IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received April 16, 2019, accepted May 3, 2019, date of publication May 13, 2019, date of current version May 30, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2916330

Real-Time Tunnel Crack Analysis System via Deep Learning

QING SONG[1], YINGQI WU[1], XUESHI XIN[1], LU YANG[1], MIN YANG[1], HONGMING CHEN[1], CHUN LIU[1], MENGJIE HU[1], XUESONG CHAI[2], AND JIANCHAO LI[2]

[1]Pattern Recognition and Intelligent Vision Laboratory, Beijing University of Posts and Telecommunications, Beijing 100089, China
[2]China Academy of Railway Sciences, Beijing 100081, China

Corresponding author: Qing Song (priv@bupt.edu.cn)

This work was supported in part by the Beijing University of Posts and Telecommunications, and in part by the China Academy of Railway Sciences.

ABSTRACT Cracks in the tunnel become an unavoidable problem in tunnel construction and tunnel using. Cracks will affect the stability of the tunnel and have a negative impact on the operation of the train. It is a crucial part of rail safety as well as rail defects and train defects. Therefore, cracks in the tunnel must be identified and repaired in time. At present, the detection of tunnel cracks in the domestic railways relies on the manual inspection mainly. It is difficult to satisfy the requirements of the rapidity and the accuracy of railway inspection by the manual inspection due to the subjective judgment of the inspection personnel. At the same time, tunnel images have some complex situations such as water stains, scratches, structural seams, uneven illumination, and a lot of noise, which have brought bottlenecks to the development of traditional image processing methods. It is necessary to adopt more effective methods to detect the tunnel cracks in time. This paper builds the first tunnel crack dataset with semantic segmentation annotation and proposes an objective and fast tunnel crack identification algorithm using semantic segmentation in computer vision to construct a complete tunnel crack identification and analysis system. The system applies advanced semantic segmentation to the railway tunnel image analysis to achieve precise segmentation of tunnel crack locations, thereby saving the railway department a lot of manpower and material resources and improving efficiency.

INDEX TERMS Deep learning, semantic segmentation, tunnel crack analysis system.

I. INTRODUCTION

Rail transport is a main way of transportation in society. It is important to guarantee the safety of railway to keep the personal security and property security. Repairing the rail defects is an urgent demand in railway safety including rail defects, train defects, and tunnel cracks. Our system focuses on the tunnel cracks mainly. Cracks in tunnels become an unavoidable problem which must be identified in time [26]. But the current detection methods are basically some methods with low efficiency and few technical skills. The traditional tunnel crack detection is mainly realized by manual which means the technician builds scaffolding on site [4], and then the technician observes the crack through the human vision or takes a photo of the whole tunnel for looking for cracks in the photo. It is very inefficient and labor intensive, and the accuracy of the inspection depends largely on the experience and quality of the technician. On the other hand, due to a large amount of data, the method of manual identification loses timeliness and costs a lot of money.

With the development of computer science and digital image processing technology, the use of image processing to detect cracks has attracted more and more attention. It has the advantages of non-contact, high efficiency, convenience, and directness, and has gradually become the main direction of research and obtained a large number of research results [26]. However, tunnel images have some complex situations, such as water stains, pollution, structural seams, uneven illumination, numerous noises, and irregular distribution, which have brought bottlenecks to the development of traditional image processing methods. Due to the disturbance of these situations, traditional image processing methods cannot get a good performance to solve the problems.

In recent years, the rapid development of artificial intelligence and deep learning has brought revolutionary development in computer vision. Object image classification [9], [14], [20], [22], [32], [33] and segmentation [5], [10], [11], [13] are very active tasks in computer vision. Given a picture,

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

the object classification is to estimate the type of object in the picture. The object segmentation can be approximated as an upgraded version of the classification problem, which is a pixel-level classification. For each picture, each pixel is discriminated and classified, and finally, the region of the target object is segmented, and attributes such as position information and object shape can be acquired. Therefore, this paper aims to propose an objective and fast tunnel crack identification algorithm using semantic segmentation in computer vision to construct a complete tunnel crack identification and analysis system.

In this paper, we propose a tunnel crack identification and analysis system to do the segmentation of tunnel cracks including image acquisition system, vehicle-control system, and crack identification and management system. The research significance of the tunnel crack identification and analysis system using deep learning technology is mainly reflected in its practical use value. At present, in tunnel image crack detection, the artificial method is obviously not suitable. The development of recognition technology based on image processing in the non-destructive testing method has inevitable defects, and a new fast and effective algorithm is demanded. The application of image semantic segmentation technology based on deep learning to tunnel crack identification can quickly realize crack recognition and improve the recognition rate. At the same time, it can also obtain complete information about the crack, including its location, shape, size and other attributes, so that staffs can do the relevant analysis and processing in time. Through this method, the efficiency of railway maintenance related work will be greatly improved, the cost will be reduced, and the automation of related work will be realized, which has great application value and significance. Our main contributions can be summarized into four-fold:

- We build the first tunnel cracks datasets with semantic segmentation annotation including single-annotation datasets for cracks, named *tunnel crack datasets* and multi-annotation datasets for cracks, structural seams, water stains, and scratches, named *augmented tunnel crack datasets*.
- We propose a complete system to do the tunnel crack identification and analysis including image acquisition system, vehicle-control system, and crack identification and management system.
- We introduce semantic segmentation into the tunnel crack identification via deep learning and get a good performance for the follow-up repair measures.
- We introduce a lightweight model for semantic segmentation to speed up for the real-time identification which does not lose the accuracy but gets the fast speed for the system.

II. RELATED WORK

At present, the widely used crack detection methods include ultrasonic detection method, optical fiber sensing detection method, acoustic emission detection method and image processing detection method in addition to manual methods [27]. Among them, the image detection method has become the main direction of research with the development of computer science and digital image processing technology and has obtained some achievements. In recent years, through the joint efforts of many domestic and overseas researchers, digital image processing and detection technology has obtained some achievements in bridge crack detection and pavement crack detection [29].

In China, L and H *et al.* [16] proposed an image segmentation algorithm based on the Sobel operator and maximum entropy method. C [6] proposed a crack feature extraction algorithm based on a gray image and its texture characteristics. Wang, Feng *et al.* proposed a multi-image and multi-resolution pavement crack detection method, which uses image fusion technology, and where the multi-scale method preserves the collection characteristics of the image and greatly improving reliability and accuracy of crack detection [25]. Dong *et al.* used the similarity of the gray value to extract cracks [7]. Chen *et al.* mapped the two-dimensional image into the three-dimensional space according to the gray value and detected the crack by searching for the "*ditch*" in the three-dimensional image [30]. Xu *et al.* detected the cracks on the pavement according to the saliency of the image [8].

In foreign countries, crack detection algorithms based on network analysis and based on the minimum path crack detection algorithm have proposed. Such methods are not suitable for processing images contaminated by severe noise. Yusuke *et al.* proposed a two-step processing algorithm that effectively removes noise caused by uneven illumination, shadows, and stains in the image [28]. Later, a crack classification algorithm using statistical classification of morphological processing and logistic regression was proposed. The accuracy of crack extraction was over 80%, but the algorithm would miss some small cracks and the calculation amount was large. Haris *et al.* in the United States evaluated the effectiveness of using various threshold segmentation algorithms to separate lesions from pavement fracture images [21]. Cuhadar *et al.* [1] developed an automatic segmentation algorithm, which applied the wavelet algorithm to the segmentation recognition of pavement crack images.

III. SYSTEM IMPLEMENTATION

Based on the analysis of domestic and foreign researches, it can be found that the existing tunnel crack detection systems generally have the following problems [1]:

- In view of the selection, adjustment and illumination of the image sensor, the image has many bad states, such as uneven illumination, serious background noise, unclear image, severe local exposure, and tailing.
- The image acquisition speed cannot meet the demand. At present, the speed of the commercial detection system that has appeared in a foreign country is generally 5-20km/h. The domestic railway engineering department usually performs manual detection. Due to the

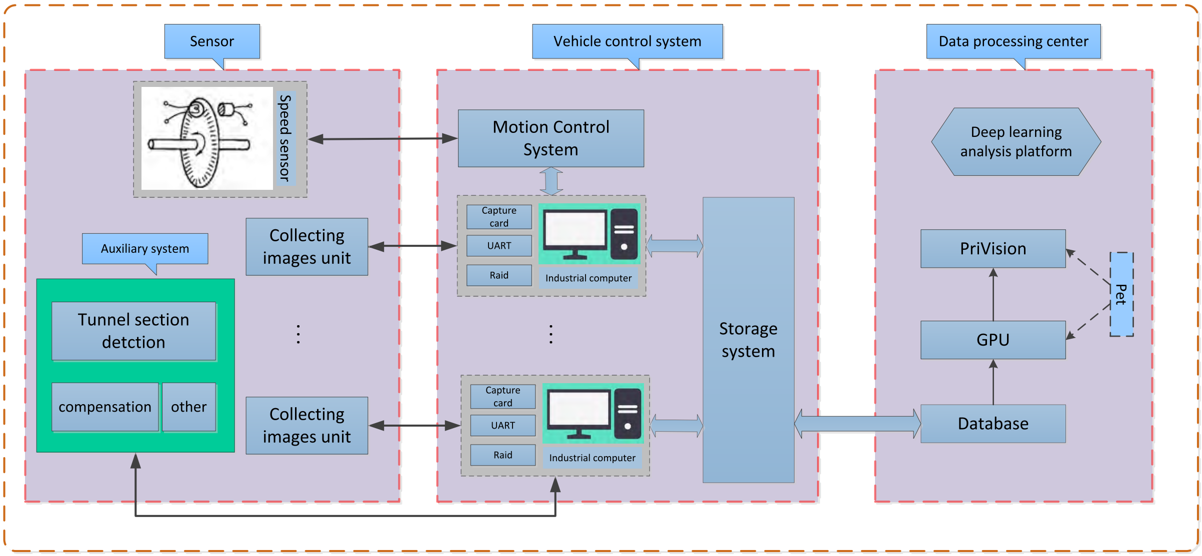


FIGURE 1. The overview of system framework containing image acquisition system, vehicle-control system and crack identification and management system.

limitation of the sunroof time, it is impossible to do the large-scale and high-frequency inspection.

- Crack identification accuracy and efficiency are low. The current tunnel crack detection system mostly uses traditional digital image processing, manually assisted methods, and even manual browsing to identify cracks. For massive data, it is difficult to meet the requirements of fast and high-frequency inspection.
- There are obstacles in engineering application. Some universities have successively developed engineering prototypes. Wang from Tongji University used two cameras for inspection. It is necessary to scan the entire tunnel lining surface for 9 times. For long tunnels, it is not enough to satisfy the actual testing requirements, and it still stays in the prototype stage. The market has not seen the application of engineering.

In order to meet the requirement of detection accuracy and speed and consider the needs of practical engineering applications, the tunnel crack detection system designed in this paper consists of image acquisition system, vehicle-control system and crack identification and management system, as shown in Fig 1.

A. IMAGE ACQUISITION SYSTEM

The image acquisition system is responsible for the acquisition of the crack image. The camera and the computer installed on the inspection vehicle realize the image acquisition.

The image acquisition system consists of a high-definition image acquisition unit, camera arrangement and synchronization, and auxiliary system [1].

1) HD ACQUISITION UNIT

For the acquisition of two-dimensional images, the usual methods are area-array cameras and linear array cameras [1], both of which have their own advantages and disadvantages. Line-array camera has advantages of large dynamic range, small image distortion, and high-speed image acquisition which is not easy to tail, and so linear array camera has been widely used in the image acquisition. Linear array camera sensors benefit from a single line of sensitive element scanning, making high scan frequencies and high resolution possible. In view of the high speed and high-resolution requirements of railway tunnel lining, this paper selects industrial linear array CCD camera and industrial grade lens for high-definition image acquisition.

In addition, the quality of the light source system affects the quality of image acquisition directly, thus determining the effect of crack recognition. Due to the harsh environment and poor lighting conditions in the tunnel, after multiple trials and comparisons, this paper uses laser illumination. Laser illumination is minimally affected by the outside world, with the advantages of long life, stability, and easy integration. We select the appropriate power laser to integrate with linear array camera to ensure the brightness and ensure the safety through precise control in this paper.

The high-definition acquisition unit is the core module of the frontend equipment in the tunnel lining detection system, which determines the quality of the image and determines the effect of crack recognition indirectly. There are many factors that affect image quality and recognition accuracy, such as camera, lens, light source, depth of field, effective pixels, and working distance. In order to obtain high-quality images, the high-definition image acquisition unit adopts the integration of a linear array CCD sensor, an industrial lens, and a laser light source.

2) CAMERA PLACEMENT AND SYNCHRONIZATION

In order to get an efficient collection of tunnel lining images, the number and layout of the cameras designed in this paper should cover the tunnel section completely. For single-line tunnels, the detection should be completed in one operation. Using multiple computers to work together and writing tunnel lining scanning program can automate the acquisition, transmission, and storage of large data volume.

The synchronization between multiple cameras adopts the pulse external trigger mode. This paper designs an incremental photoelectric encoder to realize synchronous pulse output [12]. After adjusting the signal processing module, multiple cameras are synchronized to work. In addition, the control computer simultaneously acquires the encoder pulse output to calculate the mileage information accurately, as shown in Fig 2.

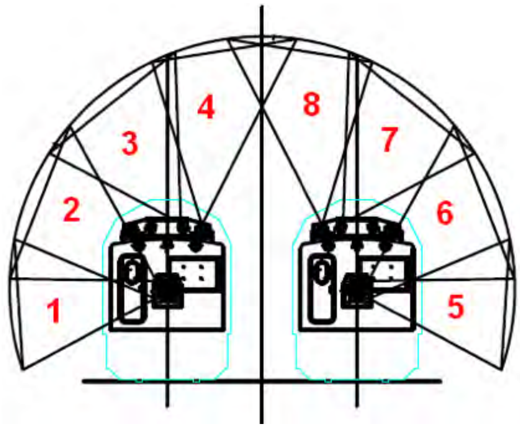


FIGURE 2. The placement of cameras.

3) AUXILIARY SYSTEM

The auxiliary system collects tunnel headroom data, mileage information, and compensation information, for the registration and correction of image data, so that the crack geometry information and position are more accurate [17].

The text design of the headroom detection system implements the following functions:

- The working distance of the linear array CCD camera from the tunnel lining is obtained by laser ranging, thereby determining the resolution of the images at different positions, providing a basis for calculating the crack geometry. Meanwhile, the precise positioning of the crack is realized by calculating the circumferential direction of each image along the tunnel lining with position information.
- Auxiliary detection identifies the location of the tunnel section where large deformation may occur.
- Detecting the clearance size of the tunnel provides basic data for tunnel boundary management.
- Provide a sign of exit and entrance for the tunnel inspection system for automatic detection.

The motion compensation system is installed on the left and right sides of the bogie to collect the vibration during the running of the vehicle for accurately locating the circumferential position data of the tunnel section. The mileage system obtains the encoder pulse count, calculates the real-time mileage and speed, and realizes the longitudinal location of image, clearance, and compensation.

B. VEHICLE-CONTROL SYSTEM

The captured road image is digitally processed and transmitted to the computer for storage in the vehicle-control system.

Since multiple cameras are used for simultaneous detection, high-speed storage and copying of large data volumes become the key technology. It is necessary to study and determine a reasonable acquisition and storage strategy and a cost-effective storage solution while taking into account the reliability of the solution and the convenience of daily use. This paper adopts Raid card and DAS direct-attached storage scheme [31], which intelligently realizes high-speed real-time writing of multi-channel data while satisfying the demands of fast transfer and data recovery.

As shown in the middle of the vehicle-control system module in Fig 1, the vehicle-control industrial computers are connected to the storage array in the SAS mode directly and are set to the Raid0 mode to satisfy the high-speed writing respectively. When the detection system is working, the data is stored in real time to the high-speed disk array. After the test is completed, the disk can be removed and connected to the crack identification and management system after the ground.

C. CRACK IDENTIFICATION AND MANAGEMENT SYSTEM

The crack identification system is responsible for crack identification and feature calculation. Crack identification is the core of the system, including image preprocessing, crack identification and segmentation, and crack performance evaluation. Finally, the segmentation result is output according to the crack evaluation index. The system contains the frontend and the backend based on PytorchEveryThing(Pet) which is a framework of deep learning researched and developed by us.

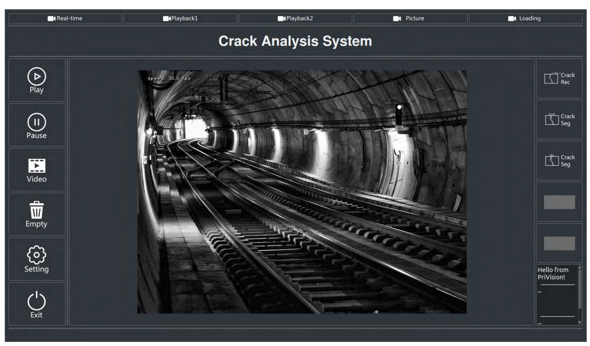


FIGURE 3. The UI of crack analysis system.

1) FRONTEND

The frontend of Pet is also a visual interface, as shown in Fig 3 below. The interface basically includes all the crack analysis content, mainly displayed in the three toolbars on the left side, the top side and the right side.

Firstly, the top toolbar contains five optional modes, real-time mode, playback mode 1, playback mode 2, picture mode, and loading mode. Real-time mode is to call the external camera to pass in the crack data directly and analyze the crack visualization result and the analysis result immediately. Playback mode 1 and playback mode 2 can view the results of real-time mode again. Loading mode uses the data acquisition device to collect the video that needs to be analyzed firstly and then passes it to the database in the frontend for analysis. Picture mode is a special case of the loading mode, that is, the input data is a picture data set.

Then there are the toolbars on the left and right sides. On the left are some state settings, such as start, pause, video, and so on, which can assist the analysis results better. There are function settings on the right side. At present, there are two kinds of functions, crack classification and crack segmentation, which realize the classification and segmentation of the crack image respectively.

2) BACKEND

The backend of Pet is the core algorithm of all tasks, mainly based on Pytorch, a Python-based scientific computing package designed to serve two types of scenarios: replacing numpy with GPU potential and providing an experimental learning platform with efficiency and high-degree flexibility.

At present, the Pet backend has very powerful functions which can realize many tasks, not only the segmentation of crack images, but also the classification, detection, segmentation, re-identification and posture analysis of images, and so on. It is perfect and practical support of backend, as shown in Fig 4.

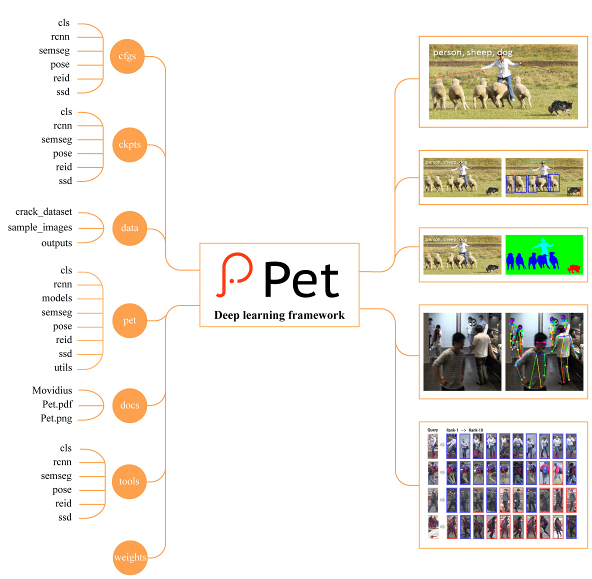


FIGURE 4. The overview of Pet framework which shows the structure of files contained and the executable tasks.

Fig 4 shows the tasks that Pet can accomplish and the relevant algorithm implementation mainly. The image on the right is the performance of each task, and the left side corresponds to the core algorithm that implements the corresponding task.

Among them, *cfgs* is the configuration file of the training and testing model, and all the *yaml* files of the model training are stored in the directory. When training and testing in the Pet framework, different configuration files determine different algorithm model structures and training and testing parameters. Pet gives a lot of example *yaml* files that work well on public datasets. In general, we can make changes to our actual situations. *ckpts* contains models generated during the training process, including the latest model and the current state model. If the current training is interrupted, the current model parameters can continue training the model without starting from scratch. *data* is a database file that holds the data sets needed for the training model. It also stores the results generated during the network testing phase. *Pet* contains the core algorithm of the different tasks, including the various algorithm modules for different tasks, which is the core directory of the entire project. *tools* stores some extension function scripts in Pet mainly, such as *train_net.py* for the training process, *val_net.py* for the verification process, and *test_net.py* for the testing process. These files will call the algorithms of Pet which is also the beginning of the training model. *weights* stores the pre-training models.

IV. CRACK SEGMENTATION

A complete segmentation process includes inputting image, preprocessing image, computational inference, and outputting results. The segmentation network can segment the images in pixel-level and identify which pixels belong to the crack and which pixels do not belong to the crack to obtain the information of all the pixels. The segmentation network contains two phases: the training phase and the inference phase. In the training phase, the input image is the data with the segmentation annotation from the dataset, and the deep learning framework is used to train the segmentation network. The segmentation network can identify each pixel in the image and determine if there is a crack in the image. In the inference phase, the new tunnel image is sent to the segmentation network without annotation, and the segmentation network completes the segmentation of the crack so that the staff can do the analysis and take corresponding measures in time. The segmentation process is shown in Fig 5.

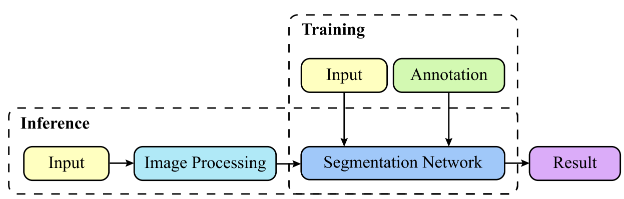


FIGURE 5. The segmentation process contains training process and inference process.

In this paper, we decouple the network into two parts. We consider that each part of the network plays a different role and their optimization directions are different. We adopt ResNet18 [14] as the encoder and use ASPP [35] as the decoder which will be the baseline in the following experiments about datasets. The encoder aims at extracting the feature of tunnel cracks which focuses on the feature extraction and the inference time and the decoder aims at doing the semantic segmentation which focuses on the receptive field for the semantic segmentation and the better performance on segmentation. And we introduce other lightweight networks, MobileNet-v1 [34] and CrossNet as the encoder, to do the experiments and then optimize the network to accelerate for the real-time system. MobileNet-v1 uses simpler network structure design and depthwise convolutions which outperforms many networks with a lot of computations. Considering efficiency, the depth of lightweight networks should not be too large. In this paper, we use ResNet with 18 layers and MobileNet-v1 with 28 layers. And the depth of CrossNet is 58. Though the depth is larger, it will not affect the speed a lot. CrossNet introduces a Pod structure where the feature map of depthwise convolutions can be reused by the cross-connection which leads to less computation.

ASPP used in our experiments refers to DeepLab-v3 [35] which is proposed in [36]. ASPP uses multiple parallel atrous convolutional layers with different sampling rates to extract feature for segmentation.

V. EXPERIMENTS

A. DATABASE

In deep learning, the preparation of datasets is a crucial part, and it is related to the performance of model training. At the same time, the feature of tunnel cracks is very different from most public datasets, and there are domain differences. In order to solve the problem of tunnel crack segmentation, a special dataset called *tunnel crack dataset* is prepared. To the best of our knowledge, our tunnel crack dataset with semantic segmentation annotation is the first crack dataset for segmentation. The camera scans the acquired tunnel image in the real tunnel continuously and collects the whole 7 tunnels in different scenarios. It takes a lot of manpower to screen the crack in the image one by one to do the annotation. Finally, the image data of 7 complete tunnels are selected. A total of 643 crack images are used as the original crack images of the dataset. These crack images cover different light intensities, different sections and different types of tunnels. Since the resolution of the crack image captured by the camera is very high, 4096×4096, which is a high-resolution image, and such an image cannot be an input into the neural network. Therefore, the original high-resolution image needs to be cropped. The size of the cropped image is 512×512. This size is similar to the image size of Pascal VOC [19] and MSCOCO [23] dataset where the length of the side is kept between 400 and 600 basically. Because of the cropping, some images do not contain cracks which need to be filtered. After cropping and filtering, the total amount of dataset is 15718.

For these cracked images, professional data annotators also need to annotate all cracks in each tunnel crack image. The annotation of the crack needs to use a drawing software brush to cover the location of the crack. The width of the brush is 5 to 8 pixels more than the width of the crack to obtain the global information of the crack better, which is beneficial to improve the final segmentation accuracy, as shown in Fig 6. Then, using the image preprocessing method, the annotation of each marked tunnel crack image is generated.

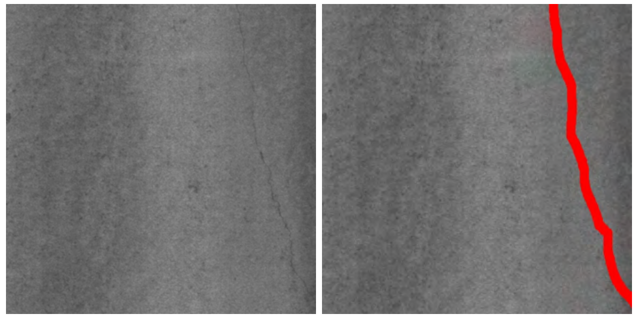


FIGURE 6. The original image (left) and the image with crack annotation (right).

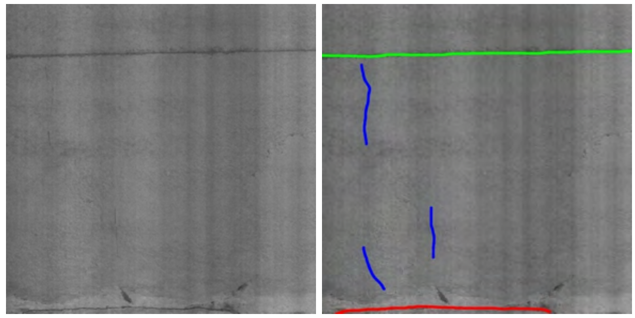


FIGURE 7. The original image (left) and the image with crack, structural seam, water stain and scratch annotations (right).

At present, there are 15718 crack pictures after image pre-processing with the size of 512×512. The crack dataset includes the crack images and their corresponding ground truths. The training set has a total of 8702 images and the verification set is 7016 images.

At the same time, a large number of follow-up experiments have shown that the structural seams have a certain correlation with the water-stained edges, and these factors will greatly affect the performance of the experimental results. In other words, the structural seams and water-stained edges are identified as cracks, which reduces the accuracy of the cracks. For the special case of the tunnel crack dataset, this paper adds an augmented dataset called *augmented tunnel crack dataset*. Based on the original tunnel crack dataset, the structural seams, water stains, and scratches are added into the dataset, and the others are unchanged. The image and annotation of augmented dataset is shown in Fig 7.

Image preprocessing is also an important step of deep learning, and it will affect the performance of the final result to some extent. Due to noise, illumination, and other external environmental factors or the device itself, the quality of the original digital image obtained is usually not very high [24]. Therefore, before image classification, image detection, and image segmentation, the original image needs to be preprocessed to be more clear and the image features are highlighted so that the image can be identified and analyzed well. In deep learning, there are many ways to augment images, such as cropping, flipping, rotating, shaking, normalization, and so on. In this paper, the preprocessing of the original images is to crop the original images and the corresponding annotations.

The ground truth of the original crack image is the replacement of the specific element value by the labeled image that was previously annotated by the professional annotator. For example, the pixel of the crack in the annotated image is replaced with 1, and the other background is replaced by 0. The annotation of multiple categories is the same.

B. EVALUATION CRITERIA

In semantic segmentation task, the evaluation criteria for measuring the segmentation performance are Pixel Accuracy (PA, Pixel Accuracy), Mean Pixel Accuracy (MPA, Pixel Accuracy), Mean Intersection over Union (MIoU), and Frequency Weighted Intersection over Union (FWIoU) [2]. For ease of explanation, it is assumed that there are a total of $k+1$ classes (from l_0 to l_k, including an empty class or background). p_{ii} indicating that the number of pixels belonging to class i is also predicted to be class i, that is, the true positive. p_{jj} is the same. p_{ij} and p_{ji} are interpreted as false positives and false negatives respectively.

- Pixel accuracy is the simplest evaluation criterion, which refers to the proportion of pixels with the correct label to the total pixels, as shown in the following equation.

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}}, \tag{1}$$

- Average pixel precision is a simple improvement of PA. Calculate the proportion of correctly classified pixels in each class, and then find the average of all classes, as shown in the following equation.

$$MPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}}, \tag{2}$$

- Mean intersection over union is the standard evaluation of semantic segmentation. It calculates the intersection of the two sets and the union of the two sets. In semantic segmentation, the two sets are ground truth and predicted segmentation respectively. IoU is calculated on each class, and then averaged as shown in the following equation.

$$MIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}, \tag{3}$$

- Frequency weighted intersection over union is an improvement of MIoU. This method sets weights according to the frequency of occurrence of each class, as shown in the following equation.

$$FWIoU = \frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k} p_{ij}}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}, \tag{4}$$

Among all the above metrics, MIoU is the most commonly used metric because of its simplicity and representativeness. Most researchers use it to report their results. However, this subject is only about the category of cracks, which belongs to a single category. Adding other categories only serves as a competition of inter-class. It is not an important object of this subject, so MPA, MIoU, and FWIoU are not suitable for the evaluation criteria of this subject. The subject selected PA and IoU of MIoU as the performance evaluation criteria of the experiments finally.

C. EXPERIMENTS

This paper uses the ResNet18 [14] as encoder, and the decoder uses the ASPP [35] which is the baseline in experiments. The segmentation quality measures all categories by IoU scores and Pixel Acc. IoU calculates the ratio of the intersection and union between two sets. In semantic segmentation, these two sets are ground truth and predicted segmentation. Firstly IoU in each category is calculated. Similarly, after obtaining the IoU for each category, then the average is calculated to be MIoU. The accuracy of a pixel is the ratio of the number of correctly classified pixels to the total number of pixels. All the implementation is based on Pet.

1) THE SEGMENTATION PERFORMANCE COMPARISON BETWEEN SINGLE-ANNOTATION AND MULTI-ANNOTATION

First, both the tunnel crack dataset and its augmented dataset include the crack images and their corresponding annotation images. Except for the annotation information, the other parameters of the two datasets are unchanged. Then we do some experiments on these two different datasets. The other parameters in the experiments are the same.

Detailed experimental results for each category, MIoU and Pixel Acc are shown in Table 1. For the tunnel crack dataset, we experimented with only one crack category with an IoU of 0.3716, a MIoU of 0.6836, and a Pixel Acc of 99.57%. As we can see, Pixel Acc gets the highest result than other evaluation criteria. Because Pixel Acc calculates all the categories in the images including the background which can get really high accuracy due to it covers a huge area in images. On its augmented dataset, we obtained a crack class with an IoU of 0.4088, a MIoU of 0.4092, and a Pixel Acc of 98.15%. Compared to the original crack dataset, we can see that we improved the IoU (3.72%) of the crack category on its augmented dataset significantly, which demonstrates the effectiveness of our approach. Meanwhile, it should not be overlooked that the IoU of the structural seam category reached 0.5239 on its augmented dataset.

In addition, Fig 8 shows a comparison of the results before and after the addition of structural seams, scratches and water stains annotations. For (a), the first line shows the predictions before adding structural seams, scratches, and water stains annotations; the second line shows another. At the same time, the image on the left in (a) is the original image of the crack; the image in the middle is the annotation image; the image

TABLE 1. Results of experiments in two kinds of datasets.

Dataset	Input Size	Classes	IoU(Background)	IoU(Crack)	IoU(Scratch)	IoU(Structural seam)	IoU(Water stain)	MIoU	Pixel Acc.(%)
Tunnel-Crack	512 × 512	2	0.9957	0.3716	–	–	–	0.6836	**99.57**
Tunnel-Crack-aug	512 × 512	5	0.9820	**0.4088**	0.1148	0.5239	0.0167	0.4092	98.15

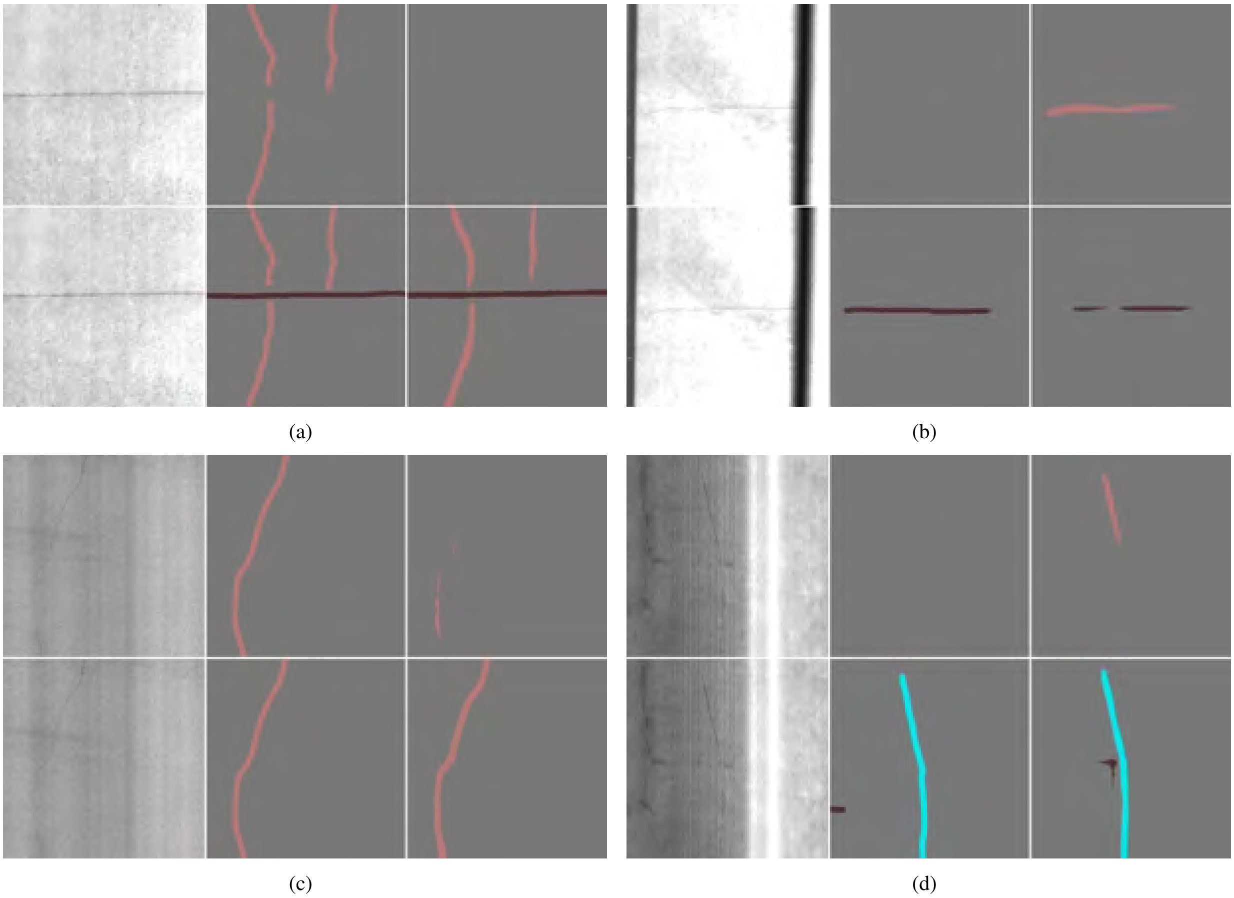


FIGURE 8. The comparison between the original dataset and the augmented dataset.

on the right is the predicted image. Others are similar. The first rows in (a) and (c) in Fig 8 use the original dataset and the second rows use augmented dataset with the addition of structural seams, scratches and water stains annotations. It is obvious to observe that after the annotation information is added, the more pronounced cracks are segmented correctly. Meanwhile, for (b) and (d) in Fig 8, some structural seams and scratches are divided into cracks erroneously before the annotation information is added. Through the above analysis, when training the segmentation network, competition of inter-class occurs after the annotation information is added. This makes the crack's confidence (possibility) higher, so that the crack is accurately segmented, and because of the addition of these annotation information, the segmentation network learns the features of these categories and learns the feature differences between these categories and cracks, thereby reducing misjudgment of cracks, such as (b) (d). It indicates that the annotation plays a supervisory role during training [18], thereby reducing the false positive rate of cracks.

2) THE SEGMENTATION PERFORMANCE COMPARISON BETWEEN DIFFERENT WEIGHTS

In the augmented dataset, there are four categories of tunnel images, cracks, structural seams, scratches, and water stains. These categories vary greatly in location, quantity, and length, especially in terms of quantity. At the same time, there is a certain correlation between them. In order to improve the performance of the crack category better, we set different weights for each of the four categories to do the experiment, so that the loss function of the crack category [3] occupies the main gradient, and the experimental results are shown in Table 2.

Based on the analysis of the crack image, it can be easily observed that the percentage of pixels in the structural seam category is the highest except for the background.

TABLE 2. Results of experiments with different weights of different cracks in augmented dataset.

Dataset	Input Size	Weights	IoU(Background)	IoU(Crack)	IoU(Scratch)	IoU(Structural seam)	IoU(Water stain)	MIoU	Pixel Acc.(%)
Tunnel-Crack-aug	512 × 512	–	0.9820	0.4088	0.1148	0.5239	0.0167	0.4092	**98.15**
		[1,4,1,1,1]	0.9815	**0.4216**	0.1325	0.5250	0.0135	0.4148	98.05
		[1,8,1,1,1]	0.9809	0.4064	0.1021	0.5217	0.0151	0.4053	97.99
		[1,16,1,1,1]	0.9796	0.3771	0.1266	0.5180	0.0020	0.4007	97.84

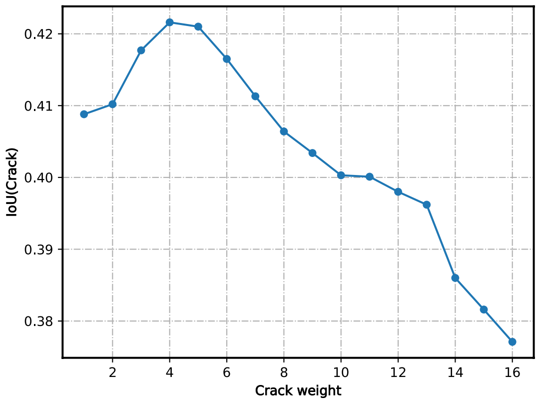


FIGURE 9. The IoUs of crack with different crack weights.

Therefore, to balance these two categories, we set the weight ratio to 4:1 and all other categories to 1 by default. At the same time, we adjust the weight of the crack category constantly. As can be seen from Table 2., when the weight of the crack is 4, the IoU of the crack category is the highest, which is 0.4216. It also increases by 0.0128 or 1.28% compared to the group of experiment with no weight, which demonstrates the effectiveness of our proposed method.

In addition, Fig 9 shows the results of different crack weights. As can be seen from Fig 9, the curve takes the form of a convex function. When the weight coefficient of the crack is 4, the maximum value is obtained, which is the best result. When the optimal crack weight coefficient, 4, becomes smaller or larger, the corresponding crack IoU decreases, that is, the segmentation performance of the tunnel crack will decrease. When the optimal crack weight coefficient becomes smaller, the dominant positive sample (crack) in the training set is reduced. Structural seams become the main learning part of the segmentation network, so the IoU of the cracks decreases. On the contrary, when the optimal crack weight coefficient becomes large, the surveillant effect of the structural seams becomes weak. The special case is that the structural seam category does not have any surveillant effect, which is equivalent to the segmentation of the cracks and the background. When the crack weight coefficient is 4, data equalization is realized. Data equalization can greatly optimize the performance of the training network, which is also very important in deep learning.

Fig 10 shows the crack image segmentation performance predicted by the baseline model and the final optimized model. For (a), the first row shows the results of the baseline model; the second row shows the results of the final optimized model. Meanwhile, the left side of image (a) is the original image of the crack; the middle image is the annotation image; the image on the right is the predicted image. Other images are similar.

As can be seen from the first line of (a), relatively obvious tunnel cracks are not segmented; the tunnel cracks in (b)(c) are not very obvious. They come from different illumination environment, different brightness in the tunnel. As can be seen from the first line of the corresponding figure, they are not segmented. Further, (d) shows that there is a plurality of cracks in the tunnel image. However, after using the baseline model, the crack was not completely segmented. It is apparent from the first line of (e)(f) that the structural seams and scratches in the tunnel image are identified as cracks. The above situation leads to poor crack segmentation performance. For the second line of all images, our final optimized model solves the problems and get good performance.

For Fig 11, it shows the final crack segmentation performance. For (a), it is the same as the distribution form of (a) in Fig 10. (a) (b) shows the segmentation of cracks in the horizontal direction in various complicated environments; (c) (d) shows the segmentation of cracks in the vertical direction in various complicated environments.

3) THE PERFORMANCE FOR REAL-TIME SEGMENTATION

In this section, we use some lightweight networks to do the real-time experiments to improve the speed of crack segmentation for analyzing system. MobileNet-v1 is a classic lightweight network as a backbone which has lightweight parameters and fast speed in inference time. Its details are introduced in Sec IV. The results are shown in Table 3.

In this experiment, we use the best crack weight [1,4,1,1,1] to verify the effectiveness of the real-time system. The size of the input image is 512 × 512. Other evaluation criteria in Table 3 is the same as the following. But there is a new evaluation criterion, *speed*, which means the testing time of one image for one GPU. As we can see, the IoU of crack with MobileNet-v1, 0.4649, is the best one, which has no large difference with the result with CrossNet. The most important in this experiment is the speed. In this table, the speed of the experiment with MobileNet-v1 is the fastest one which is 42.6 ms/image about 23 fps. It can prove that the accuracy of crack segmentation will not be affected with the improvement of speed which is crucial in the real-time experiment.

The results demonstrate that our crack segmentation algorithm in this system can almost get the real-time performance and our algorithm can be applied in our real-time analysis system. And in our system, we adopt MobileNet-v1 as the encoder and ASPP as the decoder to do the crack segmentation finally.

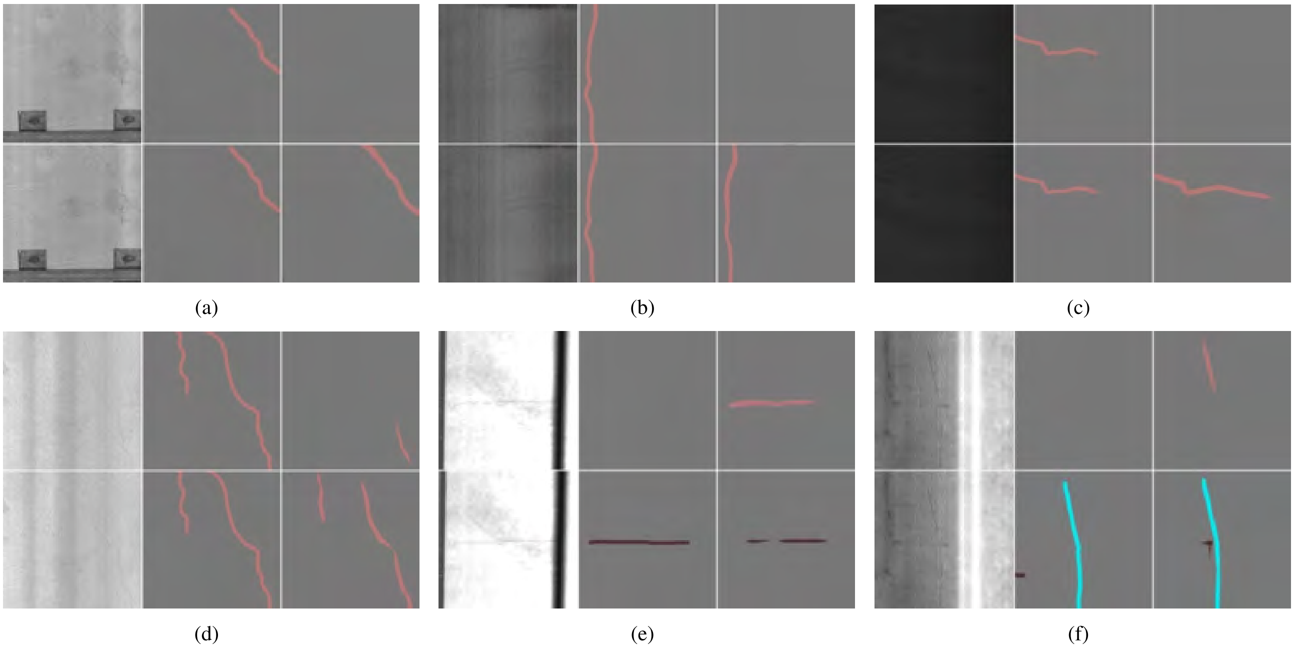


FIGURE 10. The comparison between the original model and the optimized model.

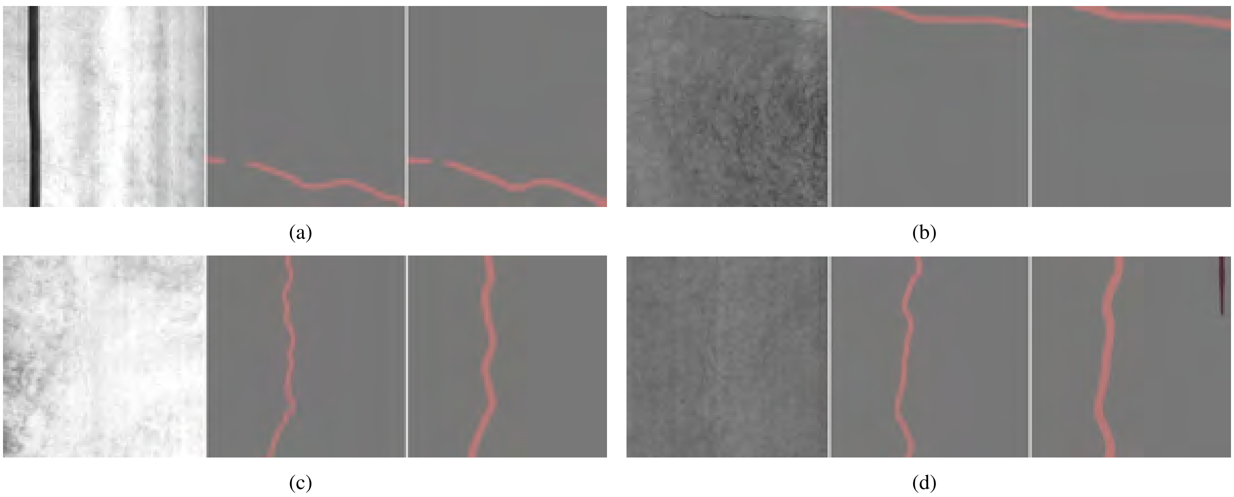


FIGURE 11. The final segmentation results.

TABLE 3. Results of real-time experiment with different backbone with crack weight [1,4,1,1,1]. And the image size is 512 × 512. *speed* in the table means the time for testing one image in one GPU in the inference stage.

Backbone	IoU(Background)	IoU(Crack)	IoU(Scratch)	IoU(Structural seam)	IoU(Water stain)	MIoU	Pixel Acc.(%)	speed(ms)
ResNet18	0.9815	0.4216	0.1325	0.5250	0.0135	0.4148	98.05	51.2
MobileNet-v1	0.9760	**0.4649**	0.0917	0.5029	0.0102	0.4111	**98.19**	**42.6**
CrossNet	0.9801	0.4521	0.1082	0.5017	0.0111	0.4106	98.05	44.7

VI. CONCLUSION

In this paper, we propose a tunnel crack identification and analysis system based on the deep learning algorithms. Meanwhile, we build the first tunnel crack datasets with semantic segmentation annotation for the research and the development of the system. This system adopts MobileNet-v1 network structure as the encoder, builds the crack segmentation network with reference to DeepLabv3 [35] framework, makes many adjustments according to the crack segmentation problem, improves the final model, and does a lot of experiments to optimize the segmentation network, and finally gets 46.49% of IoU on the segmentation verification set and gets 42.6 ms/image about 23 fps of speed.

At present, the number of images in the dataset is small, and there is still a certain limit on the learning ability of the model, and the scale of the dataset can be improved further.

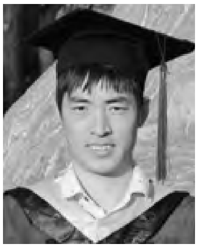
Due to the small dataset, it takes more manpower to do more annotations for segmentation to build a larger dataset, which is also a crucial problem affecting the development of image semantic segmentation. At the same time, the speed of image semantic segmentation is far from the speed of classification, which is the direction of future work.

REFERENCES

[1] A. Cuhadar, K. Shalaby, and S. Tasdoken, "Automatic segmentation of pavement condition data using wavelet transform," in *Proc. IEEE CCECE Can. Conf. Elect. Comput. Eng.*, May 2002, pp. 1009–1014.
[2] M. A. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *Proc. Int. Symp. Vis. Comput.* Springer, 2016, pp. 234–244.
[3] B. De Brabandere, D. Neven, and L. Van Gool. (2017). "Semantic instance segmentation with a discriminative loss function." [Online]. Available: https://arxiv.org/abs/1708.02551
[4] B. Bai, "Crack recognition algorithm of the surface image of subway tunnel," (in Chinese), Beijing Jiaotong Univ., Beijing, China, Tech. Rep., 2015.
[5] 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. 2016.
[6] Y. Chu, "Crack feature extraction based on grayscale image and texture features," (in Chinese), Highway, Tech. Rep., 2010.
[7] A. Dong and M. Liang, "Crack detection algorithm based on gray correlation," *Appl. Res. Comput.*, Oct. 2013.
[8] Y. Wang and Y. Cui, "Skeleton blur removing algorithm based on direct chain code," *J. Comput. Appl.*, 2013.
[9] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, 2017, pp. 4700–4708.
[10] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1925–1934.
[11] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2017, pp. 2881–2890.
[12] J. Jiang, D. Jiang, and H. Gao, "High reliability incremental photoelectric encoder interface circuit design," (in Chinese), *Meas. Control Technol.*, vol. 28, no. 2, pp. 1–4, 2009.
[13] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[14] 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.
[15] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 6, pp. 1904–1916, Sep. 2015.
[16] G. Li and Y. He, "A new method for detecting and classifying pavement cracks with uneven illumination," (in Chinese), *Acta Photon. Sin.*, 2010.
[17] J. Li, C. Zhang, and X. Chai, "Crack detection system of tunnel lining based on image recognition technology," (in Chinese), *Railway Eng.*, 2018.
[18] L. Ye, Z. Liu, and Y. Wang, "Learning semantic segmentation with diverse supervision," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2018, pp. 1461–1469.
[19] S. Vicente, J. Carreira, L. Agapito, and J. Batista, "Reconstructing PASCAL VOC," in *Proc. CVPR*, Jun. 2014, pp. 41–48.
[20] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. CVPR*, Jun. 2017, pp. 1492–1500.
[21] H. N. Koutsopoulos, I. El Sanhouri, and A. B. Downey, "Analysis of segmentation algorithms for pavement distress images," *J. Transp. Eng.*, vol. 119, no. 6, pp. 868–888, 1993.
[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.
[23] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in *Proc. ECCV*, 2014, pp. 740–755.
[24] X.-F. Ou, C.-Q. Xiang, L.-Y. Guo, B. Tu, J.-H. Wu, and G.-Y. Zhang, "Research of recognition of digital characters on vehicle license based on caffe deep learning framework," *J. Sichuan Univ.*, May 2017.
[25] X. Wang, X. Feng, and J. Dang, "Pavement crack detection method based on multi-image and multi-resolution," *J. Lanzhou Univ. Technol.*, Jan. 2013.
[26] Y. Wang, Z. Yu, and B. Bai, "Subway tunnel crack recognition algorithm based on image processing," (in Chinese), *Chin. J. Sci. Instrum.*, vol. 35, no. 7, 2014.
[27] X. Li, "Analysis of the causes of cracks in tunnel lining," *Railway Construct. Technol.*, pp. 103–106, 2005.
[28] Y. Fujita, Y. Mitani, and Y. Hamamoto, "A method for crack detection on a concrete structure," in *Proc. 18th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2006, pp. 901–904.
[29] S. Yang, L. Shao, and X. Guo, "Concrete crack image recognition algorithm based on skeleton and fractal," *Chin. J. Sci. Instrum.*, vol. 33, no. 8, 2012.
[30] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, 1984.
[31] L. Z, "Research on raid," Nat. Univ. Defense Technol., Tech. Rep., 2002.
[32] S. Zagoruyko and N. Komodakis. (2016). "Wide residual networks." [Online]. Available: https://arxiv.org/abs/1605.07146
[33] M. D. Zeiler and R. Fergus. (2013). "Visualizing and understanding convolutional networks." [Online]. Available: https://arxiv.org/abs/1311.2901
[34] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," in *Proc. CVPR*, 2017.
[35] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam. (2017). "Rethinking atrous convolution for semantic image segmentation." [Online]. Available: https://arxiv.org/abs/1706.05587
[36] L.-C. Chen, G. Papandreou, I. Kokkinos, and A. L. Yuille. (2016). "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs." [Online]. Available: https://arxiv.org/abs/1606.00915

QING SONG received the Ph.D. degree from Tianjin University, Tianjin, China, in 2006. She is currently a Scientific Researcher with the Beijing University of Posts and Telecommunications (BUPT), where she is currently involved in computer vision technology study. She is the Founder of the Pattern Recognition and Intelligent Vision Laboratory (PRIV). She has led the PRIV Team to the championship of COCO2018 DensePose Challenge. She has published over 70 academic papers in international journals and conferences. She is in charge of many national, provincial and ministerial projects, and enterprise cooperation projects.

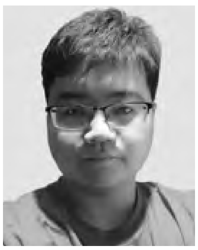
YINGQI WU received the B.Sc. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2017. She is currently pursuing the master's degree with the Pattern Recognition and Intelligent Vision Laboratory (PRIV), BUPT. Her research interests include face detection, face recognition, adversarial attack, and generative adversarial networks.



XUESHI XIN received the B.Sc. degree from the Beijing University of Posts and Telecommunications and the M.Sc. degree from the University of Beijing University of Posts and Telecommunications. He is currently pursuing the Ph.D. degree with the Beijing University of Posts and Telecommunications. His research interests include the target tracking of video sequences, face detection, and human pose estimation.

LU YANG received the bachelor's degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2012. He is currently pursuing the Ph.D. degree with the Automated Institute, BUPT. He has been involved in research work with the Pattern Recognition and Intelligent Vision laboratory, since 2012. He has participated in the MSCOCO 2018 Challenge, one of the top competitions in the field of computer vision, and won the DensePose champion. His research interests include the fields of artificial intelligence, computer vision, and machine learning.

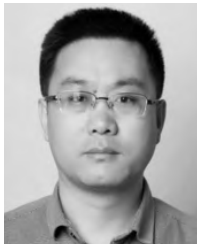
MIN YANG is a master's degree with the Pattern Recognition and Intelligent Vision Laboratory (PRIV), BUPT, Beijing, China. His research interests include image retrieval, object detection, and object tracking.



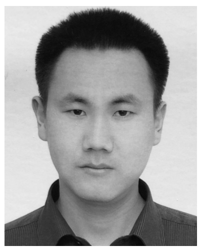
HONGMING CHEN was born in Shanxi, China, in 1995. He received the B.Sc. degree from the Taiyuan University of Science and Technology. He is currently pursuing the M.Sc. degree with the Beijing University of Posts and Telecommunications. His research interests include image classification, image detection, and image semantic segmentation in computer vision.

CHUN LIU received the Ph.D. degree from the University of Kassel, Germany, in 2014. She is currently a Lecturer with the Beijing University of Posts and Telecommunications (BUPT). Her main research interest includes primarily in intelligent computation and optimization, especially evolutionary algorithms in solving optimization problems such as planning and scheduling.

MENGJIE HU received the Ph.D. degree from Beihang University, Beijing, China, in 2017. She is currently a Lecturer with the Beijing University of Posts and Telecommunications. Her current research interests include primarily in computer vision and machine learning, especially object detection, visual tracking, and visual geometry.



XUESONG CHAI received the Ph.D. degree from the China Academy of Railway Sciences, Beijing, China, in 2002, where he is currently a Scientific Researcher and involved in railway wheel/rail interaction research and railway construction condition monitoring technique research. He has published over ten academic papers and two professional books. He has developed wheel tread damage inspection systems for high-speed vehicle and rapid inspection systems for tunnel surface defects with his team. He was in charge of many national, provincial, and ministerial projects.



JIANCHAO LI received the master's degree from the Beijing Information Science and Technology University, in 2013. He is currently a Scientific Researcher with the China Academy of Railway Sciences, where he is involved in image detection technology and laser detection technology research and related monitoring product development work. As a core technical staff, he participates in the completion of key topic: rapid inspection systems for tunnel surface defects. At the same time, it successfully developed a variety of testing equipment such as railway freight inspection hd detection systems, CRTSIII track board rapid detection systems, and high-speed train number image recognition systems. He has cooperated to apply for two invention patents, five utility model patents, and one software copyright.