering, Hunan University, Changsha, Hunan 410082, China

Correspondence should be addressed to Zhugeng Duan; duanzg@csuft.edu.cn

Received 24 October 2022; Revised 11 February 2023; Accepted 16 February 2023; Published 15 March 2023

Academic Editor: Antonio Lazaro

Copyright © 2023 Lingxiao Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pavement disease detection and classification is one of the key problems in computer vision and intelligent analysis. This is an automated target detection technology with great development potential, which can improve the detection efficiency of road management departments. The research based on the convolutional neural network is aimed at realizing asphalt pavement disease detection based on low resolution, occlusive interference, and complex environment. Considering the powerful function of the convolutional neural network and its successful application in object detection, we apply it to asphalt pavement disease detection, and the detection results are used for subsequent analysis and decision-making. At present, most of the research on pavement disease detection focuses on crack detection, and the detection of multiclass diseases is less, and its detection accuracy and speed need to be improved, which does not meet the actual engineering application. Therefore, a rapid asphalt pavement disease detection method based on improved YOLOv5s was proposed. The complex scene data enhancement technique was developed, which is used to enhance and extend the original data to improve the robustness of the model. The improved lightweight attention module SCBAM was integrated into the backbone network, which can enhance the feature extraction ability and improve the detection performance of the model for small targets. The spatial pyramid pooling was improved into SPPF to fuse the input features, which can solve the multiscale problem of the target and improve the reasoning efficiency of the model to a certain extent. The experimental results showed that, after the model is improved, the average accuracy of pavement disease reaches 94.0%. Compared with YOLOv5s, the precision of the improved YOLOv5-pavement is increased by 3.1%, the recall rate is increased by 4.4%, the F1 score is increased by 3.7%, and the mAP is increased by 3.8%. For transverse cracks, longitudinal cracks, mesh cracks, potholes, and repaired pavement, the detection accuracy of pavement disease detection method based on YOLOv5-pavement is improved by 3.4%, 3.1%, 4.0%, 7.5%, and 4.8%, respectively, compared with that based on YOLOv5s. The proposed method provides support for the detection work of pavement diseases.

1. Introduction

Asphalt pavement is durable, comfortable to drive on, smooth, and has other characteristics, that make it widely used in road construction. During the service of asphalt roads, it is easy to produce diseases due to vehicle load and water intrusion, which affect the comfort of automobile driving and road traffic safety. Therefore, to ensure the normal service of roads, timely monitoring, evaluation, and repair of road diseases are very important and necessary. At present, the detection methods of pavement diseases mainly include visual inspection, camera survey, and ground pene-

trating radar [1, 2]. Visual inspection relies on human eye observation or video to observe road disease, which is inefficient and has large errors. Camera measurement requires the use of professional detection vehicles driving along the lane and taking images of the road surface, and the use of post-processing software to identify the types of pavement diseases. This method has high detection accuracy, but the detection cycle is long and the processing links are too many, resulting in low efficiency. Ground penetrating radar (GPR) can detect most pavement diseases by transmitting an electromagnetic pulse [3], but it is difficult to apply on a large scale because of its high cost. Zhou et al. [4] put forward a

focus on intelligent road disease detection and summarized the commonly used detection equipment in intelligent road disease detection technology, including cameras, GPR, LiDAR, and IMU. The evolution and development of road disease detection technology are described systematically, which is of practical significance to the development of road disease detection technology in the future. Therefore, it is of great practical significance for developing an automated pavement disease detection method based on images and deep learning.

With the continuous development of computer technology, object detection algorithm based on deep learning have achieved rapid development and has been widely used in autonomous driving, face recognition, crop disease and pest recognition, defect detection, and other fields [5–7]. There are two main types of object detection algorithms based on deep learning: One is a two-stage target detection algorithm that divides feature extraction and target localization into two stages, such as R-CNN [8, 9] (Region Proposals for Convolutional Neural Networks), Fast R-CNN [10], and Faster R-CNN [11–13]. The second category is a one-stage target detection algorithm that integrates feature extraction and location processing, such as SSD [14] (Single Shot Multibox Detector) and YOLO [15, 16] (You Only Look Once) series. YOLOv5s is an improved version of the YOLO series [17–19], with a simple training process, small physical space occupation, and high detection accuracy that can be used for the recognition and detection of asphalt pavement diseases. YOLOv5s extracts target features through multilayer convolution and pooling operations, which will result in a large amount of feature information loss and low accuracy of the target. The enhancement module can be used to improve the feature extraction ability of the model and help with the detection task. In the industrial field, it is proposed to change the last part of the deep residual network (ResNet) structure into a deformable convolution for lightweight improvement [20]. The feature fusion structure was improved to improve the positioning accuracy and recognition rate [21]. The function detection scale was added [22], and the NMS is replaced by the introduction of Distance Intersection over Union nonmaximum suppression (DIoU NMS) [23, 24], which can suppress error detection and enhance the detection capability for small targets. The performance of the model is greatly improved by adding a small target detection layer, Squeeze and Excitation Networks (SENet) [25], introducing CIOU loss function, and using migration learning methods [26].

However, in the research of road surface disease recognition based on deep learning, existing road surface disease detection methods mainly focus on crack detection, and there are few studies on multiclass disease detection. For example, Zhu [27] proposed a pavement crack detection algorithm based on defect image segmentation and image edge detection. Wu et al. [28] proposed a novel and efficient image-processing method for extracting pavement cracks from blurred and discontinuous pavement images. And the crack detection network is based on a multiscale extended convolution module and upsampling module proposed by Song et al. [29]. Li et al. [30] describe an innovative vision-

based pavement crack detection strategy that provides a direct pavement surface condition index (PCI) for specific pavement locations. This strategy uses a convolutional neural network (CNN) algorithm to mine a database containing more than 5000 pavement damage images to classify pavement crack types, and the overall pavement damage rate detection accuracy reaches 90%. Zhang et al. [31] proposed a unified crack and seal crack detection method, which detects and separates cracks and seal cracks under the same framework. It trains the deep convolutional neural network to preclassify pavement images into cracks, sealed cracks, and background regions. A block threshold segmentation method is proposed to effectively separate crack and seal crack pixels. Finally, a curve detection method based on tensor voting was used to extract the crack/seal crack. For the detection of multiclass road diseases, Dong and Liu [32] proposed an automatic road damage detection and location method using the Mask R-CNN algorithm and active contour model. Tang et al. [33] proposed a new deep-learning framework called IOPLIN. Similarly, Zhao et al. [34] proposed DASNet, a deep convolutional neural network, which can be used to automatically identify road diseases. The network uses demorphable convolution instead of conventional convolution as the input of the feature pyramid. Before feature fusion, the same supervisory signal is added to multiscale features to reduce semantic differences. Context information is extracted through residual feature enhancement, and information loss of the top pyramid feature map is reduced. The loss function is improved according to the problem of unbalance of positive and negative backgrounds. Compared with the Faster RCNN baseline, the method achieved an improvement of 41.1 mAP and 3.4 AP. Mao et al. [35] built a framework of pavement disease recognition and perception system based on UAV, used UAV to carry out the pavement image data acquisition experiment, analyzed the pavement disease image preprocessing technology based on wavelet threshold transform, studied the pavement disease image preprocessing technology based on DPM, and proposed the pavement disease recognition method based on VGG-16 neural network model. Due to the high complexity, many parameters, and large size of these algorithms, the detection speed of many models is slow, which is difficult to meet the practical application. Therefore, it is of high research significance to improve the recognition accuracy of multiclass road diseases and ensure the lightweight model at the same time.

This research takes the four diseases of asphalt pavement, namely, transverse cracks, longitudinal cracks, mesh cracks, and potholes, and the repaired pavement, as the research object, uses the detection vehicle to collect pavement image data, improves the lightweight target detection model YOLOv5s, and develops a special fast detection model YOLOv5-pavement for various diseases of asphalt pavement. The improved attention module SCBAM is integrated into the backbone network. The spatial pyramid pooling structure is improved into SPPF; complex scene data enhancement technology is developed to enhance the training set images. The processing flow is shown in Figure 1. This method provides technical support for the accurate

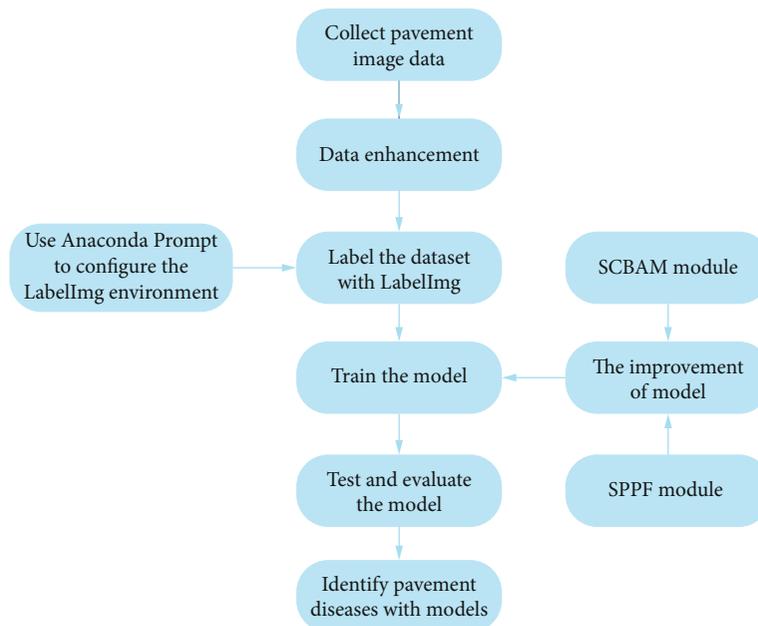


FIGURE 1: The processing flow of the pavement disease detection model.

and rapid detection of multiclass disease targets on asphalt pavement.

2. Materials and Methods

2.1. Construction of Pavement Disease Image Dataset

2.1.1. Study Area and Data Collection Method. This study takes the standard two-lane asphalt highway and four-lane asphalt highway as the research object and obtains the original image of pavement disease. The image data of asphalt roads in Changsha city of Hunan Province were collected by road detection vehicles. The image size is $1628 \text{ pixels} \times 1840 \text{ pixels}$, the pixel size is $5.5 \mu\text{m} \times 5.5 \mu\text{m}$, the sampling spacing is 5 m, the shooting height is 2 m, and the shooting time is daytime. A total of more than 50,000 images were collected in this study, and the collection method was continuous along the road. After the collection, images containing diseases in the data set should be screened out. After inspection, 7641 images in the data set were found to have diseases. Due to certain limitations, this study tested the four diseases of transverse cracks, longitudinal cracks, mesh cracks, potholes, and repaired pavement, as shown in Figure 2.

2.1.2. Image Preprocessing. The object detection model based on deep learning is implemented based on training a large amount of image data, so it is necessary to build a data set with sufficient data volume and rich types. Firstly, 764 images were randomly selected from 7641 disease image data as the test set, and the remaining 6877 images were used as the training set. The distribution details of image samples in the data set are shown in Table 1. Secondly, to improve the training efficiency of the road disease detection model, the original 6877 images of the training set are compressed,

and their length and width are compressed to 1/2 of the original image. Then, the image data annotation software “LabelImg” was used to draw the outer rectangular boxes of different road disease targets in the compressed road disease images to realize the road disease annotation. The images were labeled according to the smallest rectangle around the different pavement diseases to ensure that the rectangle contains as little background area as possible. Finally, to enrich the background information of the image data, data enhancement was performed on the training set to better extract the features of diseases belonging to different labeled categories and avoid overfitting the trained model.

Table 1 shows that there are large differences in the number of different types of diseases. The number of transverse cracks, longitudinal cracks, and mesh cracks is sufficient, and only background information enhancement can be used to build a rich data set. The number of potholes and repaired pavements is seriously insufficient, which is unbalanced with other diseases in terms of quantity, and there are many small targets of potholes, so it needs to be strengthened and expanded. In the actual detection task, complex environmental factors, such as light, shadow, and water are the key to affect the accuracy of the model. Due to the different causes of asphalt pavement diseases, transverse cracks and longitudinal cracks have strong directionality, so the traditional data enhancement methods of stretching and rotation cannot be used to enhance pavement disease data sets. To improve the detection performance of the algorithm model in a complex environment and change the problem of unbalanced samples among multiple diseases, this study referred to the possible complex environment of the road surface and proposed the complex scene data enhancement technology based on the function in the python data enhancement library “imgaug”, which changed the problem of the unbalanced number of different

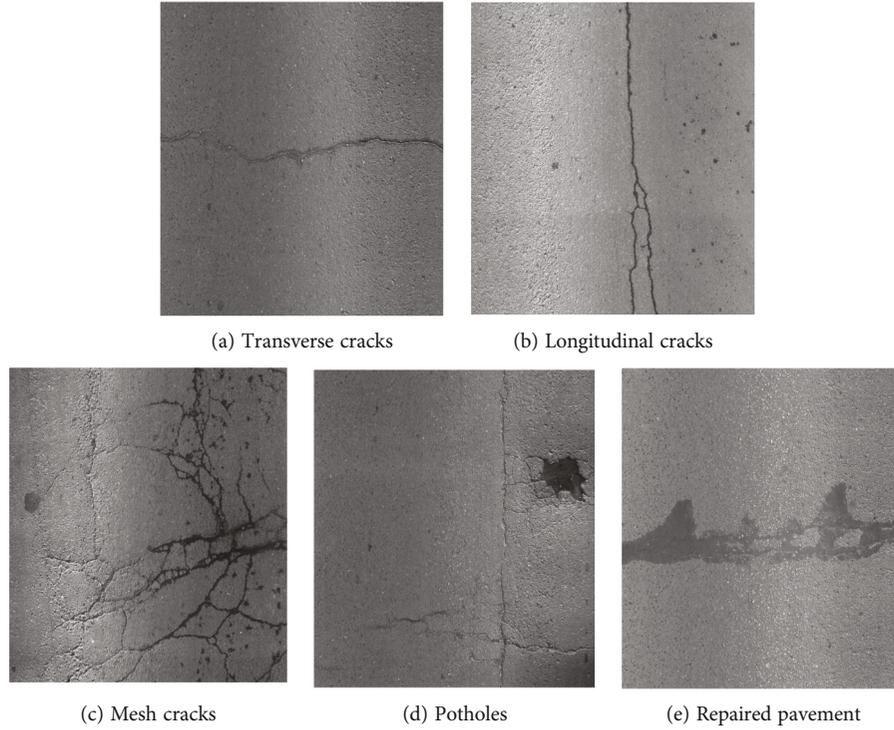


FIGURE 2: Images of different disease categories.

TABLE 1: Detailed information of images in a data set.

Type	Total	The number of training sets	The number of test sets
Image data	7641	6877	764
Target of detection	8482	7572	910
Transverse cracks	2546	2239	307
Longitudinal cracks	2773	2520	253
Mesh cracks	2549	2301	248
Potholes	237	198	39
Repaired pavement	377	314	63

pavement diseases, enriches the data set and improves the robustness of the model in the complex detection environment.

The detailed steps of complex scene data enhancement are as follows: the complex scene data enhancement is divided into two parts. In the first part, the size of the Gaussian noise, fog, rain, snow, mud, and other environmental noise is determined by calling the modules of “Imgaug” library, and then it is gradually added to the original image to imitate the road environment. The second part is to invoke the modules of motion blur, brightness adjustment, and saturation in “Imgaug” library and add them to the original image to mimic the weather environment. In the first part, 2 or 3 enhancement methods are randomly selected, and then in the second part, 1 enhancement method is randomly selected and added to the original image, respectively, to complete the image enhancement process.

The original image data were enhanced by complex scene data enhancement technology, and the amount of data in each category was enhanced to about 2500. The final

training set consisted of 12500 images, which were used as the training data set of the pavement disease detection model, including 5623 enhanced images and 6877 original images. The training set data enhanced by complex scene data is shown in Figure 3.

2.2. The Structure of YOLOv5s. YOLOv5 is the latest in the YOLO series. The network model detection accuracy is high, the inference speed is fast, and the fastest detection speed can reach 140 frames per second (FPS). The size of the YOLOv5 target detection network model is small, only 14.4 M. Therefore, the advantages of the YOLOv5 network are high detection accuracy, lightweight, and fast detection speed.

The YOLOv5s consist of four parts, as shown in Figure 4, including input, backbone, neck, and prediction. Input mainly contains the preprocessing of data. Backbone extracts image features at different levels through multilayer convolution operations. Neck network consists of two parts, including a feature pyramid network (FPN) and the path

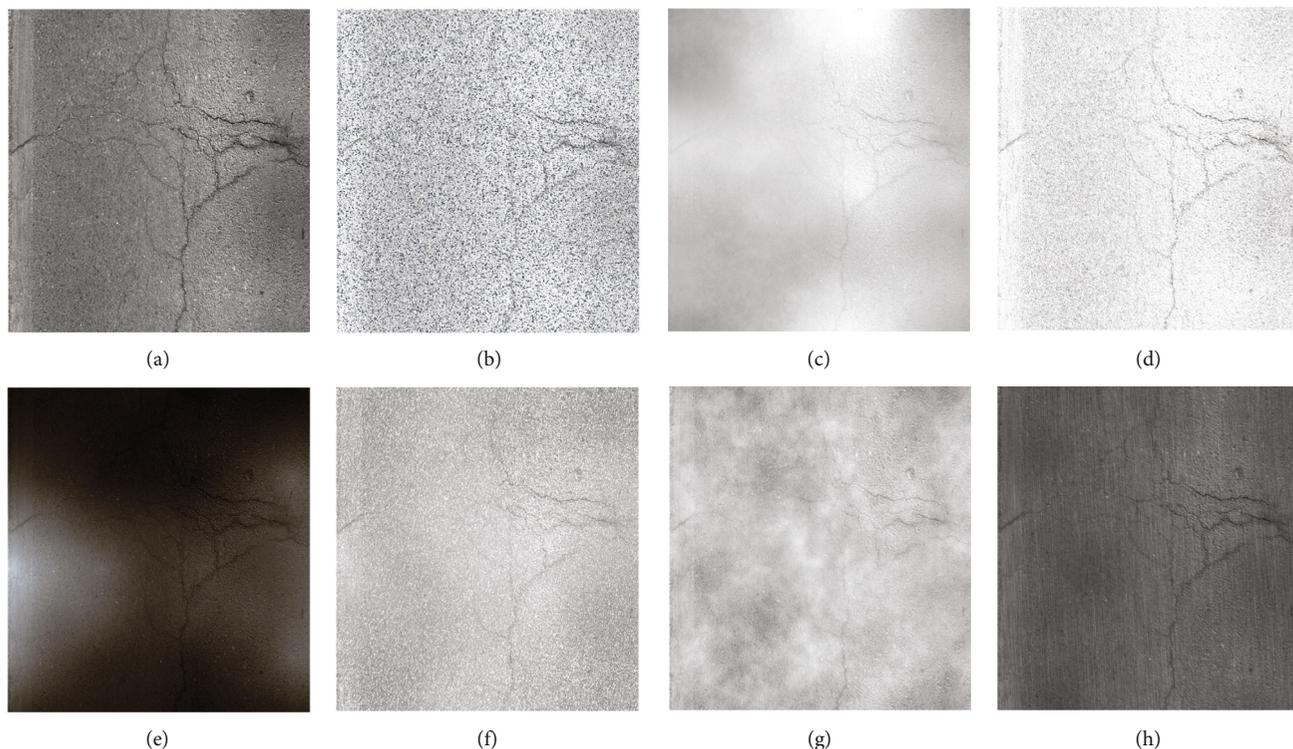


FIGURE 3: Complex scene data enhancement techniques for data sets. (a) The original image. (b) The Gaussian noise point. (c) The fuzzy. (d) High brightness. (e) Low brightness. (f) The snow. (g) The fog. (h) The rain.

aggregation network (PANet). FPN transmits semantic information from top to bottom in the Network. PANet transmits positioning information from bottom to top and integrates information from different network layers on the backbone to further improve detection capability. The main function of prediction is to predict targets of different sizes on feature maps of different sizes.

2.3. Improvement of YOLOv5s. The pavement disease detection algorithm needs to accurately identify various diseases under various circumstances in the complex pavement environment. Considering the simple structure of YOLOv5s, the weak feature extraction ability, and the complex pavement environment, these factors will lead to the low detection accuracy of the pavement disease detection algorithm. The attention mechanism has been proven to enhance the ability of model-degree feature extraction. Adding attention opportunity to the road surface disease detection model is helpful to improve the detection accuracy for complex environments and small targets. Therefore, this study improves the backbone network of YOLOv5s by adding the improved attention module SCBAM to improve the feature extraction ability of the model, which can improve the detection accuracy of the model in complex scenes and strengthen the detection ability of small targets. At the same time, in order to solve the multiscale problem of the target to a certain extent and improve the reasoning speed of the model, this paper improved the SPP module to the SPPF. In order to achieve the high reasoning speed of the model and improve

its detection accuracy, the asphalt pavement disease detection model can be optimized and improved.

2.3.1. CBAM Attention Module and Its Improvement. The convolutional block attention module (CBAM) is lightweight. CBAM processes the input feature map through the internal channel attention module and spatial attention module to simplify Feature extraction and improve the detection speed of the model. The action mechanism is weighted fusion (Equations (1) and (2)). In Equation (1), F is the input feature map. F_1 is the feature map obtained by channel attention weighting. F is the feature map obtained by spatial attention weighting. $M_c(F_1)$ is the channel attention output weight. $M_s(F_1)$ is the spatial attention output weight. The CBAM attention module and its improvement are shown in Figure 5.

$$F_1 = M_c(F) \otimes F, \quad (1)$$

$$F_2 = M_s(F_1) \otimes F_1, \quad (2)$$

The mechanism of the channel attention module is as follows: the input feature map F is processed by global max pooling and global average pooling to obtain two $1 \times 1 \times c$ feature maps. The multilayer perceptron (MLP) was fed with them, and the feature map of channel attention was generated by the addition operation of element-wise on the output feature map of MLP, which was processed by the Sigmoid activation function. Finally, M_c an input feature map F is processed, and an element-wise multiplication

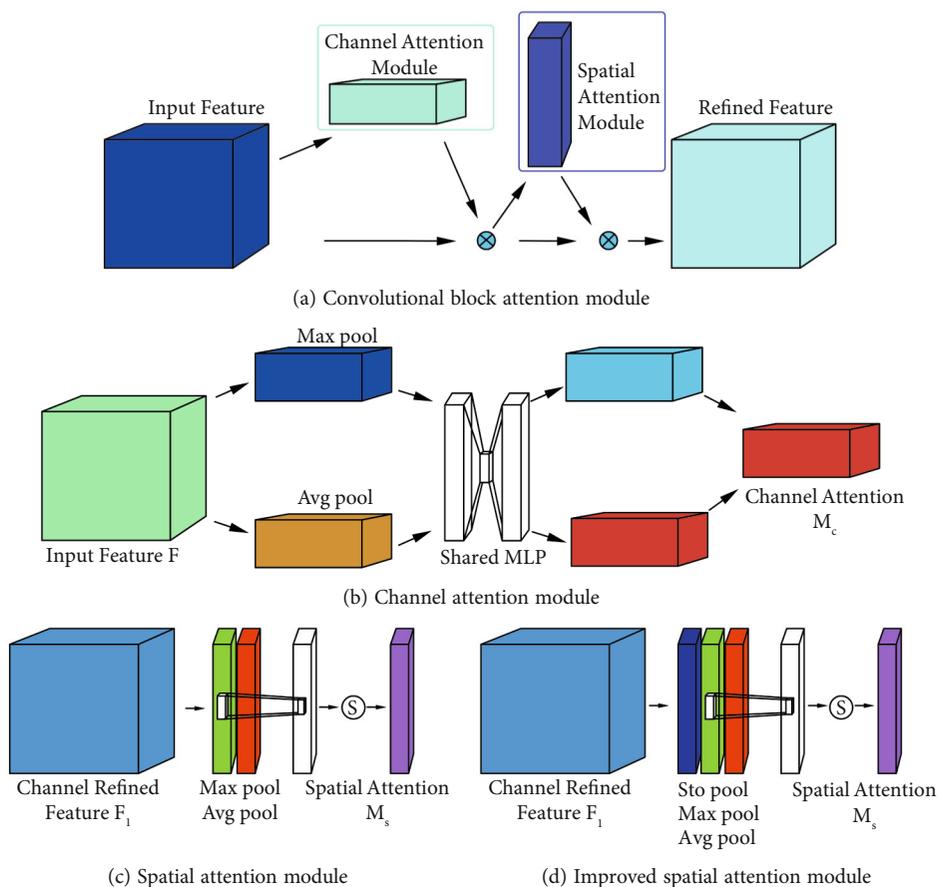


FIGURE 5: The CBAM attention module and its improvement.

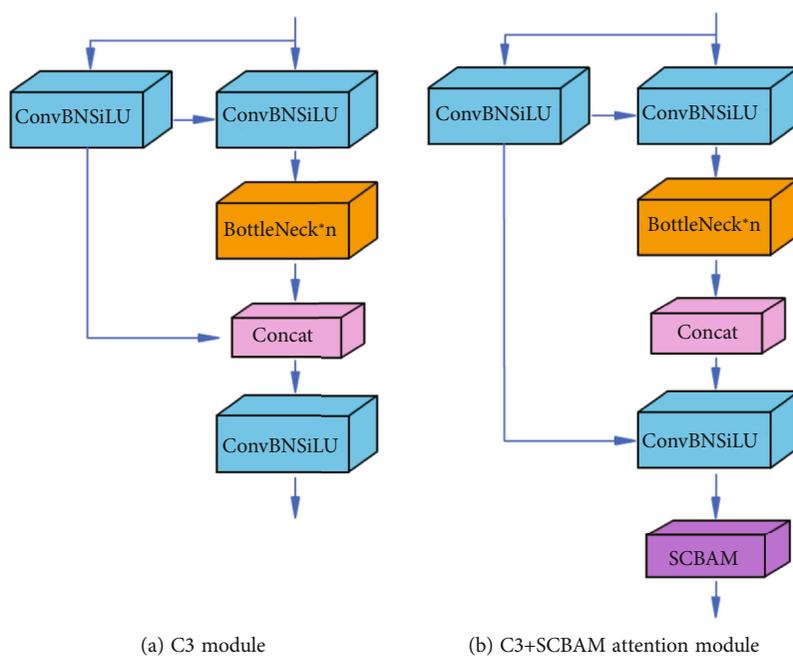


FIGURE 6: C3 module and attention module.

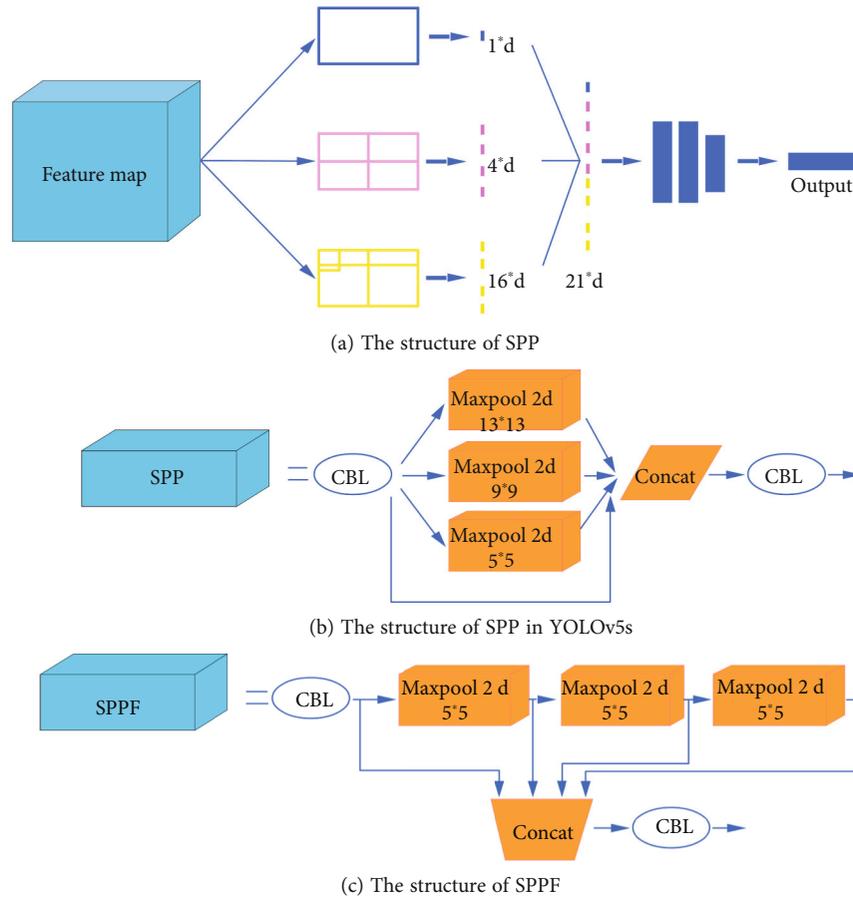


FIGURE 7: Structures of different SPP.

to give more weight to important areas. When outputting diseases, it will focus on the corresponding areas in the picture to improve the feature extraction ability of road diseases.

2.3.2. Improvement of Spatial Pyramid Pooling. Spatial pyramid pooling (SPP) was proposed by Microsoft in 2015 [36]. It can transform the input features of any scale into the same dimension and output the features after block stacking in parallel, which can solve the problem of underutilization of deep semantic information. Different SPP structures are shown in Figure 7, and the feature extraction process includes the following: firstly, the input feature map is partitioned by pooling of 1×1 , 2×2 , and 4×4 sizes, and 21 subblocks are obtained. Secondly, the maximum value is selected from the 21 subblocks, so that the input feature map of arbitrary size is converted into a fixed size of the 21-dimensional feature. Finally, the 21-dimensional features are stacked and pooled to complete the whole SPP process. Compared with the original SPP structure, the new SPP has one more convolution block (Conv, Batch Normalization, Leaky ReLU, CBL) before and after the new SPP, and the size of the pool core in the middle is 1×1 , 5×5 , 9×9 , and 13×13 , respectively.

The SPP module realizes the fusion of local features and global features, enriches the expression ability of the final

feature mAP, and thus improves the mAP. However, the parallel pooling method is time-consuming. Therefore, in this paper, the improved spatial pyramid pooling structure (SPPF) with higher efficiency is adopted for pooling processing, and the deep semantic information expression ability is enhanced by stacking features after multicore pooling. Its structure is to change the pooling mode of input feature layers from parallel to serial, and stack pooling is carried out after multiple convolutions in serial mode to increase the image receptive field and enrich feature information. SPPF uses three 5×5 serial pooling layers instead of a single 13×13 pooling layer to obtain the same processing effect, but the calculation speed is significantly improved. Under the conditions of the same basic convolution code, the number of input and output layers is 32 and 128, respectively, and the inference time of SPP is 0.5373 s, while the inference time of SPPF is 0.2078 s. SPPF takes 0.3295 s less than spp.

Based on the improvement of the YOLOv5s detection task for multiclass asphalt pavement diseases, the overall framework of asphalt pavement disease detection was proposed and named YOLOv5-pavement. The attention module was added to the backbone network, the SPP module was improved to SPPF, and the data set was enhanced by using data enhancement technology. Its overall framework is shown in Figure 8.

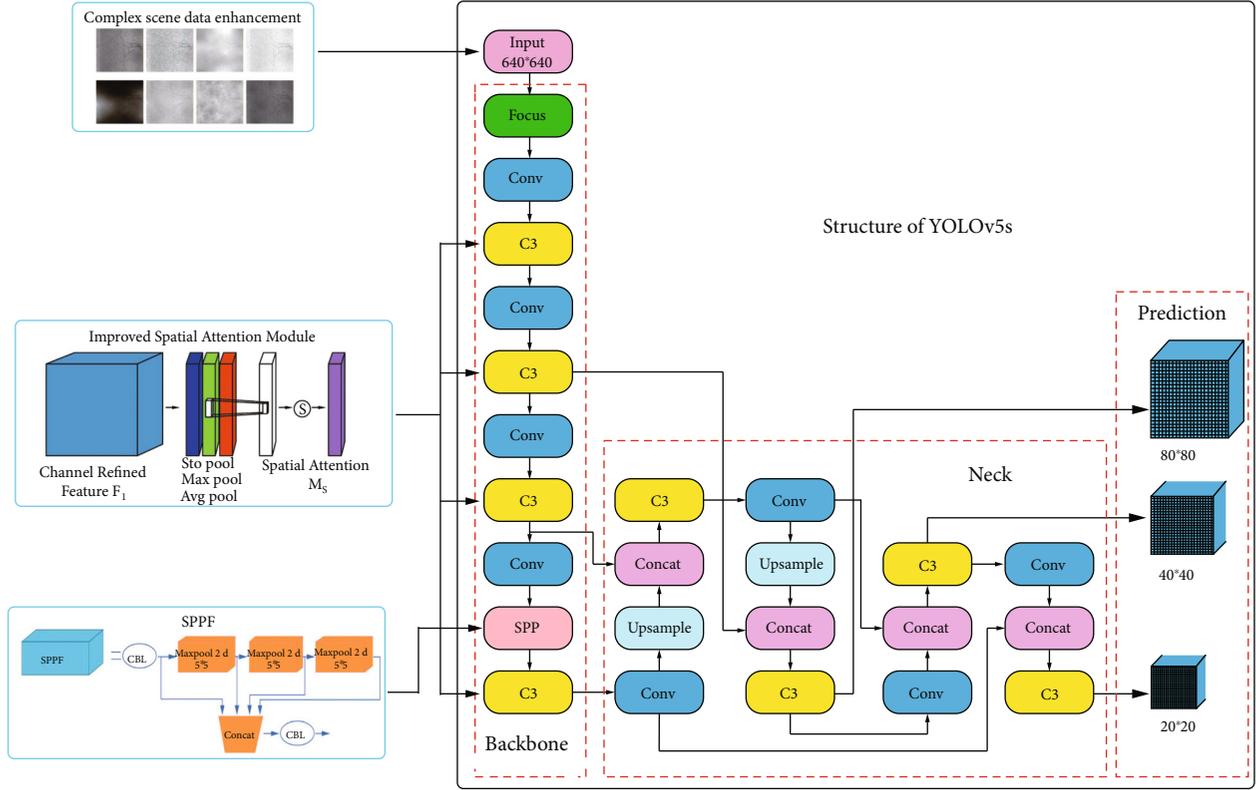


FIGURE 8: The structure of YOLOv5-pavement.

2.4. Experimental Environment Configuration. The model was trained using a graphics processing unit (GPU), and the computer is configured with an RTX3070M graphics card. The PyTorch deep learning framework is built under Windows 10 system, the program code is written in Python language and CUDA is called. Cudnn, OpenCV, and other required libraries to achieve the training and validation of pavement disease detection models.

The initial learning rate, weight attenuation coefficient, training momentum, confidence, IOU threshold, and Epoch of the pavement disease detection model training based on YOLOv5-pavement were set to 0.001, 0.0005, 0.9, 0.5, and 300, respectively. After training, the weight file of the pavement disease detection model was saved, the test set was validated, and the results were used to evaluate the performance of the model. The final output of the network is the identification of the five target boxes and the probability of belonging to a specific category.

2.5. Evaluation Indicators of Model. Precision, recall, F1 score, mean average precision (mAP), space occupied by the model, and reasoning speed were selected as the evaluation indexes of each model.

Precision and recall are important indexes for evaluating model accuracy, as shown in Equations (3) and (4).

$$P = \frac{n_{TP}}{n_{FP} + n_{TP}}, \quad (3)$$

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

where P is the precision, R is the recall, TP is the true positives, FP is the false positives, and FN is the false negatives.

The F1 score is a measure of classification problems. It is the harmonic average of the accuracy rate and recall rate, as shown in Equation (5):

$$F1 = 2 \frac{P \times R}{P + R} \quad (5)$$

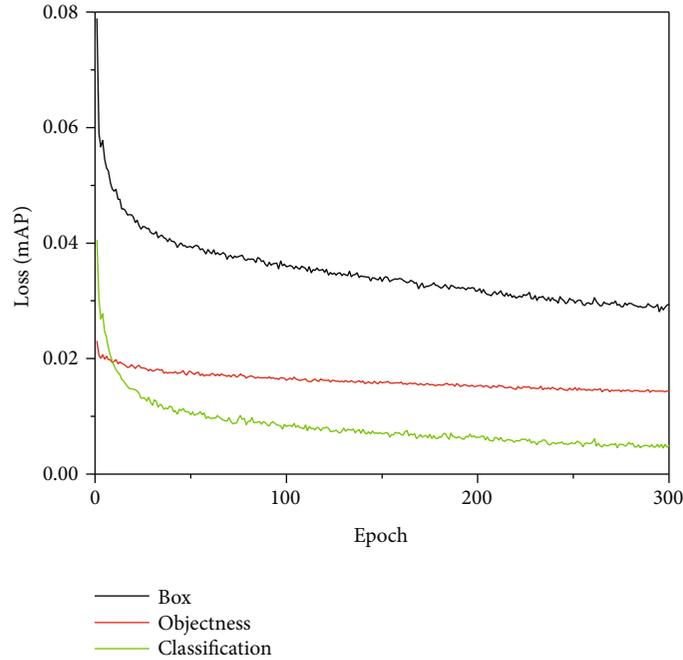
where P is the precision and R is the recall.

AP and mAP are calculated, as shown in Equations (6) and (7).

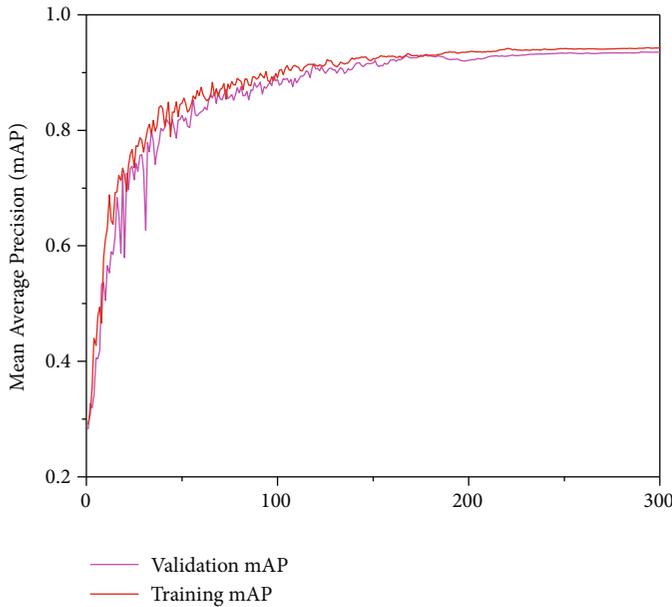
$$AP@0.5 = \frac{1}{n} \sum_{i=1}^n P_i = \frac{1}{n} P_1 + \frac{1}{n} P_2 + \dots + \frac{1}{n} P_n, \quad (6)$$

$$mAP@0.5 = \frac{1}{n} \sum_{i=1}^n AP_n, \quad (7)$$

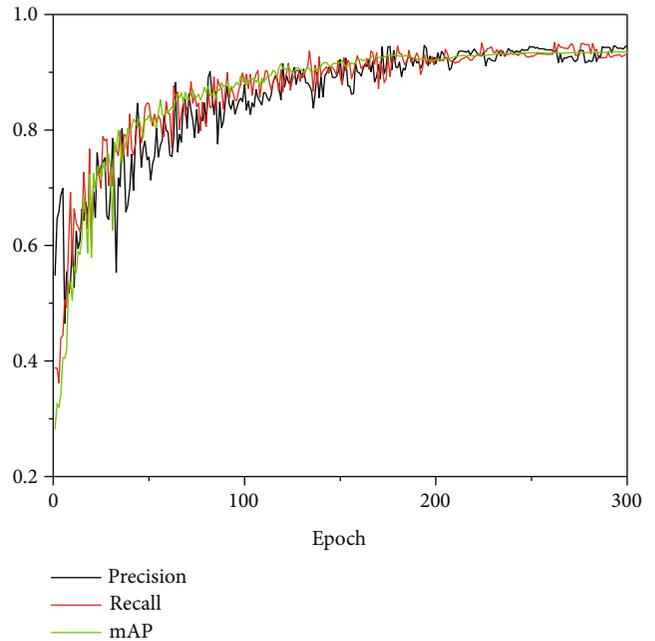
AP at 0.5 is defined as when the IOU threshold is 0.5, all detection results of a certain type of sample with n positive examples are arranged in descending order of confidence, and each additional positive example corresponds to a Precision value (P_i). The average value of P is obtained to obtain AP at 0.5, and mAP at 0.5 is the mean value of all AP classes. n is the number of detected categories. The space occupied



(a) The loss of the training process



(b) Training and validation mean average precision of the model



(c) Validation Precision-Recall and mean average precision of the model

FIGURE 9: The training process of YOLOv5-pavement.

by the model is counted by megabytes (M), and the inference speed of the model is evaluated by FPS.

3. Results

3.1. Results of Training and Validation. The change in the training loss value of YOLOv5-pavement is shown in Figure 9, where the results show that the loss value drops sharply when the number of iterations is from 0 to 20. From

20 to 300 times, the change range of loss value was stable, showing a slow downward trend. After 300 iterations, the loss value tends to be stable around 0.02, and the model reaches a relatively stable state to complete the training. The results in the figure show that the model is well trained without overfitting.

3.2. Analysis of Pavement Disease Detection. In order to verify the performance of the pavement disease detection model

TABLE 2: The validation results of YOLOv5-pavement.

	Number	Precision/%	Recall/%	mAP/%	F1/%
Total	910	91.2	92.9	93.5	92.0
Transverse cracks	307	91.2	92.2	91.9	91.7
Longitudinal cracks	253	90.1	92.1	92.0	91.1
Mesh cracks	248	91.9	94.4	96.7	94.1
Potholes	39	94.9	94.9	94.7	94.9
Repaired pavement	63	87.3	90.5	92.1	88.9

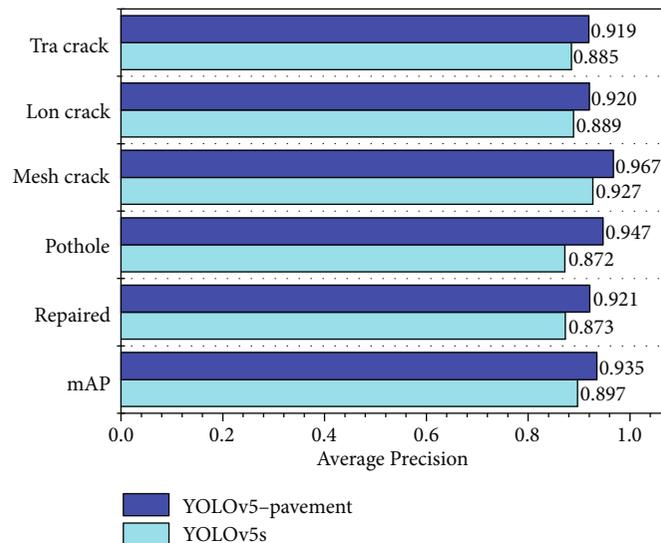


FIGURE 10: Validation mAP of pavement diseases.

based on YOLOv5-pavement proposed in this study, the recognition results of the model on 764 images in the test set were further analyzed. There are 910 targets to be detected in the 764 test set images, including 307 targets for transverse cracks, 253 targets for longitudinal cracks, 248 targets for mesh cracks, 39 targets for potholes, and 63 targets for repaired pavement. The pavement disease detection results based on YOLOv5-pavement proposed in this study are shown in Table 2 and Figure 10, including the precision, recall, and mAP of each type of detection target.

The pavement disease detection and validation results based on YOLOv5s in Table 2 and Figure 10 shows that the average precision of transverse cracks, longitudinal cracks, mesh cracks, potholes, and repaired pavement is 88.5%, 88.9%, 92.7%, 87.2%, and 87.3%, respectively, and the mAP is 89.7%. In the pavement disease detection validation based on YOLOv5-pavement, the average precision of transverse cracks, longitudinal cracks, mesh cracks, potholes, and repaired pavement is 91.9%, 92.0%, 96.7%, 94.7%, and 92.1%, respectively, and the mAP is 93.5%. In the pavement disease detection validation based on YOLOv5-pavement, the F1 score of transverse cracks, longitudinal cracks, mesh cracks, potholes, and repaired pavement is 91.7%, 91.1%, 94.1%, 94.9%, and 88.9%. The F1 score of the model is 92.0%.

Figure 10 shows that the detection average precision of YOLOv5s for transverse cracks, longitudinal cracks, pot-

holes, and repaired pavement is lower than 90%. This is because many transverse cracks, longitudinal cracks, and potholes collected in this study exist in the form of small targets, while the feature extraction ability of YOLOv5s is weak, resulting in the poor detection ability of small targets and the low average precision of potholes. The characteristics of the repaired pavement are often not obvious, and the feature extraction ability of YOLOv5s is weak, resulting in low mAP of the repaired pavement. Compared with the pavement disease detection method based on YOLOv5s, the average precision of YOLOv5-pavement for transverse cracks, longitudinal cracks, mesh cracks, potholes, and repaired pavement had been improved by 3.4%, 3.1%, 4.0%, 7.5%, and 4.8%, respectively. The detection results show that the improved scheme in this paper has a positive effect on the detection of pavement diseases, effectively improves the feature extraction ability of the pavement disease detection model and the detection ability of small targets, and significantly improves the average precision of the four types of asphalt pavement diseases and the repaired pavement. Figure 11 showed the detection results of the two models for five types of targets in the complex environment. It can be found that the pavement disease detection model based on YOLOv5s misses mesh cracks under strong light conditions and misses potholes with small areas under weak light conditions. The pavement disease detection model based on YOLOv5-pavement had not been disturbed by

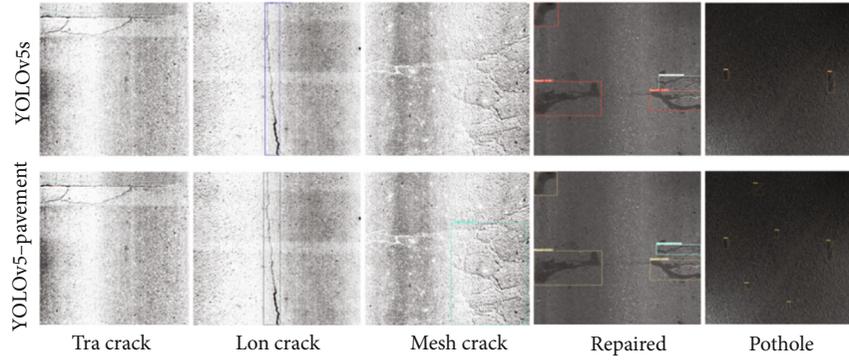


FIGURE 11: Detection results of two models in a complex environment.

TABLE 3: Comparison results of the YOLOv5 series.

	Model	Precision/%	Recall/%	F1/%	mAP/%	Param/M	FPS
A	YOLOv5s	88.1	88.5	88.3	89.7	14.4	105
B	YOLOv5s+SPPF	88.7	89.3	89.0	90.1	14.4	109
C	YOLOv5s+CBAM	90.3	91.3	90.8	91.8	14.1	78
D	YOLOv5s+SCBAM	90.9	91.7	91.3	92.2	14.1	78
E	YOLOv5-pavement	91.2	92.9	92.0	93.5	14.2	82

the environment, the mesh cracks were completely detected, and all the potholes were detected in the low-brightness environment.

3.3. Ablation Experiment and Comparison of Different Models. To verify the lifting effect of each module of the improved algorithm, ablation experiments were conducted in this paper. Based on YOLOv5s, the CBAM module, SCBAM module, and SPPF were added, respectively, to explore the influence of the improved algorithm on the model. The ablation experiment results are shown in Table 3:

The ablation experiment results show that the precision, recall, F1 score, mAP, and reasoning speed of the plan B model are improved by 0.6%, 0.8%, 0.7%, 0.4%, and 4FPS, respectively, compared with plan A by adding the SPPF module. The results show that SPPF can better integrate the feature information of the target in the multiscale target detection task, improve the detection accuracy of the algorithm, and accelerate the reasoning speed of the model to a certain extent. Compared with plan A, the precision of plan D integrated with the improved SCBAM attention module is increased by 2.8%, the recall is increased by 3.2%, the F1 score is increased by 3.0%, and the mAP is increased by 2.5%. It showed that the feature extraction ability of the model is enhanced after adding SCBAM. At the same time, compared with plan C, the accuracy of the plan D model integrated with the SCBAM is increased by 0.6%, the recall is increased by 0.4%, the F1 score is increased by 0.5%, and the mAP is increased by 0.4%. It shows that the improved SCBAM retains more target features than CBAM and strengthens the feature extraction ability of the model. The results of the pave-

ment disease detection model based on YOLOv5-pavement in plan E show that the precision-recall F1 score and mAP are increased by 3.1%, 4.4% 3.7%, and 3.8% compared with the results in plan A, respectively. The results of ablation experiments show that the improved module has a positive effect on the model.

To verify the effectiveness of the proposed method, the proposed algorithm model is compared with other object detection algorithm models, and 200 pavement images were selected for testing. Faster R-CNN, SSD, YOLOv3, YOLOv4, and YOLOv5-pavement are proposed in this paper and are selected for lateral comparison experiments. The results are shown in Table 4.

Experimental results show that the average accuracy of the pavement disease detection model based on YOLOv5-pavement was slightly lower than the two-stage algorithm faster R-CNN and was second only to YOLOv4 in the one-stage algorithm, reaching 93.5%. Compared with SSD and YOLOv3, YOLOv5 pavement saw mAP increases of 7.4% and 9.6%, respectively. In terms of the space occupied by the model, the YOLOv5 pavement is the smallest, at only 14.1 M, and the size of the model is only 7.5% of SSD and 6.0% of YOLOv3. Moreover, YOLOv5-pavement has obvious advantages in reasoning speed, with FPS reaching 82, which is 34FPS, 70FPS, 26FPS, and 18FPS faster than SSD, faster R-CNN, YOLOv3, and YOLOv4, respectively. SSD, Faster R-CNN, YOLOv3, and YOLOv4 spent 16.1s, 29.3s, 14.2s, and 14.0s, respectively, in detecting 200 pavement images, while the time of YOLOv5-pavement proposed in this paper was only 12.6s. The method in this paper has the fastest detection speed when dealing with the same detection task.

TABLE 4: Test results of different models.

Object detection networks	mAP/%	Param/M	FPS	Time
SSD	86.1	189	48	16.1 s
Faster R-CNN	95.7	108	12	29.3 s
YOLOv3	83.9	236	56	14.2 s
YOLOv4	94.9	200.3	64	14.0 s
YOLOv5 pavement	93.5	14.1	82	12.6 s

4. Discussion

The advantages of the pavement disease detection method proposed in this paper are reflected in the following aspects: first, the algorithm model can detect and identify multiclass diseases, which is not possessed by many pavement disease detection models. Secondly, the model has high stability in a complex environment and can adapt to various complex detection environments. Finally, the detection accuracy of the improved YOLOv5 pavement is significantly improved, and the high reasoning speed and lightweight are maintained, which has the potential for large-scale deployment and application.

YOLOv5 includes four different architectures (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x), whose main difference is the number of feature extraction modules and convolution at specific locations in the network. The size of the model and the number of model parameters increase in each of the four architectures. So YOLOv5 has high flexibility, according to the actual requirements to choose different structures for development and application. This study considers the precision and reasoning speed of the pavement disease detection algorithm, considering the pavement disease detection methods depend on the road detection vehicle collected along the road, with minimal YOLOv5s size, implement the deployment of the biggest potential in the mobile terminal, can be deployed on the test vehicle and in the process of acquisition directly in pavement disease detection.

The pavement disease detection method proposed in this study relies on images to detect and classify pavement diseases. This method cannot work without light at night, and there is a lot of traffic during the day, which will have a certain impact on the detection work. Due to the limitation of experimental conditions, this study only studied some possible diseases of asphalt pavement, did not detect and classify all kinds of diseases of asphalt pavement, and did not consider the possible types of diseases caused by roads of other materials, which limited its application scope. On the other hand, the method proposed in this paper can only detect road surface diseases, while the diseases generated by the internal structure of the road cannot be detected, which is another limitation of this research method.

5. Conclusion

Aiming at the problems of lack of asphalt pavement disease detection means, low accuracy of automated detection

model, poor ability of small target detection, and easy to be disturbed by environment, an asphalt pavement disease detection method based on improved YOLOv5s was proposed.

A complex scene data enhancement technique is proposed to balance different types of pavement diseases and simulate complex road environments to improve the robustness of the model. The SCBAM attention module is integrated into the backbone network to optimize the feature extraction ability of the backbone network and improve the feature extraction ability and detection accuracy of the pavement disease detection model for difficult-to-detect objects. The SPP is improved as SPPF, and the input features are further fused, which can solve the target multiscale problem and improve the reasoning speed to a certain extent. The results showed that the improved network model can effectively recognize road surface diseases. The precision, recall, F1 score, mAP, and detection speed of the model reached 91.2%, 92.9%, 92.0%, 93.5%, and 82 FPS, respectively. Compared with the detection methods based on YOLOv5s, the mAP of transverse cracks, longitudinal cracks, mesh cracks, potholes, and the repaired pavement has increased by 3.4%, 3.1%, 4.0%, 7.5%, and 4.8%, respectively, and the mAP has increased by 3.8%. The ablation experiment results show that the improved scheme proposed in this paper can improve the accuracy of the pavement disease detection method, and maintain the light weight of the model, and the reasoning speed of the model is also maintained at a high level. The results of the comparison between YOLOv5-pavement and other models show that the detection accuracy of YOLOv5-pavement has certain advantages, and the accuracy gap between YOLOv5-pavement and faster R-CNN with the highest accuracy is very small. Moreover, the reasoning speed of YOLOv5-pavement is the fastest, and the model occupies only 14.1 M. The asphalt pavement disease detection model presented in this paper has certain practical value for the detection of asphalt pavement diseases.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflict of interest.

Authors' Contributions

All the authors contributed substantially to the manuscript. Original draft preparation and formal analysis were performed by Lingxiao WU. Supervision was conducted by Zhugeng Duan. Review and editing were done by Chenghao Liang. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This work was supported by the Nature Foundation of the Department of Science and Technology of Hunan Province (research on geological hazard monitoring based on UAV Lidar and high-resolution image data) (2020JJ4941).

References

- [1] W. Xia, "An approach for extracting road pavement disease from HD camera videos by deep convolutional networks," in *2018 International Conference on Audio, Language and Image Processing (ICALIP)*, pp. 418–422, Shanghai, China, 2018.
- [2] J. Xie, F. Niu, W. Su, Y. Huang, and G. Liu, "Identifying airport runway pavement diseases using complex signal analysis in GPR post-processing," *Journal of Applied Geophysics*, vol. 192, article 104396, 2021.
- [3] I. Giannakis, F. Tosti, L. Lantini, and A. M. Alani, "Diagnosing emerging infectious diseases of trees using ground penetrating radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 2, pp. 1146–1155, 2020.
- [4] Y. Zhou, X. Guo, F. Hou, and J. Wu, "Review of intelligent road defects detection technology," *Sustainability*, vol. 14, no. 10, p. 6306, 2022.
- [5] H. Wan, L. Gao, M. Su, Q. You, H. Qu, and Q. Sun, "A novel neural network model for traffic sign detection and recognition under extreme conditions," *Journal of Sensors*, vol. 2021, Article ID 9984787, 16 pages, 2021.
- [6] H. Wang and S. Zhang, "A robust and lightweight detector for ship target with complex background in SAR image," *Journal of Sensors*, vol. 2022, Article ID 8199418, 16 pages, 2022.
- [7] Y. Tian, G. Yang, Z. Wang, E. Li, and Z. Liang, "Detection of apple lesions in orchards based on deep learning methods of CycleGAN and YOLOV3-dense," *Journal of Sensors*, vol. 2019, Article ID 7630926, 13 pages, 2019.
- [8] S. S. Ahmed, Z. Mehmood, I. A. Awan, and R. M. Yousaf, "A novel technique for handwritten digit recognition using deep learning," *Journal of Sensors*, vol. 2023, Article ID 2753941, 15 pages, 2023.
- [9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, 2016.
- [10] L. Yu, J. Shi, C. Huang et al., "An integrated rice panicle phenotyping method based on X-ray and RGB scanning and deep learning," *Crop Journal*, vol. 9, no. 1, pp. 42–56, 2021.
- [11] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, Santiago, Chile, 2015.
- [12] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater image processing and object detection based on deep CNN method," *Journal of Sensors*, vol. 2020, Article ID 6707328, 20 pages, 2020.
- [13] X. Lou, Y. Huang, L. Fang et al., "Measuring loblolly pine crowns with drone imagery through deep learning," *Journal of Forestry Research*, vol. 33, no. 1, pp. 227–238, 2022.
- [14] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multi-box detector," in *Computer Vision – ECCV 2016. ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9905 of Lecture Notes in Computer Science, pp. 21–37, Springer, Cham, 2016.
- [15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, Las Vegas, NV, USA, 2016.
- [16] H. Huang, Q. Liang, D. Luo, and D. H. Lee, "Attention-enhanced one-stage algorithm for traffic sign detection and recognition," *Journal of Sensors*, vol. 2022, Article ID 3705256, 8 pages, 2022.
- [17] D. Wang and D. He, "Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning," *Biosystems Engineering*, vol. 210, pp. 271–281, 2021.
- [18] P. Zhang and D. Li, "EPSA-YOLO-V5s: a novel method for detecting the survival rate of rapeseed in a plant factory based on multiple guarantee mechanisms," *Computers and Electronics in Agriculture*, vol. 193, article 106714, 2022.
- [19] B. Yan, P. Fan, X. Lei, Z. Liu, and F. Yang, "A real-time apple targets detection method for picking robot based on improved YOLOv5," *Remote Sensing*, vol. 13, no. 9, p. 1619, 2021.
- [20] S. Targ, D. Almeida, and K. Lyman, "ResNet in ResNet: generalizing residual architectures," 2016, <https://arxiv.org/abs/1603.08029>.
- [21] M. Gong and Y. Shu, "Real-time detection and motion recognition of human moving objects based on deep learning and multi-scale feature fusion in video," *IEEE Access*, vol. 8, pp. 25811–25822, 2020.
- [22] S. Tan, G. Lu, Z. Jiang, and L. Huang, "Improved YOLOv5 network model and application in safety helmet detection," in *2021 IEEE International Conference on Intelligence and Safety for Robotics (ISR)*, pp. 330–333, Tokoname, Japan, 2021.
- [23] O. M. Lawal, "YOLOMuskmelon: quest for fruit detection speed and accuracy using deep learning," *IEEE Access*, vol. 9, pp. 15221–15227, 2021.
- [24] H. Xie and Z. Wu, "A robust fabric defect detection method based on improved RefineDet," *Sensors*, vol. 20, no. 15, p. 4260, 2020.
- [25] D. Cheng, G. Meng, G. Cheng, and C. Pan, "SeNet: structured edge network for sea-land segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 2, pp. 247–251, 2017.
- [26] J. Yao, J. Qi, J. Zhang, H. Shao, J. Yang, and X. Li, "A real-time detection algorithm for kiwifruit defects based on YOLOv5," *Electronics*, vol. 10, no. 14, p. 1711, 2021.
- [27] Q. Zhu, "Pavement crack detection algorithm based on image processing analysis," in *2016 8th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, pp. 15–18, Hangzhou, China, 2016.
- [28] G. Wu, X. Sun, L. Zhou, H. Zhang, and J. Pu, "Research on crack detection algorithm of asphalt pavement," in *2015 IEEE International Conference on Information and Automation*, pp. 647–652, Lijiang, China, 2015.
- [29] W. Song, G. Jia, H. Zhu, D. Jia, and L. Gao, "Automated pavement crack damage detection using deep multiscale convolutional features," *Journal of Advanced Transportation*, vol. 2020, Article ID 6412562, 11 pages, 2020.
- [30] J. Li, T. Liu, X. Wang, and J. Yu, "Automated asphalt pavement damage rate detection based on optimized GA-CNN," *Automation in Construction*, vol. 136, article 104180, 2022.

- [31] K. Zhang, H. D. Cheng, and B. Zhang, "Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning," *Journal of Computing in Civil Engineering*, vol. 32, no. 2, 2018.
- [32] C.-l. Dongye and H. Liu, "A pavement disease detection method based on the improved mask R-CNN," in *2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT)*, pp. 619–623, Shenyang, China, 2020.
- [33] W. Tang, S. Huang, Q. Zhao, R. Li, and L. Huangfu, "An iteratively optimized patch label inference network for automatic pavement distress detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8652–8661, 2022.
- [34] L. Zhao, Y. Wu, X. Luo, and Y. Yuan, "Automatic defect detection of pavement diseases," *Remote Sensing*, vol. 14, no. 19, p. 4836, 2022.
- [35] Z. Mao, C. Zhao, Y. Zheng et al., "Research on detection method of pavement diseases based on unmanned aerial vehicle (UAV)," in *2020 International Conference on Image, Video Processing and Artificial Intelligence*, Shanghai, China, 2020.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.