

Research Article

Pavement Crack Detection Method Based on Deep Learning Models

Guo X. Hu ¹, Bao L. Hu,² Zhong Yang ³, Li Huang,⁴ and Ping Li¹

¹School of Software, Jiangxi Normal University, Nanchang 330027, China

²Foreign Language College, Jiangxi Normal University, Nanchang 330027, China

³College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China

⁴School of Education, Jiangxi Normal University, Nanchang 330027, China

Correspondence should be addressed to Zhong Yang; yz.nuaa@163.com

Received 25 February 2021; Accepted 23 April 2021; Published 17 May 2021

Academic Editor: Cunhua Pan

Copyright © 2021 Guo X. Hu 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.

Severe weather and long-term driving of vehicles lead to various cracks on asphalt pavement. If these cracks cannot be found and repaired in time, it will have a negative impact on the safe driving of vehicles. Traditional artificial detection has some problems, such as low efficiency and missing detection. The detection model based on machine learning needs artificial design of pavement crack characteristics. According to the pavement distress identification manual proposed by the Federal Highway Administration (FHWA), these categories have three different types of cracks, such as fatigue, longitudinal crack, and transverse cracks. In the face of many types of pavement cracks, it is difficult to design a general feature extraction model to extract pavement crack features, which leads to the poor effect of the automatic detection model based on machine learning. Object detection based on the deep learning model has achieved good results in many fields. As a result, those models have become possible for pavement crack detection. This paper discusses the latest YOLOv5 series detection model for pavement crack detection and is to find out an effective training and detection method. Firstly, the 3001 asphalt crack pavement images with the original size of 2976×3978 pixels are collected using a digital camera and are randomly divided into three types according to the severity levels of low, medium, and high. Then, for the dataset of crack pavement, YOLOv5 series models are used for training and testing. The experimental results show that the detection accuracy of the YOLOv5l model is the highest, reaching 88.1%, and the detection time of the YOLOv5s model is the shortest, only 11.1 ms for each image.

1. Introduction

Asphalt pavement is damaged by natural disasters such as long-term exposure to the sun, rain erosion, and natural weathering. And asphalt pavement is also damaged by human error such as rolling of vehicles, pavement materials, construction quality, and later maintenance level. All of these causes have different degrees of impact on pavement performance [1]. If these damaged roads cannot be found and maintained in time, the service life of the highway will be

shortened, the service level of the highway will be reduced, and even traffic accidents may be caused [2].

At present, pavement detection is mainly manual detection that has some disadvantages, such as taking a long time, requiring more manpower, obstructing the highway, intimidating the safety of inspectors, and interfering with the detection results by human factors. With the rapid development of a highway, it is difficult to meet the detection requirements of a large-scale highway and has completely failed to meet the needs of highway development [3].

In order to improve the service level of a highway and realize the automatic detection of damaged pavement, some researchers propose the research of automatic detection of pavement based on visual technology. In the past, people used digital image processing technology for crack detection [4]. This method mainly detects by comparing the gray value difference between the crack road and its background. However, the detection rate is low due to the complex background of pavement, various lighting, and the diversity of crack types [5, 6]. Therefore, this method can only be used as auxiliary detection.

In recent years, with the extensive use of machine learning, especially deep learning in the industrial field, it is possible to use deep learning models for automatic detection of crack pavement [7–9]. Cha et al. [10] proposed a crack pavement detection algorithm based on the convolutional neural network (CNN). The algorithm constructs an object classifier based on CNN and uses two sliding windows to scan the image to detect the crack area. However, due to the complex background interference of the photo image, the detection model cannot detect the internal features of the crack. To solve the problems existing in the above detection methods, Chen and Jahanshahi [11] improved the traditional convolutional neural network and proposed a crack detection algorithm combining the convolutional neural network with naive Bayesian (NB-CNN) data fusion. In this method, the convolutional neural network and naive Bayes are integrated, which greatly increases the complexity of the model and the number of parameters and makes the model more difficult to train. It can detect the target objects in different scale images because the FCN detection model [12] combines the image features of different volume layers. So, Yang et al. [13] proposed using the full convolution network (FCN) with encoding and decoding structures for pavement crack detection. FCN can collect features of different layers, where the shallow features can concentrate on spatial information and deep features can locate the objects, and finally fuse different features to achieve a damage prediction map. On this basis, many detection models are proposed for crack pavement detection [14–17]. All of the above methods can achieve good results for crack pavement with single background and simple topology, but not for complex pavement.

Crack pavement detection belongs to the field of vision-based automatic detection. In order to improve the accuracy and robustness of crack pavement automatic detection and realize the industrial automatic detection of crack pavement, we use the excellent detection models in other fields and transfer them to the detection of crack pavement. Zhou et al. [18] propose an effective method to automatically perform the recognition and location of concealed cracks based on 3D ground-penetrating radar (GPR) and deep learning models. This proposed method by paper [18] using YOLOv4 is feasible for the detection of concealed cracks. YOLOv5, officially released in June 2020, has become one of the most popular object detection models and is the state-of-the-art model in many fields [19]. To see the effectiveness of YOLOv4 in pavement detection, this paper discusses the feasibility and implementation method of the models in the detection of crack pavement based on YOLOv5 series models.

First of all, we classify the crack pavement by analyzing the damage situation and damage degree and collect the data according to this category. We collected more than 3000 samples and randomly divided them into the training set, test set, and verification set.

Then, we test the detection models of YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x in the YOLOv5 series and find that the recognition rate of all detection models is above 85%. However, as YOLOv5s is a lightweight detection model, its detection speed is far better than the other three. In terms of both recognition rate and efficiency, we think that YOLOv5s can meet the needs of practical engineering.

The paper is organized as follows. In Section 2, we introduce related works. Thereafter in Section 3, we demonstrate the detection model. Experiments are presented in Section 4. We concluded with a discussion in Section 5.

2. Related Works

2.1. Automatic Detection Based on Image Processing. The early methods [20–22] usually assumed that the crack pixel is generally darker than the surrounding and then used various threshold processing algorithms to extract the gray of the crack area. Because the threshold segmentation method is simple and fast, it has been widely used in early image segmentation. Early researchers proposed a variety of automatic detection algorithms based on threshold segmentation from different views. Li and Liu [21] proposed the threshold technique of adjacent difference histogram for automatic recognition of cracks in images. This method maximizes the difference between the two types of pixels (crack and noncrack) and achieves better experimental results than the traditional threshold method. Paper [22] detected crack pavement by processing the binary image obtained by the connected domain algorithm (directional segmentation expansion algorithm) and got good results.

Because the threshold segmentation method only considers the gray information of the image and ignores the spatial information of the image and is sensitive to noise, this algorithm is often combined with other methods to improve the segmentation accuracy. Gavilán et al. [23] found that the noncrack features in the image will present false positives. Therefore, in order to obtain the crack area, it is proposed to eliminate the false positive cracks in the noncrack image by calculating the average gray value of the pixels corresponding to the inner and outer contours of the linear object in the image. Li and Mao [24] firstly segmented the image into several complementary overlapping subimage regions. Then, the neighborhood difference histogram was used to segment and fuse the cracks in each subimage region. Lastly, the crack region is obtained in the images. This method is effective in a small range of complex fractures, but not in a large area of complex background.

The traditional threshold extraction method lacks the description of global information and is sensitive to noise. The detection effect depends on the selection of the threshold. However, in practical application, the road background

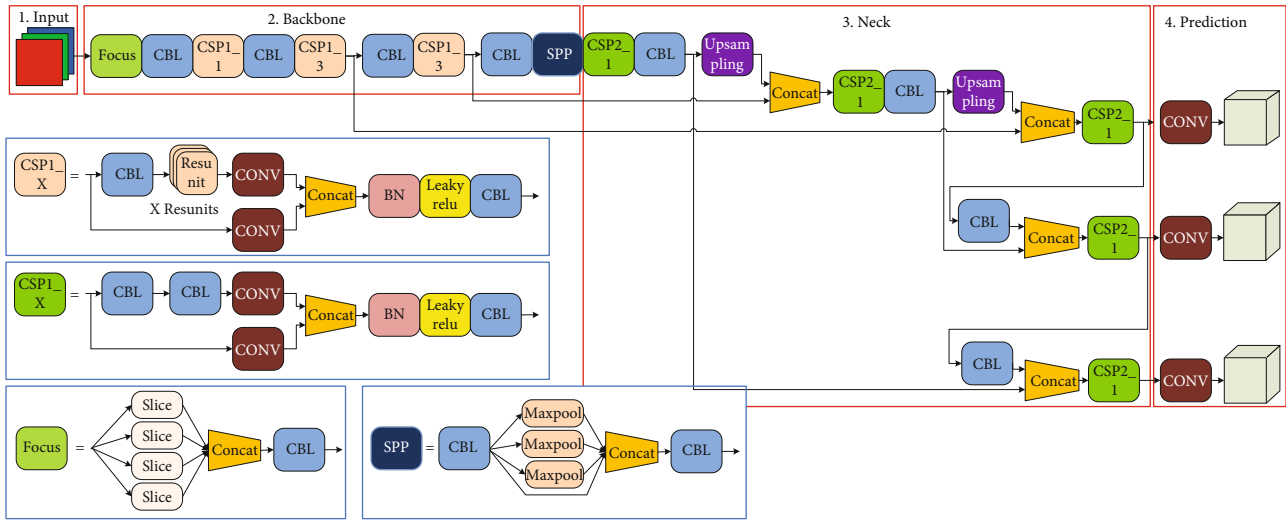


FIGURE 1: YOLOv5 structure.

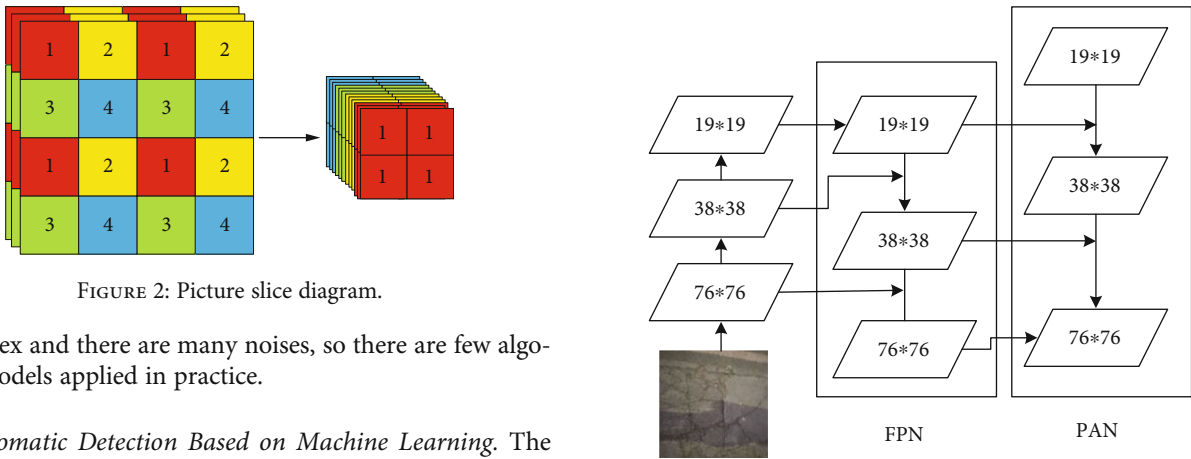


FIGURE 2: Picture slice diagram.

FIGURE 3: FPN&PAN structure.

is complex and there are many noises, so there are few algorithm models applied in practice.

2.2. Automatic Detection Based on Machine Learning. The machine learning method based on feature engineering has been successfully applied in many fields. The crack pavement has significant texture characteristics. Lots of researchers obtain the texture features of images from different views and use the classification technology of machine learning to detect the crack pavement automatically. Hu et al. [25] proposed a new automatic crack pavement detection method based on texture analysis and shape description. It is found that the texture features of the crack pavement are uneven. Therefore, that paper proposed to use six texture features and two translation invariant shape descriptors to describe the irregular texture and uneven illumination features of the image and then use the SVM classifier to classify the image into cracks and noncracks. Cord and Chambon [26] proposed a general crack detection method based on supervised learning, which can be applied to all types of defects in those images. The paper thought that the pavement with cracks presented strong texture information. The texture features of the crack pavement fluctuated in the local range and showed uniformity in the global range. Therefore, this paper uses linear and nonlinear filters to describe the image texture information. The image characteristics of different scales are analyzed by using morphological transformation, linear filtering, and nonlinear filtering. Finally, the AdaBoost classi-

fier will be used to learn and classify the above texture information, so as to obtain the pavement damage area. The experiment shows that this method can improve the performance of crack detection to a certain extent but cannot complete the fine extraction of cracks. Shi et al. [27] proposed crack forest, a method for asphalt crack pavement extraction using random structure forest. The detection framework was structured based on the representative and distinguishing overall channel features. And it combined the framework with random forest. This model can be trained in small datasets with full supervision. In addition, the proposed two feature histograms are used to express the cracks and eliminate the noise labeled as cracks. Although this method can overcome a small part of the road noise interference, it is still unable to meet the real complex background crack extraction. Hoang and Nguyen [28] used the support vector machine (SVM), artificial neural network (ANN), and random forest (RF) to train and verify the performance of the machine learning algorithm in the dataset. The feature set composed of projective integral and fracture properties can get the most ideal results.

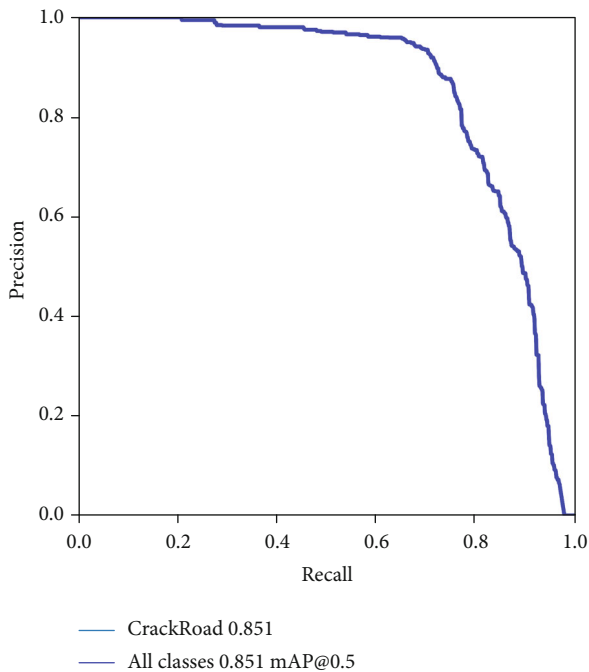
FIGURE 4: The P - R curve of the YOLOv5s model.

FIGURE 5: Crack pavement datasets.

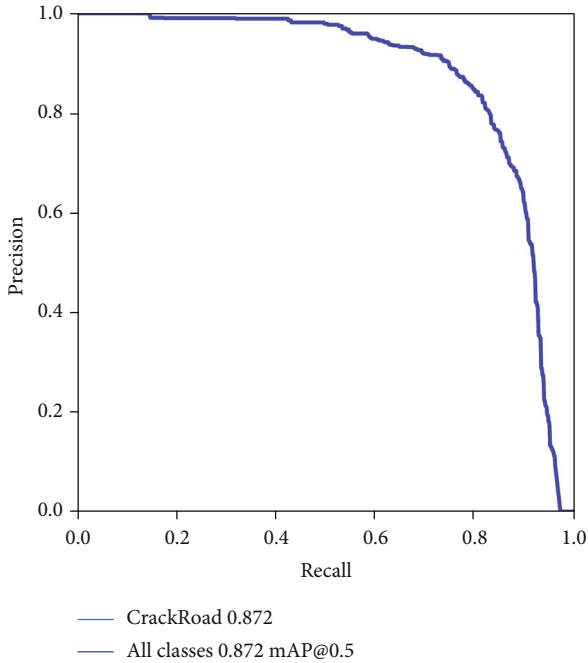
In a word, the detection model based on the machine learning algorithm greatly depends on the manually extracted image features including texture and color. And different feature extraction models need to be designed for

TABLE 1: Basic concept in the evaluation indicators.

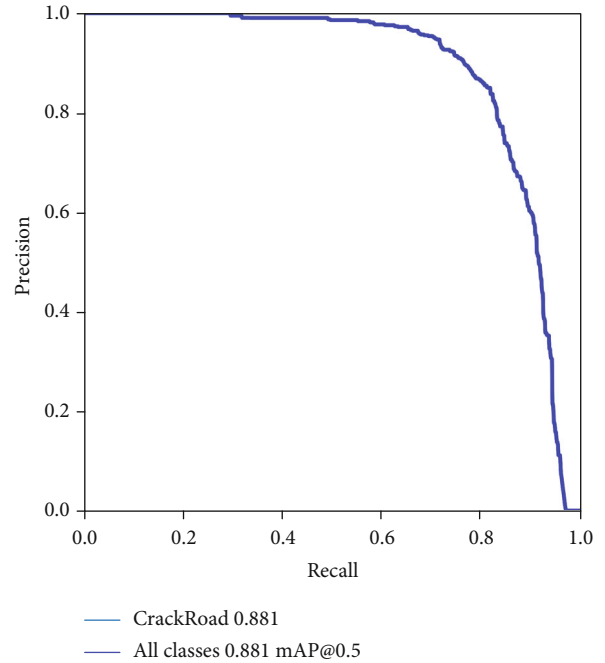
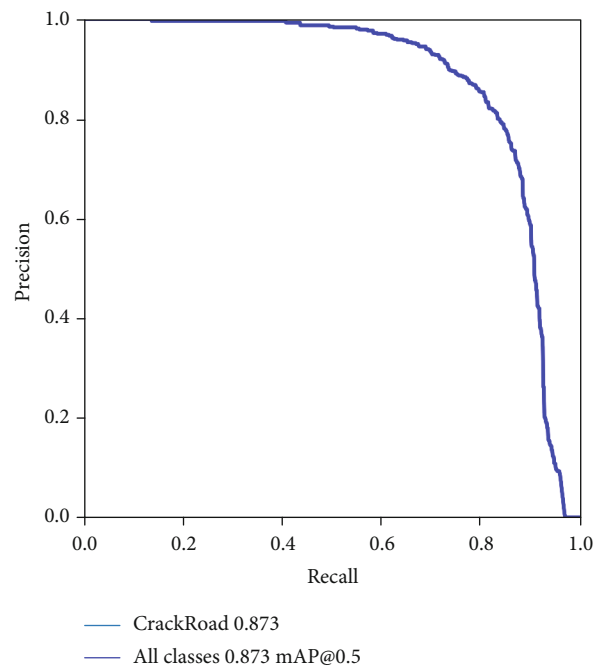
Symbol	Description
TP	TP refers to the number of cracks that are correctly classified as cracks
TN	TN refers to the number of backgrounds that are correctly classified as backgrounds
FP	FP refers to the number of backgrounds that are incorrectly identified as backgrounds
FN	FN refers to the number of cracks that are incorrectly identified as the background

different scenes, different lighting, and so on. Asphalt road is widely distributed, and the road surface includes all kinds of debris and other noises. So, it is difficult to extract effective features with a unified feature model for cracks in complex and changeable road environment, which leads to poor robustness of the detection model. All in all, the detection model based on machine learning can only be applied in a small scope, not universal.

2.3. Automatic Detection Based on Deep Learning. In recent years, with the development of deep learning technology, it is possible to detect crack pavement automatically based on the deep learning model. Combined with DCNN, by learning different crack samples, the performance of automatic crack detection has been significantly improved. Some people used the methods of object detection [29, 30] or image segmentation [31, 32] to complete the crack extraction. These methods cannot complete the pixel-level detection of cracks and also

FIGURE 6: The P - R curve of the YOLOv5m model.

cannot accurately determine the damage category and severity in the subsequent measurement and evaluation. Zhang et al. [33] proposed a pavement automatic detection system named crack net based on the convolution neural network. The system is aimed at the pixel-level extraction of cracks and realized the automatic detection of 3D asphalt crack pavement. Unlike conventional CNN, crack net did not have any pooling layer to reduce the output of the previous layer. Crack net ensured the accuracy of crack extraction by using the constant image width and height technology in all network layers. Compared with the traditional crack detection method based on machine learning, the extraction accuracy of this method is obviously better than the traditional method. Inspired by crack net, Fei et al. [34] proposed an efficient deep network named crack net-v based on crack net, which was used for pixel-level crack automatic detection of asphalt pavement 3D image. Compared with the original crack net, crack net-v had a deeper structure and fewer parameters, which improves the calculation accuracy and efficiency. Crack net-v used the same space size for all layers, so that supervised learning could be carried out at the pixel level. The efficiency of crack net-v further revealed the advantages of deep learning technology in pixel-level pavement crack automatic detection. Zou et al. [35] proposed an end-to-end trainable deep convolution neural network for automatic crack detection, named deep crack, by learning the advanced features of crack representation. The multiscale deep convolution features learned from different convolution layers are fused together to form a linear structure. The image features obtained by this method had more detailed representation characteristics in large-scale feature maps and more comprehensive representation characteristics in small-scale feature maps. The deep crack network which is constructed on the encoder-decoder architecture of SEG net

FIGURE 7: The P - R curve of the YOLOv5l model.FIGURE 8: The P - R curve of the YOLOv5x model.

fused the convolution features generated in the encoder network and decoder network with the same scale. The deep crack could complete the pixel-level crack extraction.

In a word, the detection effect of the detection model based on deep learning is much better than that of the machine learning model based on feature engineering. Combined with the current best object detection model, and applying it to the detection of crack pavement, it will greatly improve the efficiency of pavement detection.

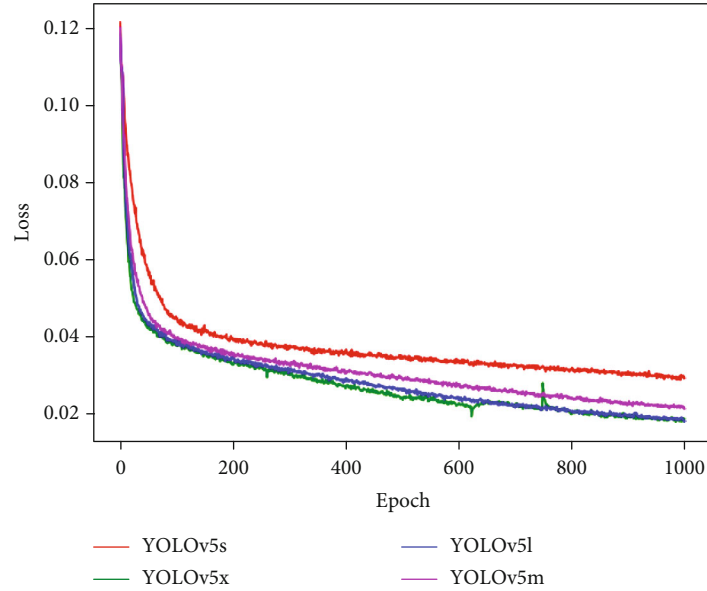


FIGURE 9: Loss figure for YOLOv5 models in training.

3. Model Introduction

At present, object detection models based on deep learning can be divided into two schools.

- (1) Two-stage model: generating candidate regions and classifying candidate regions by CNN (RCNN Series [36])
- (2) One-stage model: categorizing objects and locating them in one step (YOLO Series)

Following RCNN [36], Fast RCNN [37], and Faster RCNN [38], YOLO is another framework proposed by Ross Girshick for the problem of object detection speed. It has been updated to the fifth edition.

YOLOv5 consists of four parts: input, backbone, neck, and prediction. The details are shown in Figure 1 [39]. Compared with the well-known YOLOv3 [40], YOLOv5 has made the following improvements in the above four parts:

- (1) In the input module, mosaic data enhancement and adaptive anchor frame calculation are added
- (2) Focus and CSP structure are added in the backbone module
- (3) FPN and PAN structure is added in the neck module
- (4) GIOU_Loss is used in the prediction module

There are four versions of YOLOv5, including YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The original author thinks that the network structure of the above four versions of the model is from small to large, the corresponding detection accuracy is from low to high, but the detection speed is from fast to slow. That is to say, YOLOv5s has the smallest network, the least speed, and the lowest AP accuracy. If the

detected objects are mainly large objects, the network can quickly detect the detected objects. YOLOv5s cannot detect small objects if there are a lot of small cracks in the crack pavement.

In order to find out which model has the best combination of detection accuracy and efficiency, this paper uses the four models in YOLOv5 for road detection.

3.1. Input Module

3.1.1. Mosaic Data Enhancement. Yun et al. [41] proposed a new data enhancement method, CutMix, to enhance the diversity of samples by mixing two images and clipping them. The mosaic data enhancement method proposed by YOLOv5 mixes and stitches four images based on CutMix to produce a new image. Learning the new mosaic image is equivalent to learning four pictures at the same time, which improves the learning efficiency. Meanwhile, because the BN calculation is also four pictures at the same time, the minibatch in the training can be set to a smaller value. In a word, YOLOv5 can use one GPU to speed up the training without changing the detection accuracy.

The implementation steps of mosaic data enhancement are as follows:

- (1) Four images are randomly read from the dataset at a time
- (2) Flip left and right, zoom in and out, and change the brightness, saturation, and hue gamut of the four images
- (3) The transformed image is combined with pictures and bounding boxes. The first picture is placed on the top left, the second picture is placed on the bottom left, the third picture is placed on the bottom right, and the fourth picture is placed on the top right

TABLE 2: Performance comparison table of 4 detection models.

Model	Precision	Recall	mAP0.5	Time	Model parameters	Net layers
YOLOv5s	0.891	0.512	0.851	11.1 ms	7,063,542	283
YOLOv5m	0.908	0.556	0.872	24.2 ms	21,056,406	391
YOLOv5l	0.908	0.569	0.881	40 ms	46,631,350	499
YOLOv5x	0.9	0.582	0.873	67.7 ms	87,244,374	607

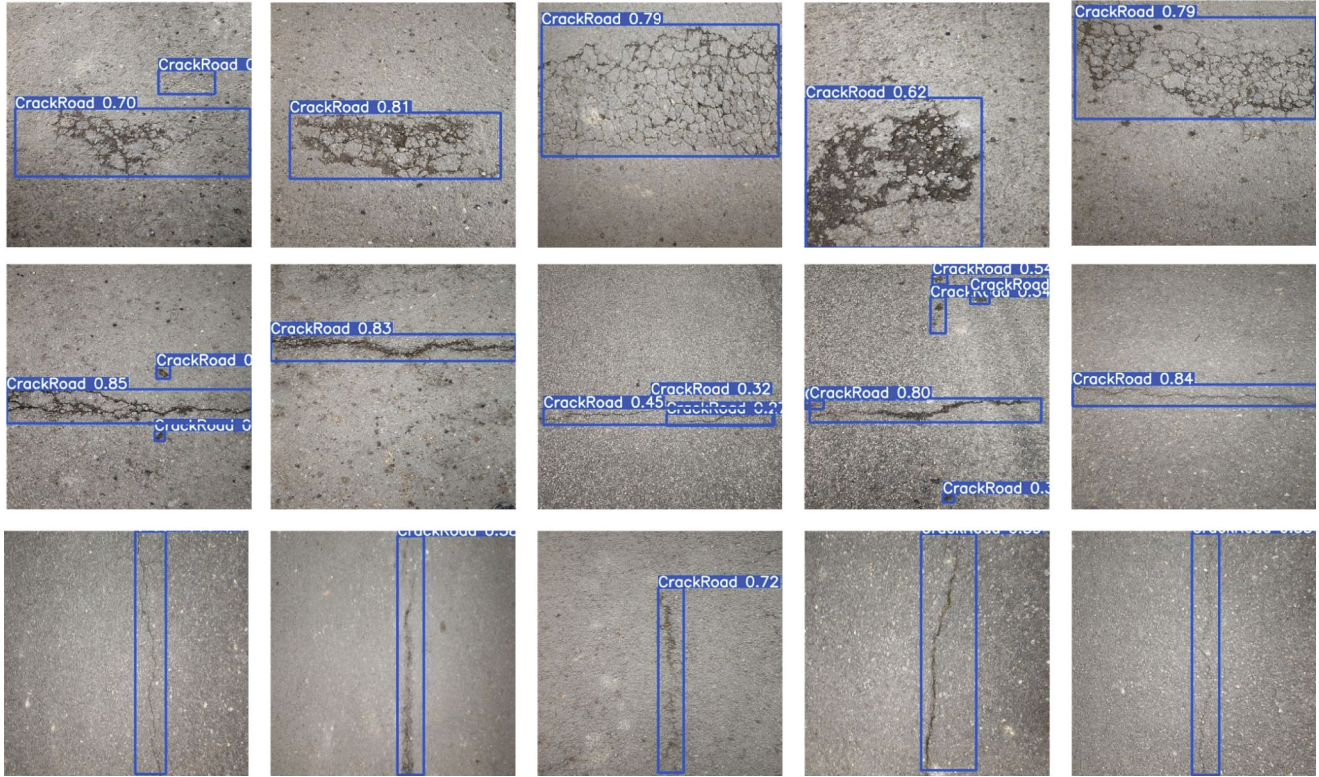


FIGURE 10: The detection chart of the YOLOv5s model.

in four directions. They are spliced into a new picture, which contains a bounding box and other contents

- (4) When the bounding box (or the picture itself) of a picture exceeds the dividing line between two pictures, we need to remove the part that exceeds the dividing line or the part of the picture for edge processing

3.1.2. Adaptive Anchor Calculation. YOLOv5 sets the initial anchor for each image. In the process of network training, the prediction bounding box is output based on the initial anchor. And then, the bounding box is compared with the real ground truth. The gap between those is calculated. Lastly, the network parameters are updated backward and iterated.

3.1.3. Adaptive Image Scaling. Since different images are different in length and width, the common way is to scale the original image to a standard size and then send it to the detection network for the traditional object detection algorithms.

The YOLOv5 model believes that many pictures have different aspect ratios. So, the sizes of black edges at both ends are different after the picture is scaled and filled. If there are more black edges filled, there will be information redundancy, which will affect the training speed.

Therefore, the author improves the traditional fixed-size scaling and adopts the adaptive method of adding the least black edges for the YOLOv5 model. The specific methods are as follows:

- (1) Calculate the scale. If the ratio of length to width of the original image is not fixed, there may be two scaling ratios; the smaller one is chosen
- (2) The length and width of the scaled image are calculated, respectively
- (3) Calculate the number of the black edge filled. YOLOv5 is filled with gray pixels instead of black pixels, that is (114,114,114). At the same time, YOLOv5 does not use the way of reducing the black edge in training but uses the traditional way of filling,

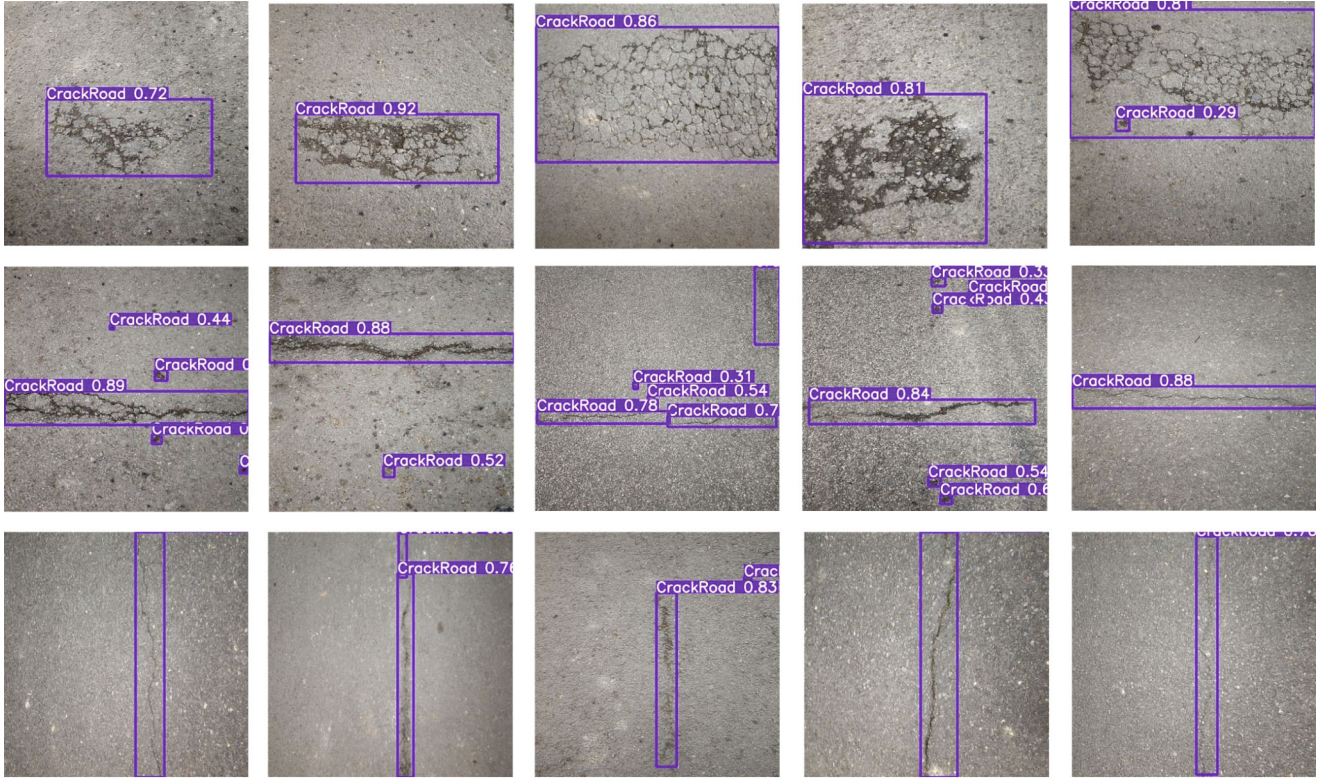


FIGURE 11: The detection chart of the YOLOv5m model.

that is, reducing image to the size of 416×416 . In the test phase, the black edge reduction method is adopted to improve the speed of object detection and reasoning when the model reasoning is used

3.2. Backbone Module

3.2.1. Focus Structure. Focus is mainly used for image slicing, as shown in Figure 2. We slice the $4 \times 4 \times 3$ image into a $2 \times 2 \times 12$ feature map.

For YOLOv5, the $608 \times 608 \times 3$ image is sliced into a $304 \times 304 \times 12$ feature map. Then, the feature map is convoluted by 32 convolution cores once and finally becomes a $304 \times 304 \times 32$ feature map.

3.2.2. CSP Structure. The author thinks that the problem of excessive reasoning is caused by the repetition of gradient information in network optimization. Cross stage partial network (CSPnet) [42] integrates the change of gradient into the feature graph from beginning to end, which can reduce the amount of calculation and ensure accuracy.

Two CSP structures are designed in YOLOv5. Taking the YOLOv5s model as an example, the CSP1_X structure is applied to the backbone network; another kind of CSP2_X structure is used in the neck. After adding CSP, the training speed of the model is improved. The main advantages of the CSP structure are as follows:

- (1) Enhance the learning ability of CNN and make it lightweight while maintaining accuracy

- (2) Reduce computing bottlenecks
- (3) Reduce memory cost

3.3. Neck Module. For object detection, in order to better extract fusion features, some layers are usually inserted in the backbone and output layer, which is called the neck. YOLOv5 adopts the FPN+PAN structure in the neck module.

The high-level convolution layer has abstract description meaning for a large object. However, due to the lack of pixel information, object features cannot be described for small objects. FPN [43] combines convolution features of different layers for images, which can not only satisfy the abstract description of large objects but also ensure the feature details of small objects. YOLOv5 uses FPN in the neck layer to ensure that objects are not missed.

In order to improve the detection accuracy of small objects, FPN uses the low-level features of the image. Although low-level features help to detect objects, it is more and more difficult to accurate the position of the object using the features combining the underlying structure feature with the top-level feature. Path aggregation network (PANet) [44] shorts the information path between low-level and top-level features through the bottom-up path enhancement and the whole feature level that is enhanced by using accurate low-level positioning signals. The model connects the feature grid with all feature layers, so that the useful information in each feature layer can be directly propagated to the following subnetwork.

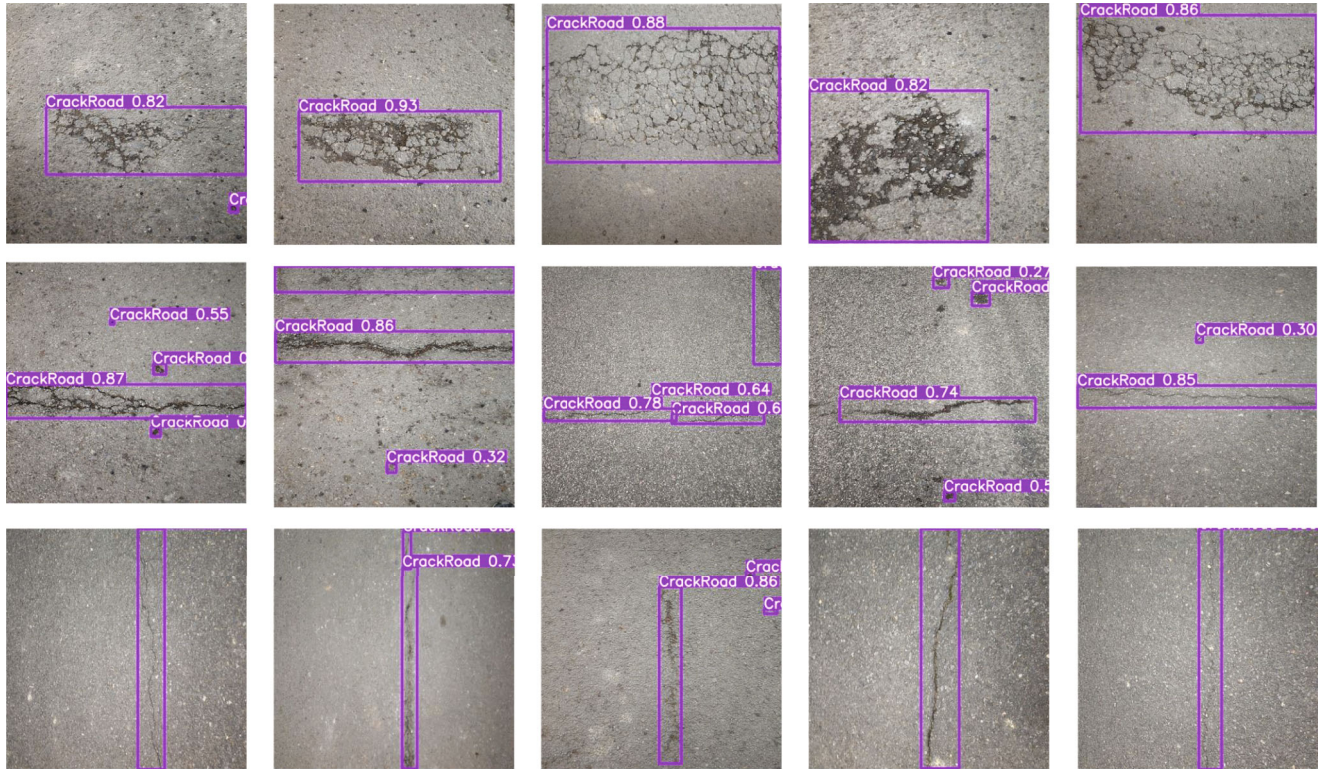


FIGURE 12: The detection chart of the YOLOv5 model.

YOLOv5 uses FPN and PANet in the neck to ensure detection accuracy and positioning accuracy. The details are shown in Figure 3.

3.4. Prediction Module. In YOLOv5, Leaky ReLU is used in the middle/hidden layer of the model. The final detection layer uses sigmoid as the activation function. In order to increase the detection of occluded overlapped objects, YOLOv5 uses DIOU-nm. The default optimization method of the model is the gradient descent method. The loss function of the model consists of three parts: object score loss, category probability loss, and bounding box regression loss. Among them, logits loss is used for object score loss, cross-entropy loss is used for category probability loss, and GIOU [45] loss is used for bounding box regression loss.

4. Experimental Analysis

4.1. Dataset Acquisition. The Federal Highway Administration (FHWA) divides pavement cracks into three categories, including fatigue crack, longitudinal crack, and transverse crack. When we take pictures of crack pavement, we pay attention to various types of sample collection. At the same time, for each type of crack degree is different, we will take samples of crack severity that is divided into high, medium, and low. In practical engineering, there are lots of interference in the pavement that is often accompanied by sign lines, fallen leaves, oil, garbage, and light. These interferences have a great impact on detection accuracy. In order to ensure that the training data samples cover all kinds of conditions as far

as possible, we also consider the above interference factors in the process of data acquisition.

Under the condition of combining all the above crack types and interference factors, we took 3001 photos of crack pavement with a Nikon camera. The picture is 2976×3968 pixels. Considering the existence of various types of crack and interference in practical application, we did not conduct separate model training for each type of crack data in the training process. The data was randomly divided into 1920 training datasets, 480 verification datasets, and 601 test datasets. The specific sample data is shown in Figure 4.

In Figure 5, the first line is fatigue crack type data, the second line is transverse crack type data, and the third line is longitudinal crack type data. For the fatigue crack, due to the large damage area, there is ambiguity in many places. So, this type of data is difficult for machine detection. For transverse and longitudinal cracks, these types of data have the characteristics of thin and long, and many damaged areas are difficult to distinguish by human eyes. Therefore, the detection model is another challenge for the detection of a small crack.

4.2. Experimental Environment. All the algorithms in this paper were programmed by Python3. The models were implemented by using the classical deep learning-based pytorch framework and trained on the Ubuntu experimental platform, Intel® Xeon® Gold 5218 CPU@ 2.30 GHz processor, 260 GB RAM, and NVIDIA Tesla P100-PCIE 16 GB GPU. Pytorch was born to carry out quick experimentation, Nvidia driver version is 4450.51.05, CUDA version is 11.0, and it is also able to quickly convert its ideas

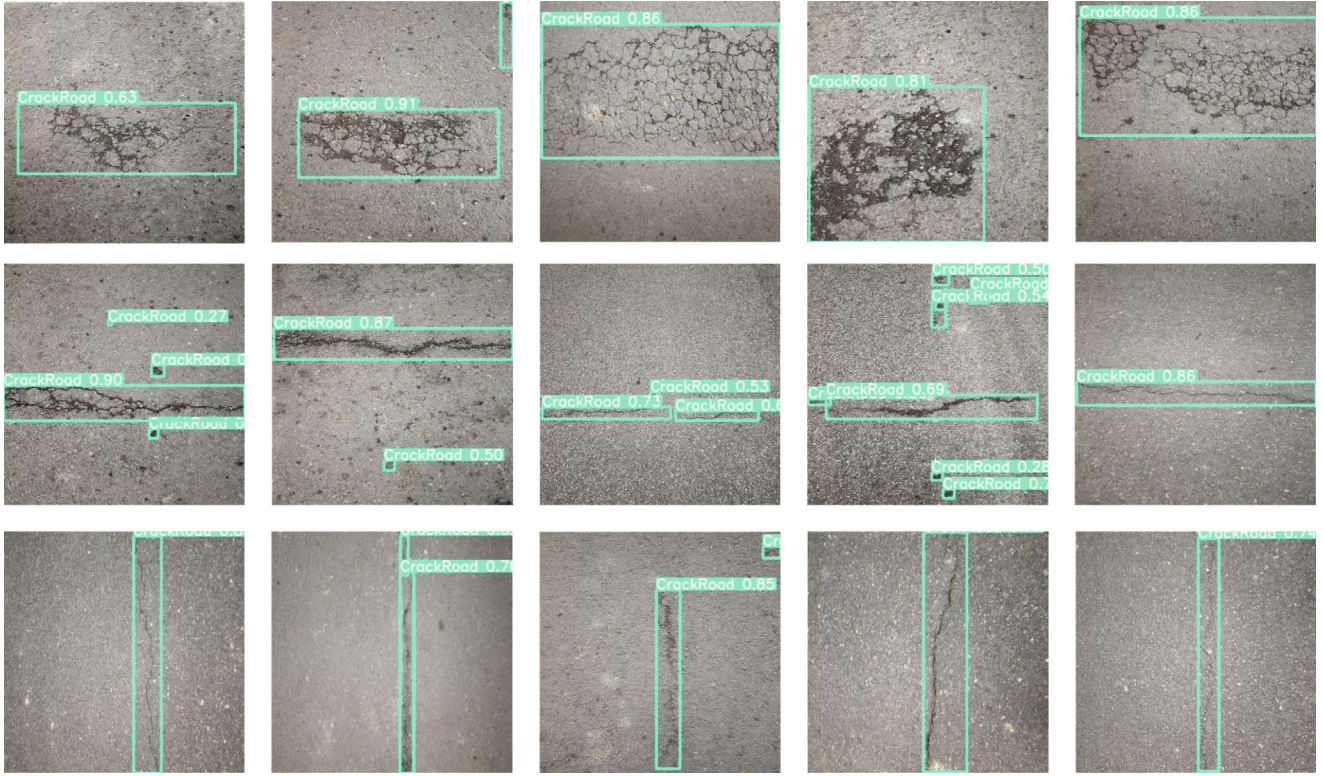


FIGURE 13: The detection chart of the YOLOv5l model.

into results. For this reason, pytorch can quickly and efficiently compare the visualization results of different crack detection algorithms.

4.3. Model Testing. In order to test the performance of the YOLOv5 model in crack pavement detection, we test it on the dataset described in Section 4.1. In this paper, accuracy, recall, and F1 score are used to quantify different crack classification algorithms. These three indicators were calculated with TP (true positive), TN (true negative), FP (false positive), and FN (false negative), respectively. The specific definitions are shown in Table 1.

The formula for precision, recall, and F1 score is as follows:

$$\begin{aligned}
 \text{Precision} &= \frac{TP}{TP + FP}, \\
 \text{Recall} &= \frac{TP}{TP + FN}, \\
 \text{F1-score} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.
 \end{aligned} \tag{1}$$

Since there are many types of crack pavement and the degree of different types of crack is different, it is easy to produce imbalance between different types of datasets under the premise of certain datasets. Both precision and recall in the *P-R* curve consider the detection rate of positive samples, which reduces the error between evaluation criteria due to data imbalance. The detection effect of the model is described by the *P-R* curve (see Figures 4 and 6–8 for details).

In the process of training, four pretraining weights (YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x) provided by YOLOv5 are used for training. In the test process, ensure that the superparameter settings and training weights are the same, such as the learning rate and momentum parameters. The training losses of the four models are shown in Figure 9. It can be seen from the figure that in the process of the same number of iterations, the model with deep layers has better training effect and faster convergence.

The network layers of v5s, v5m, v5l, and v5x, which are provided by YOLOv5, increase from small to large, and the network parameters also increase from small to large. In order to obtain the influence of different size models on the detection effect, we use four models to compare and analyze the crack datasets collected in this paper. The details are shown in Table 2.

According to Table 2, it can be seen that the model with the highest detection accuracy is YOLOv5l; mAP reaches 0.881. However, the detection accuracy between all the four models is not very different, and the maximum difference is within 0.03. The smallest network model is the YOLOv5s model, which has only 283 layers. Because the model level of YOLOv5s is the smallest, the parameters are the least among all models, only 7,063,542. Due to the different sizes of the model, in the detection process, the small model takes less time. On the contrary, the large model takes more time. Therefore, for the four models mentioned above, the least time-consuming model is YOLOv5s. For each 640 * 640 image, the detection practice only takes 11.1 ms, while YOLOv5l takes 40 ms. The most time-consuming is YOLOv5x, which reaches 67.7 ms. For the above four models,

in order to test the detection effect of the models on the actual project, we use the above four models to test in the datasets shown in Figure 5, and the effect is shown in Figures 10–13.

5. Conclusions

Pavement crack is very difficult to detect because of many categories and larger influence of the surrounding environment. The existing pavement crack detection models fail to detect the cracks because the features of objects are difficult to be extracted by many convolution and pooling operations. Combined with one of the state-of-art object detection models YOLOv5, we discuss the possibility of transferring this model to crack pavement detection. The experimental results show that the detection accuracy of the YOLOv5 series models is above 85%. The shortest time-consuming YOLOv5s model only needs 11.1 ms to detect 640×640 pixels. Therefore, if we pay attention to the detection rate in the actual project, we can choose to use YOLOv5l. If we need to consider both the detection rate and detection efficiency, we can choose the YOLOv5s model. However, no matter which model is used, it can only be used as an aid.

Automatic crack pavement detection is one of the difficult research contents in the field of object detection. The detailed features of the crack pavement have semantic relevance with its surrounding road surface. And the subsequent semantic segmentation technology can be combined to further improve the model detection accuracy. Meanwhile, YOLO is an object detection framework based on anchors. Although YOLO has done a lot of optimization in reducing the amount of computation compared with an anchor-free model, YOLO still needs a lot of computing resources in the training process. The detection rate of the traditional detection model based on anchors is higher than that of the anchor-free model. How to improve the detection rate and reduce the resources is a problem that many engineering applications need to consider. Centernet2 [46] adopts an anchor-free framework, and the detection rate of this model is higher than that of all existing models based on anchors. The proposal of centernet2 provides direction for our next work. In order to meet the requirements of a lightweight model in practical engineering, we will study the detection based on anchor-free in the future.

Data Availability

The datasets we used in this paper enable us to shoot with a Nikon camera. We took 3001 photos of crack pavement. The picture is 2976×3968 pixels. Considering the existence of various types of crack and interference in practical application, we did not conduct separate model training for each type of crack data in the training process. The data was randomly divided into 1920 training datasets, 480 verification datasets, and 601 test datasets.

Conflicts of Interest

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the Science and Technology Project of Jiangxi Provincial Department of Education under Grant nos. GJJ200305 and GJJ191689 and the Natural Science Foundation of Jiangxi Province under Grant no. 20202BABL202016.

References

- [1] V. P. Tran, T. S. Tran, H. J. Lee, K. D. Kim, J. Baek, and T. T. Nguyen, "One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects," *Journal of Civil Structural Health Monitoring*, vol. 11, no. 1, pp. 205–222, 2020.
- [2] X. Z. Xiang, Y. Q. Zhang, and A. El Saddik, "Pavement crack detection network based on pyramid structure and attention mechanism," *IET Image Processing*, vol. 14, no. 8, pp. 1580–1586, 2020.
- [3] H. N. Nguyen, T. Y. Kam, and P. Y. Cheng, "An automatic approach for accurate edge detection of concrete crack utilizing 2D geometric features of crack," *Journal of Signal Processing Systems*, vol. 77, no. 3, pp. 221–240, 2014.
- [4] Z. Y. Yang, "The design of glass crack detection system based on image preprocessing technology," in *2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*, pp. 39–42, Chongqing, China, 2014.
- [5] Y. Yi, J. Wang, W. Zhou, Y. Fang, J. Kong, and Y. Lu, "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognition*, vol. 92, pp. 258–273, 2019.
- [6] Y. Yi, Y. Chen, J. Wang, G. Lei, J. Dai, and H. Zhang, "Joint feature representation and classification via adaptive graph semi-supervised nonnegative matrix factorization," *Signal Processing-Image Communication*, vol. 89, article 115984, 2020.
- [7] D. Ma, H. Y. Fang, B. H. Xue, F. M. Wang, M. A. Msekh, and C. L. Chan, "Intelligent Detection Model Based on a Fully Convolutional Neural Network for Pavement Cracks," *CMES-Computer Modeling in Engineering & Sciences*, vol. 123, no. 3, pp. 1267–1291, 2021.
- [8] Y. Wang, K. Song, J. Liu, H. Dong, Y. Yan, and P. Jiang, "RENet: rectangular convolution pyramid and edge enhancement network for salient object detection of pavement cracks," *Measurement*, vol. 170, article 108698, 2021.
- [9] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Non-negative matrix factorization with locality constrained adaptive graph," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 427–441, 2020.
- [10] Y. J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.
- [11] F. C. Chen and M. R. Jahanshahi, "NB-CNN: deep learning-based crack detection using convolutional neural network and naïve Bayes data fusion," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, 2018.
- [12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440, Boston, MA, USA, 2015.

- [13] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1090–1109, 2018.
- [14] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, and H. Ling, "Feature pyramid and hierarchical boosting network for pavement crack detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1525–1535, 2020.
- [15] H. Chen, H. Lin, and M. Yao, "Improving the efficiency of encoder-decoder architecture for pixel-level crack detection," *IEEE Access*, vol. 7, pp. 186657–186670, 2019.
- [16] Z. Fan, C. Li, Y. Chen et al., "Ensemble of deep convolutional neural networks for automatic pavement crack detection and measurement," *Coatings*, vol. 10, no. 2, p. 152, 2020.
- [17] X. Z. Xiang, Y. Q. Zhang, and A. E. Saddik, "Pavement crack detection network based on pyramid structure and attention mechanism," *IET Image Processing*, vol. 14, no. 8, pp. 1580–1586, 2019.
- [18] S. Zhou, Y. Bi, X. Wei et al., "Automated detection and classification of spilled loads on freeways based on improved YOLO network," *Machine Vision and Applications*, vol. 32, no. 2, 2021.
- [19] E. U. Rahman, Y. Zhang, S. Ahmad, H. I. Ahmad, and S. Jobaer, "Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs," *Sensors*, vol. 2, no. 3, 2021.
- [20] Q. Li, Q. Zou, D. Zhang, and Q. Mao, "FoSA: F* seed-growing approach for crack-line detection from pavement images," *Image and Vision Computing*, vol. 29, no. 12, pp. 861–872, 2011.
- [21] Q. Q. Li and X. L. Liu, "Novel approach to pavement image segmentation based on neighboring difference histogram method," in *2008 Congress on Image and Signal Processing*, pp. 792–796, Sanya, China, 2008.
- [22] F. Liu, G. Xu, Y. Yang, X. Niu, and Y. Pan, "Novel approach to pavement cracking automatic detection based on segment extending," in *2008 International Symposium on Knowledge Acquisition and Modeling*, pp. 610–614, Wuhan, China, 2008.
- [23] M. Gavilán, D. Balcones, O. Marcos et al., "Adaptive road crack detection system by pavement classification," *Sensors*, vol. 11, no. 10, pp. 9628–9657, 2011.
- [24] Q. Q. Li and Q. Z. Mao, "Land-borne pavement rapid test and measurement," *Journal of Transport Information and Safety*, vol. 27, no. 1, pp. 7–10, 2009.
- [25] Y. Hu, C. X. Zhao, and H. N. Wang, "Automatic pavement crack detection using texture and shape descriptors," *IETE Technical Review*, vol. 27, no. 5, pp. 398–405, 2010.
- [26] A. Cord and S. Chambon, "Automatic road defect detection by textural pattern recognition based on AdaBoost," *Computer-Aided Civil and Infrastructure Engineering*, vol. 27, no. 4, pp. 244–259, 2012.
- [27] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.
- [28] N. D. Hoang and Q. L. Nguyen, "A novel method for asphalt pavement crack classification based on image processing and machine learning," *Engineering with Computers*, vol. 35, no. 2, pp. 487–498, 2019.
- [29] H. F. Li, J. P. Zong, J. J. Nie, Z. L. Wu, and H. Y. Han, "Pavement crack detection algorithm based on densely connected and deeply supervised network," *IEEE Access*, vol. 9, pp. 11835–11842, 2021.
- [30] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection using deep neural networks with images captured through a smartphone," 2018, <http://arxiv.org/abs/1801.09454>.
- [31] X. Wang and Z. Hu, "Grid-based pavement crack analysis using deep learning," in *2017 4th International Conference on Transportation Information and Safety (ICTIS)*, pp. 917–924, Banff, AB, Canada, 2017.
- [32] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, no. 10, p. 3452, 2018.
- [33] A. Zhang, K. C. P. Wang, B. Li et al., "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 10, pp. 805–819, 2017.
- [34] Y. Fei, K. C. P. Wang, A. Zhang et al., "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 273–284, 2020.
- [35] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "Deep-Crack: learning hierarchical convolutional features for crack detection," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2019.
- [36] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, 2014.
- [37] R. Girshick, "Fast R-CNN," in *The 15th IEEE International Conference on Computer Vision (ICCV'15)*, pp. 1440–1448, Santiago, Chile, 2015.
- [38] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [39] D. B. Jiang, "Easy to understand Yolo series: Yolov5 core basic knowledge complete explanation," <https://blog.csdn.net/nan355655600/article/details/107852353>.
- [40] R. Joseph and F. Ali, "YOLOv3: an incremental improvement," 2018, <http://arxiv.org/abs/1804.02767>.
- [41] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, and J. Choe, "Cut-Mix: regularization strategy to train strong classifiers with localizable features," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 6022–6031, Seoul, Korea, 2019.
- [42] C.-Y. Wang, H.-Y. M. Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: a new backbone that can enhance learning capability of CNN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1571–1580, Seattle, WA, USA, 2020.
- [43] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936–944, Honolulu, HI, USA, 2017.
- [44] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8759–8768, Salt Lake City, UT, USA, 2018.

- [45] H. Rezatofighi, N. Tsoi, J. Y. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 658–666, Long Beach, CA, USA, 2019.
- [46] X. Y. Zhou, V. Koltun, and P. Krähenbühl, "Probabilistic two-stage detection," 2021, <http://arxiv.org/abs/2013.07461>.