IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 4, 2020, accepted March 8, 2020, date of publication March 11, 2020, date of current version March 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980086

Semi-Supervised Semantic Segmentation Using Adversarial Learning for Pavement Crack Detection

GANG LI[1,3], **JIAN WAN**[1], **SHUANHAI HE**[2], **QIANGWEI LIU**[1], **AND BIAO MA**[1]

[1]School of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China
[2]Key Laboratory for Old Bridge Detection and Reinforcement Technology of Ministry of Transportation, Chang'an University, Xi'an 710064, China
[3]Key Laboratory of Road Construction Technology and Equipment of MOE, Chang'an University, Xi'an 710064, China

Corresponding authors: Gang Li (15229296166@chd.edu.cn) and Jian Wan (1041949906@qq.com)

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant 300102329203, in part by the CHD, in part by the Science and Technology Bureau of Xi'an Project under Grant 201805045YD23CG29(5), and in part by the Key Industry Chains of Shaanxi Technology Committee under Grant 2020ZDLGY-LIGANG.

ABSTRACT Regular inspection of pavement conditions is important to guarantee the safety of transportation. However, current approaches are time-consuming and subjective, which requires the technician to annotate each training image exactly pixel by pixel. To ease the workload of the inspector and lower the cost of acquiring the high-quality training dataset, a semi-supervised method for the pavement crack detection is proposed. Firstly, unlabeled pavement images can be used for the model training in our proposed algorithm, our model can generate a supervisory signal for unlabeled pavement images, which makes up for the deficiency of image annotation. Secondly, an adversarial learning method and a full convolution discriminator are adopted, which can learn to distinguish the ground truth from segmentation predictions. To improve the accuracy of pavement crack detection, the adversarial loss is coupled with the cross-entropy loss in discriminator. Thus, the quality of the training model is no longer dependent on the quantity of the labeled dataset and the accuracy of the labeled. Compared with existing methods that can only employ labeled images, our method utilizes unlabeled images to improve the pavement crack detection accuracy. Moreover, our model is validated on the CFD dataset and AigleRN dataset, the experimental results show that the proposed algorithm is effective. Compared with existing methods, not only can our method detect different types of cracks, but also be particularly effective when only a few labeled are available: when using 118 crack images with a resolution of 480 × 320, using only 50% of the labeled data, the detection accuracy of our model can reach 95.91%.

INDEX TERMS Adversarial learning, crack detection, semi-supervised learning, semantic segmentation.

I. INTRODUCTION

In road maintenance processes, to ensure the safety of the road and the highway, one of the most important tasks is to realize timely, accurate and automatic detection of pavement cracks. In order to maintain the traffic safety, it is an important responsibility for the traffic maintenance department to locate and repair the cracks. However, manual detection of pavement cracks is very time-consuming and requires expertise in related fields. In order to reduce the workload of technicians and facilitate the road inspection, it is necessary to realize the automatic detection of cracks.

Over the past few decades, various algorithms to automatically detect cracks on the pavement have been proposed. Deep-learning-based approaches (such as FCN-based) have accomplished marvelous achievement, but they need an excessive amount of training data. Different from the target detection, the task of semantic segmentation requires accurate pixel annotation of the training image, which needs a lot of time and cost. To reduce the cost of obtaining high-quality training datasets, semi-supervised semantic segmentation is adopted. These methods generally assume additional annotations at the image level [1]–[5] or the point level [6].

This paper proposes a semi-supervised algorithm for pavement crack detection, which can efficiently learn from the annotation-free images. The recent success of Generative

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

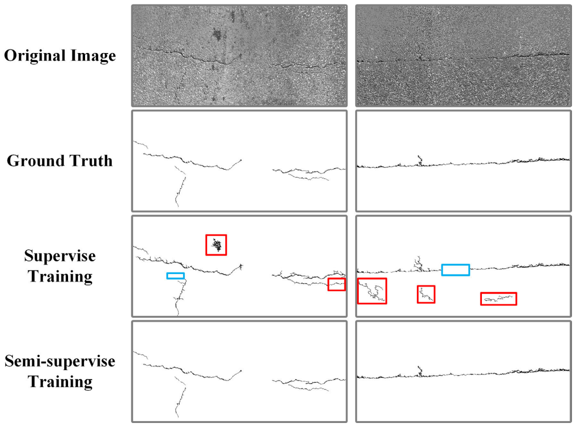


FIGURE 1. The first two lines of images are from the AigleRN dataset and its ground-truth segmentation mask. The third line is the prediction results obtained by using the supervised learning method and using all the images labeled to train the model. The last line is the prediction results obtained by training the model with 50% labeled images and 50% unlabeled images using the semi-supervised learning method.

Adversarial Networks (GANs) [7] has promoted the development of unsupervised and semi-supervised learning in the field of semantic segmentation. A general GAN is composed of two subnetworks which are optimized alternately during the training stage. A sample vector is an input to the generator, and output the predicted results at the pixel level of the corresponding sample, while the discriminator distinguishes the predicted results from the target sample. Then we train the generator to generate samples similar to the target distribution, to achieve the purpose of confusing the discriminator. We use the same idea to detect pavement cracks and replace the generator in the traditional GAN framework with a more efficient segmentation network. In this case, we train the segmentation network to make the output as close to the ground truth one-hot mapping as possible.

Our design is based on the observation of two limitations of supervised learning in the training process. The traditional method mainly has the following two problems, see the third line of Fig.1.

(1) The data of supervised learning training must be labeled, so that, the quality of the trained model depends on the number of marked datasets and the accuracy of the marks, and over-fitting may occur in some cases (like the red box in Fig.1).

(2) Traditional supervised learning is insensitivity to low-resolution pixel information, resulting in the inability to identify the crack area of the original image (like the blue box in Fig.1).

To solve these problems and reduce the inspector's workload of annotation for each training image and improve the accuracy of crack detection under low resolution, we apply an adversarial learning method and a full convolution discriminator which can learn to distinguish the real label from the prediction label of the segmentation network output. Inspired by [8], we use an adversarial loss encouraged segmentation network to generate the prediction probability maps, which is close to the ground truth.

The approach is related to the probability graphical model like Conditional Random Fields (CRFs) [9]–[11], but such method has no additional post-processing module in the course of the test stage. Furthermore, in the test phase, the discriminator is not needed during the detection of cracks, so the proposed model will not add any computational load in this phase. Based on the adversarial learning method, we make further efforts to explore the proposed model under the semi-supervised field. To realize semi-supervised learning, two specific loss functions are applied in our model, it will be detailed in Section III. First of all, the output of the discrimination network is used as the supervisory signal, which points out which pixels in the prediction result are close to the ground truth, then utilize the specific loss function and prediction results to train the segmentation network. Secondly, in the semi-supervised training of the network, we use the antagonistic loss function, which will make the model predict the picture close to the ground truth, thus increasing the discrimination difficulty of the discriminator.

The remaining of the paper is formed as follows: Section II provides a brief review of crack detection methods and our contribution; In Section III, we describe in detail our propose module, and provides a brief review of crack detection method; In Section IV based on CFD [12] dataset, AigleRN [13] dataset SDNET2018[14] dataset, we conducted a series of experiments with our model and other popular models, and presented the corresponding results and analysis; Lastly, Section V sum up the main work presented in this article.

II. RELATED WORK

In this section, we will briefly introduce the application of traditional methods, machine learning-based methods and deep learning-based methods for crack detection.

A. TRADITIONAL METHOD

Early studies like [15]–[17] find that the cracks are darker than the background in a pavement image; thus, the crack can be extracted by setting a threshold. However, the difficulty with this approach is how to select the appropriate threshold to fit most of the crack images. On the other hand, this method is sensitive to the illumination and noise of the image, which results in poor stability of the algorithm.

Edge detection algorithm [18]–[20] can improve the accuracy of crack detection compared with the threshold method when the contrast between the crack edge and the background is obvious. However, this method has limited effectiveness in detecting cracks in low contrast or noisy pavement images.

Gabor filters [21], [22], wavelet transform [23], [24] adopted manually designed feature descriptor's performance significant progress in detecting cracks with a single background image, but this method cannot fit for detect complex and diverse cracks. Besides, choosing suitable parameters is also commonly difficult and time-consuming.

The method in [25] use a combination of filtering, edge detectors, morphological operation and texture analysis to detection surface crack in concrete structures, the efficiency and accuracy of this method are significantly improved compared with other methods. However, the parameter selection of the filter in this method is very subjective, resulting in poor crack detection in complex scenes, and shadows, oil stains, paint with dark color are often misidentified.

B. MACHINE LEARNING

With the advancement of machine learning, the following methods focusing on feature extraction and pattern recognition have been successfully applied in crack detection: [26] use AdaBoost to select texture feature's descriptors, which extract by numerous linear and nonlinear filters, the texture features can describe crack images; [27] design crack features by considered the pavement surface as a textured surface, and then utilize the support vector machine (SVM) to distinguish crack from the non-crack; [28] can classify multiple spatial image features by the random forest method. However, the method is restricted to detecting learned crack images and finds it hard to detect new crack images.

CrackForest [12] applied a new descriptor by using random structured forests to describe cracks, which performance well to identify various complex cracks. These methods are very dependent on the quality and quantity of the manually designed features. Moreover, because of complicated pavement conditions, it is hard to design universal features effective for all pavements.

C. DEEP LEARNING

Lately, deep learning has made significant breakthroughs in the domain of computer vision. The accuracy of classification based on the convolutional neural network model has greatly exceeded the precision of traditional methods [29]–[32] and even that reach the human level [32]. References [33], [34] predict the location of cracks in pavement images based on deep learning object detection methods. References [35], [36] use a sliding window to divide the pavement images into smaller image blocks and utilizing CNN to predict whether the block contains cracks or not. In spite of these methods can exactly to locate the crack, the method can only detect patch level cracks without considering the pixel level. References [37], [38] utilize CNN to detect whether the pixels of the pavement image block appertain to the crack, which not only accomplishes pixel-level detection but also reached high accuracy. However block-based detection is time-consuming, and the small blocks cannot provide adequate context information for prediction.

References [39], [40] used the FCN network for crack detection and obtained high accuracy and detection speed. However, this method required substantial labeled pavement images, and the labeled image should be at the pixel-level (i.e. each pixel of training images must be annotated) assigning a label to each image pixel and needs a significant number of pixel-level annotated data, which is often expensive and time-consuming. Moreover, this method cannot leverage a large amount of available unlabeled pavement images.

Reference [41] proposed a new hybrid crack detector by combining the DCNN and the edge detector, which uses the output of deep learning models for crack detection using edge detector. This method can effectively reduce the noise rate of crack recognition and improve the recognition accuracy compared with other deep learning methods. However, because the number of layers of the convolutional network is too deep, there are many parameters to be recorded, which makes model training very time-consuming.

D. CONTRIBUTION

This study proposes a new model for pavement crack detection, which integrates the idea of full convolution network and adversarial learning. The contributions and novelties of our paper are summarized as follows:

(1) A semi-supervised learning framework is proposed that employs unlabeled pavement images in the model training, which greatly reduced the workload of manual annotation.

(2) Dense connection mode is added to our network structure, which requires fewer parameters than traditional crack detection algorithm.

(3) The proposed method improves the information flow and gradient in the network and makes the model easy to detect low-resolution pixel information.

Experiments were carried out on two public pavement crack datasets: CFD dataset, AigleRN dataset and SDNET2018 dataset. The distinct detection accuracy and training speed improvements on three datasets show that this method is superior to other advanced crack detection methods, and verifies the effectiveness of semi-supervised learning framework in pavement crack detection.

III. METHOD

The proposed architecture for the semi-supervised semantic segmentation using adversarial learning for pavement crack detection, as shown in Fig. 2. The model is composed of two parts: the segmentation network and the discriminator network. The part of segmentation can be an arbitrary network used for semantic segmentation, such as FCN [40], FC-DenseNet [42]. Take a pavement image with dimension of $H \times W \times 3$, as the network input, then get the image probability maps by the segmentation network, the size of the output is $H \times W \times C$, we set $C = 2$, which represents the number of semantic categories.

In the course of our experiment, the algorithm is described as four steps:

Step 1: We trained the FCN-based model as our discriminator network, which takes pavement image segmentation result or the ground truth label one-hot maps as the input, the one-hot map will be covered in detail in Section III. D, and then outputs the confidence maps of size $H \times W \times 1$, the value of p for each pixel in the confidence maps represents whether the pixel came from the segmentation network ($p = 0$) or the ground truth label ($p = 1$).

Step 2: Take a fully-convolutional discriminator network that can accept arbitrary sizes of input, compared with typical

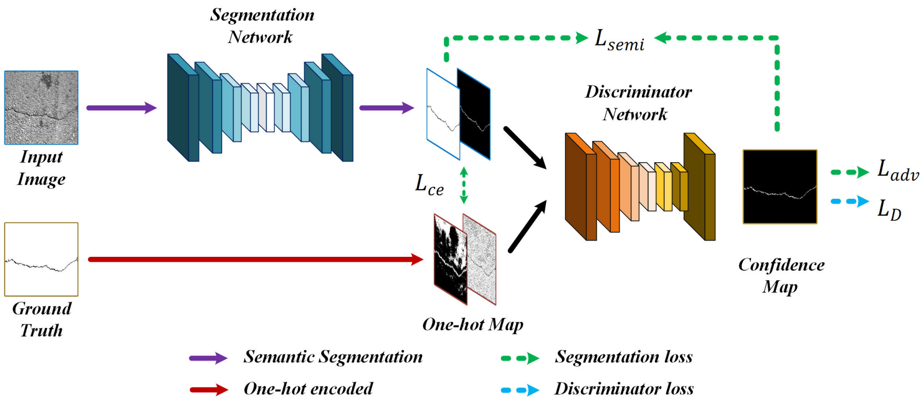


FIGURE 2. The architecture of the proposed system for pavement crack detection by semi-supervised semantic segmentation. Firstly, using the loss $\mathcal{L}_D$ to train the fully-convolution discriminator network; then, we use three losses functions to optimize the segmentation network during the training process: cross-entropy loss $\mathcal{L}_{ce}$ of segmentation output and one-hot map, fool the discriminator by adversarial loss $\mathcal{L}_{adv}$, and the semi-supervised loss $\mathcal{L}_{semi}$.

GAN discriminators network. Our framework extends the typical GAN which takes fixed-size input images and outputs a single probability value for pixel-level prediction and application in semantic segmentation. What's more important, we prove that this transformation is necessary for the proposed semi-supervised learning based pavement to the crack detection scheme.

Step 3: When training the model, labeled and unlabeled pavement images are used in the semi-supervised process. For the labeled pavement images, the cross-entropy loss $\mathcal{L}_{ce}$ and the adversarial loss $\mathcal{L}_{adv}$ is applied to train the segmentation network, this process of training segmentation network is supervised learning. What needs to be pointed out is that we train the discriminator network only make use of the labeled pavement images.

Step 4: For the unlabeled data pavement images, the semi-supervised method is proposed to train the segmentation network. We first obtain the segmentation prediction of the unlabeled pavement images from the segmentation network, and then we treat the prediction of segmentation network as the input of the discriminating network, we can acquire a confidence map from the output of this network. The output of the discriminant network is used as the supervisory signal, and we use cross-entropy loss $\mathcal{L}_{semi}$ to train the segmentation network. The output of the discriminant network reveals the quality of the segmentation results, for this reason, the segmentation network can trust the map during the training process.

In the following sections, we will describe more detail on how the proposed method implements pavement crack detection based on the semi-supervised learning. Firstly, we introduced the structure details of FC-DenseNet used in the segmentation network, such as the formulation of the densely connected convolutional network (DenseNet) [43]. Secondly, show the architecture of the segmentation network FC-DenseNet and discriminator network. Thirdly, the loss function used in our proposed method is introduced in detail. Finally, how the method can be realized to use the unlabeled and fake pavement images in a semi-supervised segmentation way is introduced, and explain how the typical GAN model is modified to apply in pavement crack detection based on semi-supervised learning.

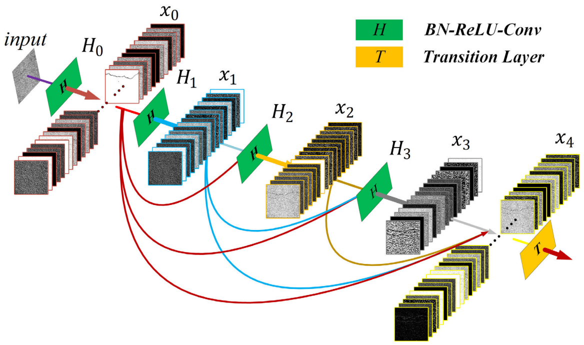


FIGURE 3. The architecture of denseblock.

A. DENSELY CONNECTED CONVOLUTIONAL NETWORKS

In order to further utilize the information flow of each layer, we adopt a different feedforward structure. This model is DenseNet, the idea is that each layer is directly connected to all subsequent layers in a feedforward manner so that each layer receives additional input from all the previous layers and passes its own property mapping to all the subsequent layers, thus maintaining the feedforward property on the model. Fig. 3 shows the schematic diagram of the DenseNet structure.

Compared with CNN-based, FCN-based methods, there is no need to relearn the redundant feature maps, this dense connection mode requires fewer parameters than traditional convolutional networks. In the traditional feedforward structure, the information flow is only passed between layers,

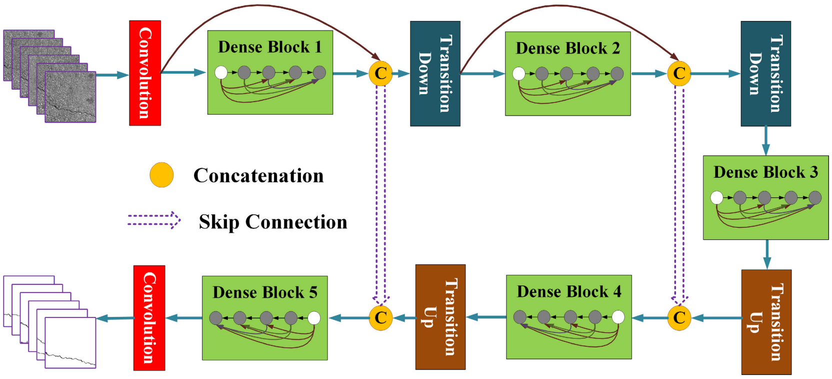


FIGURE 4. Fully convolutional DenseNets architecture.

each layer reads the state from its previous layer and writes to the next layer, changing the state while passing the information that needs to be retained. The DenseNet structure clearly distinguishes the information added to the network from the information retained, and keeps the remaining feature maps unchanged. The final classifier makes decisions according to all feature maps in the network. Another advantage of DenseNet is that it improves the flow of information and gradients across the network, which makes it easier to train, which helps train the network architecture at a deeper level. In addition, the dense connection has regularization effect, which can reduce the over-fitting of tasks with a small training set.

In order to enhance the information flow between layers, a different connectivity pattern is applied in our model: Denseblock introduces connections from the previous layer to subsequent layers. Fig. 3 illustrates the layout of the Denseblock schematically. Next, we will build a Denseblock using four steps:

Step 1: On the basis of the traditional convolutional neural network, the input to the ℓ^{th} layer of the DenseBlock is changed to the feature-maps from all the previous layers $[x_0, x_1, \ldots, x_{\ell-1}]$, the formula is expressed as follows:

$$x_\ell = H_\ell\left([x_0, x_1, \ldots, \mathrm{x}_{\ell-1}]\right) \tag{1}$$

in the formula, $[x_0, x_1, \ldots, x_{\ell-1}]$ represents the concatenation of the output feature maps in the $0, \ldots, \ell-1$ layer. In the process of experiment, the multiple inputs of $H_\ell()$ in Eq.(1) are concatenated into a single tensor.

Step 2: Deterministic the composite function $H_\ell()$, which plays an important role in the training process. In a Denseblock, this function contains three operations: batch normalization (BN) [44], rectified linear unit (ReLU) [45] and a 3×3 convolution (Conv).

Step 3: Since the concatenation operations applied in Eq.(1) require the same size of feature maps for each layer, In order to facilitate the downward sampling in the network, the network is divided into several densely connected blocks, shown in Fig. 3. We refer to layers between blocks as transition layers, which do convolution and pooling. In our experiments, the transition layers consist of a batch normalization layer, a 1×1 convolutional layer and 2×2 average pooling layer.

Step 4: Determine the hyper-parameter k. If each function H_ℓ produces k feature maps, it follows that the ℓ^{th} layer has $k_0 + \mathrm{k} \times (\ell - 1)$ input feature-maps, where k_0 is the number of channels in the input layer. The hyper-parameter k determines how much new information each layer adds to the global state. Once the network structure is determined, each layer can be applied everywhere within the network.

B. SEGMENTATION NETWORK

In view of DenseNets has the following characteristics: (1) parameter efficiency, (2) implicit deep supervision, (3) feature reuse, and it naturally generates skip join and multi-scale supervision. So, DenseNet is very suitable for pavement crack detection.

The DenseNet is extended to Fully Convolutional DenseNet(FC-DenseNet) by adding an up-sampling path to recover full input resolution, Fig. 4 shows the overview of the FC-DenseNet architecture. In order to revivification the input spatial resolution, FC-DenseNet implemented this action by transition up. Transition up modules can up-samples the former feature maps. By the skip connection, we can concatenate the up-sampled feature maps together to construct a new dense block (shown in Fig. 3). The last dense block in accordance with the same resolution and integrates all the information contained in all the previous dense blocks. So, we can use all the available feature maps at a given resolution to compute the dense blocks of the up-sampling path Fig. 4 expounds this idea in detail.

As a supplement, we introduced the detailed structure of transition down and transition up in Fig. 5. Denseblock layer is made up of batch normalization, ReLU, a 3×3 convolution layer, and a dropout layer. Transition down is made up of batch normalization, ReLU, a 1×1 convolution layer, a dropout layer and a max-pooling layer. Transition up is made up of a transposed convolution to recover the pooling operation.

C. DISCRIMINATOR NETWORK

The structure of the discriminator network is similar to FC-DenseNet, in the training process, it is different from

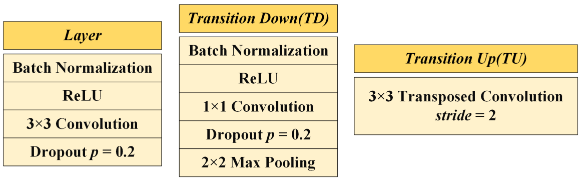


FIGURE 5. Building blocks of fully convolutional DenseNets. From left to right: layer used in the model, BN stands for Batch Normalization, TD stands for Transition Down, TU stands for Transition Up.

the segmentation network in the setting of hyper-parameters, where the growth rate k of DenseNet is set as 5, and each DenseBlock is followed by an activation function ReLU except the last block. In addition, because we set batch size as 5 in the training process, we didn't use any batch normalization layer in the discriminator network.

D. LOSS FUNCTION

Take an image X_n of size $H \times W \times 3$ as segmentation network input, we use S() to represent the segmentation network and denote the predicted result by $\mathrm{S}(X_n)$ of size $H \times W \times C$, where $\mathrm{C} = 2$ represents the number of categories (crack or non-crack). We use D() to represent the discriminator network which takes a probability map from segmentation result $\mathrm{S}(X_n)$ or one-hot encoded ground truth vector Y_n, and outputs a confidence map of size $H \times W \times 1$.

To train the discriminator network, we maximize the cross-entropy loss function $\mathcal{L}_D$ using:

$$\mathcal{L}_D = \sum_{h,w} (1 - y_n) \log \left(1 - D(S(X_n))^{(h,w)}\right) - y_n \log \left(D(Y_n)^{(h,w)}\right) \quad (2)$$

If the input takes from the segmentation network, we set $\mathrm{y}_n = 0$; if the input takes from the ground truth label, we set $\mathrm{y}_n = 1$. Moreover, $D(S(X_n))^{(h,w)}$ represents the confidence map of X at location (h,w). We obtain the ground truth one-hot map by one-hot encoding, then, the ground truth one-hot mapping of discrete labels is converted into C-channel probability mapping, if the pixel $X_n^{(h,w)}$ belongs to crack, $Y_n^{(h,w,c)}$ takes value 1 and 0 otherwise.

When training the segmentation network, we minimize a multi-task loss function to train the segmentation network, using:

$$\mathcal{L}_{\mathrm{seg}} = \mathcal{L}_{ce} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{\mathrm{semi}}\mathcal{L}_{\mathrm{semi}} \quad (3)$$

where $\mathcal{L}_{semi}$, $\mathcal{L}_{adv}$ and $\mathcal{L}_{ce}$ represent the semi-supervised loss, adversarial loss and cross-entropy loss respectively. In Eq.(3), λ_{adv} and λ_{semi} are two weights parameters.

We first discuss the situation of using labeled data, for an input pavement image X_n, Y_n and $\mathrm{S}(X_n)$ on behalf of its one-hot encoded ground truth and prediction result respectively, the cross-entropy loss function is obtained by:

$$\mathcal{L}_{ce} = -\sum_{h,w}\sum_{c \in C} Y_n^{(n,w,c)} \log \left(S(X_n)^{(h,w,c)}\right) \quad (4)$$

By receiving the output of the full convolution discriminator network D(), calculate the minimum loss function $\mathcal{L}_{adv}$ to train the counter network, the loss function is defined as follows:

$$\mathcal{L}_{adv} = -\sum_{h,w} \log \left(D(S(X_n))^{(h,w)}\right) \quad (5)$$

Based on the above function, the network training based semi-supervised learning is described as three steps:

Step 1: In Eq.(5), we train the segmentation network first and then trick the discriminator by maximizing the probability of producing the predicted results from the ground truth. For the unlabeled data pavement images, we consider the semi-supervised method to train the segmentation network. In the training process, since we have no ground truth annotation pavement image, we don't use the loss function $\mathcal{L}_{ce}$. We still apply the adversarial loss $\mathcal{L}_{adv}$, because it only acted on the discriminator network.

Step 2: In the self-taught learning network, we use unlabeled pavement images to train the discriminator network. We use the trained discriminator network to generate a confidence map $\mathrm{D}(\mathrm{S}(X_n))$, and then use the confidence map to infer the region close enough to ground truth distribution. In order to show the region of trust more clearly, we use threshold method to binary the confidence map.

Step 3: If $c^* = \mathrm{argmax}_c\, S(X_n)^{(h,w,c)}$, we set $\hat{Y}_n^{(h,w,c^*)} = 1$ for the one-hot encoded ground truth $\hat{Y}_n$. The semi-supervised loss function we use is defined as:

$$\mathcal{L}_{semi} = -\sum_{h,w}\sum_{c \in C} I\left(D(S(X_n))^{(h,w)} > T_{semi}\right) * \hat{Y}_n^{(h,w,c)} \log \left(S(X_n)^{(h,w,c)}\right) \quad (6)$$

In this function, $\mathrm{T}_{\mathrm{semi}}$ is a threshold used to control the self-learning process and I() are the indicator function. It's important to note that, when we train the model, we treat the value of I() and $\hat{Y}_n$ as constants, so, Eq.(6) can be regarded as a cross-entropy loss. During the experiment, we found that the learning effect of the model was better when $\mathrm{T}_{\mathrm{semi}}$ ranging between 0.15 and 0.26.

Finally, in order to explain the training process of our model more easily, we drew a schematic diagram as shown in Fig. 6, this diagram describes the optimization process of segmentation network and discrimination network loss function.

In more detail, when the parameters of the fixed segmentation network (G) as shown in Fig. 6(a) remains unchanged, the $\mathcal{L}_D$ value to the loss function on the network (D) is maximized according to Eq.(2). When D reaches the optimum, the parameters of fixed D remain unchanged and the global optimal solution is sought. Fig. 6(b) shows the optimization process of segmentation network G, and the value of $\mathcal{L}_{semi}$ is minimized according to Eq.(6) If and only if the value of model loss function L(G, D) reaches the point of P, the global optimal solution is achieved.

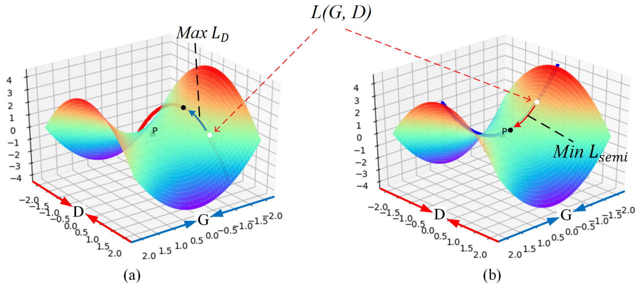


FIGURE 6. In the figure, L(G, D)represents the current loss value. The generator trains model parameters by maximizing L(G, D), and the discriminator trains model parameters by minimizing L(G, D), generator and discriminators adopt alternate optimized discriminator.

IV. EXPERIMENTS

A. IMPLEMENTATION DETAILS

We apply the Pytorch as the deep learning framework to implement our approach, and train the proposed model on an operating system of Windows 10, which has an Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz with 16GB memory and a single TitanX GPU with 12 GB memory. Stochastic Gradient Descent (SGD) optimization method is applied to train the segmentation network, we utilize the method mentioned in [10] to set the initial learning rate. For discriminator, we employ Adam optimizer [46] with the learning rate 10^{-4} and the same initial learning rate as the segmentation network. For the hyper-parameters that appear in this method, set λ_{adv} as 0.012 when training with labeled data, and set λ_{adv} as 0.005 when training with unlabeled data, and set λ_{semi} as 0.12 and T_{semi} as 0.22.

B. EVALUATION DATASETS AND METRIC

We assessment our model on three public crack datasets: CFD, AigleRN and SDNET2018. The evaluation metric applied in this paper is intersection-over-union (IoU). A definite number of categories c, predictive value o_i and the target y_i, the IoU is defined as:

$$\text{IoU}(c) = \frac{\sum_i (o_i == c \wedge y_i = c)}{\sum_i (o_i == c \vee y_i == c)} \tag{7}$$

where ∧ stands for logic and, ∨ stands for logic or.

In addition, we also considered the processing time in both training and testing phases as one of the performance metrics. In order to comprehensively evaluate the performance of each model, we introduced precision (Pr), recall (Re) and F1 score (F1), defined as:

$$\text{Pr} = \frac{TN}{TP + FP} \tag{8}$$

$$\text{Re} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{F1} = \frac{2 \times Pr \times Re}{Pr + Re} \tag{10}$$

where TP, FP , FN are the numbers of true positive, false positive and false negative respectively.

TABLE 1. Training and testing database.

Dataset	CFD		AigleRN	
Image dimension	256 × 256		256 × 256	
	Training	Testing	Training	Testing
Labeled images	6000	3000	5000	3000
Unlabeled images	12000	-	15000	-
Total images	18000	3000	20000	3000
Labeled : Unlabeled	1 : 2	-	1 : 3	-

In the CFD database, there are 118 three channels images with a resolution of 480 × 320 pixels. The images were taken by a mobile phone from the pavements of Beijing, China. These images contain a lot of interference factors, such as water stains, oil spots, shadow and other noise. The AigleRN database contains 269 gray-level images with a resolution of 991 × 462 or 311 × 462, which were collected by Aigle-RN system on French pavement. It consists of two parts: 68 images with pixel-level annotation and 201 images without pixel-level annotation. Each ground-truth is carefully annotated at the pixel level by professional engineers. We cropping the crack images into 256 × 256 image patches for the model training and testing, therefore, the training and test sets consist of cracked image patches with a resolution of 256 × 256, we embed this information in Table 1.

Compared with the CFD database, the pavement images in AigleRN contain more complex texture. For each dataset, we use labeled data with ratios of 1/8, 1/4, 1/2, 1 respectively, and we use the remaining unlabeled images as the training set.

C. ARCHITECTURE AND TRAINING DETAILS

All FC-DenseNet layers are summarized in Table 2, this architecture consists of 103 convolution layers: input layer as the first layer, the next 38 layers are the down-sampling path, then 15 layers of bottleneck and 38 in the up-sampling path. Five Transition Down (TD) and five Transition Up (TU) is applied in our model, extra convolution and transposed convolution are contained in TD, TU respectively. To provide a distribution of each class at each pixel, the final layer is a 1 × 1 convolution followed by a softmax layer.

In Table 2, Dense Block is represented by DB, Batch Normalization is represented by BN and c represents the number of categories. In our framework, we set c as 2 (crack or non-crack).

D. CFD DATASET

The CFD dataset was published in [12], and it is composed of 118 RGB images. This dataset can generally reflect the urban pavement surface condition. These images contain noises such as water stains, oil spots and shadows, these interference factors make crack detection very difficult.

1) IMPLEMENTATION DETAILS AND RESULTS

We adopt different ratios of labeled and unlabeled training sets to evaluate our approach 1/8, 1/4, 1/2, 1 represent the

TABLE 2. Architecture details of FC-DenseNet model used in our experiments.

Layer	Output size [H,W,D]
Input image	[256,256,3]
3 × 3 Convolution	[256,256,48]
DB1 (four layers) + TD	[128,128,112]
DB2 (five layers) + TD	[64,64,192]
DB3 (seven layers) + TD	[32,32,304]
DB4 (ten layers) + TD	[16,16,464]
DB5 (twelve layers) + TD	[8,8,656]
DB6 (fifteen layers)	[8,8,240]
TU + DB7 (fifteen layers)	[16,16,240]
TU + DB8 (twelve layers)	[32,32,192]
TU + DB9 (ten layers)	[64,64,160]
TU + DB10 (seven layers)	[128,128,112]
TU + DB11 (five layers)	[256,256,80]
1 × 1 Convolution	[256,256,c]
Prediction	[256,256]

TABLE 3. Crack detection results evaluation on CFD.

Semi-supervised training				
Method	Labeled Data			
	1/8	1/4	1/2	Full
FCN	55.3	68.8	72	79.7
Hybrid Crack Detector	56.2	65.3	70.6	82.3
FC-DenseNet	57.1	68.9	74.4	86.2
Ours	78.1	84.9	92.1	93.4
Supervised training				
FCN	63.3	71.8	78.0	85.9
Hybrid Crack Detector	65.1	72.9	79.2	87.1
FC-DenseNet	73.1	76.9	80.4	88.2
Ours	81.1	85.9	92.4	94.2

scale of the total training images in the dataset that are applied as labeled data, the rest of the image was applied without labels. The labeled images in the training set were randomly selected from the whole dataset, which ensured the objectivity and impartiality of the experimental results. We compare the proposed approach with FCN, FC-DenseNet and Hybrid Crack Detector [41] methods, the experimental results are shown in Table 3. When using labeled pavement image with ratios of 1/8, 1/4, 1/2, 1, the IoU of prediction reach values of 78.1%, 84.9%, 92.1%, 93.2%, respectively. Our network also performs robustly in supervised training, for which the values of IoU are 81.1%, 85.9%, 90.4%, 94.2%, respectively.

Since the images in the CFD dataset contain noises such as water stains, oil spots and shadows, which makes crack detection difficult. On the one hand, our model applied the DenseNet structure, which can maximize the information flow between layer and layer, it connects all layers directly with each other. Therefore, our method can well distinguished crack and non-crack pixels in the pavement image, even if there are many disturbing factors in the image, the crack and non-crack features can still be accurately distinguished, so, the experimental results were better than the method based on FCN. On the other hand, our proposed algorithm employs unlabeled pavement images in model training, which can generate additional supervisory signals to training model, so our method can detect low-resolution information, However, the method based on FC-DenseNet cannot use unlabeled data, which greatly limits the amount of training, therefore the experimental results were better than the method based on FC-DenseNet.

The result shown in Table 3 demonstrates that our approach outperforms the others. Fig. 7 shows the comparison of the proposed approach with FC-DenseNet, FCN and Hybrid Crack Detector. The following aspects are shown from left to right: original crack images, ground truth, prediction results from FC-DenseNet, Hybrid Crack Detector, FCN and our models. As seen from the figure, several wrong detections and missed detections occur in the images predicted by FCN, FC-DenseNet and Hybrid Crack Detector because of the small context of the block for detection. In contrast, our schema can analyze different types of cracks with sufficient contextual features, as seen in the fifth row of Fig. 7, indicating the robustness of our approach.

2) HYPER-PARAMETER ANALYSIS

The three adjustable hyper-parameters in the training process: λ_{adv} and λ_{semi} for equilibrate the learning process in Eq.(3), and T_{semi} is used to adjust the semi-supervised learning rate in Eq.(6). Table 4 shows the results of different hyper-parameters on the CFD dataset in the case of semi-supervised training.

As can be seen from the table, when the proportion of labeled pictures in the training set is 1/2, and set λ_{adv} as 0.015, the experimental results were observed by adjusting the values of λ_{semi} and T_{semi}, when we set T_{semi} as 0.23 and set λ_{semi} as 0.12, the detection precision obtain the best experimental results. Fig. 8 shows how IoU changes during the training process.

On the whole, the proposed model achieves the best IoU of 92.1%. When λ_{semi} is set as 0.12. Furthermore, we training the model with different values of T_{semi} by setting $\lambda_{adv} = 0.015$ and $\lambda_{semi} = 0.12$. With higher T_{semi}, it can be seen from the experimental results that the model only trusts the regions with the high similarity of structure, which can be used to generate supervisory signals. Fig. 9 shows image confidence maps from the discriminator.

According to the prediction results of the segmented network, the discriminant network generates a confidence map from this result, the white pixels indicate that they are closer to crack pixels. We use the white pixels as the monitoring signal of the training segmentation network. In some images, the identification of relatively fine cracks is still not very ideal

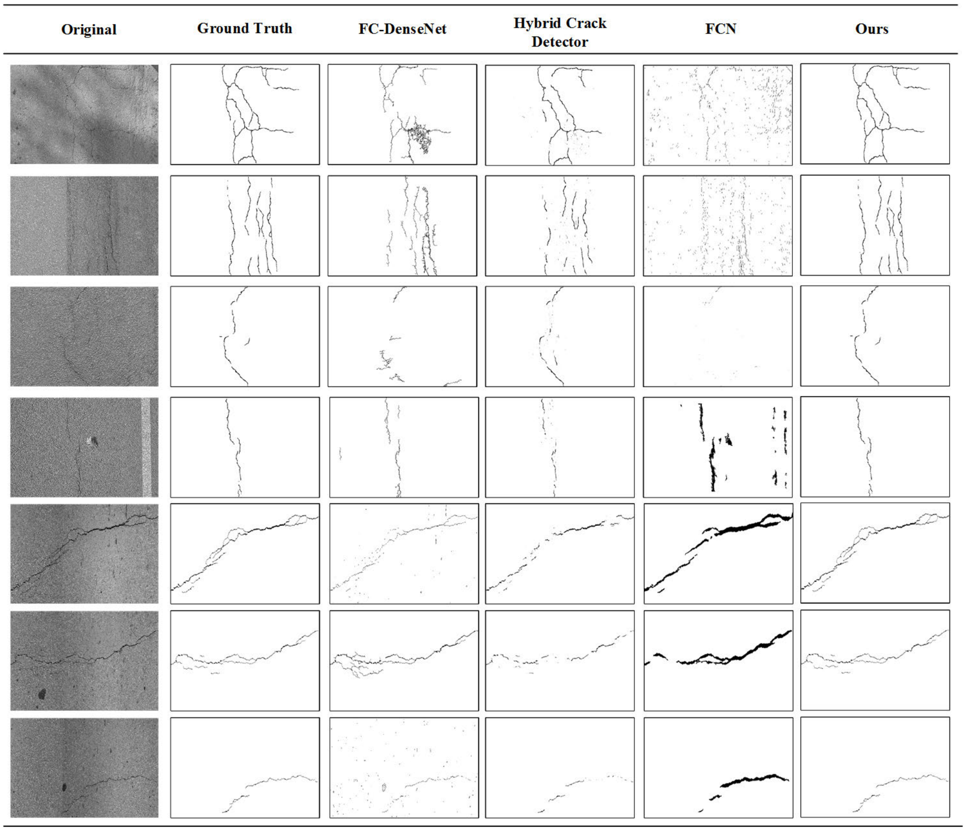


FIGURE 7. Results comparing on CFD.

TABLE 4. Hyper parameter analysis for CFD.

Data Amount	λ_{adv}	λ_{semi}	T_{semi}	IoU
1/2	0.015	0.06	0.23	90.4
1/2	0.015	0.12	0.23	92.1
1/2	0.015	0.24	0.23	91.1
1/2	0.015	0.12	0.11	90.3
1/2	0.015	0.12	0.22	92.1
1/2	0.015	0.12	0.33	91.1

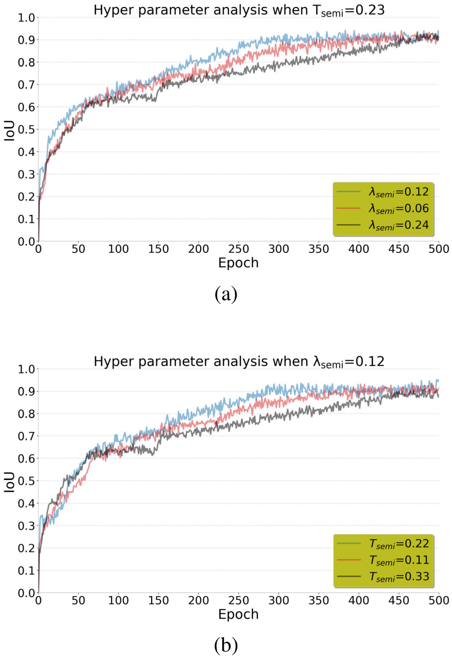
(In Fig. 9, we use the red box to identify the unrecognized area), which is what we need to further study and improve in the next step.

E. AigleRN DATASET

The AigleRN crack dataset consists of 269 grayscale pavement images, which contains four types of pavement cracks: alligator cracks, longitudinal cracks, transverse cracks and block cracks, and the conditions of each pavement image in AigleRN contain more complex texture. In addition, the image contains cracks of different length and width, which makes crack detection difficult.

1) IMPLEMENTATION DETAILS AND RESULT

As the image resolution is relatively large, we crop the image into several 256 × 256 patches with a given step length and send them to the network respectively. For the cropped image, we adopt different ratios of labeled and unlabeled



(a)

(b)

FIGURE 8. Visualization of the training process, when set $T_{semi} = 0.23$, $\lambda_{semi} = 0.06, 0.12, 0.24$ (a) and $\lambda_{semi} = 0.12$, $T_{semi} = 0.11, 0.22, 0.33$ (b) the IoU value in the training process for CFD dataset.

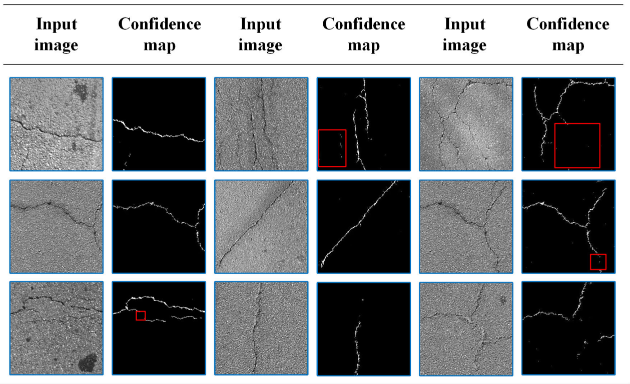


FIGURE 9. Visualization of the confidence maps.

TABLE 5. Crack detection results evaluation on AigleRN.

Semi-supervised training				
Method	Labeled Data			
	1/8	1/4	1/2	Full
FCN	46.3	68.2	74.0	83.7
Hybrid Crack Detector	49.6	70.5	75.3	87.9
FC-DenseNet	53.1	72.9	79.4	89.2
Ours	80.2	85.9	91.4	92.8
Supervised training				
FCN	55.3	69.8	76.0	84.9
Hybrid Crack Detector	60.2	71.5	79.2	88.6
FCN-DenseNet	65.1	74.9	81.4	90.1
Ours	82.1	87.4	91.5	93.2

training sets to evaluate our approach. 1/8, 1/4, 1/2, 1 represent the scale of the total training images in the dataset that are applied as labeled data respectively, the rest of the image was applied without labels. The labeled images in the training set were randomly selected from the whole dataset. We compare the proposed approach with FCN, FC-DenseNet methods and Hybrid Crack Detector, the evaluation results of the four types of cracks on the AigleRN dataset are presented in Table 5. When using labeled pavement image with ratios of 1/8,1/4,1/2,1, the IoU of prediction reach values of 80.2%, 85.9%, 91.4%, 92.8%, respectively. Our network also performs robustly in supervised training, for which the values of IoU are 82.1%, 87.4%, 91.5%, 93.2%, respectively.

Although the images in the AigleRN dataset contain different types of cracks, On the one hand, the DenseNet structure exploits feature reuse enhances the potential of the network to identify different types of features, yielding crack detection model is easy to train and suitable for complex types. Therefore, our method can well distinguish the different types of pavement cracks. On the other hand, because AigleRN dataset contains a large number of unlabeled images, our model can make full use of this unmarked data, the other two methods cannot use this unlabeled data to train the model, so our model can learn more unmarked features that other methods can't, and the experimental results were better than the method based on FCN, FC-DenseNet and Hybrid Crack Detector.

As described in Section I, the traditional method is based on supervised learning in the training process, the training dataset must be labeled, so that, the quality of the trained model depends on the number of marked datasets and the accuracy of the marks. However, the annotation image in dataset AigleRN is limited, and most of them are unlabeled data, Therefore, the traditional supervised learning method is not adapted to the dataset with a few annotations. As Fig. 10 shows, Compared with our proposed semi-supervised learning model, we observed a significant improvement in the segmentation boundary compared to the baseline model.

2) HYPER-PARAMETER ANALYSIS

Similar to the experimental process CFD dataset, when using 1/2 annotated images, fixed the parameters λ_{adv} as 0.02. Then change the values of λ_{semi} and T_{semi} for comparisons. As can be seen from Table 6, the test accuracy reaches the highest when we set T_{semi} as 0.2. Then we set λ_{semi} as 0.1 and the comparison test is performed by changing the value of T_{semi}, it can be inferred from the experimental results that the testing effect is best when λ_{semi} is 0.1 and T_{semi} is 0.2.

As can be seen from the table, when the proportion of labeled pictures in the training set is 1/2, and set λ_{adv} as 0.02, the experimental results were observed by adjusting the values of λ_{semi} and T_{semi}, when we set T_{semi} as 0.2 and set λ_{semi} as 0.1, the detection precision obtain the best experimental results. We plotted the IoU change process as shown in Fig.11.

As can be seen from the figure, the different values of the hyper-parameters only affect the training process, such as the speed of convergence, but ultimately reach the optimal accuracy, indicating the robustness of the model. Because we introduce the idea of adversarial learning into model training, the segmentation network and the discrimination network interfere with each other in the training process, which leads to the discrimination network unable to distinguish whether its input is the real label or the output of the segmentation network. Therefore, when the training epoch is enough, the influence of subtle changes of different hyper-parameters on the training results is within the controllable range. At the same time, verifies the stability of the confrontation framework.

F. SDNET2018 DATASET

In order to test the performance of our model, we use the trained model to verify on the SDNET2018 dataset, the pavement images were acquired from the roads and sidewalks on USU campus using a 16 MP Nikon digital camera. With a resolution of 256 × 256, these images contain a variety of obstructions, including shadows, scaling, edges, holes, and background debris, the interference factor make crack detection very difficult.

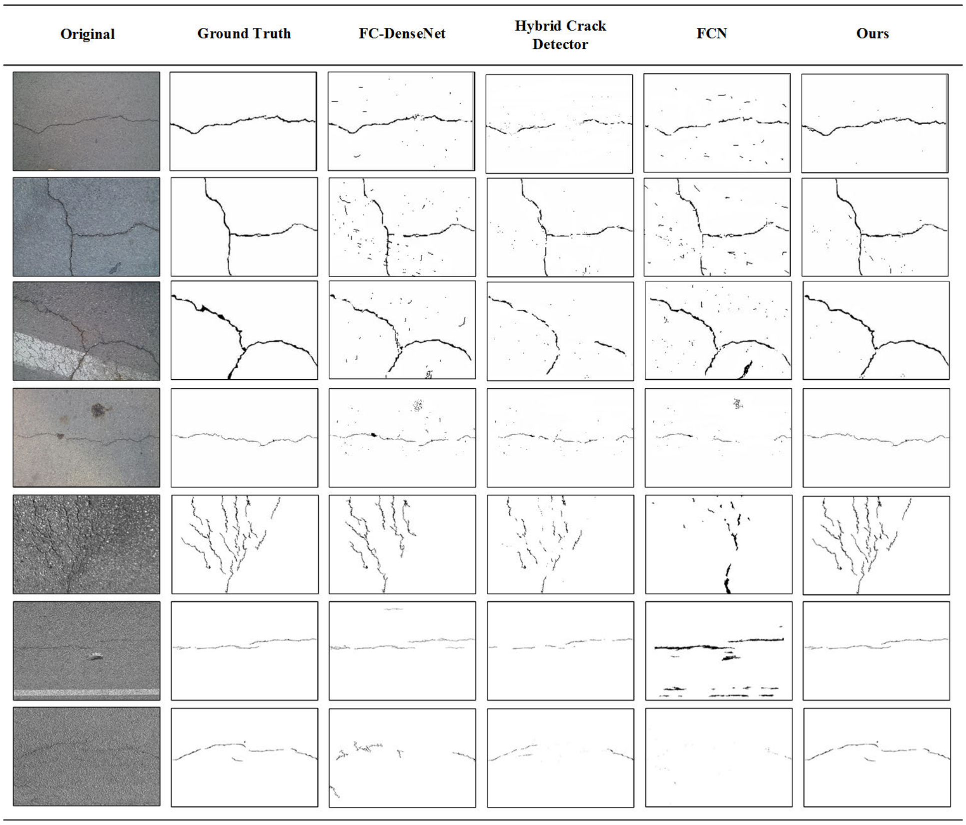


FIGURE 10. Results comparing on AigleRN.

TABLE 6. Hyper parameter analysis for AigleRN.

Data Amount	λ_{adv}	λ_{semi}	T_{semi}	IoU
1/2	0.02	0.05	0.2	90.2
1/2	0.02	0.1	0.2	91.4
1/2	0.02	0.15	0.2	91.3
1/2	0.02	0.1	0.15	90.8
1/2	0.02	0.1	0.2	91.5
1/2	0.02	0.1	0.25	91.1

Since SDNET2018 does not have pixel-level annotation for crack images, we cannot accurately obtain the detection accuracy, but we can intuitively evaluate the quality of a model by displaying the experimental results. We have selected a few typical images in SDNET2018, including images in complex environments such as shadows and water stains, Fig. 12 shows the detection results of each model.

From the figure, we can see that the model of FCN, FC-DenseNet and Hybrid Crack Detector are not sensitive to crack pixels under more complex backgrounds, the obvious cracks in the image cannot be detected, and the environmental disturbance is wrongly identified as cracks. In contrast, our model can better adapt to crack detection in complex environments, and can accurately detect crack regions in different environments. However, as can be seen from the figure, our model also has some limitations, the detection results of fine cracks in complex environments are not very ideal. Of course, other models also have such problems, which will also be the focus of our future research.

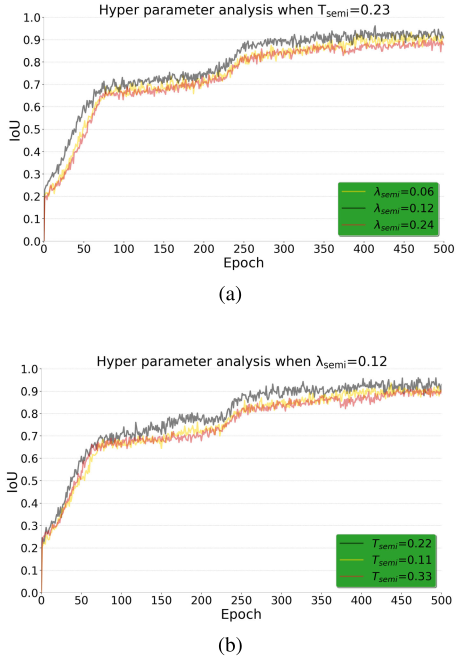


FIGURE 11. Visualization of the training process, when set $T_{semi} = 0.23$, $\lambda_{semi} = 0.06, 0.12, 0.24$ (a) and $\lambda_{semi} = 0.12$, $T_{semi} = 0.11, 0.22, 0.33$ (b) the IoU value in the training process for AigleRN dataset.

TABLE 7. Experimental results of various evaluation metrics on CFD.

Method	Evaluation metric					
	Pr	Re	F1	Parameter(M)	Training time/per image(s)	Testing time/per image(s)
FCN	0.7688	0.6812	0.6817	454.2	4.3	2.6
FC-DenseNet	0.8263	0.8410	0.8195	40.3	2.2	1.9
Hybrid Crack Detector	0.7466	0.9514	0.8318	320.5	3.2	2.4
Ours	0.9591	0.9663	0.9627	37.4	1.8	1.2

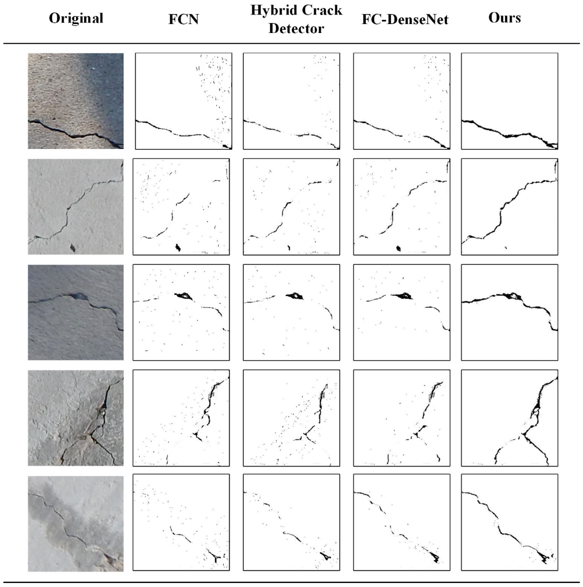


FIGURE 12. Results comparing on SDNET2018.

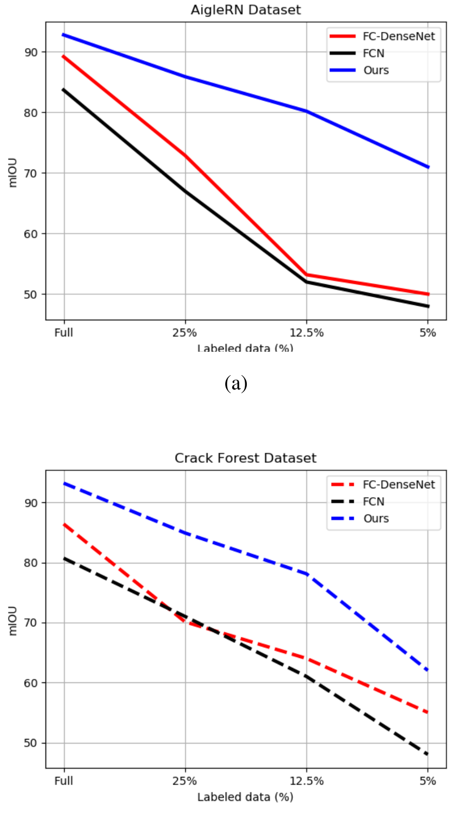
G. EXTENDS EXPERIMENT

In order to compare the comprehensive performance of each model, we used multiple evaluation metrics to analyze each model in this section, including precision, recall rate, F1 score, parameter size, training and testing time per image. We tested the above metrics on the CFD dataset, and the experimental results are shown in Table 7.

As can be seen from the results in the table, the model parameters of Hybrid Crack Detector and FCN are large, moreover, the training and testing of the model are time-consuming and the detection accuracy is not well. Although the model parameters of FC-DenseNet decreased significantly compared with the two methods, the detection accuracy still needs to be improved. Our method not only has fewer parameters, shorter training and testing time, but also has the best accuracy and recall rate compared with the current methods.

Finally, we validate the performance of our model on public CFD datasets and AigleRN dataset which include various crack types under diverse gathering conditions. Compared with existing methods, we get the best results and apply the new technology of semi-supervised semantic segmentation to crack detection.

Experimental results show that our method not only can detect different types of cracks, but also be particularly effective when only a few labeled are available: with as little as 50% labeled data, we report a great performance improvement over the state of the art (see Fig. 13). This result further shows that the method can easily take advantage of additional unannotated pavement images when these are available. It compares favorably to the existing methods operating in the same setting.



(a)

(b)

FIGURE 13. Semi-supervised Semantic Segmentation: The proposed semi-supervised learning approach improves over the baselines even when only little labeled data is available using unlabeled data, shows considerable improvement, especially with less than 5% labeled samples. Performance is shown on the AigleRN dataset (a) and CFD dataset (b).

V. CONCLUSION

In this study, an adversarial learning architecture of Semi-Supervised is proposed and applied for pavement crack detection. The following conclusions are obtained from this study:

(1) Validate our model on two challenging public datasets: CFD datasets and AigleRN dataset covering all kinds of

pavement cracks. When using labeled pavement image with ratios of 1/8, 1/4, 1/2, 1, for CFD datasets, the IoU of prediction still reach values of 78.1%, 84.9%, 92.1%, 93.2%, respectively; for AigleRN dataset, the IoU of prediction reach values of 80.2%, 85.9%, 91.4%, 92.8%, respectively.

(2) The proposed algorithm can train the model with unlabeled images, which greatly reduces the dependence on the number of labeled images and reduces the burden of manual annotation. The output of the full convolution discriminator is used as a supervisory signal, which makes up for the absence of image annotation and realizes semi-supervised learning. The significant performance and speed improvements of all datasets show that this method is superior to other most advanced crack detection methods.

(3) Adversarial framework is proposed that can improve the precision of crack detection without extra calculation load, and further improves the accuracy of pavement crack recognition by adding the unannotated pavement images

REFERENCES

[1] S. Hong, H. Noh, and B. Han, "Decoupled deep neural network for semi-supervised semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 1495–1503.

[2] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1742–1750, doi: 10.1109/ICCV.2015.203.

[3] D. Pathak, P. Krahenbuhl, and T. Darrell, "Constrained convolutional neural networks for weakly supervised segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1796–1804, doi: 10.1109/ICCV.2015.209.

[4] P. O. Pinheiro and R. Collobert, "Weakly supervised semantic segmentation with convolutional networks," in *Proc. CVPR*, 2015, vol. 2, no. 5, p. 6.

[5] X. Qi, Z. Liu, J. Shi, H. Zhao, and J. Jia, "Augmented feedback in semantic segmentation under image level supervision," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 90–105, doi: 10.1007/978-3-319-46484-8_6.

[6] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei, "What's the point: Semantic segmentation with point supervision," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 549–565.

[7] I. Goodfellow, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[8] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," 2018, *arXiv:1802.07934*. [Online]. Available: http://arxiv.org/abs/1802.07934

[9] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Apr. 2018, doi: 10.1109/TPAMI.2017.2699184.

[10] G. Lin, C. Shen, A. V. D. Hengel, and I. Reid, "Efficient piecewise training of deep structured models for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 3194–3203, doi: 10.1109/CVPR.2016.348.

[11] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. S. Torr, "Conditional random fields as recurrent neural networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1529–1537, doi: 10.1109/ICCV.2015.179.

[12] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3434–3445, Dec. 2016, doi: 10.1109/TITS.2016.2552248.

[13] S. Chambon and J.-M. Moliard, "Automatic road pavement assessment with image processing: Review and comparison," *Int. J. Geophys.*, vol. 2011, pp. 1–20, Oct. 2011, doi: 10.1155/2011/989354.

[14] S. Dorafshan, R. J. Thomas, and M. Maguire, "SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks," *Data Brief*, vol. 21, pp. 1664–1668, Dec. 2018, doi: 10.1016/j.dib.2018.11.015.

[15] J. Tang and Y. Gu, "Automatic crack detection and segmentation using a hybrid algorithm for road distress analysis," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2013, pp. 3026–3030, doi: 10.1109/SMC.2013.516.

[16] Q. Li and X. Liu, "Novel approach to pavement image segmentation based on neighboring difference histogram method," in *Proc. Congr. Image Signal Process.*, 2008, pp. 792–796, doi: 10.1109/CISP.2008.13.

[17] H. Oliveira and P. L. Correia, "Automatic road crack segmentation using entropy and image dynamic thresholding," in *Proc. 17th Eur. Signal Process. Conf.*, Aug. 2009, pp. 622–626.

[18] H. Zhao, G. Qin, and X. Wang, "Improvement of canny algorithm based on pavement edge detection," in *Proc. 3rd Int. Congr. Image Signal Process.*, Oct. 2010, pp. 964–967, doi: 10.1109/CISP.2010.5646923.

[19] R. Salim Lim, H. M. La, Z. Shan, and W. Sheng, "Developing a crack inspection robot for bridge maintenance," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 6288–6293, doi: 10.1109/ICRA.2011.5980131.

[20] R. S. Lim, H. M. La, and W. Sheng, "A robotic crack inspection and mapping system for bridge deck maintenance," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 367–378, Apr. 2014, doi: 10.1109/TASE.2013.2294687.

[21] S. Chanda, "Automatic bridge crack detection—A texture analysis-based approach," in *Proc. IAPR Workshop Artif. Neural Netw. Pattern Recognit.*, 2014, pp. 193–203: Springer, doi: 10.1007/978-3-319-11656-3_18.

[22] R. Medina, J. Llamas, E. Zalama, and J. Gómez-García-Bermejo, "Enhanced automatic detection of road surface cracks by combining 2D/3D image processing techniques," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 778–782.

[23] P. Subirats, J. Dumoulin, V. Legeay, and D. Barba, "Automation of pavement surface crack detection using the continuous wavelet transform," in *Proc. Int. Conf. Image Process.*, Oct. 2006, pp. 3037–3040, doi: 10.1109/ICIP.2006.313007.

[24] J. Zhou, "Wavelet-based pavement distress detection and evaluation," *Opt. Eng.*, vol. 45, no. 2, Feb. 2006, Art. no. 027007, doi: 10.1117/1.2172917.

[25] S. Dorafshan, M. Maguire, and X. Qi, "Automatic surface crack detection in concrete structures using OTSU thresholding and morphological operations," Tech. Rep., 2016, doi: 10.13140/RG.2.2.34024.47363.

[26] A. Cord and S. Chambon, "Automatic road defect detection by textural pattern recognition based on AdaBoost," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 27, no. 4, pp. 244–259, Apr. 2012, doi: 10.1111/j.1467-8667.2011.00736.x.

[27] Y. Hu, C.-X. Zhao, and H.-N. Wang, "Automatic pavement crack detection using texture and shape descriptors," *IETE Tech. Rev.*, vol. 27, no. 5, pp. 398–405, 2010, doi: 10.4103/0256-4602.62225.

[28] P. Prasanna, K. J. Dana, N. Gucunski, B. B. Basily, H. M. La, R. S. Lim, and H. Parvardeh, "Automated crack detection on concrete bridges," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 591–599, Apr. 2016, doi: 10.1109/TASE.2014.2354314.

[29] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105, doi: 10.1145/3065386.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.

[33] J.-H. Lee, S.-S. Yoon, I.-H. Kim, and H.-J. Jung, "Diagnosis of crack damage on structures based on image processing techniques and R-CNN using unmanned aerial vehicle (UAV)," *Proc. SPIE Sensors Smart Struct. Technol. Civil, Mech., Aerosp. Syst.*, vol. 10598, Mar. 2018, Art. no. 1059811, doi: 10.1117/12.2296691.

[34] X. Wang and Z. Hu, "Grid-based pavement crack analysis using deep learning," in *Proc. 4th Int. Conf. Transp. Inf. Saf. (ICTIS)*, Aug. 2017, pp. 917–924, doi: 10.1109/ICTIS.2017.8047878.

[35] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep learning-based crack damage detection using convolutional neural networks," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017, doi: 10.1111/mice.12263.

[36] K.-Y. Jang, B. Kim, S. Cho, and Y.-K. An, "Deep learning-based concrete crack detection using hybrid images," *Proc. SPIE Sensors Smart Struct. Technol. Civil, Mech., Aerosp. Syst.*, vol. 10598, Mar. 2018, Art. no. 1059812, doi: 10.1117/12.2294959.

[37] Z. Fan, Y. Wu, J. Lu, and W. Li, "Automatic pavement crack detection based on structured prediction with the convolutional neural network," 2018, *arXiv:1802.02208*. [Online]. Available: http://arxiv.org/abs/1802.02208

[38] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 12, pp. 1090–1109, Dec. 2018, doi: 10.1111/mice.12412.

[39] H.-W. Huang, Q.-T. Li, and D.-M. Zhang, "Deep learning based image recognition for crack and leakage defects of metro shield tunnel," *Tunnelling Underground Space Technol.*, vol. 77, pp. 166–176, Jul. 2018, doi: 10.1016/j.tust.2018.04.002.

[40] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440, doi: 10.1109/TPAMI.2016.2572683.

[41] S. Dorafshan, R. J. Thomas, and M. Maguire, "Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete," *Construct. Building Mater.*, vol. 186, pp. 1031–1045, 2018, doi: 10.1016/j.conbuildmat.2018.08.011.

[42] S. Jegou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 11–19, doi: 10.1109/CVPRW.2017.156.

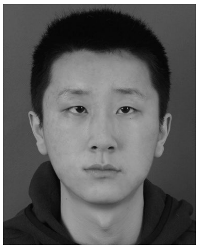
[43] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.

[44] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*. [Online]. Available: http://arxiv.org/abs/1502.03167

[45] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.

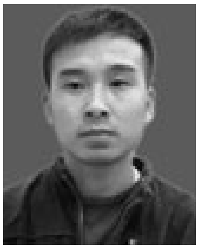
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

GANG LI received the Ph.D. degree in traffic information engineering and control from Chang'an University. He was a Postdoctoral Station of civil engineering with Chang'an University, and a Visiting Scholar with Tulane University, USA. He is currently working with the Department of Automation, School of Electronics and Control Engineering, Chang'an University. He has published more than ten articles in journals and academic conferences at home and abroad, including more than six articles in SCI/EI. He has presided over and participated in more than ten vertical and horizontal scientific research projects. His main research areas and directions are image processing and machine vision, traffic information engineering and control, machine learning and pattern recognition, and crack detection academic achievements.



JIAN WAN received the master's degree in electronic and control engineering from Chang'an University. His main research areas are deep learning, pattern recognition, and image processing. He has participated scientific research projects include the Natural Science Foundation of Shaanxi Provincial Department of Science and Technology, Identification and Positioning of MultiSensor Fusion Bridge Substructure Cracks, and the Industry-University-Research Cooperation Promotion Project.

SHUANHAI HE is currently pursuing the Doctor of Engineering (D.Eng.) degree with Chang'an University. He is also a second-level Professor, a Doctoral Tutor, and an Assistant to the President of Chang'an University. His research direction is bridge structure theory, bridge structure safety evaluation, bridge reinforcement theory and method, and so on.



QIANGWEI LIU received the master's degree in electronic and control engineering from Chang'an University, His research interests are machine learning and machine vision. The scientific research projects I have participated in and completed include the Natural Science Foundation Project of Shaanxi Science and Technology Department, Identification and Positioning of MultiSensor Fusion Bridge Substructure Cracks, and the Industry-University-Research Cooperation Promotion Project.

BIAO MA received the master's degree in electronic and control engineering from Chang'an University, His research interests are machine learning and machine vision. The scientific research projects I have participated in and completed include the Natural Science Foundation Project of Shaanxi Science and Technology Department, Identification and Positioning of MultiSensor Fusion Bridge Substructure Cracks, and the Industry-University-Research Cooperation Promotion Project.