Pixel-level Intelligent Segmentation and Measurement Method for Pavement Multiple Damages based on Mobile Deep Learning

Jiaxiu Dong1,2,3#, Zhaonan Li4#, Zibin Wang5, Niannian Wang1,2,3*, Wentong Guo1, Duo Ma1,2,3, Haobang Hu1,2,3, Shan Zhong1

1 School of Water Conservancy Engineering, Zhengzhou University, Zhengzhou 450001, China
2 National Local Joint Engineering Laboratory of Major Infrastructure Testing and Rehabilitation Technology, Zhengzhou 450001, China.
3 Collaborative Innovation Center of Water Conservancy and Transportation Infrastructure Safety, Henan Province, Zhengzhou 450001, China.
4 Faculty IV Electrical Engineering and Computer Science, Technical University Berlin, Berlin13353, Germany.
5 Guangdong Hualu Communications Technology Co.,Ltd., Guangzhou 510420, China.

# These authors contributed equally to this work and should be considered co-first authors.
Corresponding author: NiannianWang (e-mail: wnnian@163.com).

This research was funded by the National Key Research and Development Program of China (No. 2017YFC1501204), the National Natural Science Foundation of China (No. 51678536), the Guangdong Innovative and Entrepreneurial Research Team Program (2016ZT06N340), the Fellowship of China Postdoctoral Science Foundation (2020M672276, 2021T140620).

ABSTRACT Regular damage detection plays an important role in timely pavement maintenance. However, the existing detection methods struggle to efficiently and accurately identify the category and contour of the damage. Therefore, this paper proposes a Road-Mask R-CNN mobile damage detection model to automatically segment and measure multiple pavement damages. First, the optimized k-means clustering algorithm is used to intelligently determine the size and ratio of the anchor. Subsequently, the traditional nonmaximum suppression (NMS) algorithm is replaced by the distance intersection over union nonmaximum suppression (DIoU-NMS) algorithm, which improves the detection accuracy of multiple damages in the same image with a mean average precision (mAP) value of 0.934. Then, a comparative experiment with U-Net, the unimproved Mask R-CNN, MSNet and the unsupervised domain adaptation network (UDA) is carried out to verify the effectiveness of the proposed model. And combined with the segmentation and measurement results, the damage is quantitatively evaluated. Moreover, a webcam damage detection system combined with a workstation and an automatic damage detection system for smartphones is developed to quickly detect multiple types of pavement damage. In addition, on-site experiments are carried out on real pavements to verify the feasibility and effectiveness of the proposed method.

INDEX TERMS Mobile Deep Learning, Road-Mask R-CNN, Pavement Damages, Smartphones

I. INTRODUCTION

Pavement damages, such as cracks and pot holes, can directly reflect the durability and safety of a structure. Therefore, regular damage detection plays an important role in daily pavement maintenance and operations. According to the type, location and size information of the pavement damages, the internal damage and potential damage causes of the subgrade level can be inferred [1]. The most common method to detect pavement damage is manual inspection. However, due to the variety and large amount of pavement damages, the detection task is time-consuming, laborious and subjective [2]. The detect ability and experience of inspection personnel seriously affect the accuracy of pavement damage detection. Therefore, to improve the accuracy and efficiency of pavement damage detection, automatic damage detection algorithms are required.

With the development of computer vision, object detection methods based on digital image processing (DIP) have been applied to pavement damage detection. The damage detection method based on DIP includes threshold segmentation, region growth and edge detection algorithms [3], [4]. The key to threshold segmentation is to find a suitable threshold. If the pixel meets the threshold requirement, it is the target area. Oliveira H., et al. used a dynamic threshold to identify dark pixels in an image, and...
then determined the corresponding crack pixels [5]. Wang S., et al. proposed a segmentation algorithm based on a multiscale local optimal threshold to effectively realize the segmentation of pavement cracks [6]. Segmentation methods based on region growth can gather similar pixels together to form regions. Zhou Y., X., et al. proposed a segmentation method based on region growth [7]. This method uses a grid unit to detect and analyze the pavement damage pixels to realize pavement damage detection and segmentation. Muduli P., R., et al. proposed a hyperbolic tangent (HBT) filtering and canny edge detection algorithm to detect pavement cracks [8]. This method can detect the target edge more accurately. Ayenu P., R., et al. proposed a method for pavement crack detection based on two-dimensional empirical mode decomposition (BEMD) and the Sobel edge detector [9]. In this method, BEMD is used to remove the noise from pavement images, and a Sobel edge detector is used to detect cracks to improve the accuracy of crack detection. However, the methods of threshold segmentation, region growth and edge detection are not effective for image detection with complex backgrounds.

At present, the pavement damage detection method based on machine learning has become a hot topic [10]. Machine learning includes supervised learning and unsupervised learning. Supervised learning algorithms mainly include support vector machines [11], neural networks [12], etc., which require labeled data to train the model and adjust the model’s parameters. For instance, Marques A., et al. proposed a pavement crack detection algorithm based on a support vector machine [13], which can effectively extract the crack area. However, this method required high image quality and was unable to effectively extract the crack contour for fuzzy images. The difference between supervised learning and unsupervised learning is whether or not there are labels in the training and validation data [14]. Common unsupervised learning algorithms are considered as reinforcement learning (RL) and clustering. For instance, Akagic A., et al. proposed an unsupervised crack detection method for asphalt pavement based on the gray histogram and the Ostu threshold method [15], while achieving a satisfactory performance when there was a low signal-to-noise ratio.

Due to the irregular texture and shape of pavement damage, the above mentioned methods were unable to correctly extract and analyze the pavement damage feature. With the development of deep learning, it has been gradually applied to the detection and segmentation of the damages [16]-[21]. Deep learning-based damage detection methods can be divided into two categories: non pixel-level detection and pixel-level segmentation. The purpose of the non pixel-level detection method is to locate the damage and determine the damage type using the bounding box. To improve the efficiency and accuracy of detection, researchers have proposed many deep learning models, such as the CNN, YOLO [22], SSD [23] models. Mandal V. et al. proposed an automatic pavement crack detection model based on the YOLO v2 framework [24], and the accuracy and recall rate of crack detection were effective. Cha Y., J., et al. proposed a convolutional neural network with four convolutional layers to extract the features of concrete cracks [25] and achieved excellent detection results. Tong, Z., et al. proposed a crack detection method based on a convolutional neural network [26], which realized the extraction of cracks in GPR images. Ma D., et al. proposed a region-based fully convolutional neural network for intelligent recognition of pavement cracks and achieved a high recognition accuracy, as well as outstanding PR curves [27]. However, the above method could only detect pavement damage, but it could not extract the specific contour of the damage. However, the pixel-level segmentation method could label each pixel in the image to extract the specific contour of the pavement damage. Shim S., et al. proposed a multiscale and adversarial learning-based semi-supervised semantic concrete structure crack detection method [28]. This method effectively realized the segmentation of concrete structure cracks. Zou Q., et al. proposed a DeepCrack model [29], which mainly used an Encoder-Decoder to effectively segment pavement image pixels into the background and crack. Jenkins M., D., et al. proposed a segmentation network based on u-net [30], which realized the effective segmentation of pavement cracks at the pixel-level. Alipour M. et al. proposed a crack detection method based on deep fully convolutional neural networks [31]. This method could effectively detect cracks in concrete structures. The segmentation model based on Mask R-CNN was widely used in damage detection [32]-[33]. Kalfarisi R. et al. proposed a crack detection method based on a faster region-based convolutional neural network (FRCNN) and structured random forest edge detection (SRFED), and then applied Mask R-CNN for crack segmentation [34]. Hsu S., H. et al. proposed a crack detection model based on Mask R-CNN [35], which could successfully identify cracks on the concrete surface. Bai Y., S., et al. proposed the optimized Mask R-CNN +HRNet (high-resolution network) and Mask R-CNN +PANet (path aggregation network) model to achieve the segmentation of cracks, but the segmentation efficiency and accuracy needed to be improved [36]. In summary, there were two malpractices in the existing algorithms. First, there were multiple damages on an image and it was difficult to accurately detect and extract the characteristics of multiple damages. Second, the existing detection algorithm did not consider the problem with detection timeliness, making it difficult to meet the purpose of mobile detection and fast detection.

However, the above method had difficulty accurately and quickly segmenting situations where there were multiple damages in the same picture. Therefore, this paper proposes a Road-Mask R-CNN pixel-level mobile detection and segmentation method for pavement cracks and potholes. The algorithm combines optimized k-means clustering and distance intersection over union nonmaximum suppression.
(DIoU-NMS), which greatly improves the segmentation accuracy of multiple damages in the same image. Then, we completed the quantitative evaluation of road damage according to Highway Performance Assessment Standards. In addition, a webcam damage detection system combined with smartphones is developed to achieve the purpose of on-site quickly detection of pavement damage. The content of this study is organized as follows: Section 2 provides an overview of the proposed method and introduces the architecture of the Road-Mask R-CNN model. Section 3 describes the data preparation, model initialization and experimental evaluation index, which are the preparations required before model training. Section 4 explains the training, validation and test results in detail. Section 5 provides the damage measurement. Section 6 presents the on-site experiments of pavement damage detection based on the proposed mobile detection system. Section 7 concludes this study.

II. METHOD AND MODEL

At present, Mask R-CNN [37] is an effective image segmentation algorithm. To effectively segment and measure a variety of pavement damages, the Road-Mask R-CNN segmentation model is constructed in this paper. In this study, the dataset for pavement cracks, which is formed of the original images and the labeled information, is fed to the Road-Mask R-CNN. After 12000 iterations in the server terminals, the hyperparameters are fixed and the final model is obtained. According to the segmentation results of the final model, the topological features of the damages are extracted. Then, the size information of all kinds of damage is predicted. Subsequently, the pavement damage video or image captured by a smartphone is uploaded to the server through the wireless local area network (WLAN). The model trained on the server can quickly detect, segment, and measure the pavement damage. In addition, based on the TensorFlow mobile API, we integrate the trained model into the smartphone to realize the rapid speed of pavement damage detection. Figure 1 shows the overall structure of the intelligent segmentation and measurement method of pavement damages based on mobile deep learning.

FIGURE 1 Overall structure of the proposed method
A. ROAD-MASK R-CNN
The Road-Mask R-CNN structure is shown in Figure 2. The network includes a convolutional neural network, an optimized region proposal network (RPN), RoIAlign and a segmentation network. In this paper, we use the pretrained convolution network to extract features and combine the feature pyramid network (FPN) [38] to perform feature fusion from top to bottom. Then, the optimized k-means clustering algorithm is used to determine the size and ratio of the anchor. The region of interest (RoI) is generated by RPN, and DIoU-NMS is used instead of NMS to maintain an accurate RoI. Then, the reserved RoIs are mapped to the fixed dimension by RoIAlign. Subsequently, the classification and pixel-level segmentation results of the pavement damage are output by the segmentation model. Finally, the size information of the damage is obtained by further measurement.

FIGURE 2 The structure of the proposed model

1) THE CONVOLUTIONAL NEURAL NETWORK
In the model, the convolutional neural network is used to extract the features of the original damage image. ResNet101 [39] is used as a convolutional neural network in this paper, and the network structure is shown in Figure 3. The feature pyramid network (FPN) is used to fuse the bottom features with a high-resolution, low-semantic bottom features and low-resolution, and high-semantic top features from top to bottom, thereby enriching the semantic information of each size feature.

FIGURE 3 Network structure of Resnet101

2) OPTIMIZED REGIONAL PROPOSAL NETWORK
(1) Anchor size and ratio selection
The optimized region proposal network (RPN) can satisfy the input of different sizes. It is a kind of fully convolutional neural network, and can obtain the characteristics of multiple regions after a full convolutional operation. RPN takes the feature graph generated by a convolutional neural network as input to complete the task of extracting RoIs. In RPN, different aspect ratios and different scale anchors are used to slide in the feature map to generate several RoIs to achieve the goal of sampling. Then, the classification layer of the RPN recognizes that each RoI is a target area or a background area. The size and aspect ratio of an anchor play an important role in RoI. To adapt to different shapes and sizes of damage areas, this study uses the improved k-means clustering algorithm to cluster RoIs and determine the appropriate anchor size and ratio.

The k-means clustering algorithm uses Euclidean distance to measure the distance between objects. At this time, the large bounding box may produce a larger squared error than the small bounding box. However, the large bounding box may have a better overlap with the target box. Therefore, in this study, we use IoU as the standard of the k-means clustering algorithm. To obtain a more accurate anchor size and ratio, the final standard formula is shown in Equation (1):

\[ L(\text{box, anchor}) = 1 - \text{IoU(box, anchor)} \]  

(1)

It can be seen from the above equation that when the box and the anchor are similar, the IoU value is larger, and L(box, anchor) is closer to 0.

The steps of the improved k-means clustering algorithm are as follows: First, K bounding boxes are randomly selected as initial anchors. Second, we calculate the IoU value and assign each box to the anchor with the highest overlap. Then, the average width and height of all boxes are calculated, and the anchor is updated. Finally, we repeat the above two steps until the anchor no longer changes or the maximum number of iterations is reached.

The final anchor size and ratio are (32×32, 64×64, 128×128, and 256×256) and (1:1,1:2, and 2:1).

(2) RoI selection

To determine whether the RoI extracted is the target area, Mask R-CNN adopts the method of calculating the Intersection over Union (IoU) with the target area, as shown in Equation (2). In this paper, we compare the segmentation effects when the IoU value is 0.90, 0.75, 0.65 and 0.55. The results show that when the IoU value is 0.65, it has an effective segmentation effect. Therefore, the RoI is reserved when the IoU value of the target area is greater than 0.65.

\[ \text{IoU} = \frac{|A \cap B|}{|A \cup B|} \]  

(2)

In the above equation, A represents the real target area, and B represents the target prediction area.

However, many RoIs are obtained, and the target area in the actual image is far less than the number of RoIs. The original Mask R-CNN uses non maximum suppression (NMS) to eliminate redundant RoIs generated by RPN. The calculation method is shown in Equation (3).

\[ C_i = \begin{cases} 1, & \text{IoU}(g, b) < N \text{, } \text{if} \text{ IoU}(g, b) \leq N \text{, } \text{if} \text{ IoU}(g, b) > N \end{cases} \]  

(3)

In the above equation, \( C_i \) represents the final RoI retained, \( G \) represents the RoI with the highest IoU value, and \( b_i \) represents other RoIs. \( N \) represents the threshold of IoU.

In the original Mask R-CNN, the non maximum suppression method first arranges the retained RoI of each classification in descending order and selects the target frame with a high score. If the overlapping area of the remaining RoI and the region is greater than a certain threshold, the region frame will be deleted. However, when two similar targets are close, missed detection can easily occur. As shown in Figure 4, it is easy to miss the crack in the lower right corner.

When eliminating redundant RoIs, we should consider not only the overlap area, but also the distance between two RoI centers. In this paper, DIoU-NMS [40] is used instead of NMS. It can effectively improve the segmentation accuracy of multiple damages in the same image. The calculation method is shown in Equation (4).

\[ R_{\text{DIoU}} = \frac{\rho(a,b)}{c} \]  

(5)

In the above equation, \( \rho \) represents the Euclidean distance, \( a \) and \( b \) represent the center points of the target real area \( A \) and the target prediction area \( B \), respectively. \( c \) represents the diagonal length of the smallest enclosed area covering the two areas.

**FIGURE 4** An example of two RoIs with large overlap, the IoU values are 0.85 and 0.88.

3) ROIALIGN

The RoI generated and retained by the RPN has a corresponding relationship with the feature map in position, and these RoIs need to be mapped to the corresponding feature map. This process is called RoI mapping. The RoI pooling is used in Faster R-CNN [41]. However, RoI pooling uses two quantization operations, which not only causes pixel-to-pixel alignment failure but also affects the accuracy of the segmentation mask. Mask R-CNN uses RoIAlign.
Instead of RoI pooling, which cancels the quantization operation and uses the bilinear interpolation algorithm to accurately position the features. This improves the accuracy of the segmentation mask. Then all corresponding eigenvectors are fixed to a certain dimension.

4) SEGMENTATION NETWORK

The fixed size feature map is put into the segmentation network to realize the location, classification and segmentation of pavement damages. The segmentation network consists of three parts. The boundary box of the RoI is corrected by bounding box expression, and then the damage is classified by the classification branch of the RoI. Finally, the prediction mask of the pavement damage is output in the Mask Branch to realize the pixel-level segmentation of the pavement damage.

B. LOSS FUNCTION

In training the Road-Mask R-CNN model in this paper, the loss function is mainly composed of classification loss, detection loss and segmentation loss, and is defined as follows:

\[
L = L_{cls} + L_{box} + L_{mask}
\]  

(6)

In the above equation, \(L_{cls}\) is the classification loss of the bounding box, and the softmax function is applied. \(L_{box}\) is the detection loss of the bounding box. At present, the common regression loss functions include the \(L1\) loss function, \(L2\) loss function and Smooth \(L1\) loss function. When the error is small, the \(L1\) loss function has difficulty converging to effective accuracy. For the \(L2\) loss function, when the error is greater than 1, the robustness is poor at this time, and the phenomenon of gradient explosion may occur. However, the Smooth \(L1\) function can effectively solve the above problems, and the Smooth \(L1\) function is applied, as shown in Equation (7). \(L_{mask}\) is the mask segmentation loss, which applies the sigmoid function to each pixel and takes the average of the binary cross loss entropy of all pixels on each RoI as the mask segmentation loss.

\[
\begin{align*}
&\left| t - d(P) \right| \\
&\left| t - d(P) \right| \leq 1 = \text{Smooth } L1
\end{align*}
\]

(7)

In the above equation, \(t\) represents the real box of the model, and \(d(P)\) represents the predicted box of the model.

III. MATH DATASET AND EXPERIMENTAL CONFIGURATION

A. DATA SOURCES

The data set required for the experiment in this paper is 9,000 damage images taken on the highway using a high-definition camera. Among them, 4,500 crack images, and 4,500 pot hole images were obtained. According to the ratio of 3:1:1, the data are divided into a training set, validation set, and test set. The detailed numbers of the training set, validation set, and test set of the three types of damage are shown in Table 1. The training set is used to train the model proposed in this paper, the validation set is used to verify the trained model, the hyperparameter selection is performed for the model, and the test set is used to evaluate the final performance of the model.

In the process of image collection, image noise interference may affect the extraction of image features. Therefore, Gaussian smoothing filter is required. After model testing, 170 images can be filtered in 1 second, and it takes approximately 52.94 seconds to filter the DATASET of 9000 images. To effectively adapt to the model proposed in this paper, we need to preprocess the collected data before labeling. The collected data are cut to 1024 × 1024.

**TABLE 1**

| Damage type | Training | Validation | Testing |
|-------------|----------|------------|---------|
| Crack       | 2700     | 900        | 900     |
| Pot holes   | 2700     | 900        | 900     |

B. DATASET LABELING

We need to label the data as mask labels, so open source labelme software is used in this paper. First, the labeled image is imported, and the target is labeled by manual tracing to form a closed curve. Two types of damage are recorded: cracks and pot holes. The background, crack, and pot hole areas are recorded as 0, 1, and 2 respectively.

C. MODEL INITIALIZATION

When training the proposed model, we use transfer learning as the initial model, which can improve the efficiency of model training. According to the strategy of transfer learning, the weights of the model are initialized using pre-trained coco.h5. Subsequently, we use the error back propagation algorithm and the stochastic gradient descent to update the weight parameters of the model proposed in this paper.

D. EXPERIMENTAL EVALUATION INDEX

At present, many evaluation indices have been proposed for image segmentation. In current pixel-level image segmentation, average precision (AP) and the mean average precision (mAP) are usually used to evaluate the effectiveness of the algorithm. First, the precision is calculated using Equation (8) and the recall is calculated using Equation (9). The calculation methods of AP and mAP are shown in Equations (10) and (11):

\[
P = \frac{TP}{TP + FN}
\]  

(8)

\[
R = \frac{TP}{TP + FN}
\]  

(9)

\[
AP = \frac{\sum_{R} P(R) \times R}{R}
\]  

(10)

\[
mAP = \frac{\sum_{R} AP_{R}}{n}
\]  

(11)
In the above equation, $TP$ is the number of true positives, $FP$ is the number of false positives, and $FN$ is the number of false negatives. The value of $AP$ is the area of the P-R curve. The value of $mAP$ represents the mean value of all categories. $AP$ and $mAP$ comprehensively consider the impact of both accuracy and recall. Therefore, this evaluation index is used in this study.

E. EXPERIMENTAL ENVIRONMENT

Big data processing requires the support of a high-performance computer. The work of this paper is based on the TensorFlow framework in the Linux system and is equipped with a Nvidia 2080Ti GPU and 32 GB RAM Intel i9 CPU. To further accelerate the speed of image processing, CUDA and CUDNN are used in this paper. The computer language used is Python 3.6.

IV. EXPERIMENTS AND DISCUSSION

A. TRAINING AND VALIDATION

1) THE IMPACT OF HYPERPARAMETERS ON THE MODEL

In neural networks, many hyperparameters need to be set artificially, such as the learning rate, number of iterations, weight decay, and momentum. To train the neural network, the learning rate and the number of iterations are greatly affected. The weight decay, batch_size and momentum are set to 0.0002, 128 and 0.9 respectively. If the number of iterations is too small, it will lead to insufficient training and under fitting of the model. When the number of iterations is too large, over fitting is prone to occur. If the learning rate is small, the learning speed will be too slow and the training time will be longer. When the learning rate is large, the convergence of the loss function will be affected. According to the needs of model training in this paper, we should choose a larger learning rate at the beginning of model training, and the loss will decrease faster. After training for a period of time, the step length of parameter updating should be reduced, meaning that the learning rate needs to be attenuated. Therefore, the choice of the initial learning rate is very important. The closer the loss is to convergence, the smaller the learning rate should be. In this paper, the maximum number of iterations is set to 12000, three sets of initial learning rates are selected as 0.01, 0.001, and 0.0001, and exponential decay is adopted. The learning rate decays by an order of magnitude every 4000 iterations. The best iteration times and initial learning rate are selected by comparative experiments.

The global optimal loss is the minimum loss value in the entire training sample. We need to train the model to find the global optimal loss. The model training loss value change curve under different initial learning rates is shown in Figure 5. It can be seen from Figure 5 that when the initial learning rate is 0.01, the model loss decreases the fastest, and when 8000 iterations are performed, the final model loss converges to 0.15, but the global optimal loss cannot be achieved. When the initial learning rate is 0.001, this curve is well realized, and its loss decreases rapidly at the beginning and slowly at the end. When iterating 10,000 times, the loss is the smallest, reaching the global optimal loss, which is 0.065. When the initial learning rate is 0.0001, Figure 5 shows that the model converges slowly.

In this paper, the model is saved every 1000 iterations, and the model is verified on the 1800 validation set. Under three groups of different initial learning rates, the mAP value change line graph is shown in Figure 6. It can be seen from the figure that the model has the largest mAP value on the verification set when the initial learning rate is 0.001 and the number of iterations is 10,000, which can reach 0.935. Therefore, in this paper, the initial learning rate is 0.001, and the number of iterations is 10,000 for pavement damage detection and segmentation.

2) THE IMPACT OF DATA PREPROCESSING ON THE MODEL

To verify the impact of image filtering on the accuracy and efficiency of pavement damage segmentation, the original dataset, Gaussian smoothing filter and average filter datasets are trained separately. Validation is performed on the untrained validation set, and the results are shown in Table 2. The results show that the model trained on the Gaussian smoothing filter dataset has more effective segmentation accuracy and efficiency.

Graphic analysis is intuitive. The PR curve is used as the graphical analysis method. Precision and recall are expected to be as high as possible. This means that the PR curve is
closer to the top right, and the performance is better. As shown in Figure 7, the red curve of the Gaussian smoothing filter is closest to the top right. Therefore, this leads to the same conclusion as mAP that the performance of the Gaussian smoothing filter is accurate and comprehensive. Because mean filtering may lose image edge information, some features will be lost. However, the Gaussian smoothing filter is a linear filter that can retain the information of image features and edges, which will increase the efficiency of feature extraction and make the extracted features more effective. Therefore, the segmentation accuracy and efficiency of the model can be improved.

TABLE 2
The effect of image filtering on model segmentation performance

| Dataset            | mAP   | Average segmentation time |
|--------------------|-------|--------------------------|
| Original           | 0.928 | 2.5                      |
| Gaussian smoothing | 0.937 | 1.8                      |
| Average filter     | 0.926 | 2.4                      |

FIGURE 7 PR curve

B. TESTING RESULTS
In this paper, we need to use 1800 images of pavement damages in the test set (without training and testing models) to test the performance of the model. The test result is shown in Figure 8. Figure 8 shows a typical example of cracks and pot holes predicted by the model in this paper. From left to right in the figure, the original image, the real area, and the prediction area are displayed. Figures 8(a) and 8(b) display the segmentation results of the pavement transverse cracks and longitudinal cracks. The trained model can effectively achieve segmentation. In the same way, the model can also effectively realize the segmentation of pavement pot holes, as shown in Figure 8(c). As shown in Figure 8(d), the model can achieve satisfactory segmentation of cross cracks, but because the texture features of the damage are more complex, the segmentation effect is not as good as cracks and pot holes.

Noise may affect the detection and segmentation effects of the model. To test the noise robustness and generalization ability of the model, this paper selects images with noise (such as complex backgrounds, debris, and multiple damages) to test the model. The test results of pavement damage with a complex background are shown in Figure 8(e). Figure 8(f) shows the pavement damage detection and segmentation effect containing debris, such as sand and soil. The test results of two damages including cracks and pot holes, are shown in Figure 8(g). The figure shows that when the above-mentioned noise is contained, the model trained in this paper can still detect and segment the damaged area from the background correctly and effectively.

To quantitatively analyze the detection and segmentation performance of the training model, Table 3 lists the mAP values of the types in Figure 8. It can be seen from the table that cracks and pot holes have the best segmentation accuracy. Cross crack damage can also be effectively detected and segmented, and the segmentation accuracy can reach 0.876. When there are various noises, the detection accuracy of this model decreases, though it can still reach above 0.916. Therefore, the model proposed in this paper has an effective segmentation effect and good noise robustness.

TABLE 3
mAP values of the model proposed in this paper under various conditions

| Type                | Number of images | mAP   |
|---------------------|------------------|-------|
| All                 | 1800             | 0.934 |
| Transverse cracks   | 356              | 0.947 |
| Longitudinal cracks | 341              | 0.943 |
| Potholes            | 900              | 0.958 |
| Cross cracks        | 203              | 0.876 |
| Complex background  | 256              | 0.938 |
| Damage with debris  | 214              | 0.929 |
| Multiple damages    | 156              | 0.916 |
FIGURE 8 Examples of the prediction results of this model:
(a) transverse cracks, (b) longitudinal cracks, (c) pot holes, (d) cross cracks, (e) complex background damages, (f) damage with debris, (g) multiple damages

C. COMPARATIVE RESEARCH

Other deep learning algorithms, such as U-Net [30], have achieved good results in the field of image segmentation. To further demonstrate the segmentation performance of the proposed model, the Road-Mask R-CNN in this paper is compared with classic U-Net [30] and Mask R-CNN [35] on the testing set. It is also compared with the latest MSNet [42] proposed by Zhu, Xiaoyu and others in 2020, as well as the latest unsupervised domain adaptation network (UDA) [43] proposed by Suhyeon Lee and others in 2021. Table 4 shows the performance of the above models based on the segmentation mAP value and the segmentation time of the pavement damage. In Table 4, the mAP value of the Road-Mask R-CNN is slightly better than that of the U-Net, Mask R-CNN and MSNet models in the segmentation of pavement cracks and pot holes. The segmentation effect is also much better than that of the UDA network. However, it has more obvious advantages in the segmentation of pavement cross cracks and multiple damages. Because this paper replaces NMS with DIoU-NMS, the effective RoI of multiple targets in the same image is preserved, thereby improving the segmentation accuracy of the multiple damages. In addition, for the average segmentation time, the Road-Mask R-CNN is better than U-Net and not as good as Mask R-CNN, although the difference is small. The Road-Mask R-CNN is more complex than the UDA network. The Road-Mask R-CNN retains more RoIs than Mask R-CNN, and these RoIs can effectively solve the problem of missed detection in the target area. The model needs to process more RoI feature maps. The time cost is higher, but it can also meet the requirements of fast segmentation. Therefore, according to the above results, the Road-Mask R-CNN can provide effective segmentation results for subsequent pavement damage topological feature extraction and damage size measurement.

TABLE 4
Comparison of model segmentation performance

| Network           | mAP     | Time |
|-------------------|---------|------|
|                   | Crack   | Pot holes | Cross crack | Multiple damages |     |
| Road-Mask R-CNN   | 0.943   | 0.958    | 0.893       | 0.926         | 1.7 s |
| U-Net [30]        | 0.917   | 0.943    | 0.823       | 0.872         | 3.2 s |
| Mask R-CNN [35]   | 0.926   | 0.936    | 0.856       | 0.834         | 1.5 s |
| MSNet [42]        | 0.907   | 0.913    | 0.852       | 0.804         | 2.6 s |
| UDA [45]          | 0.853   | 0.846    | 0.828       | 0.818         | 1.3 s |
V. PAVEMENT DAMAGE MEASUREMENT AND EVALUATION

In most of the current studies, the segmentation and measurement of the damaged area have been completed. A quantitative evaluation of the damage measurement results is not carried out [17], [26], [44]. Therefore, because we realize the accurate segmentation of pavement damage, the quantitative measurement of damage size information is carried out in this paper. According to Highway Performance Assessment Standards [45], the degree of damage is evaluated. The measurement process of pavement damage size is shown in Figure 9. First, to prevent the real size error caused by image distortion, we place a triangle ruler of known real size next to the damage and take a photo. If the image is distorted, it can be rectified by the triangle ruler. Second, the image is imported into the segmentation model trained in this paper, and the pavement damage is segmented. At the same time, the edge detection algorithm is used to process the image, and the triangle ruler area is extracted. Then, the image of the damage is superimposed with the triangle ruler image, and the topological features of the superimposed image are extracted. Based on this, the pixel-level size of the side length of the triangle ruler is counted by the traversal algorithm [46]. The pixel-level size of the length and the area of pavement damage is calculated. Furthermore, the ratio between the true size of the side length of the triangle ruler and the pixel level size is calculated, and the ratio is marked as $K$. $K$ represents the real size of the unit pixel level size. The calculation method is shown in Equations (12).

$$K = \frac{\text{True size}}{\text{Pixel size}}$$

The average crack width is calculated by using the predicted real size of the area and the predicted real size of the length. Figure 10 shows five examples of damage measurements: (a) transverse cracks, (b) longitudinal cracks, (c) pot holes, (d) cross cracks, and (e) small scale damage. From left to right in the figure are the original image, the topological feature image, the predicted pixel size and the real size.

FIGURE 9 Measurement process of pavement damage size

Figure 10 shows that this paper effectively extracts the topological features of various damages. The effective extraction of topological features is conducive to correct statistics of damaged pixel information, is combined with the Ratio $K$ to further calculate the true size of the damage, and uses the same method to measure the size of the original image. The highway performance assessment standards' provisions on pavement damage evaluation are shown in Table 5. The original size, predicted real size and evaluation information in Figure 10 are shown in Table 6. Because the area of potholes and cross cracks can be calculated to meet the engineering requirements, “null” in the table means that no measurement has been performed. Table 6 shows that the measurement error of small scale damage is greater than that of ordinary cracks. Because the small scale damage is more similar to the background area pixels, it is difficult to extract the features. However, it can also achieve effective segmentation and measurement, as well as meet the actual engineering needs. In addition, a scientific quantitative evaluation of the degree of damage can effectively provide assistance to the highway maintenance department.
### TABLE 5
Assessment standards [43]

| Type                  | Standard                                           |
|-----------------------|----------------------------------------------------|
|                       | Severe                                             |
|                       | Mild                                               |
| Transverse cracks     | Average width>3 mm                                 |
|                       | Average width≤3 mm                                 |
| Longitudinal cracks   | Average width>3 mm                                 |
|                       | Average width≤3 mm                                 |
| Pot holes             | Area<0.1 m² (100000 mm²)                           |
|                       | Area≥0.1 m² (100000 mm²)                           |
| Cross cracks          | Area<0.1 m² (100000 mm²)                           |
|                       | Area≥0.1 m² (100000 mm²)                           |

### TABLE 6
Measurement results of pavement damage size information

| Number | Ratio K | Original size | Predicted real size | Error ratio | Damage Level |
|--------|---------|---------------|---------------------|-------------|--------------|
| Number |         | Length (mm)   | Area (mm²)          | Average width (mm) | Length (mm) | Area (mm²) | Average width (mm) |         |
| (a)    | 0.1356  | 356           | 5243               | 13           | 384         | 5653       | 14           | 7.82% | Severe |
| (b)    | 0.2936  | 519           | 6522               | 11           | 558         | 6936       | 12           | 7.50% | Severe |
| (c)    | 0.4098  | null          | 108149             | null         | null        | 102742     | null         | 5.00% | Severe |
| (d)    | 0.1597  | null          | 10869              | null         | null        | 12185      | null         | 12.11% | Mild |
| (e)    | 0.1514  | 1098          | 3283               | 3            | 1205        | 3635       | 3            | 10.71% | Mild |

Note: “null” means not to be measured.

![FIGURE 10](image)

**FIGURE 10** Example of measurements of the pavement damage size: (a) transverse cracks, (b) longitudinal cracks, (c) pot holes, (d) cross cracks, (e) small scale damage

To further prove the validity of the predicted size in this paper, the predicted size of the damaged area of 356 transverse cracks, 341 longitudinal cracks, 900 pot holes and 203 cross crack images are compared with the original size, as shown in Figure 11. Figure 11 (a), (b), (c), and (d) are the comparative curves of transverse cracks, longitudinal cracks,
pot holes and cross cracks, respectively. The calculated error values of the four kinds of damage are 7.82%, 7.43%, 4.97% and 11.95%.

It can be seen from Figure 11 and the error value that the prediction method of pavement damage size proposed in this paper has excellent prediction effects in transverse cracks, longitudinal cracks and pot holes. Because the texture features of cross cracks are more complex, the prediction effect of cross crack size is not as good as the above three damages.

To further verify the effectiveness of the damage measurement method in this paper, 78 cracks, 64 potholes and 37 cross crack images are compared with the threshold segmentation size measurement method. The error value was calculated, and the results are shown in Table 6. This proves that the measurement method in this paper is more effective.

![Comparison of predicted size and original size: (a) transverse cracks, (b) longitudinal cracks, (c) pot holes, (d) cross cracks](image)

**FIGURE 11** Comparison of predicted size and original size: (a) transverse cracks, (b) longitudinal cracks, (c) pot holes, (d) cross cracks

**TABLE 6**

| Method               | Error ratio |
|----------------------|-------------|
|                      | Crack       | Pot holes | Cross cracks |
| The method proposed  | 7.12%       | 5.23%     | 11.78%       |
| Threshold segmentation| 11.35%     | 9.87%     | 17.56%       |

**VI. EXPERIMENTS OF THE DAMAGE DETECTION SYSTEM BASED ON MOBILE DEEP LEARNING**

At present, smartphones have been rapidly popularized, and the number of intelligent phones in the world has exceeded 4 billion. Smartphones have been widely used as sensing devices in various fields. Using mobile phones for on-site detection is very effective and convenient. In this study, we propose a new pavement damage detection method based on mobile smartphones. To verify the performance of the detection system, on-site experiments were performed.

A. PAVEMENT DAMAGE DETECTION SYSTEM BASED ON A MOBILE WEBCAM

Mobile smartphones have a strong capacity for data collection, processing, and communication. Due to the camera and communication capabilities of a mobile smartphone, it can be regarded as a mobile webcam. To more conveniently and quickly detect pavement damage, the mobile webcam and the trained model are combined in this study to realize fast detection, segmentation and measurement.

The pavement damage detection system based on a mobile web camera is shown in Figure 12. The on-site experiment is...
performed on the asphalt pavement of Nanyang Road in Zhengzhou City. First, the pavement damage video or image captured by the smartphone is uploaded to the server through the wireless local area network (WLAN). The model trained on the server can quickly detect, segment, and measure pavement damage. Figure 12 shows examples of pavement cracks and pothole damage detection. It should be noted that the fixed-size triangle ruler in the figure is used as a calibration object when the pixel-level size is converted to the real size. In addition, the triangle ruler can also be replaced with other calibration objects with a fixed size. The experimental results prove the effectiveness and practicability of the road damage detection system based on the mobile webcam in intelligent pavement damage detection. In addition, the mobile webcam can be replaced by any other camera (such as a fixed webcam) for the long-term detection of pavement damage. By uploading the image to the server, it can provide big data support for deep learning and improve more basic information for road maintenance.

**FIGURE 12** The detection system and the results based on mobile webcams

**B. THE PAVEMENT DAMAGE DETECTION SYSTEM BASED ON MOBILE SMARTPHONES**

This paper uses Android studio to develop a mobile intelligent terminal. The OKHTTP framework is constructed, using multi-threaded access to the server to obtain data, and thousands of MB of data can be downloaded in milliseconds. The RecyclerView framework can realize vertical data scrolling and horizontal scrolling, storing invisible data as visible data, and improving the running speed of the application, so that pictures can be loaded quickly.

In addition, in this study, the central processing unit (CPU) of the smartphone is used to deal with the images obtained by the smartphone camera. Based on the TensorFlow mobile API, we integrate the model trained in Section 4 into the damage detection system of the smartphone, thus, we realize the fast damage detection of pavement damage detection by using the smartphones. We verify the detection performance of the smartphone through on-site experiments on Nanyang Road in Zhengzhou. The on-site detection images are shown in Figure 13. In this study, we conducted field experiments under normal lighting, dark lighting, and interference with debris.

**FIGURE 13** On-site detection experiment: (a) normal lighting, (b) dark lighting

Figure 14 shows the results of smartphone on-site detection under normal lighting. Figure 14 shows that the pavement damage detection system based on mobile smartphones can be used to identify, segment and measure the crack and pot hole damage of pavement structures. There is one crack in Figure 14(a), Figure 14(b) and Figure 14(c), and the crack is successfully identified by segmentation and measurement. The measurement of crack size information is realized. In addition, one pot hole in Figure 14(d), Figure 14(e) and Figure 14(f) also demonstrates successful segmentation and measurement.

To further verify the effectiveness and robustness of the method, this study also used a smartphone to detect pavement damage in two special situations of dark lighting.
and interference with debris. The segmentation and measurement results are shown in Figure 15 and Figure 16, respectively, which indicate that all cracks and pot holes of the pavement are successfully identified and segmented in two special cases, and the measurement of the size information is further realized. It can be seen that the method is also effective for pavement damage detection with special interference, and it is proven that the method has good generalization ability and robustness against noise. The triangle ruler in Figure 14, Figure 15, and Figure 16 is also the calibrator for calculating the real size of pavement damage.

![Figure 14](image1)

**FIGURE 14** On-site detection experimental results under normal lighting: (a) (d) images of on-site detection, (b) (e) screenshots of smartphone segmentation, and (c) (f) screenshots of smartphone measurement

To further prove the effectiveness of the on-site experiments, this paper measured 48 sets of segmentation mAP values and averaged the segmentation time for on-site experiments. The mAP value can still reach 0.929, and the average segmentation time is 2.14 s longer than that of the server. Because there is a gap between the computing power of the smartphone and the server, the segmentation time is increased, but rapid segmentation can still be achieved.

In summary, the pavement damage detection method based on mobile smartphones proposed in this study can provide a powerful supplement for road damage detection. More importantly, in some situations that cannot be easily detected by humans, using this method can effectively improve the efficiency and accuracy of detection.

![Figure 15](image2)

**FIGURE 15** On-site detection experimental results under dark lighting: (a) (d) images of on-site detection, (b) (e) screenshots of smartphone segmentation, and (c) (f) screenshots of smartphone measurement

![Figure 16](image3)

**FIGURE 16** On-site detection experimental results under interference with debris: (a) images of on-site detection, (b) screenshots of smartphone segmentation, and (c) screenshots of smartphone measurement

**VII. CONCLUSION**

Based on mobile deep learning, this paper proposed an intelligent pavement damage detection model, which was used for the pixel-level intelligent fast segmentation and measurement of pavement cracks and pot holes. A total of 9000 images of pavement damages were collected by a high-definition camera, and the images were preprocessed and marked manually to build a dataset for model training, validation and testing.

The first contribution of this paper was that the optimized k-means clustering algorithm was used to intelligently determine the size and ratio of the anchor. Applying DloU-
NMS instead of the non-maximum suppression algorithm (NMS) retained the accurate region of interest (RoI) to improve the accuracy of multiple damage segmentation. Transfer learning was leveraged to initialize the model weight parameters. To optimize the hyperparameters such as the initial learning rate and the maximum number of iterations, through multiple comparison experiments, the optimal hyperparameter combination was selected for model training. Based on the best training effect, the model was verified using the validation set, and the mAP value reached 0.934. The model performance was tested using testing images that were not used for training and validation. The experimental results show that the trained model has a satisfactory segmentation effect for transverse cracks, longitudinal cracks and pot holes. Cross cracks can also achieve effective segmentation, but the segmentation accuracy is not as good as the above three damages. Images with complex backgrounds, debris and multiple damages were selected to analyze the model’s robustness. The results show that the model has good noise robustness. In addition, the Road-Mask R-CNN was compared with the U-Net, Mask R-CNN, MSNet and UDA networks. The comparison results show that the Road-Mask R-CNN has a better performance in the intelligent segmentation of pavement damage.

The second contribution was to extract the topological features of the damage and realize the automatic measurement of the pavement damage size information. According to the effective segmentation results, the topological features of four kinds of pavement damage, including transverse cracks, longitudinal cracks, pot holes and cross cracks, were extracted, and the pixel size information of the damage was further predicted. Then, combined with the Ratio K calculated in this paper, the true size information of the damage was predicted. The validity of the predicted results was proven. In addition, the quantitative evaluation of pavement damage has been completed to provide assistance of the highway maintenance department.

Most importantly, the pavement damage detection method based on mobile smartphones proposed in this study and an on-site experiment was carried out on Nanyang Road, Zhengzhou City. First, the pavement damage detection system based on a mobile webcam was connected to the server via WLAN and the segmentation of road damage was realized. Second, the pavement damage detection system based on mobile smartphones realized intelligent segmentation of pavement damage without other equipment.

In the future, we will focus on the research of cross cracks and network cracks. Further research on the evaluation of damage measurement results will provide more effective help for highway maintenance.

REFERENCES:
[1] Yang X. C., Li H., Yu Y. T., et al., “Automatic pixel-level crack detection and measurement using fully convolutional network.” Computer-Aided Civil and Infrastructure Engineering, 33.12: 1090-1109, 2018.
[2] Li S. Y., Zhao X. F., Zhou G., Y., “Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network.” Computer-Aided Civil and Infrastructure Engineering, 34.7: 616-634, 2019.
[3] Cao W., M., Liu Q., F., He Z., Q., “Review of pavement defect detection methods.” IEEE Access, 8: 14531-14544, 2020.
[4] Zou Q., Cao Y., Li Q., Q., et al., “CrackTree: Automatic crack detection from pavement images,” Pattern Recognition Letters, 33.3: 227-238, 2012.
[5] Oliveira H., Correia P. L., “Automatic road crack segmentation using entropy and image dynamic thresholding,” 2009 17th European Signal Processing Conference, IEEE, 2009.
[6] Wang S., C., Tang W., S., “Pavement crack segmentation algorithm based on local optimal threshold of cracks density distribution,” International Conference on Intelligent Computing. Springer, Berlin, Heidelberg, 2011.
[7] Zhou, Y. X., “Seed-based approach for automated crack detection from pavement images,” Transportation Research Record, 2589.1: 162-171, 2016.
[8] Muduli, P. R., Pati, U. C., “A novel technique for wall crack detection using image fusion,” International Conference on Computer Communication & Informatics IEEE, 2013.
[9] Ayenu P. A., Attoh O. N., “Evaluating pavement cracks with bidimensional empirical mode decomposition,” EURASIP Journal on Advances in Signal Processing, (2008): 1-7, 2008.
[10] Ashkan S., M., Elson R., A., Yolibeth M., et al., 2020. “Estimating Stripping of Asphalt Coating Using k-Means Clustering and Machine Learning-Based Classification[J],” Journal of Computing in Civil Engineering, 34(1), 2020.
[11] Jahanshahi M., R., Masri S. F., “Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures,” Automation in Construction, 22.Mar.: 567-576, 2012.
[12] Larsson A., “Demarcating good solutions in system biology computer models using artificial neural networks,” Physics & Astronomy, 2010.
[13] Marques A., Correia P. L., “Automatic Road Pavement Crack Detection using SVM,” 2012.
[14] Barlow H., B., “Unsupervised Learning,” Neural Computation, 1.3: 295-311, 2014.
[15] Akagic A., Buza E., Omanovic S., et al., 2018. “Pavement crack detection using Otsu thresholding for image segmentation,” Pavement Crack Detection Using Otsu Thresholding for Image Segmentation IEEE, 2018.
[16] Lau S., Chong E., Yang X., et al., “Automated Pavement Crack Segmentation Using U-Net-based Convolutional Neural Network,” IEEE Access, vol. 8, pp. 114892-114899, 2020.
[17] Li S., Zhao X., “Automatic Crack Detection and Measurement of Concrete Structure using Convolutional Encoder-Decoder Network,” IEEE Access, 134602-134618, 2020.
[18] Zhang, Allen, Wang, et al., “Deep Learning-Based Fully Automated Pavement Crack Detection on 3D Asphalt Surfaces with an Improved CrackNet[J],” Journal of Computing in Civil Engineering, 2018.
[19] Song Q., Wu Y., Xin X., et al., “Real-time Tunnel Crack Analysis System via Deep Learning[J],” IEEE Access, vol. 7, pp. 64186-64197, 2019.

[20] Kim H., Kim H., Hong Y., et al., “Detecting Construction Equipment Using a Region-Based Fully Convolutional Network and Transfer Learning," Journal of Computing in Civil Engineering, 32(2):04017082, 2018.

[21] Yang J., Wang W., Lin G., et al., “Infrared Thermal Imaging-Based Crack Detection Using Deep Learning[J],” IEEE Access, vol. 7, pp. 182060-182077, 2019.

[22] Redmon J., Divvala S., Girshick R., et al., “You only look once: Unified, real-time object detection,” Proceedings of the IEEE conference on computer vision and pattern recognition, 779-788, 2016.

[23] Liu, W., Anguelov D., Erhan D., et al., “SSD: Single Shot MultiBox Detector,” European Conference on Computer Vision Springer, Cham, 2016.

[24] Mandal V., Lan U., Adu-Gyamfi Y., “Automated road crack detection using deep convolutional neural networks,” 2018 IEEE International Conference on Big Data (Big Data). IEEE, 5212-5215, 2018.

[25] Cha Y. J., Choi W., Buyukozturk O., “Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks,” Computer Aided Civil & Infrastructure Engineering, 32.5: 361-378, 2017.

[26] Tong Z., Gao J., Zhang H., “Recognition, location, measurement, and 3D reconstruction of concealed cracks using convolutional neural networks,” Construction and Building Materials, 146: 775-787, 2017.

[27] Ma D., Fang H., Xue B., et al., “Intelligent Detection Model Based on a Fully Convolutional Neural Network for Pavement Cracks,” Computer Modeling in Engineering and Sciences, 123.3: 1267-1291, 2020.

[28] Shim S., Jin K., Cho G., et al., “Multiscale and Adversarial Learning-Based Semi-Supervised Semantic Segmentation Approach for Crack Detection in Concrete Structures,” IEEE Access, vol. 8, pp. 170939-170950, 2020.

[29] Zou Q., Zhang Z., Li Q., et al., “Deeppcrack: Learning hierarchical convolutional features for crack detection,” IEEE Transactions on Image Processing, 28.3: 1498-1512, 2018.

[30] Jenkins M., D., Carr T. A., Iglesias M., L., et al., “A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks,” 2018 26th European Signal Processing Conference (EUSIPCO). IEEE, 2120-2124, 2018.

[31] Alipour M., Harris D., K., Miller G., R., “Robust Pixel-Level Crack Detection Using Deep Fully Convolutional Neural Networks,” Journal of Computing in Civil Engineering, 33(6):04019040.1-04019040.14, 2019.

[32] Pan Z., Yang J., Wang X., et al., “Surface Scratch Detection of Monolithic Glass Panel Using Deep Learning Techniques,” International Conference on Computing in Civil and Building Engineering. Springer, Cham, 133-143, 2020.

[33] Doğru A., “Using Convolutional Neural Networks to Automate Aircraft Maintenance Visual Inspection,” Aerospace, 7.12: 171, 2020.

[34] Kalfarisi R., Wu Z., Y., Soh K., “Crack Detection and Segmentation Using Deep Learning with 3D Reality Mesh Model for Quantitative Assessment and Integrated Visualization,” Journal of Computing in Civil Engineering, 34(3):04020010, 2020.

[35] Hsu S., H., Chang T., W., Chang C., M., “Concrete Surface Crack Segmentation Based on Deep Learning,” European Workshop on Structural Health Monitoring, Springer, Cham, 2020.

[36] Bai Y., Sehen Z., Yilmaz A., “End-to-end Deep Learning Methods for Automated Damage Detection in Extreme Events at Various Scales,” arXiv preprint arXiv:2011.03098, 2020.

[37] He K., M., Gkioxari G., Dollár P., et al., “Mask R-CNN,” 2017 IEEE International Conference on Computer Vision (ICCV) IEEE, 2017.

[38] Lin T., Dollar P., Girshick R., et al., “Feature Pyramid Networks for Object Detection,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[39] He K., Zhang X., Ren S., et al., “Deep residual learning for image recognition,” In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778, 2016.

[40] Zheng, Z., Wang P., Liu W., et al., “Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression,” AAAI Conference on Artificial Intelligence, 12993-13000, 2020.

[41] Ren, S., He K., Girshick R., et al., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis & Machine Intelligence, 39.6:1137-1149, 2017.

[42] Lee S, Hyun J, Seong H, et al, “Unsupervised Domain Adaptation for Semantic Segmentation by Content Transfer” arXiv preprint arXiv:2012.12545, 2020.

[43] Zhu X., Y., Liang J., W., Alexander Hauptmann, “Msnets: A multilevel instance segmentation network for natural disaster damage assessment in aerial videos,” Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021.

[44] Song, E. P. S. H. Eem , H. Jeon ., “Concrete crack detection and quantification using deep learning and structured light,” Construction and Building Materials, 252.5(2020):119096.

[45] “Highway Performance Assessment Standards ” JTG 5210-2018.0.

[46] Lai Y., Chung Y., “An Efficient Rasterization Unit With Ladder Start Tile Traversal in 3-D Graphics Systems,” IEEE Transactions on Magnetics, 50(7 Part1):1-4, 2014.

Jiaxiu Dong was born in Henan Province, China, 1992. He received the M.S. degree in computer technology from School of Information Engineering, Zhengzhou University, China, in 2019. He is currently pursuing the Ph.D. degree in School of Water Conservancy Engineering, Zhengzhou University, China.

His research interest is the cross integration of deep learning and civil engineering. Now he is mainly engaged in the research of intelligent detection technology of engineering infrastructure.

Zhaonan Li was born in Henan Province, China, received double B.Eng. degrees from Chinesisch-Deutsche Hochschule für Angewandte Wissenschaften, Tongji University, China, in 2018. He is currently pursuing M.S. degree with major in automotive systems, Technical University Berlin, Germany.

His research interest includes automatic image analysis, artificial intelligence of damage detection and classification. Now he is engaged in intelligent vehicle control and teleoperation.
Zibin Wang was born in Guangdong Province, China. At present, he is a engineer of Guangdong Hualu Communications Technology Co., Ltd. Now he is mainly engaged in the research of intelligent detection technology of Highway Damage.

Shan Zhong was born in Hunan Province, China, 2000. He is currently pursuing the B.S degree in School of Water Conservancy Engineering, Zhengzhou University, China. His research interests are deep learning and engineering protection. Now he is mainly engaged in the intelligent detection of damage for high highway.

Niannian Wang was born in Henan Province, China, 1989. She received the Ph.D. degree in structural engineering from Dalian University of Technology, China, in 2019. At present, she is a professor in School of Water Conservancy Engineering, Zhengzhou University, China.

Her research interest includes intelligent detection of engineering structures, damage identification and evaluation, and big data analysis.

Wentong Guo was born in Shanxi Province, China, 1999. He is currently pursuing the B.S degree in School of Water Conservancy Engineering, Zhengzhou University, China.

His research interests are deep learning and engineering protection. Now he is mainly engaged in the intelligent detection of damage for high highway.

Duo Ma received the B.S. degree in hydraulic engineering from Zhengzhou University, Henan Province, China, in 2017. He is currently pursuing the Ph.D. degree in School of Water Conservancy Engineering, Zhengzhou University, China.

His research interests include the integration of artificial intelligence, big data and civil engineering. His main research direction is the intelligence detection for engineering disease based on deep learning.

Haobang Hu was born in Henan Province, China, 1995. He received the B.S. degree in engineering from Zhengzhou University, Zhengzhou, China, in 2019. He is currently pursuing the M.S degree in engineering at Zhengzhou University, Zhengzhou, China.

His research interest includes the nondestructive detection based on ground penetrating radar and image recognition based on deep learning.