

## Research Article

# Pavement Crack Detection and Segmentation Method Based on Improved Deep Learning Fusion Model

Xiaoran Feng,<sup>1</sup> Liyang Xiao ,<sup>1</sup> Wei Li,<sup>1</sup> Lili Pei,<sup>1</sup> Zhaoyun Sun,<sup>1</sup> Zhidan Ma,<sup>1</sup> Hao Shen,<sup>1</sup> and Huyan Ju<sup>2</sup>

<sup>1</sup>School of Information Engineering, Chang'an University, Xi'an 710064, Shaanxi, China

<sup>2</sup>Centre for Pavement and Transportation Technology (CPATT), Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

Correspondence should be addressed to Liyang Xiao; [liyang\\_xiao@163.com](mailto:liyang_xiao@163.com)

Received 30 June 2020; Revised 11 October 2020; Accepted 24 November 2020; Published 10 December 2020

Academic Editor: Mariko Nakano-Miyatake

Copyright © 2020 Xiaoran Feng 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 damage is the main factor affecting road performance. Pavement cracking, a common type of road damage, is a key challenge in road maintenance. In order to achieve an accurate crack classification, segmentation, and geometric parameter calculation, this paper proposes a method based on a deep convolutional neural network fusion model for pavement crack identification, which combines the advantages of the multitarget single-shot multibox detector (SSD) convolutional neural network model and the U-Net model. First, the crack classification and detection model is applied to classify the cracks and obtain the detection confidence. Next, the crack segmentation network is applied to accurately segment the pavement cracks. By improving the feature extraction structure and optimizing the hyperparameters of the model, pavement crack classification and segmentation accuracy were improved. Finally, the length and width (for linear cracks) and the area (for alligator cracks) are calculated according to the segmentation results. Test results show that the recognition accuracy of the pavement crack identification method for transverse, longitudinal, and alligator cracks is 86.8%, 87.6%, and 85.5%, respectively. It is demonstrated that the proposed method can provide the category information for pavement cracks as well as the accurate positioning and geometric parameter information, which can be used directly for evaluating the pavement condition.

## 1. Introduction

Pavement distress is the main factor affecting road performance. Timely and accurate detection of pavement damages is a crucial step in pavement maintenance. Cracks are the initial manifestation of various types of pavement diseases. Pavement cracks will not only affect pavement appearance and driving comfort but also can easily expand to cause pavement structural damage and shorten the overall service performance and life of the pavement [1, 2]. Therefore, early crack detection and timely maintenance of the cracked pavement can reduce the economic cost of pavement repairing and ensure the safety of vehicles and drivers transiting on the pavement.

Early pavement detection and maintenance mainly rely on manual detection, which is not only time-consuming and

laborious but also has low detection accuracy and some associated risks [3–5]. Scholars from across the world, taking advantage of recent science and technology developments, have carried out a series of extensive and in-depth research to accurately and efficiently extract crack information from images [6–8]. In 2014, Wang et al. [9] proposed a pavement crack extraction method based on the valley bottom boundary; it uses a series of image processing algorithms to obtain the crack detection results. In 2015, Liang et al. [10] proposed a pavement crack connection algorithm based on the Prim minimum spanning tree, which obtains the crack structure by filling the fracture.

The disadvantages of these traditional crack detection methods are obvious. Each method is designed for a specific database or scenario, but the crack detector will fail if the dataset or scenario changes. With the rapid development of

artificial intelligence, convolutional neural networks (CNNs) have been widely applied in the field of image recognition [11–14]. In recent years, deep learning methods have been increasingly applied to pavement crack detection and segmentation [15, 16]. Combining deep learning methods with pavement crack detection techniques considerably improves the efficiency and accuracy of pavement crack detection [17].

In 2016, Zhang et al. [18] proposed a crack detection method based on deep learning. They trained a deep CNN based on supervised learning, proving the feasibility of combining deep learning with pavement crack recognition. In 2017, Zhao et al. [19] proposed a pavement crack detection method based on a CNN using images of different scales and taken at different angles for training, achieving the detection of cracks of various shapes. However, owing to road surface interference and noise, the detection accuracy of this system peaked at 82.5%. In 2017, Markus et al. developed the open dataset GAPs for the training of deep neural network and evaluated the pavement damage detection technology for the first time, which is of great significance [20, 21]. In 2018, Nhat-Duc et al. [22] established an intelligent method for the automatic recognition of pavement crack morphology; this study constructs a machine learning model for pavement crack classification that included multiple support vector machines and an artificial swarm optimization algorithm. Using feature analysis, a set of features is extracted from the image projection integral, which can significantly improve the prediction performance. However, the algorithm is complex and programming it becomes very difficult. In 2020, Zhaoyun Sun et al. [23] proposed a method to detect pavement expansion cracks with the improved Faster R-CNN, which can achieve accurate expansion crack location detection through the optimization model. The aforementioned studies only detect and classify pavement cracks and their location but cannot quantify certain crack characteristics, such as crack width and area. On the other hand, there are also many studies on crack segmentation. In 2018, Zhang and Wang [24] proposed CrackNet, which is an efficient architecture based on CNN to predict the class of each image pixel, but its network structure is related to input image size, which prevents the generalization of the method. In the same year, Sen Wang et al. [25] proposed to use the full convolutional networks (FCNs) to detect cracks and built the Crack-FCN model taking into account the shortcomings of the FCN model in the crack segmentation experiment and obtained a complete crack image. However, the highest accuracy obtained by their method is only 67.95%; thus, segmentation performance needs to be improved. In 2019, Piao Weng et al. [26] proposed a pavement crack segmentation method based on the VGG-U-Net model. It solves the problem of fracture in the crack segmentation result in complex background, but its training time is slightly longer and its efficiency is low. In 2020, Zhun Fan et al. [27] proposed an encoder-decoder architecture based on hierarchical feature learning and dilated convolution (U-HDN) detects cracks in an end-to-end manner. The U-HDN method can extract and fuse different context sizes and different levels of feature mapping, so it

has high performance. In the same year, Zhun Fan et al. [28] proposed an ensemble of convolutional neural network based on probability fusion for automatic detection and measurement of pavement cracks, and the predicted crack morphology is measured by skeleton extraction algorithm. In summary, these previous studies only use the segmentation method, which cannot achieve accurate crack classification and location determination.

Although all of these methods were able to recognize crack diseases from pavement images to a certain extent, some problems remain: (1) Most algorithms still extract crack information using feature extraction, which is relatively complex and programmatically difficult [6–8]. (2) Although deep learning-based pavement crack recognition algorithms already exist, they still do not eliminate the specificity requirement of “one device, one algorithm.” If the source image is taken by different devices or on different road sections, using a single dataset leads to inaccurate results. Therefore, their adaptability is poor [18, 19]. (3) Complex environmental factors affect the stability and accuracy of crack identification algorithms [29]. (4) Although there is already a pavement crack recognition method based on CNNs, the model has a single function and most of the cracks are classified and roughly positioned through the detection box and cannot be directly used for evaluating road conditions [23–26].

Given the abovementioned problems in pavement crack identification, this paper proposes a method based on a deep convolutional neural network fusion model for pavement crack identification, which is applicable in many crack detection cases (including detector vehicle and smartphone). By training on a learning image data having a variety of sources and sizes, the method can effectively identify cracks, and recognition accuracy can be guaranteed. At the same time, a detected crack can be segmented, and the segmented binary image can be used to calculate the geometric parameters of the crack. Therefore, the proposed model is of great significance for intelligent pavement detection and it can also achieve detection and segmentation simultaneously, thereby significantly improving model efficiency.

## 2. Methodology

In this paper, a crack identification method based on a deep CNN fusion model is proposed. First, the image dataset is established, and the image noise in the dataset is filtered out to increase the contrast between road cracks and background. Next, the processed images are provided as input into an improved single shot multibox detector (SSD) crack detection model and an improved U-Net crack segmentation model for training. Then, the binary image of a crack obtained by the segmentation model is used to calculate the geometric parameters of the crack. By integrating the advantages of the two models, this pavement crack identification method can effectively overcome the single-model limitations of inaccurate positioning and imperfect information. The overall process flowchart is shown in Figure 1. The details of each step are discussed in Section 2.1.

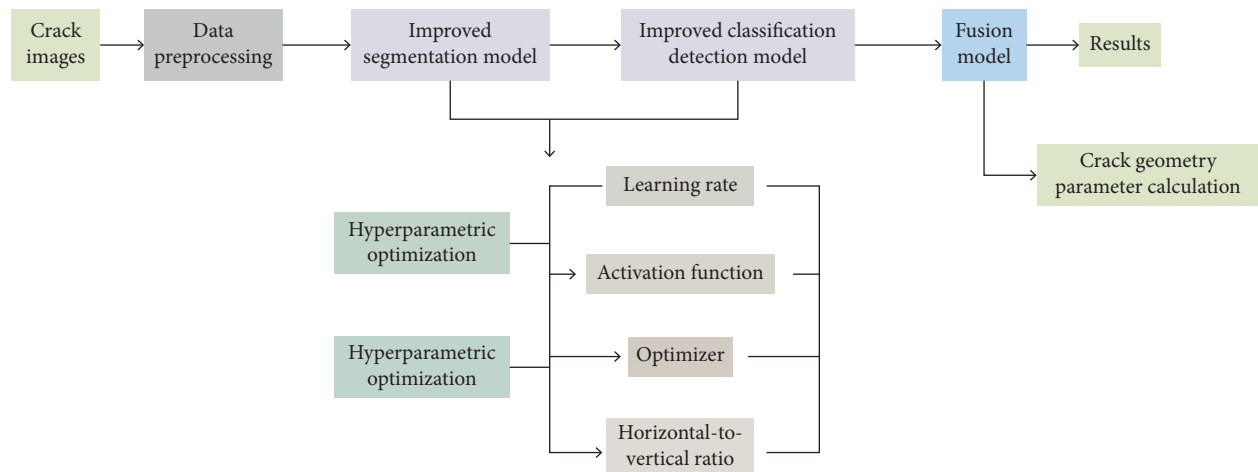


FIGURE 1: Flow chart of the proposed methodology.

**2.1. Image Collection and Preprocessing.** In order to be applicable to crack detection in a variety of scenarios, the proposed method uses a detector vehicle (Figure 2(a)) and a smartphone (Figure 2(b)) to collect crack images. The pixel of the image captured by the detector vehicle is  $1024 \times 960$ , and the pixel of the image captured by the smartphone is  $2560 \times 1024$ .

The pavement crack images are preprocessed before network training to reduce the noise in the images and improve the prediction accuracy. Preprocessing consists mainly of augmenting and denoising the pavement crack images.

First, the number of images needs to be increased. As it is difficult to distinguish the effects of rotation by using an actual crack image, a black-and-white double-arrow picture is used here to exemplify how to increase the number of images (Figure 3). By horizontal reflection, vertical reflection, and clockwise rotation of the image by  $45^\circ$ ,  $90^\circ$ , and  $180^\circ$ , the training image dataset can be expanded eightfold.

The images, as taken by the camera, are seriously affected by discrete pulse noise and zero-mean Gaussian noise. Therefore, median and bilateral filters are used to denoise the road images. Then, contrast enhancement is performed to increase the difference between crack information and road background, which improves the quality of the sample images. As seen in Figure 4, the processed crack information is more prominent than in the original image.

**2.2. Classification and Detection Using Improved SSD Model.** The SSD [30] network model used in this study was proposed in 2016; it combines the characteristics of the You Only Look Once (YOLO) model [31], which provides fast speed, and the Faster R-CNN (Region-CNN) model [32], which provides accurate recognition. We use Tensorflow and Keras framework to implement the SSD network.

The characteristic feature of the SSD network model is its capability of performing multiscale feature map detections. The model adds some convolutional feature layers at the end

of the feature extraction network, and the feature maps extracted from these convolutional layers have the feature of decreasing in size. Image prediction is carried out by means of fusion of the multiscale detection results. In Figure 5, the feature fusion of conv4\_3 and fc7, two convolution layers, is given as an illustrative example. conv4\_3\_norm\_priorbox sets each point to generate four preselected boxes. The sample dataset used in this study contains three categories; thus, the value of the conv4\_3\_norm\_mbox\_conf channel is 12 ( $\text{num\_priorbox} \times \text{num\_class} = 4 \times 3 = 12$ ). Each preselected box returns four position transitions, so the value of the conv4\_3\_norm\_mbox\_loc channel is 16 ( $4 \times 4 = 16$ ). fc7 generates six preselected boxes per point, in addition to the others. Finally, mbox\_conf and mbox\_loc are merged, mbox\_conf behind reshapes, and then softmax classification is performed. As can be seen from the above example, each convolution feature layer will produce corresponding prediction results, and finally, the prediction results on different scales will be fused to obtain the best fracture prediction results.

Increasing the number of network layers can improve the accuracy of the network in identifying pavement cracks. Therefore, in this study, the feature extraction network structure Visual Geometry Group 16 (VGG16) in the SSD network model was replaced with a deep residual network to improve the pavement crack identification accuracy. The deep residual network [33, 34] solves this problem by fitting a residual map instead of the original map and by adding multiple connections between layers.

In order to convert the basic VGG16 network of the SSD model into a deep residual network, it is necessary to connect the convolution layer feature map output size in the residual network with the matching VGG16 convolution layer. Table 1 shows the names and outputs of convolution layers matching the outputs of the feature maps of the two feature extraction networks.

The structure of the improved crack classification and detection network is shown in Figure 6. The convolutional layer of different feature map sizes contains two kinds of  $3 \times 3$  convolution kernels. One is used for position



FIGURE 2: Pavement crack image acquisition and collection by (a) detector vehicle and (b) smartphone.

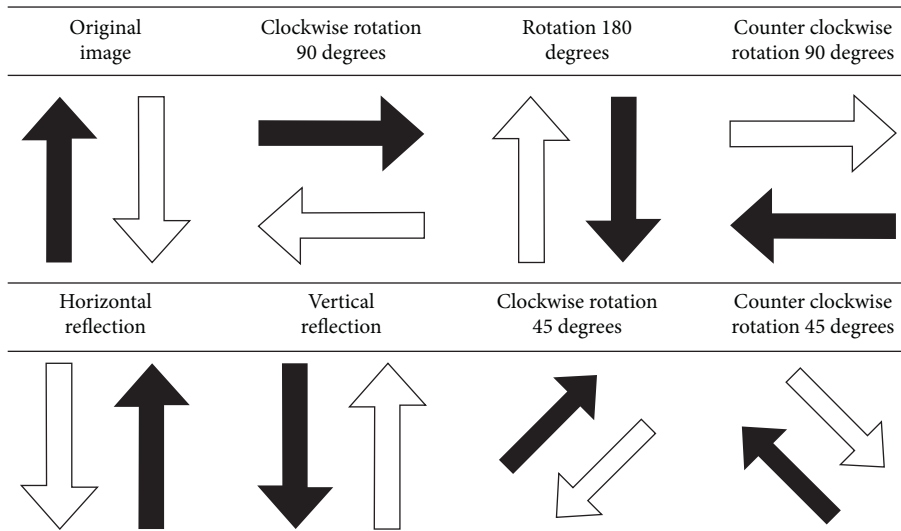


FIGURE 3: Example of how to increase the number of images.

regression of the prediction box, and the other is used for crack classification.

In order to improve the robustness of the model, all predicted crack boxes of different sizes at all positions of the feature map are combined to form a diversified prediction set. When the pavement crack images were input for training, with settings  $steps\_per\_epoch=100$  and  $final\_epoch=100$  (100-step iterations), 10,000 training iterations were conducted. The test set was used for testing after each

step. If  $val\_loss$  was reduced, the weight file was saved to continue training until training was complete.

Figure 7 shows the loss function curves obtained by training on the pavement crack dataset in the improved model. The red line represents the training set loss curve, and the blue line represents the val set loss curve. When  $final\_epoch=90$ , the model is effectively stable.

The reason behind improving the model is to improve crack detection accuracy and to generate a better detection

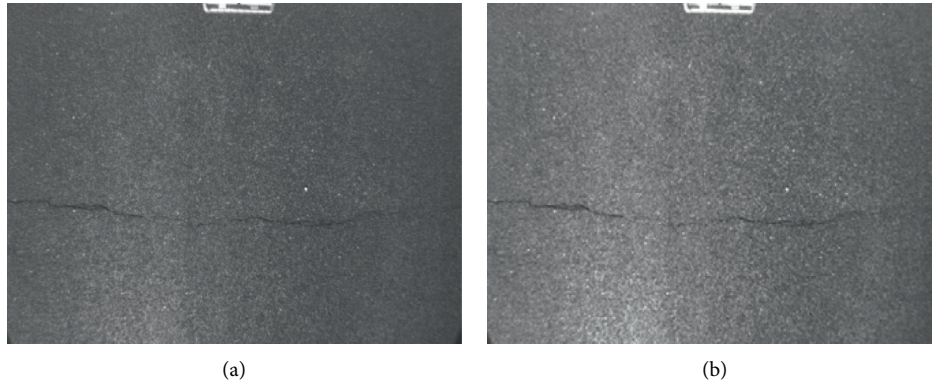


FIGURE 4: Effect of filtering: (a) original image; (b) image after filtering.

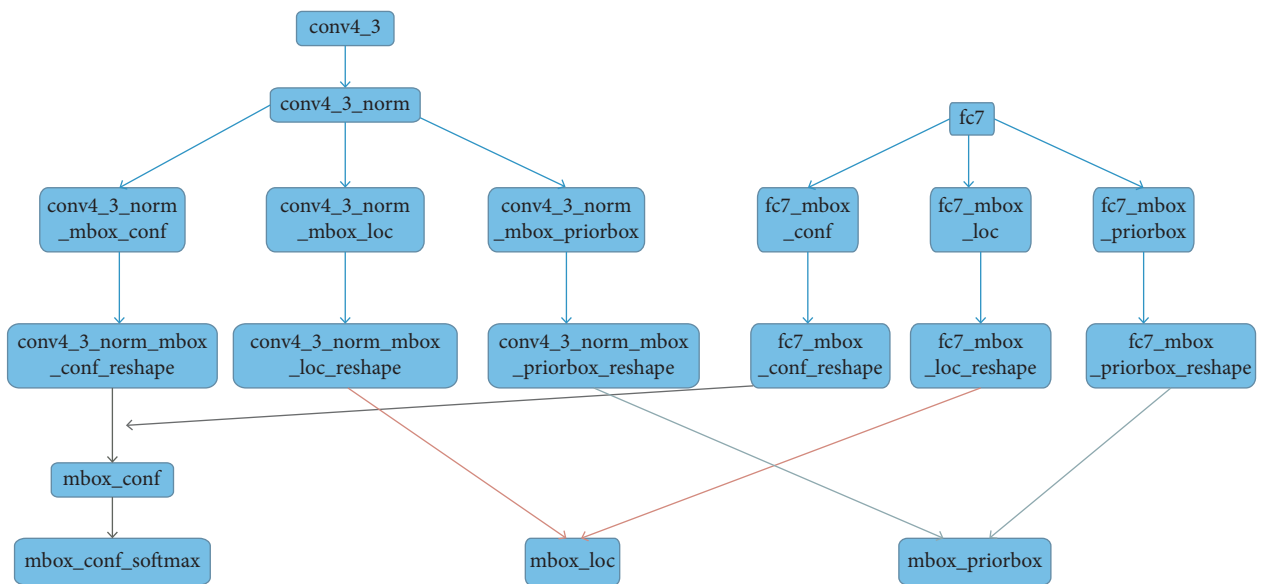


FIGURE 5: Single-shot multibox detector (SSD) multiscale feature detection example.

TABLE 1: SSD convolution feature layers and feature map outputs.

VGG16 layer (type)	Resnet layer (type)	Output shape
conv4_3	conv2d_24	(64, 64, 512)
fc7	conv2d_43	(32, 32, 1024)
conv6_2	conv2d_53	(16, 16, 512)
conv7_2	conv2d_63	(8, 8, 256)
conv8_2	conv2d_73	(4, 4, 256)
conv9_2	conv2d_83	(2, 2, 256)
conv10_2	add_28	(1, 1, 256)

box to surround the cracks. In order to ascertain whether the model prediction results were improved after improving the crack detection model, model results before and after the improvement were compared using the test set (Table 2).

By constantly adjusting the hyperparameters in the model and comparing the accuracy of crack classification and detection, the precision–recall (PR) curves for the recognition of three types of pavement cracks (transversal, longitudinal, and alligator) were constructed (Figure 8).

Three types of cracks were randomly selected for testing, and the test results are shown in Figure 9. By comparing the prediction results of the model before and after the improvement for the same category and the same picture, it can be seen that the improved crack classification and detection model provides a higher degree of confidence for identifying the crack category in the pavement image, and the prediction results are more accurate, which demonstrates the effectiveness of model improvement and optimization.

By replacing the feature extraction structure of the original SSD network with the deep residual network, the network accuracy and recall rate in predicting pavement cracks were substantially improved. This analysis of the experimental results shows that the proposed method achieves good results in the classification and detection of cracks. From the prediction effect, however, a classification by the pavement crack detection method based on the single SSD crack location model is incomplete, is not conducive to subsequent crack geometry parameter computation steps, and will produce larger calculation errors. Thus, as the

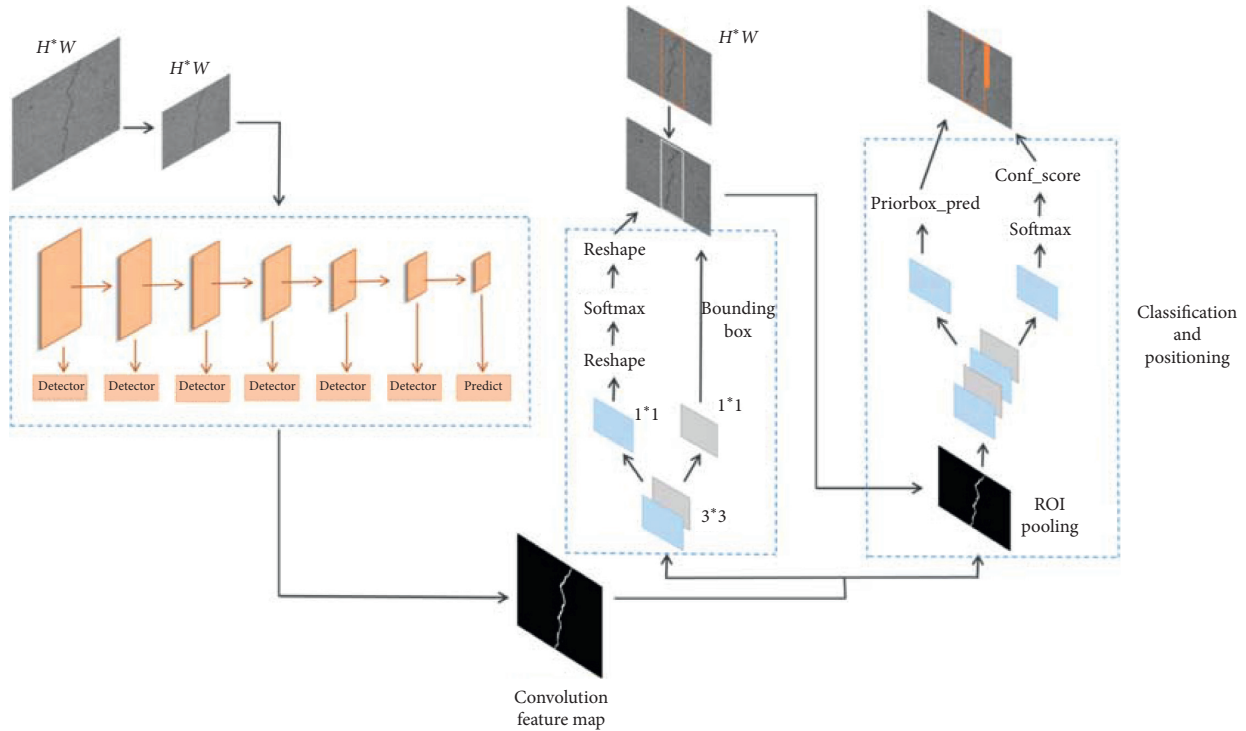


FIGURE 6: Improved crack classification detection network.

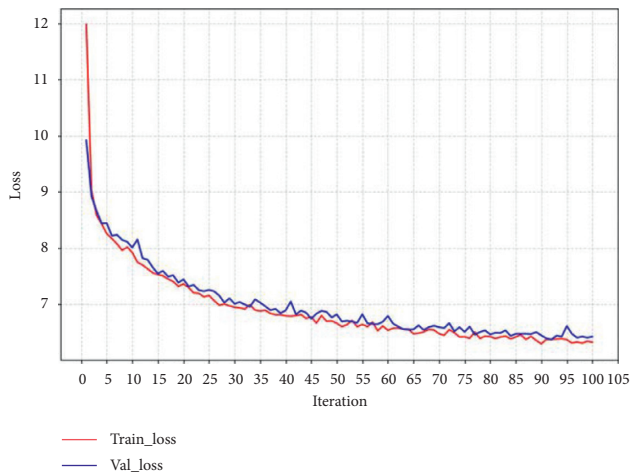


FIGURE 7: Training and val loss curves for the crack classification detection model.

TABLE 2: Comparison of crack classification detection network prediction results.

Crack type	Precision		Recall		Average precision (AP)	
	Before	After	Before	After	Before	After
Transverse	0.773	0.805	0.780	0.811	0.776	0.810
Longitudinal	0.771	0.827	0.782	0.836	0.779	0.834
Alligator	0.694	0.811	0.721	0.824	0.718	0.825
Mean average precision (mAP)					0.758	0.823

practical application value is still lacking at this point, this study adopted the fusion segmentation model approach to address this problem.

**2.3. Crack Segmentation Using Improved U-Net Model.** In the proposed method, the U-Net [35] model is used as the pavement crack segmentation model. We use Tensorflow and Keras framework to implement U-Net network. As shown in Figure 10, the U-Net structure is divided into two main parts, the first part for feature extraction and the second part for the upsampling operation; this structure is also known as the encoder-and-decoder structure. U-Net uses splicing for feature fusion and splicing features on channel dimensions to form richer features, thus facilitating the network's learning of crack features.

The structure of the U-Net network is simple, and the original U-Net network has crack segmentation accuracy problems. Therefore, the feature extraction network of the U-Net crack segmentation model was also replaced with a deep residual network to fully extract crack features and ensure crack segmentation accuracy. The specific improvement steps are similar to those of the SSD model. As shown in Table 3, the two basic network feature graph outputs match the network layers. After adjusting the corresponding layers of the feature extraction network, it is still necessary to adjust the network parameters through continuous training to optimize the crack segmentation effect. The improved crack segmentation model is shown in Figure 11.

In the training of U-Net crack segmentation model, the ReLU function is used as the activation function and the input data samples are regularized many times.

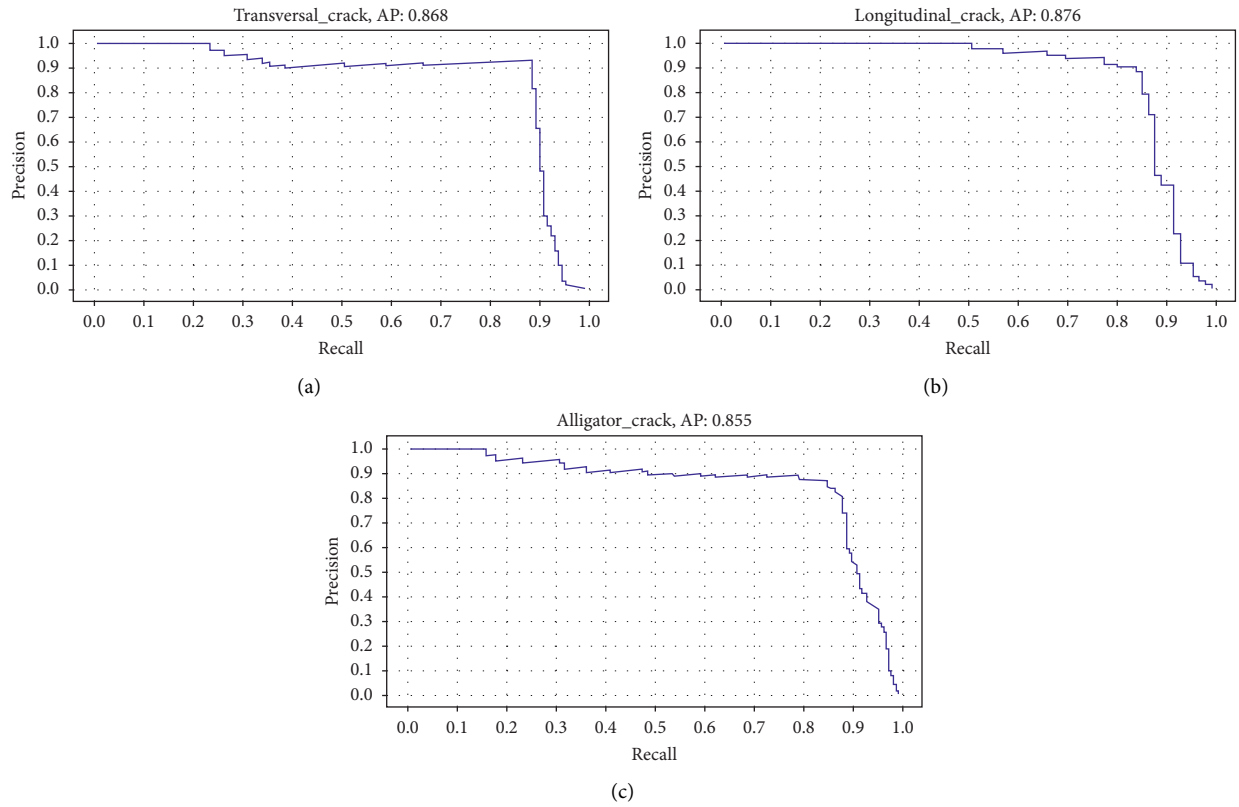


FIGURE 8: Classification network precision–recall (PR) curves for three types of pavement crack recognition: (a) transverse crack (average accuracy: 0.868); (b) longitudinal crack (average accuracy: 0.876); and (c) alligator crack (average accuracy: 0.855).

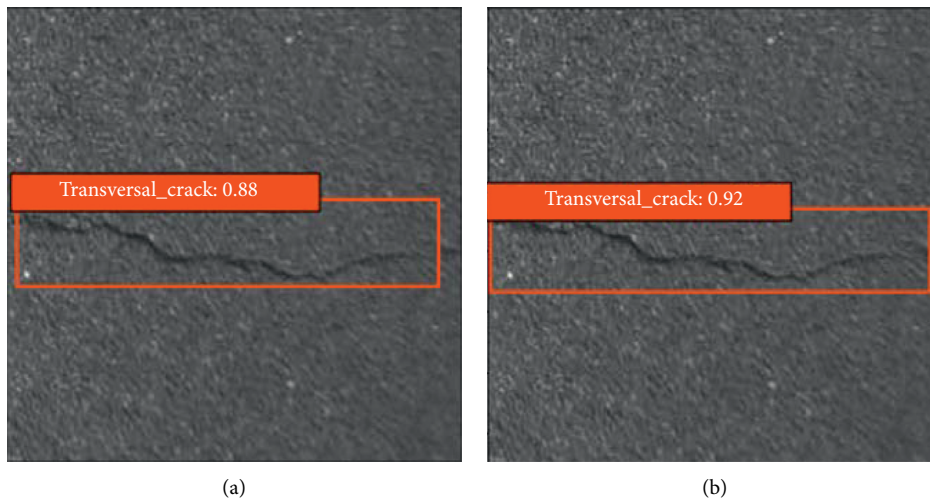


FIGURE 9: Continued.

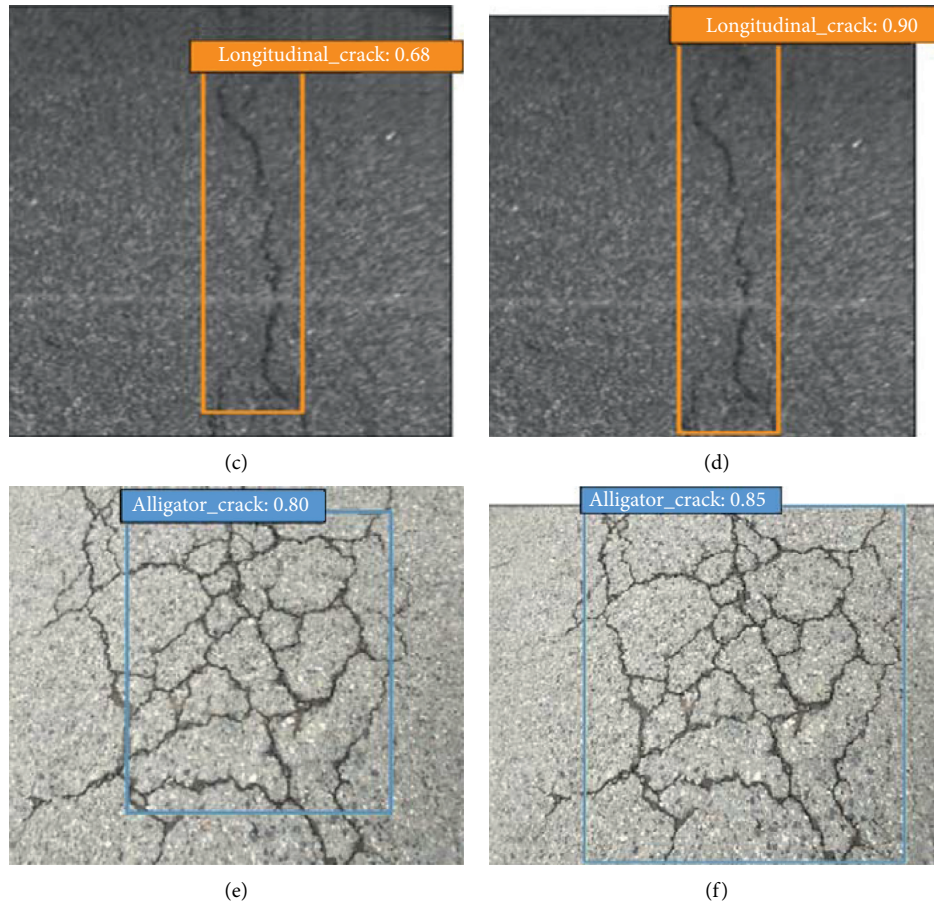


FIGURE 9: Comparison of crack classification and detection. Top: transverse crack detection before (a) and after (b) improvement. Center: longitudinal crack detection before (c) and after (d) improvement. Bottom: alligator crack detection before (e) and after (f) improvement.

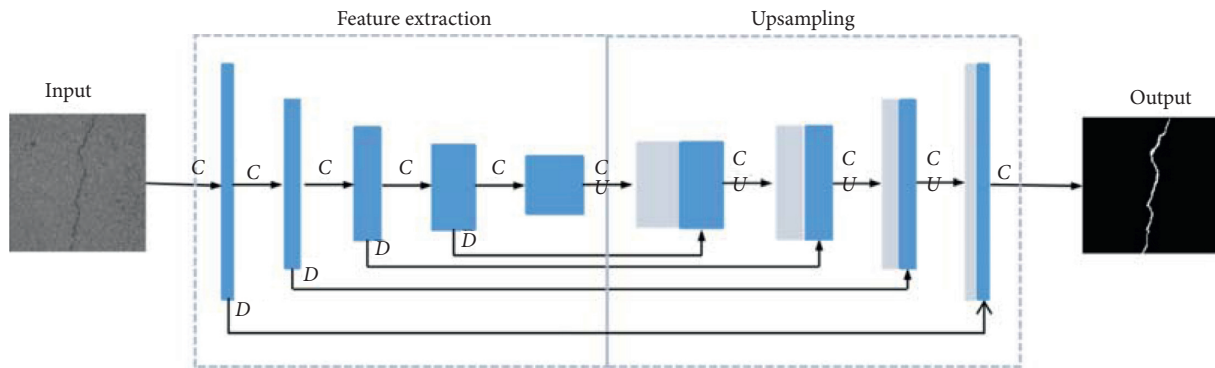


FIGURE 10: Structure of U-Net.

TABLE 3: U-Net convolution feature layers and feature map outputs.

VGG16 layer (type)	Resnet layer (type)	Output shape
block1_conv2	leaky_re_lu_1	(512, 512, 64)
block2_conv2	leaky_re_lu_4	(256, 256, 128)
block3_conv3	leaky_re_lu_9	(128, 128, 256)
block4_conv3	leaky_re_lu_26	(64, 64, 512)
dropout_1	leaky_re_lu_43	(32, 32, 1024)

Regularization adjusts the output value of each convolutional network layer to the same distribution, thereby avoiding a deviation or change in the distribution of feature vectors caused by network deepening. The segmentation model uses the upsampling method. That is, the feature map with the new size is obtained by the convolution inversion operation, and the feature map with the size corresponding to the convolution layer is added as the upsampling result.



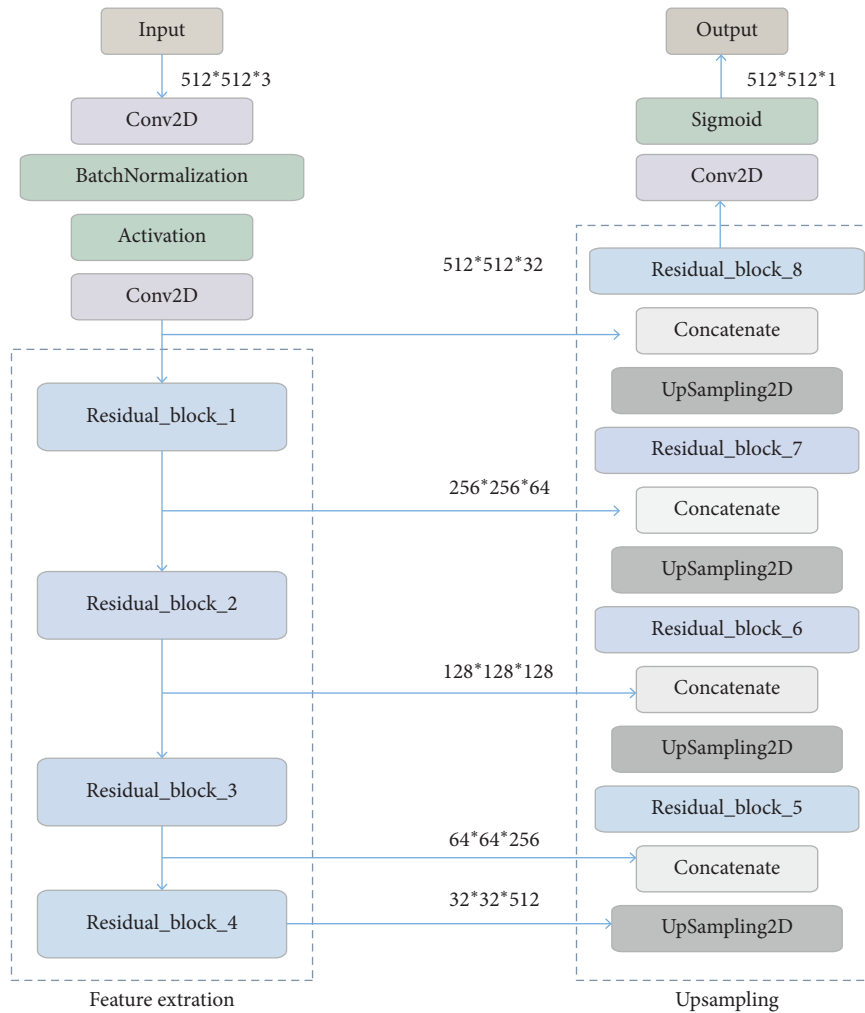


FIGURE 11: Structure of improved crack segmentation model.

The segmentation network performs upsampling of the feature extraction network feature maps with sizes of  $20 \times 20$ ,  $40 \times 40$ ,  $80 \times 80$ , and  $160 \times 160$ , and the upsampling process combines the feature extraction network feature maps with sizes of  $40 \times 40$ ,  $80 \times 80$ ,  $160 \times 160$ , and  $320 \times 320$ ; this improves the segmentation network accuracy through multilevel joint learning.

Figure 12 shows the loss function curve obtained by training the pavement crack dataset in the improved model. The red line represents the training loss curve of the training set, and the blue line represents the val loss curve of the test set. According to the loss curve, the model training effectively reached stability by the 50<sup>th</sup> epoch. The detection results for the test set before and after model improvement are shown in Table 4.

The model was improved to optimize the hyperparameters in the network. After comparing the segmentation effect under different hyperparameter settings, the activation function was set as the sigmoid function and the SGD optimizer was selected to optimize the network training. The PR curves of the pavement crack segmentation model before and after improvement for the three types of

crack images are shown in Figure 13. As can be seen, the upper right convexity of the PR curve after improvement is more evident than before the improvement, indicating a better performance by the improved model.

Figures 14–16 show the segmentation effect of the three crack segmentation models for transverse, longitudinal, and alligator cracks, respectively. It can be seen that before improvement, the crack information obtained by the model is significantly wider and somewhat distorted when compared with the ground truth. The improved model, by contrast, can obtain more accurate segmentation results. Therefore, the improved method is more suitable for crack segmentation.

The crack segmentation model uses the cascade mode of multiple residual elements, which can effectively extract the morphological characteristics of the pavement crack and improve the learning effectiveness of the neural network on the crack characteristics. A single crack segmentation method based on U-Net can provide the crack pixel location information, but it cannot classify the crack [35]. Therefore, in this study, a fusion of two models was adopted to identify pavement crack images and to obtain the crack category, location information, and geometric parameters, thereby

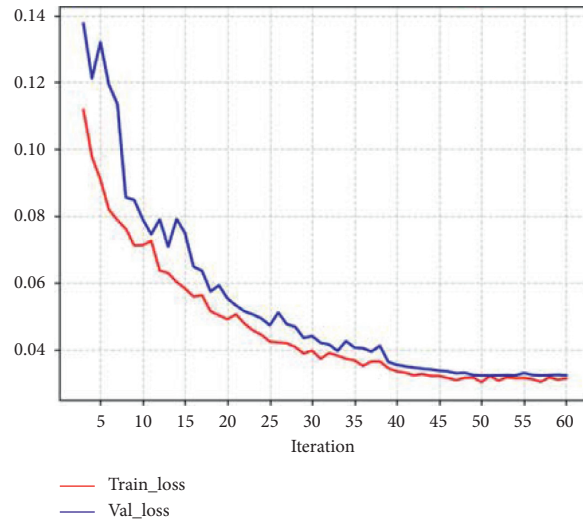


FIGURE 12: Training loss curve for the crack segmentation model.

TABLE 4: Comparison of prediction results of segmentation networks.

	Precision	Recall	F1-score
Before	0.654	0.763	0.699
After	0.723	0.795	0.752

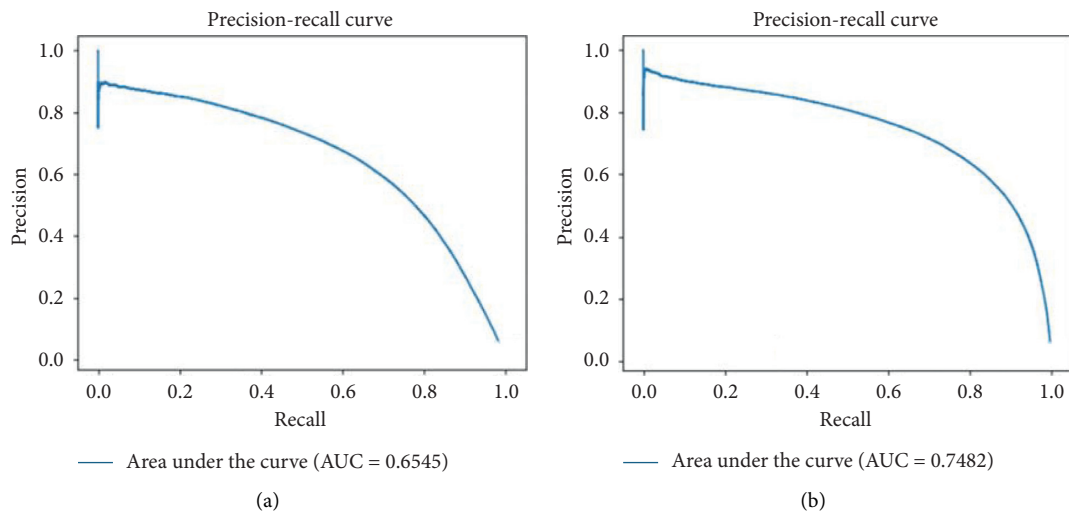


FIGURE 13: Pavement crack segmentation model PR curves before (a) and after (b) improvement.



FIGURE 14: Transverse crack segmentation model results: (a) original image; (b) before improvement; (c) improved result; (d) ground truth.



FIGURE 15: Longitudinal crack segmentation model results: (a) original image; (b) before improvement; (c) improved result; (d) ground truth.

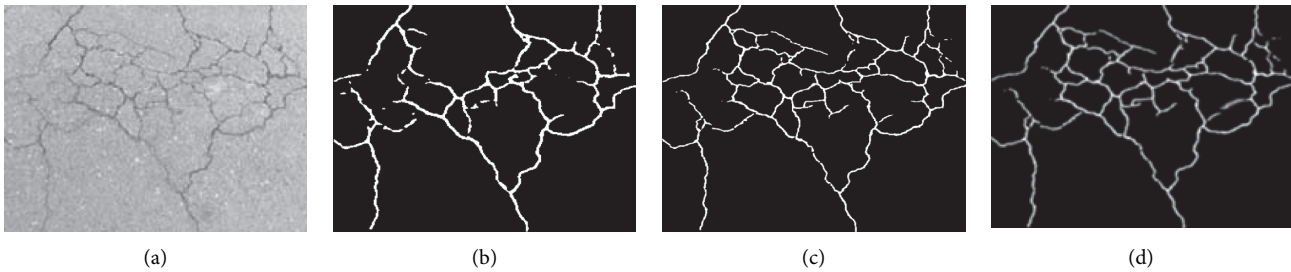


FIGURE 16: Alligator crack segmentation model results: (a) original image; (b) before improvement; (c) improved result; (d) ground truth.

facilitating accurate quantification and evaluation of pavement cracks.

#### 2.4. Proposed Fusion Model

**2.4.1. Fusion Model Design.** As shown in Figure 17, when only the SSD detection network is used, the number of cracks can be accurately obtained, but the crack width cannot be quantified. Use of only the U-Net segmentation model will lead to misjudgment of the number of cracks; e.g., a fractured crack may be misidentified as multiple cracks. The fusion of the detection and segmentation networks can avoid this phenomenon and ensure that the crack is identified as a single crack. Thus, the advantage of the fusion model is that it can accurately identify the number of cracks and ensure that cracks are quantified correctly.

The proposed fusion model adopts the following order: (1) detection, (2) segmentation. Because the detection network will obtain the result of a detection box, which contains one crack, and then segment that crack. This method can not only ensure crack number accuracy but also quantify the crack width information, which can prevent the misjudgment of crack number caused by using only the segmentation model. If the order is changed, i.e., segmentation first and detection second, in the presence of fractures in the cracks, the segmentation result will be misjudged into multiple cracks. Moreover, the segmented image is a binary image, which is not suitable for detection.

The proposed fusion model can improve pavement crack identification accuracy and provide a good foundation for measuring crack parameters. The fusion model structure is

shown in Figure 18. First, the input crack image enters the feature extraction network for feature extraction and learning. Convolutional feature maps with sizes of  $64 \times 64$ ,  $32 \times 32$ ,  $16 \times 16$ ,  $8 \times 8$ ,  $4 \times 4$ ,  $2 \times 2$ , and  $1 \times 1$  are used for multiscale fusion classification, and feature maps with sizes of  $512 \times 512$ ,  $256 \times 256$ ,  $128 \times 128$ ,  $64 \times 64$ , and  $32 \times 32$  are used for upsampling. After continuous calculation of losses and updating of the weights, the crack classification and position results are obtained. Then, the geometric parameters of the crack are calculated according to the segmentation results, and the identification results of the final fusion model are obtained.

The specifications of the fusion model process are shown in Figure 19. First, the standardized pavement crack image is input into the crack classification and detection model for classification. If there are cracks in the image, the categories and confidence values of the cracks are obtained. Next, the image of a detected crack and its category information are input into the crack segmentation model to obtain the precise pavement crack location, and the geometric parameters of the crack are calculated according to the segmented binary image of the pavement. Finally, the classified detection results are fused with the segmentation results, and the crack's geometric parameters are calculated. The result is restored to the corresponding original image for display, and the final pavement crack identification result is obtained.

**2.4.2. Feature Extraction and Analysis.** The deep convolutional network is trained by means of supervised learning. By extracting the morphological features of the cracks and

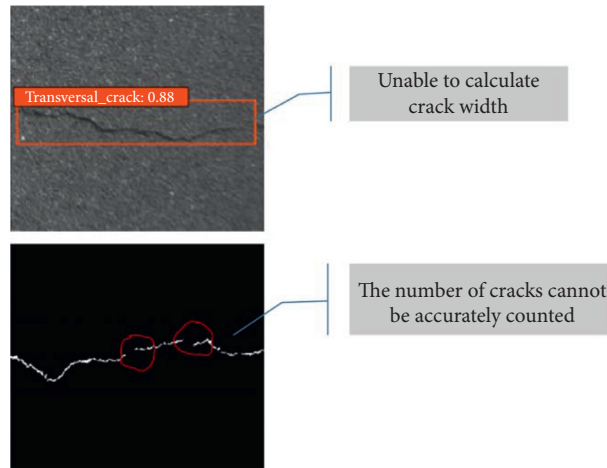


FIGURE 17: Motivation for using the fusion model.

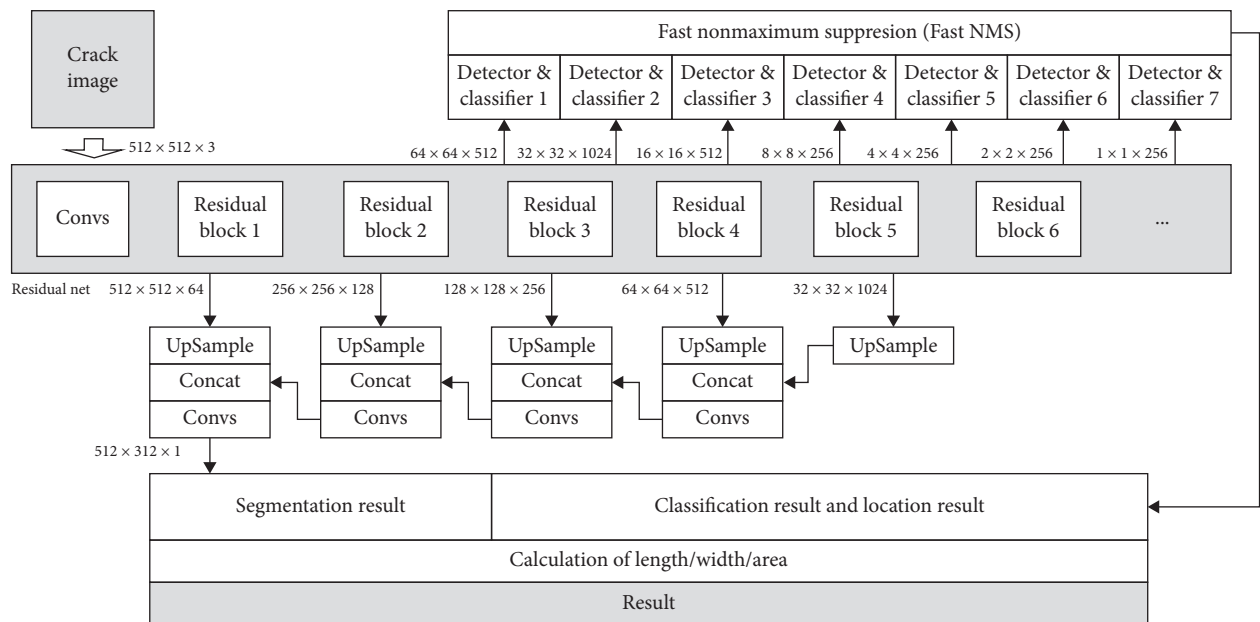


FIGURE 18: Structure of the fusion model.

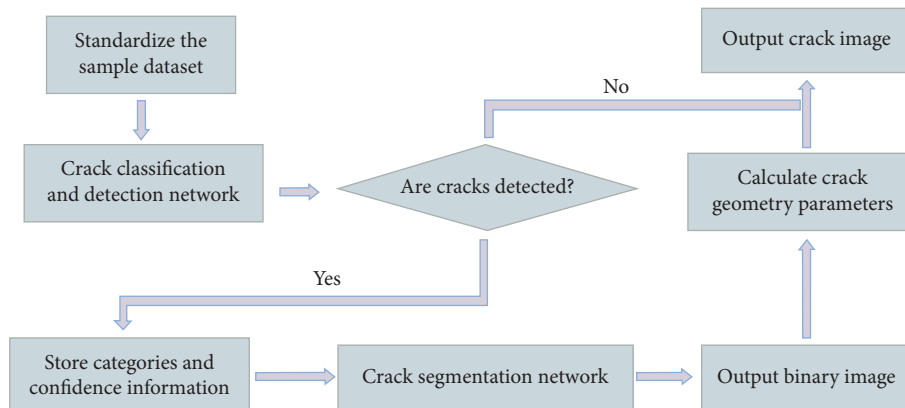


FIGURE 19: Flow chart of the fusion model.

continuously comparing them with the tag values to calculate the loss, the parameters of each network layer are continuously adjusted to finally reach the state of small loss and accurate judgment. With the deepening of convolutional layers in the network model, the model has more powerful feature extraction capabilities and can therefore identify and detect more abstract information. At this stage of the experiment, the crack characteristics learned by the deep convolutional network are analyzed and studied, and the characteristics of the three crack types at different depths are displayed by means of feature visualization.

Figure 20(a) shows original transverse, longitudinal, and alligator crack images (left to right) that were provided as input into the model for feature visualization. Owing to the large number of feature maps generated by the convolution of each layer, it will cause confusion to display all the feature maps, so we can analyze the single feature map generated by each convolution kernel. Figure 20(b) shows a single feature map for each of three kinds of crack on a convolutional layer with an output size of  $128 \times 128$ . It can be seen from the feature map that the morphological features of the cracks are still relatively evident. The model has learned the curve shapes and directions of the three types of fractures, which can be discerned by the human eye. However, these features can be somewhat unclear. Figure 20(c) shows a single feature map for each of three kinds of crack on a convolutional layer with an output size of  $64 \times 64$ . Compared with the  $128 \times 128$  crack feature map, the pavement crack feature obtained by this layer is less clear, and the location feature of the crack is magnified and brighter. However, the basic linear morphological features of the three crack types can still be distinguished by eye. Figure 20(d) shows a single feature map for each of three kinds of crack on a convolutional layer with an output size of  $32 \times 32$ . Compared with the other two convolutional layer feature maps, the brightness of the feature map of this layer is considerably improved, whereas the morphological features of the three crack types can no longer be distinguished, and the learned features are more complex.

From the above analysis, it can be seen that crack features extracted by different convolutional layers in the deep convolutional network are not the same. Meanwhile, with the deepening of network layers, crack features extracted by the deep convolutional neural network evolve from low-order features to high-order features. Therefore, in order to perform a comprehensive study of crack features and improve the identification accuracy of pavement cracks, it is necessary to fully extract their features by increasing the convolutional network depth.

**2.5. Calculation of Crack Parameters.** As shown in Figure 21, linear cracks (transverse and longitudinal cracks) and alligator cracks have different morphological characteristics; the former need two parameters, length and width, to be calculated, whereas the latter needs area parameters to be calculated. Therefore, these two types of morphological cracks are considered separately when the geometric parameters of pavement cracks are calculated. Each image is a

two-dimensional image, in which each point represents a pixel. There may be differences in the size of the image captured by different cameras, but the size of the pixel points is fixed. According to the corresponding relationship between image size and the pixel points, the real crack parameter value (unit: centimeter) can be calculated; e.g., if the width of the image is 1000 pixels and the camera's shooting range is 4 meters wide, a pixel point represents  $4 \text{ m}/1000 \text{ px} = 0.4 \text{ cm}$ .

In the binary image output by the pavement crack segmentation model, the pavement background is in black (a pixel value of 0) and the crack information is in white (a pixel value of 1). The steps for calculating the geometric parameters of a crack are as follows.

### 2.6. Calculation of Alligator Crack Area

Step 1: set the initial size (in pixels) of the pavement crack as  $S_x$ , and scan the pixels in the image from left to right and top to bottom. A pixel value of 1 means that it belongs to the crack area, and the area increases by 1:  $S_x = S_x + 1$ .

Step 2: if the pixel value is 0, it is part of the background, and the scan moves to the next pixel.

Step 3: the value of  $S_x$  obtained after all pixels have been scanned is the crack area, denoted as  $S$ .

### 2.7. Calculation of Linear Crack Length

Step 1: first, set the parameter  $L_x$ , denoting crack length (in pixels), to an initial value of 0. Scan the pixels in the image columns from left to right. When a column is found to contain a pixel value of 1,  $L_x = L_x + 1$ . Continue until all columns are scanned, and retain the final  $L_x$  value.

Step 2: scan the pixels in the rows from top to bottom according to the same principle as in the previous step, and set the parameter as  $L_y$ . If the value of any pixel in a row is 1,  $L_y = L_y + 1$ . Continue scanning all the rows, and keep the final  $L_y$  value.

Step 3: according to the Pythagorean theorem, crack length (in pixels) can be calculated as follows:

$$L = \sqrt{L_x^2 + L_y^2}. \quad (1)$$

**2.8. Calculation of Linear Crack Width.** The width of the crack can be obtained by dividing the linear crack area by its length, as follows:

$$W = \frac{S}{L}. \quad (2)$$

Through these steps, the length, width, and area of the crack can be expressed in pixels. Some pavement images were randomly selected for calculation of their geometric parameters, and the results are shown in Table 5.

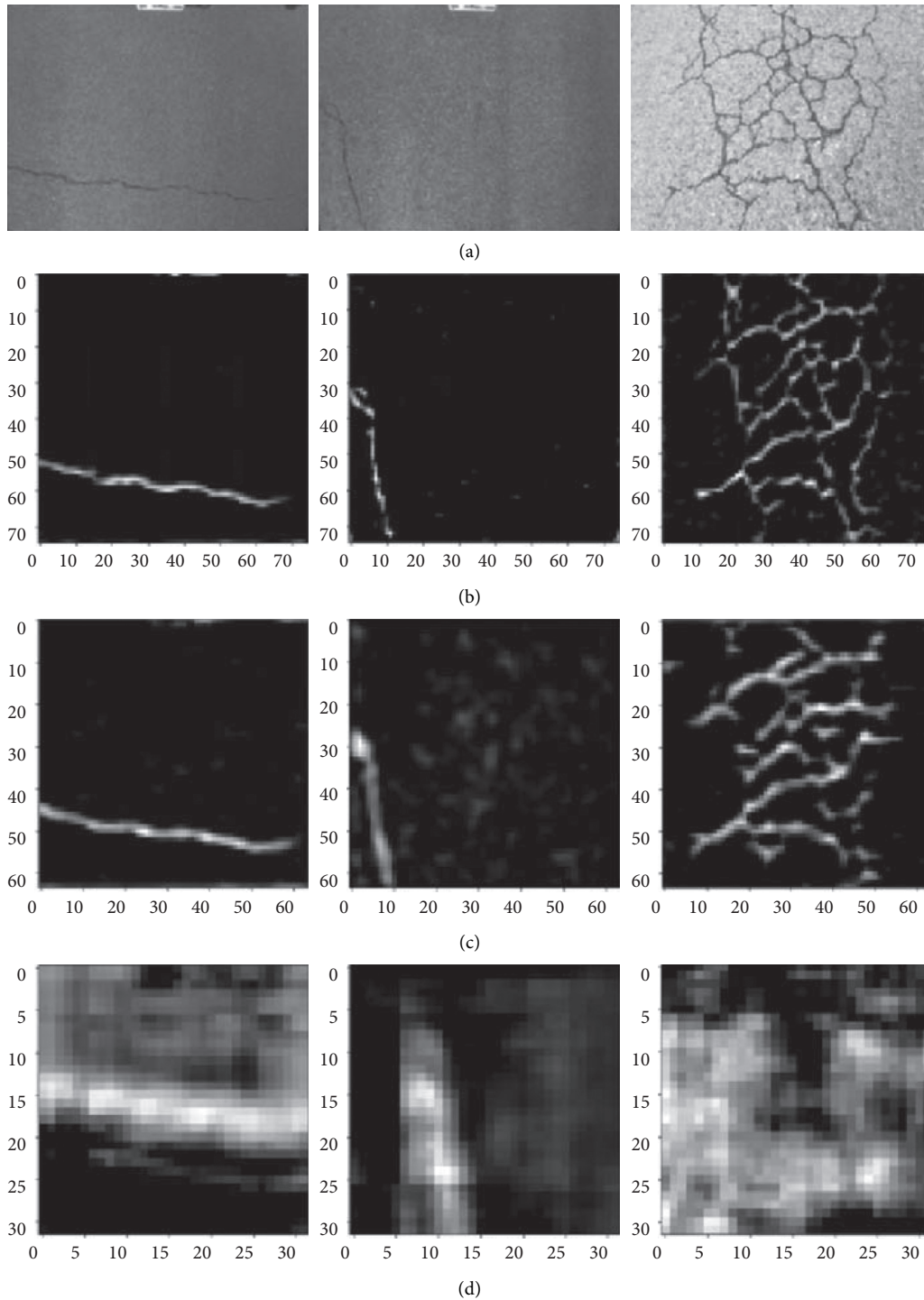


FIGURE 20: Different sizes of feature visualization.

### 3. Implementation Details and Results

**3.1. Data Preparation and Environment Setup.** In this research, a detector vehicle and a smartphone were used to collect crack images. The test sample dataset contained 8000 crack images, including 2800 transverse, 2800 longitudinal, and 2400 alligator crack images. With a ratio of 6:2:2, the sample data were divided into a training set (4800),

verification set (1600), and test set (1600) for training and testing the pavement crack identification model. The size of each image is  $1024 \times 960$  pixels. Figure 22(a) shows the proportional distribution of the numbers of the three crack types, and Figure 22(b) shows the proportional distribution of the dataset partitions for training. Details of the experimental equipment and software used in this study are given in Table 6.

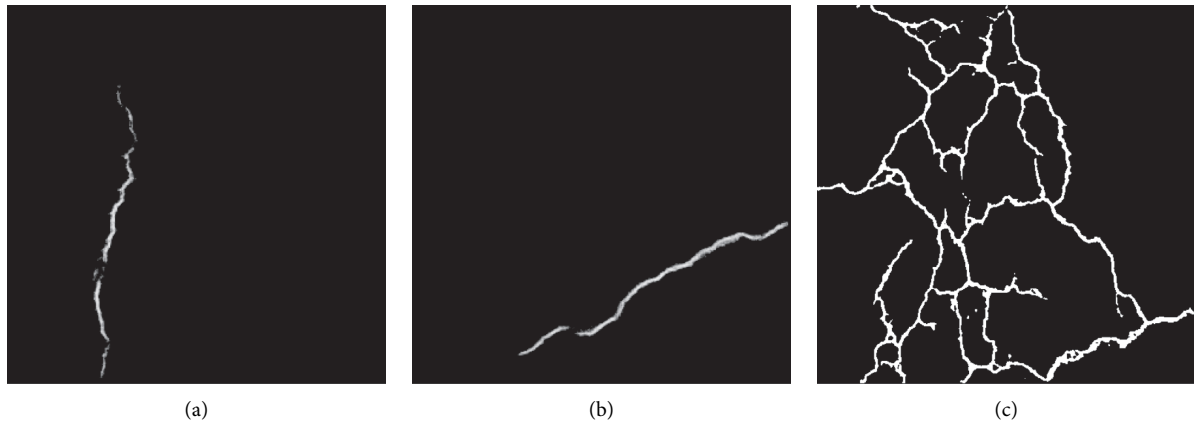


FIGURE 21: Binary images of three types of crack output by the segmentation network.

TABLE 5: Calculation results of geometric parameters of cracks (unit: pixels).

Image name	Crack type	Linear crack length	Linear crack width	Alligator crack area
000010.jpg	Transverse crack	534	5	—
000394.jpg	Transverse crack	496	4	—
001581.jpg	Longitudinal crack	3	512	—
002944.jpg	Longitudinal crack	4	557	—
003911.jpg	Transverse crack	533	3	—
005864.jpg	Longitudinal crack	6	490	—
020107.jpg	Alligator crack	—	—	13803
021323.jpg	Alligator crack	—	—	21837
022144.jpg	Alligator crack	—	—	16366
...	...	...	...	...

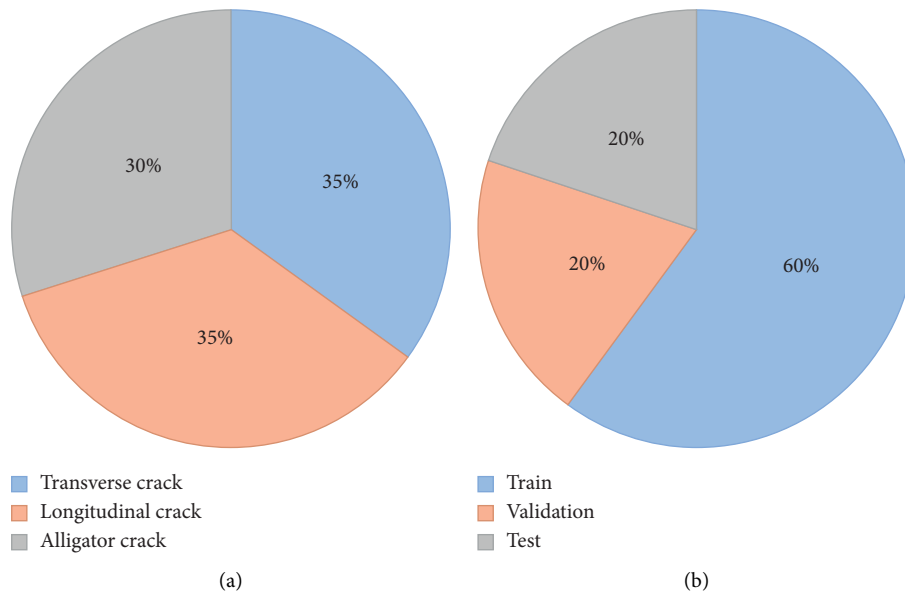


FIGURE 22: Proportional distributions: (a) percentages of the three crack types; (b) data set partitions for training.

### 3.2. Comparison of Each Individual Model Optimization

3.2.1. *Learning Rate.* For this study, a dynamic learning rate was adopted to adjust the learning rate of the model, thereby

optimizing the learning rate efficiently and improving the network training efficiency. The initial learning rate can be set to be relatively large and can then be gradually reduced as the number of iterations increases. Figure 23 shows the

TABLE 6: Experimental environment configuration details.

Device name	Version
PC operating system	Windows7 64 bit professional
Processor	Intel (R) core (TM) i7-8700 CPU @ 3.20 GHz
GPU	1080Ti
CUDA	cuda_9.0.176_windows
CUDNN	Cudnn-7.0-windows
TensorFlow	1.8.0
Keras	2.2.4
Python	3.0
Anaconda	Anaconda3-5.0.1-windows-x86_64

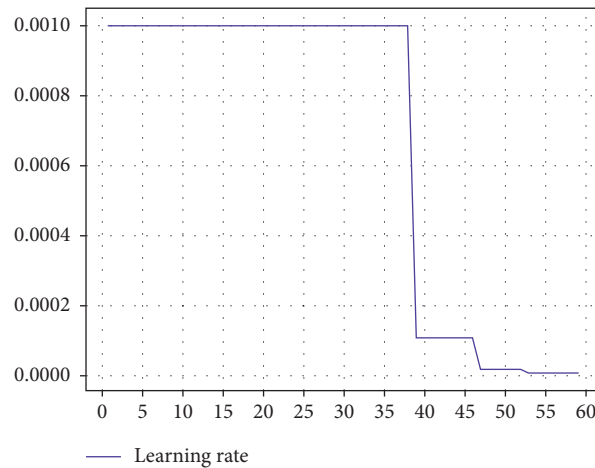


FIGURE 23: Learning rate during the training process.

learning rate of the pavement crack segmentation model during the training process. As the number of iterations increases, the learning rate is dynamically adjusted, becoming increasingly smaller.

**3.2.2. Activation Function.** The ReLU function is a linear function in the positive direction [36, 37]. In this study, the ReLU, tanh, and sigmoid functions were compared, as shown in Figure 24. When the SGD algorithm is used, the ReLU function converges faster than the other two functions, with the additional advantage of low computational complexity. It does not need to perform an exponential calculation; it only needs to set an activation threshold. Moreover, it is more suitable for back propagation, which can maintain a constant gradient and avoid the occurrence of gradient dispersion. For these reasons, the ReLU function was selected as the activation function of the network model proposed in this paper.

**3.2.3. Optimizer.** The role of the optimizer in the deep neural network is to update and calculate the network parameters that affect model training and model output so as to make them approximate an optimal value and to minimize the loss function. For deep convolutional neural networks, choosing an appropriate optimizer plays a decisive role in the final recognition accuracy of the model. In this study, Adam, SGD, AdaGrad, AdamW, and Nadam were used to train and test the model. The test results are shown in Table 7. When training on crack data,

although the Adam optimizer has a relatively fast training speed, its crack prediction accuracy is not the best. Whereas differences in training times were small, accuracy values clearly differ. Therefore, SGD was selected as the model optimizer.

**3.2.4. Transverse and Longitudinal Ratios.** Owing to the linear shape of transverse and longitudinal cracks, there is a major difference between length and width. Considering the particularity of pavement crack morphology, this study modified the transverse-to-longitudinal ratio parameter in the model in an attempt to improve pavement crack identification accuracy. Results for the modified model tests are shown in Table 8. Modifications of the transverse-to-longitudinal ratio for training do not directly influence the model, but when the gap between the transverse and longitudinal ratios is set too large, the prediction accuracy for the alligator cracks is substantially reduced, reducing, in turn, the average precision of the model. Therefore, the transverse-to-longitudinal ratio parameter was set as (1.0, 2.0, 0.5, 3.0, 1.0/3.0).

### 3.2.5. Optimization Results

#### (a) Crack classification and detection network

As shown in Table 9, after optimizing the feature extraction network and hyperparameters of the crack classification and detection model, the model accuracy is improved.



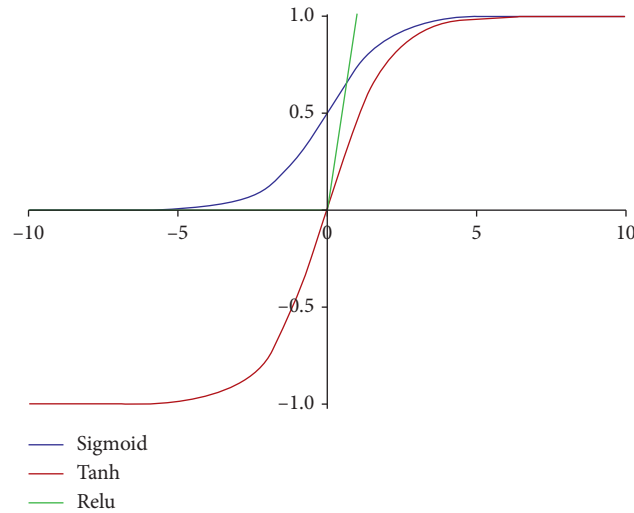


FIGURE 24: Comparison of three activation functions.

TABLE 7: Model speed and accuracy with different optimizers.

Optimizer	Training time (s)	Transverse cracks	Longitudinal cracks	Alligator cracks	Average accuracy
Adam	25	0.801	0.813	0.796	0.803
SGD	37	0.828	0.839	0.828	0.832
Adagrad	30	0.799	0.810	0.773	0.794
AdamW	27	0.806	0.817	0.802	0.808
Nadam	32	0.822	0.832	0.818	0.824

TABLE 8: Model speed and accuracy with different transverse and longitudinal ratios.

Transverse-to-longitudinal ratios	Training time (s)	Transverse cracks	Longitudinal cracks	Alligator cracks	Average accuracy
1.0, 2.0, 0.5, 3.0, 1.0/3.0	26	0.833	0.841	0.847	0.840
2.0,1.0,0.5,1.0/3.0,1.0/4.0	27	0.820	0.817	0.785	0.807
1.0,0.5,1.0/3.0,1.0/4.0,1.0/5.0	27	0.828	0.799	0.769	0.799
2.0,1.0,0.5,1.0/3.0,1.0/3.0	26	0.819	0.805	0.697	0.774

TABLE 9: Detection accuracy before and after optimization.

	Transverse crack		Longitudinal crack		Alligator crack	
	After	After	Before	After	Before	After
Accuracy	0.810	0.868	0.834	0.876	0.825	0.855

According to these pavement crack identification results, the improved and optimized method developed in this study is effective for classifying and locating cracks in pavement images.

(b) Crack segmentation network

After optimizing the feature extraction network and hyperparameters of the crack segmentation network, the segmentation accuracy of the model for pavement cracks is also improved. Results are shown in Table 10.

model-identified parameter information for each crack is shown in Table 11. The confusion matrices for the training and testing phases are shown in Figure 25.

The effectiveness of crack positioning is displayed visually in the form of images (Figures 26–28). As can be seen in the figures, the accuracy of crack identification and positioning is considerably improved by image segmentation. The length, width, and area of pavement cracks can be calculated more accurately by using the segmented binary image. At the same time, the crack information can be restored to the original crack image to accurately cover the crack area.

3.3. Fusion Model Identification Results and Analysis. In order to demonstrate the effectiveness of the model proposed in this paper, three sample images of each pavement crack type were randomly selected from the test set, and the

3.4. Comparison of Different Datasets. To illustrate the reliability of this experiment, we used three kinds of different datasets (our dataset, CFD dataset, and AigleRN dataset) to

TABLE 10: Segmentation accuracy before and after optimization.

	Before optimization	After optimization
Accuracy	0.752	0.787

TABLE 11: Fusion model prediction results (unit: pixels).

Sample number	Crack type	Confidence	Length	Width	Area
1	Transverse	0.70	496	4	—
2	Transverse	0.98	529	3	—
3	Transverse	0.71	519	3	—
4	Longitudinal	0.92	4	299	—
5	Longitudinal	0.87	3	512	—
6	Longitudinal	0.70	5	535	—
7	Alligator	0.91	—	—	21837
8	Alligator	0.88	—	—	13227
9	Alligator	0.88	—	—	20818

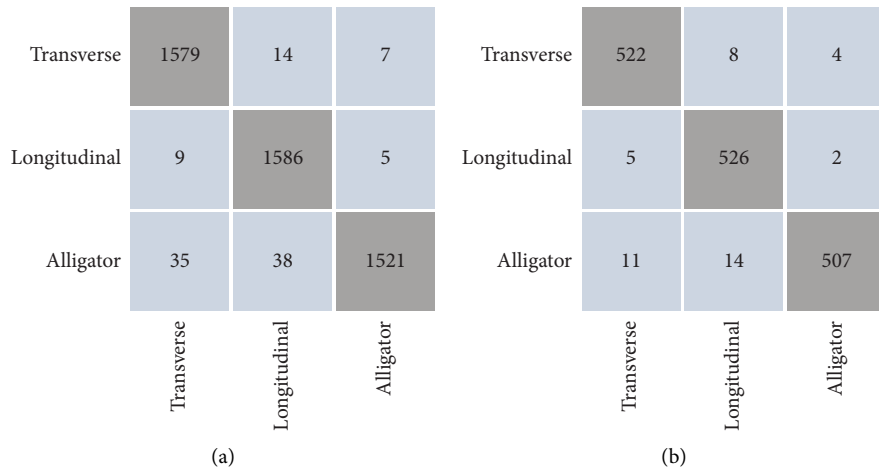


FIGURE 25: The confusion matrices for the training and testing phases. (a) Training set; (b) test set.

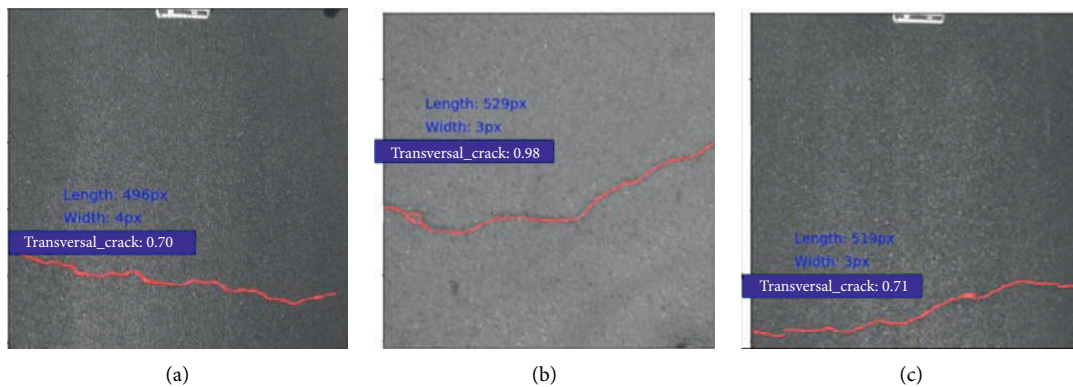


FIGURE 26: Transverse crack identification results.

carry out the experiment, respectively. Figure 29 shows some images of three different datasets. CrackForest Dataset (CFD) is an annotated road crack image database which can

reflect urban road surface condition in general. Its website is <https://github.com/cuilimeng/CrackForest-dataset>. The AigleRN dataset contains 38 images with pixel level

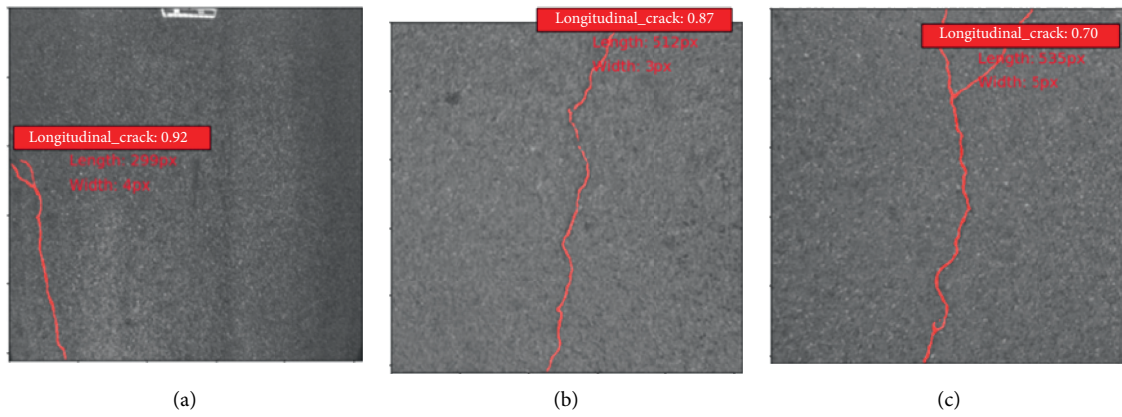


FIGURE 27: Longitudinal crack identification results.

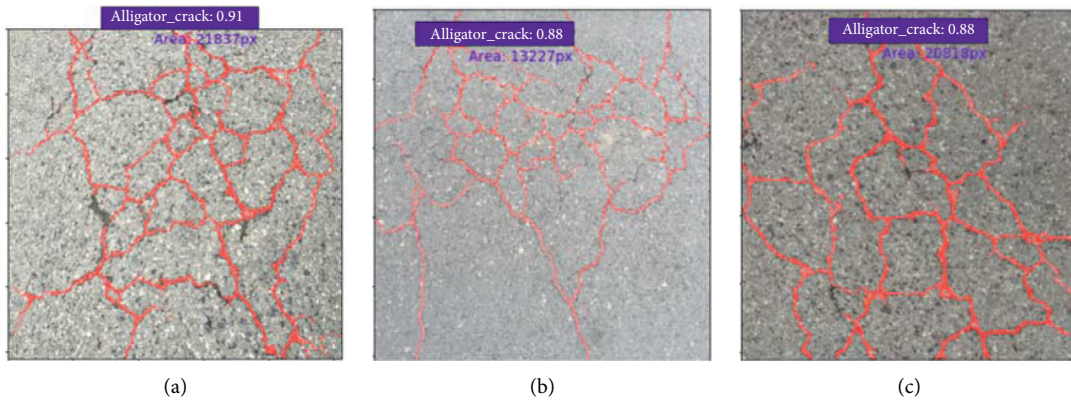


FIGURE 28: Alligator crack identification results.

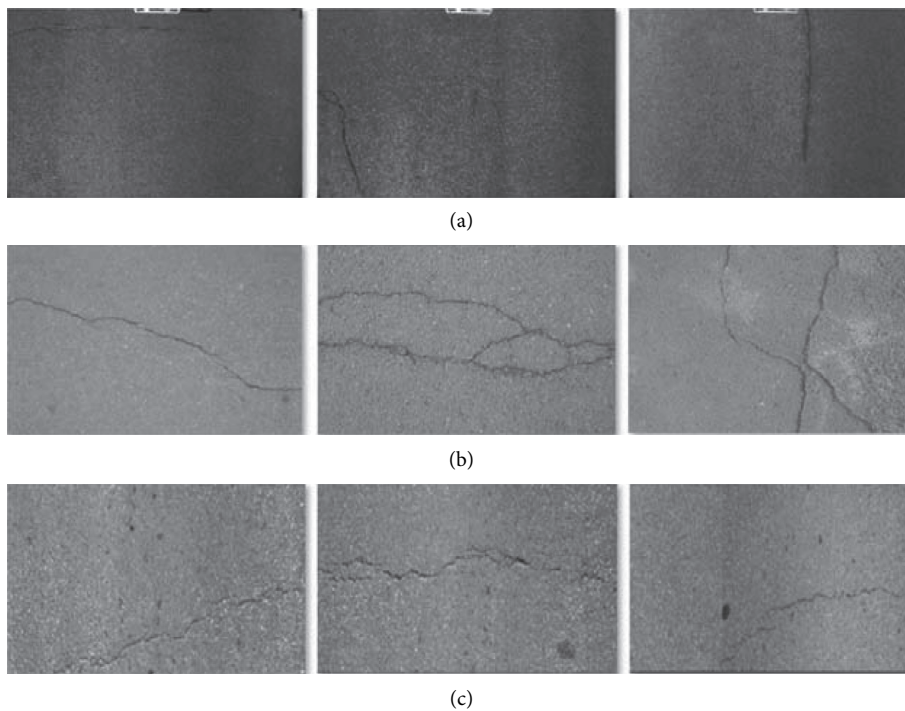


FIGURE 29: Three different datasets: (a) our dataset; (b) CFD; (c) AigleRN.

TABLE 12: Accuracy of different dataset.

	Our dataset	CFD	AigleRN
Accuracy	78.7%	80.1%	75.9%

annotation, which was obtained at driving speed, and the French road condition was regularly monitored using the Aigle RN system. Its website is <https://www.irit.fr/~Sylvie.Chambon/AigleRN.html>. Table 12 shows the accuracy of the proposed segmentation model in the three datasets. Experimental results show that the model has good performance on three datasets.

#### 4. Conclusions

In this research, a crack identification method based on a deep convolutional neural network fusion model is proposed. The strategy for model optimization was carried out through repeated experiments, and the model hyperparameters were optimized, effectively improving its pavement crack identification accuracy. In summary, the following conclusions can be drawn from this research:

- (1) To achieve accurate crack classification and segmentation, we propose a fusion model. The SSD network is used as the detection model and the U-Net network is used as the segmentation model, which can achieve crack classification and segments the cracks in the detection box at the same time. Moreover, crack length, width, and area parameters are calculated using the crack-segmented binary image. This method not only ensures the accuracy of crack number but also calculates crack parameters, which can prevent the misjudgment of crack number caused by using only a single model.
- (2) The SSD network model was proposed as the pavement crack classification and detection network. We improved the SSD model by replacing the VGG16 feature extraction network with the deep residual network. Experimental results show that the mAP of the improved model is 6.5% higher than that of the former model, which indicates that the classification and detection level of pavement cracks can be improved by optimizing the network. We made the same improvement to the U-Net model. By joining the segmentation model behind the detection model, we can solve the problem of inaccurate crack location in the pavement classification detection network. Results show that the precision of the improved segmentation model is 6.9% higher than that of the former. Therefore, the proposed fusion model has value in the field of pavement crack identification and classification.
- (3) Compare and select the learning rate, activation function, optimizer, and other parameters of the model. Experimental results demonstrate that the proposed model not only improves performance when compared to the original model but also

achieves higher accuracy, which has certain practical application value.

In summary, the method developed in this study offers an intuitive display and accurate reference for pavement crack identification for future road surface maintenance and for automatic pavement crack repair based on the calculated parameter information it provides.

#### Data Availability

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. Images from all sample sets used in this paper can be obtained from the corresponding authors after publication. Models and code used in this paper can also be obtained from the corresponding authors, if the requirements are reasonable.

#### Conflicts of Interest

The authors declare that they have no conflicts of interest.

#### Authors' Contributions

LX, XF, and WL conceptualized the study; LX and LP curated the data; LX, XF, and ZM performed the formal analysis; LX, XF, and ZM were involved in the methodology; LX and WL were responsible for the project administration; XF, LX, and HJ collected the resources; WL was responsible for the funding acquisition; LX, ZS, and XF investigated the study; LX, ZS, XF, and HS validated the study; LX, LP, and ZM wrote and prepared the original draft; XF, LX, WL, LP, ZS, ZM, and HJ writing, reviewed, and edited the manuscript.

#### Acknowledgments

This research was funded by the National Key Research and Development Program "comprehensive transportation and intelligent transportation" special project—"road infrastructure intelligent perception theory and method" (Grant number: 2018YFB1600202), National Natural Science Foundation of China (Grant number: 51978071), and Fundamental Research Funds for the Central Universities, CHD (Grant numbers: 300102249301 and 300102249306), China. This research was also supported by the Norman W. McLeod Chair in Sustainable Pavement Engineering, Centre for Pavement and Transportation Technology (CPATT), University of Waterloo, Waterloo, Ontario, Canada. Therefore, they are highly acknowledged.

#### References

- [1] D. Zhang and Q. Li, "Overview of the development of highway pavement rapid detection technology," *Surveying and Mapping Geographic Information*, vol. 40, no. 1, pp. 1–8, 2015.
- [2] D. Zhang and Q. Li, "Overview of rapid detection technology development of highway pavement," *Mapping Geographic Information*, vol. 40, no. 1, pp. 1–8, 2015.

- [3] B. Menghe, "Research on the development of expressway in China," *Northern Economy*, vol. 7, no. 3, pp. 60–61, 2013.
- [4] Ministry of Transport of the People's Republic of China, "2017 statistical bulletin on the development of highway and waterway transportation industry," 2018, [http://www.gov.cn/xinwen/2018-03/30/content\\_5278569.html](http://www.gov.cn/xinwen/2018-03/30/content_5278569.html), 2018-03-30.
- [5] J. Y. Cao, *Research on the Technology of Automatic Pavement Crack Detection Based on Digital Image Processing*, Chang'an University, Xian, China, 2014.
- [6] Z. Liu and Y. Zhang, "Detection technology of damaged road surface based on PDE and wavelet analysis," *Microcomputer and Application*, vol. 31, no. 8, pp. 35–37, 2012.
- [7] D. An-Guo, L. Miao-Miao, S. O. Science et al., "Crack detection algorithm based on gray correlation," *Application Research of Computers*, vol. 30, no. 10, pp. 3120–3121, 2013.
- [8] H. Oliveira and P. L. Correia, "Crack-IT an image processing toolbox for crack detection and characterization," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 798–802, Paris, France, October 2014.
- [9] W. Wang and L. Wu, "Pavement crack extraction based on fractional integral valley bottom boundary detection," *Journal of South China University of Technology (Natural Science Edition)*, vol. 42, no. 1, pp. 117–122, 2014.
- [10] L. Ren, X. U. Zhigang, X. Zhao et al., "Pavement crack connection algorithm based on Prim minimum spanning tree," *Computer Engineering*, vol. 41, no. 1, pp. 31–36, 2015.
- [11] Z. Anjun, X. Yang, L. Jia et al., "SRAD-CNN for adaptive synthetic aperture radar image classification," *International Journal of Remote Sensing*, vol. 40, no. 9, pp. 1–25, 2018.
- [12] S. Zhou, Y. Liang, J. Wan et al., "Facial expression recognition based on multi-scale CNNs," in *Proceedings of the Chinese Conference on Biometric Recognition*, October 2016.
- [13] Z. Zeng, W. Xie, Y. Zhang, and Y. Lu, "RIC-unet: an improved neural network based on Unet for nuclei segmentation in histology images," *IEEE Access*, vol. 7, pp. 21420–21428, 2019.
- [14] Y.-J. Cha, W. Choi, and O. Büyükköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 5, pp. 361–378, 2017.
- [15] G. Xu, J. Ma, F. Liu, and X. Niu, "Automatic recognition of pavement surface crack based on BP neural network," in *Proceedings of the 2008 International Conference on Computer & Electrical Engineering*, pp. 19–22, Phuket, Thailand, December 2008.
- [16] Q. Zou, Z. Zhang, Q. Li et al., "DeepCrack: learning hierarchical convolutional features for crack detection," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2018.
- [17] W. R. L. D. Silva and D. S. d. Lucena, "Concrete cracks detection based on deep learning image classification," *Proceedings*, vol. 2, no. 8, p. 489, 2018.
- [18] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proceedings of the 2016 IEEE International Conference on Image Processing*, pp. 3708–3712, IEEE, Phoenix, AZ, USA, September 2016.
- [19] S. Zhao and H. Ning, "Detection of pavement cracks based on convolutional neural network," *Sensors and Microsystems*, vol. 36, no. 11, pp. 135–138, 2017.
- [20] M. Eisenbach, R. Stricker, D. Seichter et al., "How to get pavement distress detection ready for deep learning? a systematic approach," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, Anchorage, AK, USA, May 2017.
- [21] R. Stricker, M. Eisenbach, M. Sesselmann, K. Debes, and H.-M. Gross, "Improving visual road condition assessment by extensive experiments on the extended gaps dataset," in *Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, July 2019.
- [22] H. Nhat-Duc, N. Quoc-Lam, and T. B. Dieu, "Image processing-based classification of asphalt pavement cracks using support vector machine optimized by artificial bee colony," *Journal of Computing in Civil Engineering*, vol. 32, no. 5, Article ID 04018037, 2018.
- [23] Z. Sun, L. Pei, W. Li, H. Xueli, and C. Yao, "Pavement encapsulation crack detection method based on improved Faster R-CNN," *Journal of South China University of Technology (Natural Science Edition)*, vol. 48, no. 2, pp. 84–93, 2020.
- [24] S. Wang, X. Wu, Y. Zhang et al., "Full convolution network image crack detection based on deep learning," *Journal of Computer Aided Design and Graphics*, vol. 30, no. 5, pp. 115–123, 2018.
- [25] 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 & Infrastructure Engineering*, vol. 32, no. 5, pp. 1–17, 2018.
- [26] P. Weng, *Research on Pavement Crack Segmentation Algorithm in Complex Environment*, Zhengzhou University, Zhengzhou, China, 2019.
- [27] Z. Fan, C. Li, Y. Chen et al., "Automatic crack detection on road pavements using encoder-decoder architecture," *Materials*, vol. 13, no. 13, p. 2960, 2020.
- [28] 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.
- [29] B. Qian, Z. Tang, X. Shen et al., "Pavement crack detection based on multi feature manifold learning and matrix decomposition," *Journal of Instrumentation*, vol. 37, no. 7, pp. 1639–1646, 2016.
- [30] W. Liu, D. Anguelov, D. Erhan et al., "SSD: single shot multiBox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 1–17, Amsterdam, The Netherlands, October 2016.
- [31] R. Joseph, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 5, Las Vegas, NV, USA, June 2016.
- [32] 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.
- [33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE 2016 Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [34] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proceedings of the European Conference on Computer Vision*, pp. 630–645, Springer, Amsterdam, The Netherlands, October 2016.
- [35] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing & Computer-Assisted Intervention*, pp. 234–241, Munich, Germany, October 2015.
- [36] P. Luo and H. Li, "Research on quantum neural network algorithm based on Tanh multilayer function and its

- application,” *Computer and Digital Engineering*, vol. 40, no. 1, pp. 4–39, 2012.
- [37] Z. Qu and X. Hu, “Research on convolutional neural network based on improved activation function,” *Computer Technology and Development*, vol. 27, no. 12, pp. 77–80, 2017.