IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 10, 2019, accepted November 24, 2019, date of publication November 27, 2019,
date of current version December 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2956191

Automatic Pavement Crack Detection and Classification Using Multiscale Feature Attention Network

WEIDONG SONG[1], GUOHUI JIA[1], DI JIA[2], AND HONG ZHU[3]
[1]School of Geomatics, Liaoning Technical University, Fuxin 123000, China
[2]School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China
[3]College of Ecology and Environment, Institute of Disaster Prevention, Beijing 101601, China

Corresponding author: Guohui Jia (jgh6080@163.com)

This work was supported in part by the Public Welfare Research Fund in Liaoning Province, China, under Grant 20170003, in part by the Key Natural Science Plan Fund of Liaoning Province, China, under Grant 20170520141, in part by the Department of Education of Liaoning Province, China, under Grant LR2016045, and in part by the National Natural Science Foundation of China under Grant 41871379 and Grant 61601213.

ABSTRACT Pavement crack detection and characterization is a fundamental part of road intelligent maintenance systems. Due to the high non-uniformity of cracks, topological complexity, and similar noise from crack texture, the challenge arises in this domain with automated crack detection and classification in a complex environment. In this work, an overarching framework for a universal and robust automatic method that simultaneously characterizes the type of crack and its severity level was developed. For crack detection, we propose a novel and efficient crack detection network that captures the crack context information by establishing a multiscale dilated convolution module. On this foundation, an attention mechanism is introduced to further refine the high-level features. Moreover, the rich features at different levels are fused in an upsampling module to generate more detailed crack detection results. For crack classification, a novel characterization algorithm is developed to classify the type of crack after detection. The crack segment branches are then merged and classified into four types: transversal, longitudinal, block, and alligator; the severity levels of cracks are assessed by calculating the average width and distance between the crack branches. The proposed crack detection method effectively detects crack information in a complex environment, and achieves the current state-of-the-art accuracy. Compared to manual classification results, the classification accuracy of transversal and longitudinal cracks is higher than 95%, and the classification accuracy of block and alligator is above 86%.

INDEX TERMS Pavement crack detection, crack classification, convolutional neural network, multiscale feature extraction, attention mechanism.

I. INTRODUCTION

Automatic detection and classification of pavement cracks is an important part of intelligent transportation systems and acts as a primary rapid analysis of pavement distresses. The implementation of a fast and accurate automatic pavement crack detection system is essential for maintaining and monitoring complex transportation networks, and is an effective way to improve the road service quality [1]. Pavement crack automatic detection and characterization systems perform three primary tasks: data acquisition, crack detection, and crack classification. With the development of mobile mapping technology and hardware storage devices, fast acquisition devices are becoming more widely used in pavement distress screening [2] as they can quickly obtain road distress data. Fig. 1(a) shows a road surface image acquisition device installed on a roof, whereas Fig. 1(b) is a pavement image taken vertically, which can be used to measure the crack location and for qualitative analysis. In recent years, a numerous experts and scholars have devoted themselves to researching automatic detection of pavement cracks, and have obtained promising research results [3], [4]. At present, the research on automatic detection of pavement cracks is roughly divided into three methods: traditional image

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

(a) (b)

FIGURE 1. Road surface distress acquisition image: (a) road surface image acquisition device installed on the roof; (b) road surface distress image data after collection.

processing methods, machine learning methods, and deep learning methods.

In the traditional methods, the crack region was usually detected using the threshold approach [5]–[7]. These algorithms can quickly detect the results in the input image by setting different thresholds. Ideally, cracks can be detected easily, as cracks always absorb more light than other areas and typically appear as darker areas in the image. However, when there is a certain amount of noise, the pixels with intensity is lower than that of the cracked pixels seriously degrade the overall detection performance. These methods lack a description of the global information, are sensitive to noise, and rely primarily on the choice of thresholds. Other researchers use artificially designed feature descriptors to detect cracks in images. For example, Gabor filters [8] and wavelet transforms [9] show significant progress in detecting simple cracks. However, due to complexity, diverse topologies, arbitrary shapes and widths, as well as oil spots, weeds, stains, and other strong disturbances on the road, the performance is still limited.

Upon further development, the machine learning method became widely adopted in the field of crack detection. The improved active contour model and greedy search-based Support Vector Machine (SVM) have been used to study the detection of bridge cracks [10]. Ai proposed [11] an SVM-based approach to calculate probability maps using information from multi-scale neighborhoods. Through the fusion algorithm, multiple probability maps obtained from the Probabilistic Generation Model (PGM) and SVM methods are merged into a fusion map, which can detect cracks with higher precision than any original probability map. Prasanna et al. [12] classified multiple spatially adjusted visual features using the random forest method. However, these detection methods are limited to detecting learned cracks, and therefore have difficulty detecting new cracks. To overcome the above problems, CrackForest [13] was proposed based on randomly structured forest for automatic crack detection; it effectively suppresses noise by selecting crack features manually and learning the internal structure. However, it fails to consider the different categories of damage under the complex situation of crack extraction. As traditional methods simulate cracks by setting color or texture features manually, the features set manually can only satisfy crack extraction in some specific situations. The main weakness of these methods is their failure to address robust detection in the changeable environment. Therefore, manual design features are inefficient for extracting cracks from different road images in complex situations.

Recent theoretical developments have revealed that deep learning can solve complex problems by learning features at different levels automatically [14]. The rich hierarchical features of Deep Convolutional Neural Network (DCNN), and the end-to-end trainable framework, have made significant progress in pixel-level semantic segmentation tasks [15]–[17]. Recently, several crack detection methods based on object detection [18], [19] and image block segmentation [20]–[22] and utilizing deep learning have been proposed. However, because these rough estimate methods fail to extract cracks at the pixel-level, they cannot accurately characterize crack classification and severity level assignment in the subsequent step. Huang et al. [23] proposed a solution to this problem that uses the FCN [17] network for pixel-level crack extraction. However, this method did not consider that cracks with different widths and topologies require different sizes of context information. Moreover, in this method, the different contributions of crack features to crack detection were ignored, and all crack features treated in the same manner. Several studies in the literature [24], [25] have proposed 3D crack detection networks based on DCNN for automatic pixel-level crack detection from 3D asphalt pavement. However, as the network uses a convolutional layer with uniform convolution kernels, this can lead to confusion between the target and the context. Zou et al. [26] implemented the DeepCrack network on the encoder-decoder architecture of SegNet, and merged the convolution features generated in the encoder and decoder network in pairs on the same scale to achieve pixel-level crack detection. However, based on the SegNet network structure, the characteristics of learning in the encoding-decoding stage are relatively simple, and most of the spatial information that is lost during the upsampling process cannot be restored through shallow layers. Song et al. [27] developed a crack segmentation network with the DeepLabv3 [28] framework to achieve pixel-level precise segmentation of tunnel cracks. Although this method makes full use of the Atrous Spatial Pyramid Pooling (ASPP) [29] module to obtain multi-scale information, it fails to fully acknowledge the significance of the upsampling operating for refining detection results. In general, deep-learning based methods produce better results than traditional methods. However, there is still a lack of research on robust pixel-level crack detection for trainable DCNN models that utilize rich semantic information. The above crack detection based on DCNN methods does not consider crack classification and damage severity levels.

After crack detection, there is a further task to classify and assign severity levels of cracks in the second stage. Ouyang et al. [30] used a beamlet transform-based algorithm to analyze the crack image of the road surface, and connected the crack segments. Their classification algorithm is very robust and can detect four different cracks types classified

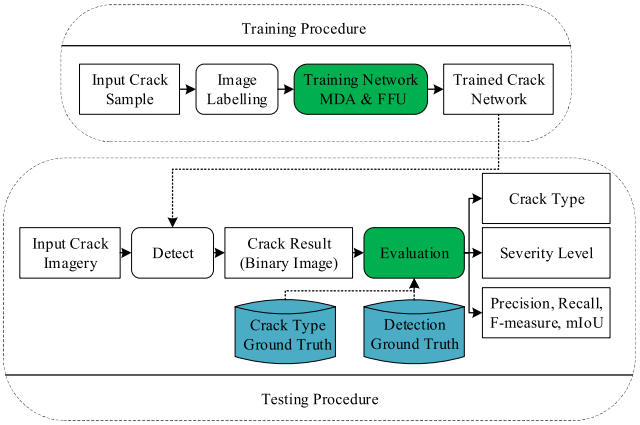

FIGURE 2. Proposed crack detection system structure.

according to the crack direction. However, their method does not consider the severity levels of cracks after classification. Oliveira and Correia [31] use the average width of the cracks to quantify the severity, which can be estimated from the total number of pixels in both the crack and crack skeleton. This method can better complete the severity assignment of linear cracks; however, the density factor and the severity levels of miscellaneous cracks are not considered. Subsequently, Wang et al. [32] proposed a Laplace equation-based crack width definition method and formulas for continuous, unambiguous, and more accurate measurement of crack width. Cubero-Fernandez et al. [33] evaluated multiple cracks in one image as a whole, without considering the spatial distribution relationship between multiple individual cracks. This method also did not consider the severity levels of cracks. In summary, few studies have yielded a universal and robust automatic method to simultaneously characterize the type and severity levels of cracks.

In summary, the existing automatic detection and classification algorithms for pavement cracks mainly focus on pixel-based analysis of the road pavement image to generate a global image analysis result, and then classify the type of crack detected in the image. However, because of the complexity of the road surface, and the variability in the data acquisition process, the automatic detection and classification of pavement cracks still presents challenges. This is especially true for the classification of cracks, which is relatively simple and cannot satisfy the requirements of automatic crack assessment. Given the above problems, in this study it was of the upmost importance to investigate the automatic detection and classification of pavement cracks towards the following two goals: 1) pixel-level robust crack detection in the complex background; 2) crack classification and severity levels assignment.

The overall structure of the automatic crack detection and classification system proposed in this study is shown in Fig. 2. For the first goal, we propose deep multi-scale convolutional features to achieve automatic detection of pavement cracks. The multiscale dilated attention (MDA) module is used in the encoding stage to achieve accurate detection of high-level crack features. Secondly, by combining low-level features and high-level semantic information with attention information, the feature fusion upsampling (FFU) module is used to restore the crack spatial resolution. Then, the MDA and FFU modules are integrated into the DCNN for crack detection. For the second goal, we investigate the crack classification and severity level assignment after detecting the crack from the complex background. The crack is divided into transversal, longitudinal, block, and alligator types by using the connected component labeling algorithm and the spatial distribution of cracks. Moreover, the severity of each type of crack is evaluated by using the average distance and width between the crack branches.

The contributions of this research are as follows:

(1) A novel trainable pixel-level crack detection network that fully exploits the semantic information of hierarchical convolution features from complex backgrounds is proposed.

(2) For feature detection, a multi-scale feature extraction module with a channel-wise attention mechanism to capture rich context information is presented. For spatial information recovery, the features at different levels are fused in the upsampling module to generate more detailed results.

(3) A fully integrated classification and severity level assignment system for crack characterization is presented.

II. MATERIALS AND METHODS

In this section, the novel automatic pixel-level crack detection network is first developed according to the MDA module, and the FFU module. Secondly, a fully automated crack classification and severity levels assignment strategy is developed to characterize the detected cracks in detail.

A. CRACK DETECTION NETWORK

1) MULTI-SCALE DILATED ATTENTION MODULE

Receptive fields are critical to the performance of pixel-level prediction tasks. To increase the receptive field, PSPNet [34], ASPP [29] and Large Kernels [35] have been proposed to capture enough receptive fields. However, the features produced by receptive fields of different scales have discriminating power at different scales, which leads to erroneous results. To overcome this problem, some networks use attentional mechanisms to guide the learning of feedforward networks using advanced information [36], [37], generating more discriminative feature representations. Based on this observation, we propose a multi-scale feature detection module with a channel-attention mechanism to enhance the feature expression ability of small targets. The proposed MDA module is shown in Fig. 3.

In Branch1, a channel-wise attention module is constructed for high-level feature maps to capture the rich context features. The high-level feature $f^h \in \mathbb{R}^{W\times H\times C}$ is represented as $f^h = [f_1^h, f_2^h, \ldots, f_C^h]$, where $f_i^h \in \mathbb{R}^{W\times H}$ is the i-th slice of feature f^h, C is the total number of channels, H and W are the height and width of the high-level feature, respectively. A global maximum pooling (GMP) is used to capture the global semantic information and reduce the number of

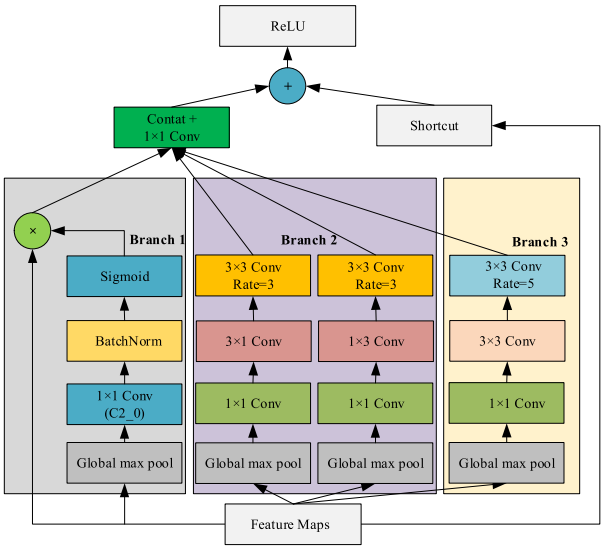

FIGURE 3. Multiscale dilated attention module.

model operations. When the feature map has a higher resolution, GMP will select the maximum response point as the global representation of the response feature map [38]. The maximum pooling operation is performed for each feature to capture global semantic information, and the high-level channel aspect feature vector is obtained as follows:

$$V^h = F_{gmp}(f^h) = \arg\max_{x_C(i,j)} f(f^h(i,j)) \tag{1}$$

where F_{gmp} represents the GMP operation. Afterwards, the 1×1 convolution is used to increase the nonlinear characteristics of the network without losing the resolution. Then we utilize batch normalization to balance the feature scale. Next, the normalization processing is used to measure the encoded channel-wise feature vector mapped to [0, 1] by using the sigmoid operation:

$$W_{CA} = F(V^h, W) = \sigma(y) \tag{2}$$

where σ refers to the sigmoid operation. The module finally outputs the feature map $f^{h'}$ by weighting the f^h with W_{CA}:

$$f^{h'} = W_{CA} \cdot f^h. \tag{3}$$

In Branch2 and Branch3, the standard convolutional layer of differently sized convolution kernels is used to obtain accurate localization mapping, so that the features with strong target features are highlighted. A two-way convolution operation is used in Branch2 to divide the 3×3 convolution operation into 3×1 and 1×3 addition calculations to reduce the computational complexity of the model. Similarly, the use of two consecutive 3×3 convolutional layers in Branch 3 represents a 5×5 convolution operation, to further reduce the model parameters. To further capture more discriminant features, convolution operations with different dilation rates are used after standard convolutional feature maps to increase the receptive field. The multi-scale dilated convolution module thus constructed obtains the crack semantic information at multiple scales, and then fuses the semantic information at different levels to obtain a global prior, achieving the purpose of merging different features. In the two-dimensional signal, the dilated convolution can be expressed as follows:

$$y[i] = \sum_{k=1}^{K} x[i + r \cdot k] w[k] \tag{4}$$

where $x[i]$ is the input feature map, $y[i]$ is the output signal, $w[k]$ is the filter of length K, and parameter r refers to the dilation rate that corresponds to the input signal sampling step, which is equivalent to convolving the input x into each spatial dimension. Then we insert r-1 zero values between the two consecutive filter values to increase the effective receptive field of the kernel. In standard convolution, $r = 1$. In Branch 2 and Branch 3, the dilation rates are 3 and 5, respectively.

Finally, the feature maps generated by the three branches are concatenated, the input features are directly connected to the shortcut connection, and the output features are calculated by the ReLU activation function, as follows:

$$f = \max(0, concat(f_1, f_2, f_3) + W_s i) \tag{5}$$

where f_1, f_2, and f_3 are the feature maps of the final output of the three branches. f and i are the output and input feature maps of the module respectively, and w_s is the linear projection used to match the two input dimensions.

2) FEATURE FUSION UPSAMPLING MODULE

Although the MDA module in the encoding stage could capture rich semantic features from the input image, these features have a coarse spatial resolution [29] and the purpose of this upsampling module is to restore these features to the input image resolution. Inspired by the decoder network in Deeplabv3+ [39], the upsampling module proposed in this study mainly contains two inputs: low-resolution features with discriminative semantic information generated by the MDA module, and high-resolution features in shallow layers. We therefore use different scales of extracted features to provide local and global context information. The upsampling module first concatenates the low-level and high-level features, then uses batch normalization to balance the feature scales. Secondly, the weighted feature vector is calculated by using the attention mechanism similar to that in the MDA. This weight vector can re-select and combine the features, further try to refine the merged features, and improve the feature representation ability. Finally, the two 3×3 convolutions are continuously used to improve the feature representation and restore to that of the original input pavement image. The upsampling module can use high-level and low-level hierarchical features to restore the positioning of crack pixels.

3) NETWORK ACHITECTURE

According to the proposed MDA and FFU modules, a crack detection network is developed as shown in Fig. 5. Given the input crack image, the ResNet [40] pre-training model is first used to extract the crack features. After the high-level

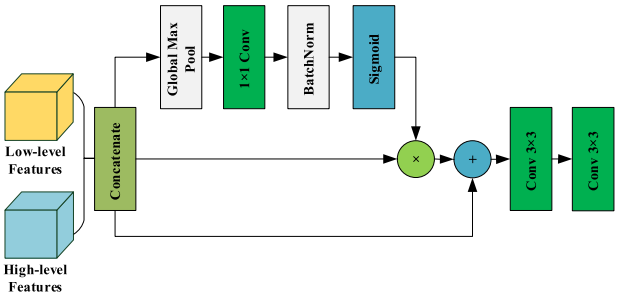

FIGURE 4. Feature fusion upsampling module.

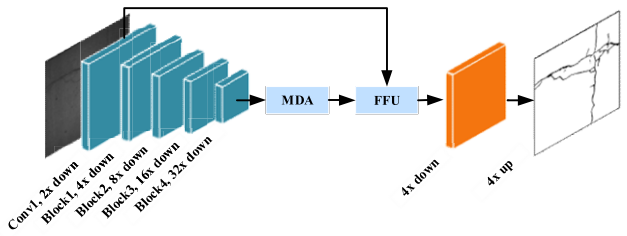

FIGURE 5. Crack detection network with the MDA and FFU modules.

features of deep neural networks are extracted, the MDA is employed to extract the crack features of multiple sizes under multi-scale. Then the semantic information at different levels is merged to obtain the global prior, which is taken as the high-level feature of the network. Next, through fusing the low-level features generated in a shallow layer by the FFU module, the feature map size of the network output is consistent with the input image resolution, and finally, the probability that each pixel belongs to a crack or a non-crack is calculated.

There are only two classes in crack detection, which can be seen as a binary classification problem. Therefore, the Generalized Dice Loss (GDL) function is used as the loss function of the crack detection network in the training process:

$$GDL = 1 - 2\frac{\sum_{l=1}^{2} w_l \sum_n r_{\ln} p_{\ln}}{\sum_{l=1}^{2} w_l \sum_n r_{\ln} + p_{\ln}} \tag{6}$$

where w_l is the weight provided for different label set properties. In this paper, we use $w_l = 1/(\sum_{n=1}^{N} r_{ln})^2$, and r_n refers to the reference foreground segmentation voxel value, while p_n refers the prediction probability map of the foreground target.

B. CRACK CLASSIFICATION AND SEVERITY LEVELS

After detecting the pixel region containing cracks, the characterization of crack types based on the connected component labeling and the spatial distribution of each crack joint branch is investigated. The crack type is divided into transversal, longitudinal, block, and alligator types in this paper. We investigate the assignment of crack severity levels based on the average width of crack pixels and the distance between branch spaces. From this we propose a new algorithm for severity level assignment in this section.

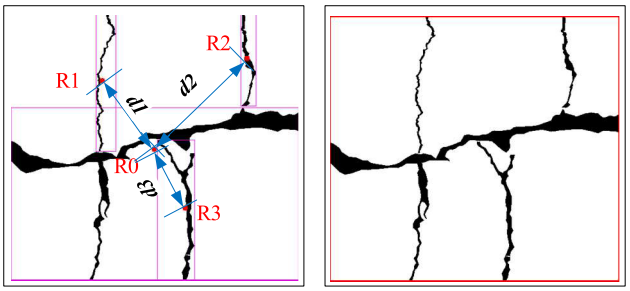

FIGURE 6. Classification of cracks. (a) Example of calculating the centroid distance between crack branches. (b) Result of merger of the crack targets.

1) CRACK CLASSIFICATION

Different from a simple transversal and longitudinal crack, the space distribution of netted type cracks is more complicated. To classify cracks as a whole, we merge the single extracted adjacent crack branches into a new target and then characterize the crack type as a whole. In our method, the crack connection analysis and classification are mainly divided into the following steps: firstly, the connected component labeling is performed on the extracted crack binary images, and the cracks are divided into independent objects. To analyze the relationship between adjacent cracks, we generate a Minimum Enclosing Rectangle (MER) for each crack target. Each MER record consists of the target coordinates (x and y), width and height. Secondly, the centroid coordinate of each crack rectangle is calculated to obtain the distance between adjacent cracks. Then, these MERs are merged into a new one by determining whether the MERs are adjacent or intersecting, and then the distance between the two branches of centroids and the number of branches are calculated. An example of calculating the distance between the centroids is as shown in Fig. 6(a), and R0 is adjacent to R1, R2, and R3, respectively. The average distance of the crack branches included in the generated new rectangle can be expressed as follows:

$$d = \frac{\sum_{n=1}^{N} d_n}{N} = \frac{d_1 + d_2 + d_3}{3} \tag{7}$$

where d_n refers to the distance between contiguous branches. Finally, if the number of branches is smaller than the branch threshold, the angle between the diagonal of the rectangle and the horizontal direction is determined as a transversal or a longitudinal crack. When the number of branches is greater than the branch threshold, the merged target is determined as a netted crack. Then, the block and alligator types are classified based on the average distance of recorded centroids. The characteristic threshold of the crack classification is shown in Table 1. The example of merged results of crack branches is shown in Fig. 6(b).

TABLE 1. Crack classification feature threshold.

Crack Types	Angle	Numbers of Branches	Average Distance
Alligator	-	n>=3	d<0.5m
Block	-		d>=0.5m
Longitudinal	α >=45	n<3	-
Transverse	α <45		-

TABLE 2. Type of pavement distresses and definitions.

Type	Severity level	Definitions	Weight
Alligator	Light	W<=2mm, 0.2m<D<0.5m	0.6
	Medium	2mm<W<5mm, D<=0.2m	0.8
	Heavy	W>=5mm, D<=0.2m	1.0
Block	Light	1mm<W<2mm, D>=1.0m	0.6
	Heavy	W>=2mm, 0.5<D<1.0m	0.8
Longitudinal	Light	W<=3mm	0.6
	Heavy	W>3mm	1.0
Transverse	Light	W<=3mm	0.6
	Heavy	W>3mm	1.0

2) SEVERITY LEVEL ASSIGNMENT

According to different types of distresses, the damage severity of cracks can be divided into different levels. The severity of alligator cracks can be divided into light, medium, and heavy levels. The severity of block, transversal, and longitudinal cracks can be divided into light and heavy. Detailed classification criteria are shown in Table 2. The main feature for judging the damage degree is the average crack width W, and the distance D between the main crack branches, where D can be calculated by (7).

As the crack width can be measured at different locations, it is difficult to quantify its width. Similar to Oliveira and Correia [31], the average width of the crack is calculated at the pixel level. We can calculate the average width of the crack as follows:

$$W_{cs} = \frac{W_c}{W_s} \cdot R_c \tag{8}$$

where W_c is the total number of cracked pixels in the image and W_s is the total number of cracked pixels in the skeleton. Then, the average width W_{cs} of the crack is calculated according to the spatial resolution R_c of the pavement images. As shown in Fig. 7, the two marked black areas are detection results, and the number of pixels is 473 and 4313, respectively; whereas the number of pixels after skeleton detection is 101 and 223, respectively. Based on the known image spatial resolution (for the data set herein, one pixel corresponds to an actual distance of 0.91 mm), from which the average width of the crack is calculated to be 4.3 mm and 17.6 mm, respectively.

Finally, the pavement crack damage rate DR can be calculated by the following formula:

$$DR = 100 \times \frac{\sum_{i=1}^{i_0} w_i A_i}{A} \tag{9}$$

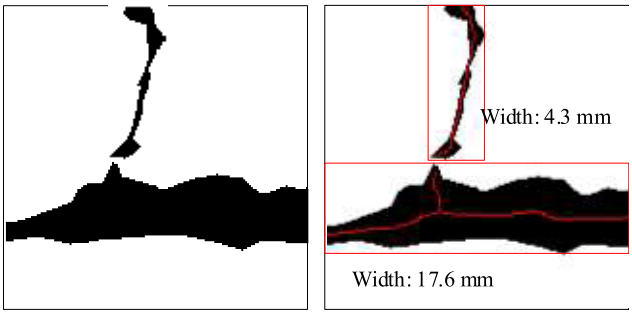

FIGURE 7. Crack width measurement: (a) Crack detection result; (b) computed width.

where A_i is the area of the crack type i (m^2), A is area of the surveyed pavement surface (m^2), and w_i is the weight of i-th crack type, which is valued as described in Table 2. i refers to the crack types, which contain severity levels (light, medium, and heavy), and i_0 is the total number of crack types.

III. EXPERIMENT AND ANALYSIS

A. EXPERIMENTAL SETUP

1) DATASET

CrackDataset: The dataset consists of pavement survey images from 14 cities in the Liaoning Province, China. The data employed in this study were primarily from plane array and charge-coupled device (CCD) cameras, which cover most of the road surface conditions and contains images from different roads and illuminations. The pavement image from the sensors have resolutions of 2330 × 1750 and 3120 × 2048. As the large size of pavement crack images, training our network with them would require a large amount of memory, resulting in overburdening of the training process. Additionally, the crack areas occupy only a small proportion of the whole image, and the remaining background areas are useless for the training process. Therefore, we divided the original road crack images into several small blocks with a size of 256 × 256 pixels. Subsequently, a subset was manually labeled by human experts and taken as ground truth. The ground truth in the dataset provides two types of labels: cracks and non-cracks. We also divided the dataset into three parts, in which the training set and the validation set comprised 4736 and 1036 crack images, respectively, and the test set contained 2416 images. Furthermore, the dataset contained 300 hand-marked classification labeling results from human experts. Example of the crack images are shown in Fig. 8, where some of the images are accompanied by noise such as shadows, oil spots, and water stains; the cracks contained in the same image also have more complex topologies.

2) IMPLEMENTATION DETAILS

We implemented our proposed crack detection network using TensorFlow, which is an open source platform for deep learning. To improve the robustness of the model, several transformations were applied to the data, including

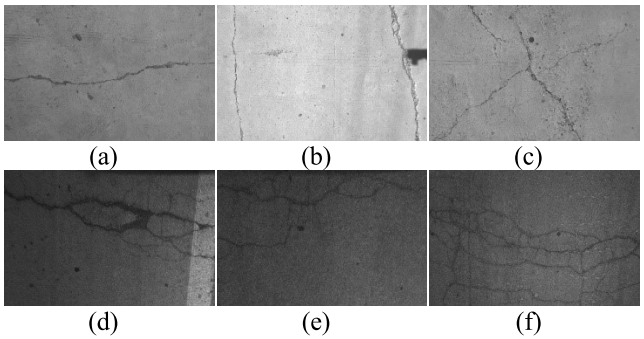

FIGURE 8. Example of a road surface data image in which some images are affected by noise such as shadows, oil spots, and water spots.

random flip, color enhancement, and enlargement. We also utilized the Adam optimizer to converge the network. The network was trained with an initial learning rate of 0.0001, and the momentum and weight decay were set to 0.9997 and 0.0005, respectively. All experiments in our work were performed using an NVIDIA GTX 1080 GPU and 8 GB of onboard memory.

3) EVALUATION METRICS

Accuracy refers to the consistency and difference between crack regions generated at the pixel-level measurement and those in the reference mask. In crack accuracy evaluation, cracked pixels and non-cracked pixels are considered as two separate types. In this study, Precision (P), Recall (R), F-Score, and mIoU were used as the accuracy indicators for quantitative performance evaluation. These four indicators can be calculated as follows:

$$P = \frac{TP}{TP + FP} \tag{10}$$

$$R = \frac{TP}{TP + FN} \tag{11}$$

$$F - Score = \frac{2P \cdot R}{P + R} \tag{12}$$

$$mIoU = \frac{1}{N} \sum_{k=1}^{N} \frac{TP_k}{TP_k + FP_k + FN_k} \tag{13}$$

where TP represents the number that is correctly divided into positive examples, FP represents the number that is incorrectly divided into positive examples, and FN represents the number that is incorrectly divided into negative examples. TP_k, FP_k, and FN_k represent true positive, false positive, and false negatives, respectively; these are measured by the class throughout the test set.

B. CRACK DETECTION RESULT

The automatic crack detection network proposed in this study was compared with other state-of-the-art deep learning semantic segmentation models, including SegNet [17], U-Net [16], PSPNet [34], DeepLab v3+ (DL-v3+) [39] and Discriminative Feature Network (DFN) [37]. Table 3 shows the quantitative comparison results of the test in the crack data set. Compared to the other deep learning-based

TABLE 3. Comparison of results on crack test data sets between our method and other detection methods.

Method	P (%)	R (%)	F (%)	mIoU (%)	Average Time
SegNet [17]	96.77	97.08	96.92	70.56	0.84s
U-Net [16]	96.99	97.09	97.04	71.49	1.16s
PSPNet [34]	96.90	96.88	96.89	69.63	0.63s
DL-v3+ [39]	97.01	97.64	97.32	71.77	1.30s
DFN [37]	96.97	97.43	97.20	71.58	0.69s
Ours	98.74	98.05	98.39	74.81	0.71s

segmentation methods, the crack detection network achieves the highest performance with Precision 97.70%, Recall 98.00%, F-score 97.34%, and mIoU 75.24%. In detecting images with a resolution of 2330×1750, our network detected cracks at a rate of 0.71 s per image. PSPNet and DFN are faster, at 0.63 and 0.69 s per image, respectively. Conversely, SegNet, U-Net, and DL-v3+ detect cracks at slower speeds of approximately 0.84 s, 1.16 s, and 1.30 s, respectively.

The crack detection visualization comparison results were performed on the six crack images (a-f) in Fig. 8, in which some images are affected by noise such as shadows, oil spots, and water stains; the cracks that are contained in the same image have complicated topologies and are shown in Fig. 9. From a (1) to f (1), in which the results of crack detection using the SegNet network are shown, a (1) and b (1) are incorrectly extracted with more shadows and stains, indicating that the method is sensitive to noise. As the network does not have contextual information to acquire diverse types of crack topological structures, an apparent discontinuity occurs in e (1), where the position of crack width changes dramatically. As can be seen from a (2)–f (2), the results obtained by U-Net detection are better than those of SegNet, but they still retain considerable noise. Since most of the spatial information lost during the pruning process cannot be effectively recovered, more fractures occur. It can be seen from Fig. 9 a(3)–a (4) that PSPNet effectively eliminates the influence of noise, but it loses more detailed information, such as d (3) and e (3). DeepLabV3+ performs better in extracting light cracks and can eliminate most of the noise interference in a (4)–f (4), but the large dilation rate produces non-existent cracks, and its single convolution kernel size may cause loss of crack information. The high-level features introduced in the channel-attention block in DFN [37] provide semantics that guide the selection of low-level features, which results in more discriminative feature selection than SegNet, U-Net, and PSPNet. As can be seen from the detection results in a (5)–f (5), DFN effectively detects cracks on a complex background, which indicates that the attention mechanism can also improve the precision of crack detection. However, the network does not consider the multi-scale feature extraction and effective upsampling operation, which limits its performance in terms of detail information recovery.

As shown in Fig. 9 a (6)–f (6), our proposed method has fewer false detection and missed detection results, indicating that the proposed method is more robust. The performance improvement is mainly due to the

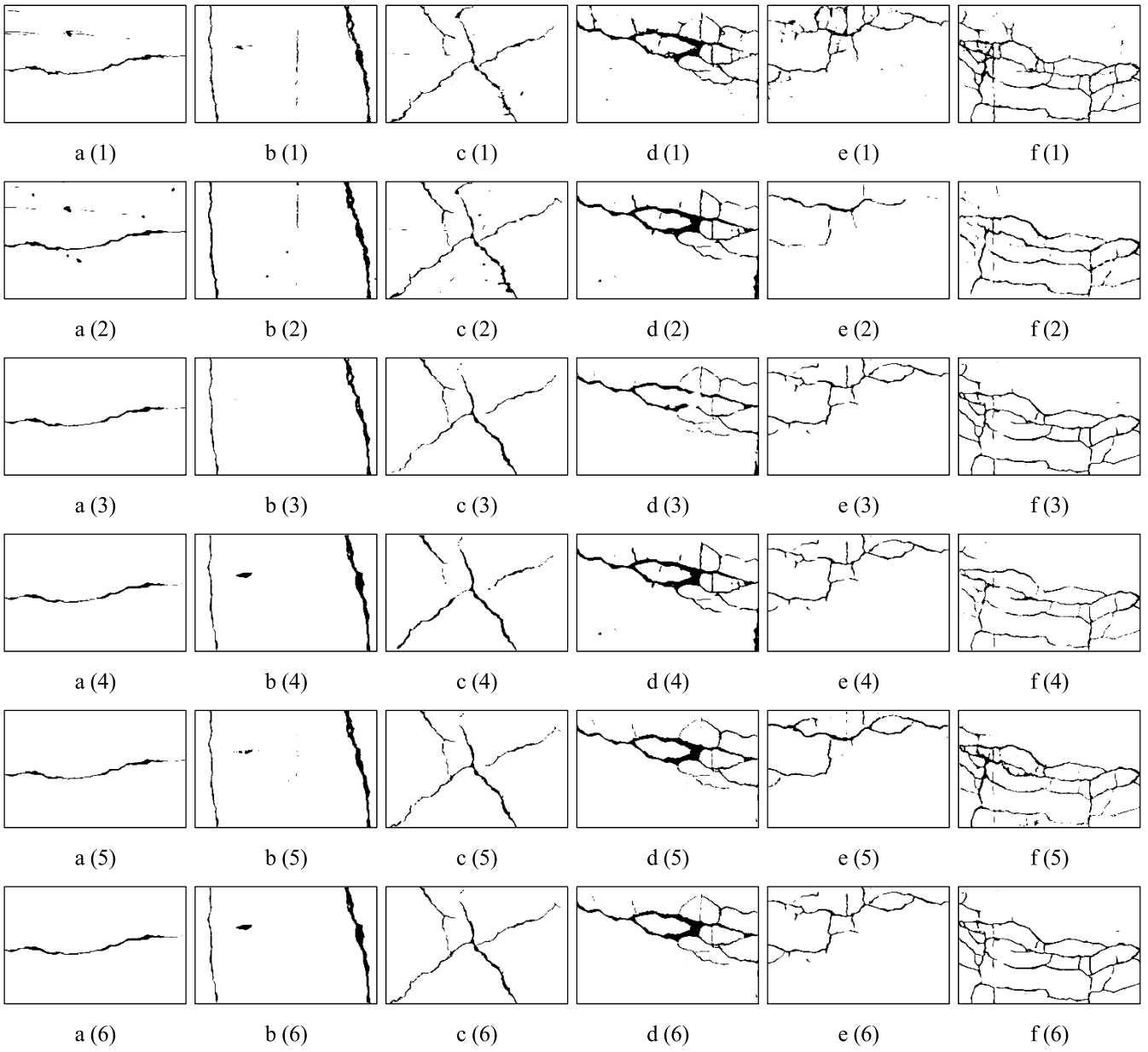

FIGURE 9. Crack detection results (from top to bottom: SegNet, U-Net, PSPNet, DeepLab v3+, DFN and the proposed method).

integration of the multi-scale dilated convolution module, and the attention mechanism in the encoding phase that captures multi-scale context information for accurate feature extraction. In addition, the fusion of high-level semantic information and low-level high-resolution features by the FFU module helps to recover the boundary information of the crack, enabling the network to more accurately extract cracks from complex backgrounds.

C. CRACK CLASSIFICATION AND SEVERITY LEVELS

This paper provides more detailed qualitative and global assessment results than other published crack classification methods. It is possible to divide multiple crack types simultaneously in one image and assign corresponding severity levels to them. In [31], although the crack type can be effectively divided, the classification criteria are relatively simple, and only the influence of the crack width on the severity is considered. In [41], according to the angle between the crack and the horizontal, the classification of longitudinal and transversal cracks can be classified. If there is a crack branch in the extracted image, no matter what angle it is, it will be considered a block crack. In addition, the method does not include assignment of severity levels, and the corresponding category cannot be assigned a corresponding weight.

In this study, according to the spatial distribution of cracks after the binary detection, the multiple crack targets in each image are discriminated, and the corresponding crack types are assigned: transversal, longitudinal, block, and alligator cracks. Crack classification and severity level assignment are mainly based on the number of crack branches, the average

TABLE 4. Detailed crack classification results.

Image	Crack ID	Branch Number	Angle	Average Distance(m)	Width(mm)	Crack Type	Severity Level	Time
(a)	1	1	6.55	-	8.8	Transversal	Heavy	0.31s
(b)	1	1	86.04	-	7.9	Longitudinal	Heavy	0.33s
	2	1	78.19	-	7.9	Longitudinal	Heavy	
	3	1	16.65	-	20.7	Transversal	Heavy	
	4	1	76.36		20.6	Longitudinal	Heavy	
(c)	1	3	-	1.36	8.3	Block	Light	0.36s
(d)	1	5	-	0.25	7.4	Alligator	Light	0.43s
(e)	1	11	-	0.86	7.0	Block	Light	0.41s
(f)	1	9	--	0.19	7.4	Alligator	Heavy	0.40s

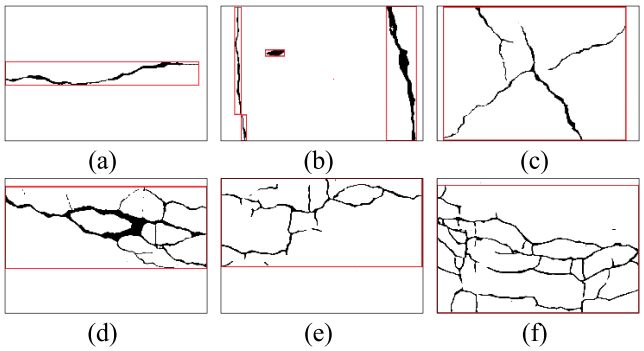

FIGURE 10. The results of automatic classification. Even in the case of fracture, the overall analysis of cracks can still be achieved.

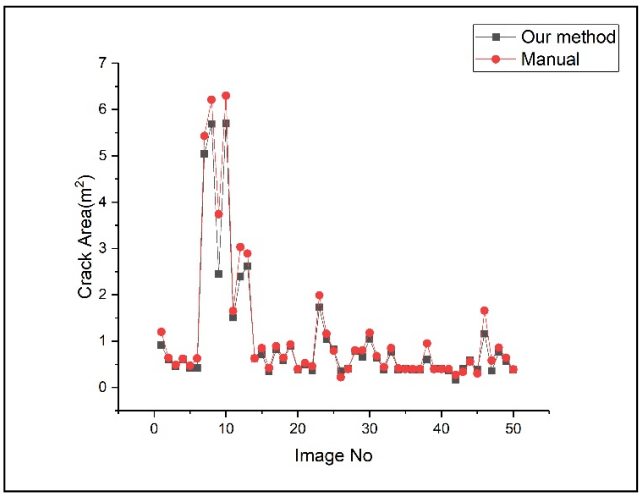

FIGURE 11. Comparisons between manual and automatic measurements of crack area.

crack width, and the distance between branches. The classification accuracy is evaluated in the crack type test dataset. The crack classification results of the six images in Fig. 8 are shown in Fig. 10, and the MER is calculated for the combined cracks into independent crack types. The detailed classification results are shown in Table 4. In classifying crack types from 2330×1750 resolution binary images, our classification method can perform crack evaluation at an approximate rate of 0.5 s per image.

To validate the influence of weight on the area measurement in the severity of crack characterization, 50 crack images were randomly selected in the crack classification test dataset and compared to the area of the hand-drawn crack area. The comparison results are shown in Fig. 11. These results illustrate that the area evaluated by the automatic detection method in this paper has a high degree of uniformity with the manually measured area, demonstrating that the crack classification method is highly reliable.

IV. DISCUSSION

In this section, the crack detection and classification results are discussed in detail for each module proposed. In this experiment, ResNet50 was used as the backbone network, and the effectiveness of the method was evaluated given the crack data set proposed in this paper.

A. CRACK DETECTION

1) MULTI-SCALE DILATED ATTENTION MODULE

To compare the influence of the MDA module on the crack extraction more clearly, the feature map was up-sampled 16 times by simple bilinear interpolation in the upsampling stage to obtain the final prediction result. In the experiment, the ResNet-50 network structure was used as the backbone network for verifying the multi-scale dilated convolution module, and several variant experiments were carried out on the multi-scale dilated convolution module. The experimental results are shown in Table 5, the mIoU of the baseline model using ResNet50 as the feature detection network is only 65.07%, the ASPP [29] module improves the performance of baseline by 1.25%, indicating that the dilated convolution improves the crack detection result. The dilation rate introduced here is an important hyperparameter that enables changing of the size of the receptive field of the MDA module in the network.

To verify the ability of the dilated convolution mechanism to capture rich features, the dilation and convolution operations of three different dilated rates groups—{6,12}, {2,4} and {3,5}—were performed on the final high-level features, and the size of the receptive field increased. The experimental results show that the multi-scale dilated convolution module increased the mIoU by 1.40%, 1.23% and 2.05%, demonstrating that the advanced features with different dilation rates and convolution kernel sizes at multiple scales have stronger characterization capabilities and are better able to help locate

TABLE 5. Comparison results of different dilation rates.

Method	P (%)	R (%)	F-score (%)	mIoU (%)
Baseline	97.32	95.61	96.46	65.07
ASPP	97.41	96.29	96.85	66.32
MD {6,12}	97.64	96.37	97.00	66.47
MD {2,4}	97.53	96.31	96.92	66.30
MD {3,5}	97.85	97.41	97.63	67.12
With Attention	98.41	97.87	98.14	69.30

TABLE 6. Different upsampling features and structural comparison results.

Method	P (%)	R (%)	F-score (%)	mIoU (%)
DL-v3+ [39]	98.68	97.93	98.30	72.68
With attention	98.84	98.05	98.39	74.81

cracked pixels during the encoding process. Although dilated convolution with a larger dilation rate, {6,12}, has a larger receptive field, it introduces other unrelated regions while capturing crack characteristics, which affects the final crack identification outcome. However, an overly small dilation rate, {2,4}, cannot effectively increase the receptive field. With the dilation rate {3,5}, better optimal convergence and better detection effect can be obtained in model training.

To further improve the feature detection performance, this study specially designed a channel-wise attention optimization module that contains a GMP for encoding output functions as vectors. The attention vector was then calculated using convolution, batch normalization, and ReLU units. The original high-level features were re-weighted by the attention vector. For the original features of the input, global context information can be easily captured without the need for complex upsampling operations. The multi-scale dilated module with the attention mechanism increased the performance of mIoU from 67.12% to 69.30%. The proposed MDA module can effectively capture the cracks with discriminative feature extraction capacity and attention mechanisms, directing the network to focus on crack objects.

2) UPSAMPLE MODULE

It can be observed from the results in Table 5 that the mIoU obtained from the simple bilinear interpolation upsampling method is only 69.30%. To further improve the crack detection performance, fusion of high-level semantic information and low-level features by the FFU module was introduced. The MDA features were used as the advanced input of the upsampling module, as it has a stronger discrimination ability. The low-level features in the network have a higher spatial resolution in which crack edge detail information is preserved. After the low-level semantic information was merged with the high-level features with discriminative power, the convolution operation is then used in the upsampling module to restore to the original size for obtaining finer segmentation results. The experimental results are shown in Table 6. The selection of different convolution times had a great influence on the final crack detection results of the model. The best effect was obtained by using two [3 × 3, 256] convolutions in Deeplabv3+ [39]; the mIoU increased by 3.38% compared to the direct upsampling method, indicating that the combination of advanced features and low-level features can greatly improve the detection performance. To further refine the fused features, the attention mechanism is used to refine the feature representation in the feature fusion stage. Compared to the simple feature fusion method in Deeplabv3+ [39], the mIoU value is increased by 2.13%. In summary, the FFU module is used to combine the shallow crack information with deep powerful semantic information, helping to fuse the multi-level features of the cracks and improve the overall crack detection accuracy.

TABLE 7. Crack classification results.

Crack type	Correct	Error	Accuracy	Severity level (No)
Transversal	132	7	95.0%	Light (20) Heavy (119)
Longitudinal	149	6	96.1%	Light (47) Heavy (108)
Block	51	7	87.9%	Light (17) Heavy (41)
Alligator	56	9	86.2%	Light (13) Medium (28) Heavy (24)

B. CRACK TYPE LABELING AND SEVERITY ASSIGNMENT

For simple linear cracks, assigning the severity level of the detected crack segment depends on the measurement of the crack width, which is calculated as the ratio of the crack area to the number of cracked pixels in the skeleton. The severity level of damage is assigned as a light crack with a width of less than 3 mm, and a heavy crack with more than 3 mm width. The crack classification algorithm was verified in 300 crack classification samples; and Table 7 shows the crack classification system evaluation results.

The netted cracks include block and alligator cracks. In general, they have different crack densities and widths. However, due to the crisscrossing of the netted cracks, the width was difficult to calculate and measure, and the more intensive density of block cracks often exceeded the alligator cracks. Therefore, it was difficult to estimate the width and density of each block and alligator crack; thus, it is impossible to characterize the type and severity level accordingly. By further observing and comparing, it was found that the key difference between block and alligator cracks is that both of them break the road surface into pieces, but the former has fewer pieces and the distance between the branches is larger, whereas the latter has more blocks and the distance between the crack branches is smaller. From the perspective of intuitive perception and machine recognition, the block and alligator cracks can be effectively distinguished according to the branch distribution between the cracks. They also have practical physical meanings, such as the smaller the distance between the crack branches, the higher the severity level.

It can be seen from Table 7 that the proposed algorithm classifies four crack types (transversal, longitudinal, block, and alligator) and has competitive performance. The classification accuracy of the method for transversal and longitudinal cracks is 95% and 96.1%, respectively. For the classification

of netted cracks, the distance between the cracks and the average width are considered comprehensively. The classification accuracy of block and alligator cracks is 87.9% and 86.2%, respectively. The classification accuracy of netted cracks is significantly lower, the reason for this confusion is that influence is exerted on the threshold for selecting the average distance and the average width when characterizing the netted crack. Nonetheless, the crack type classification method proposed has significant benefits in terms of a deal with the adjacent and intersecting crack targets, and it can still be treated as a whole when discontinuity occurs between cracks. It can be seen that the classification method is highly consistent with the definition of cracks. At the same time, thanks to accurate calculation of the size distribution between cracks, the quantitative characterization criteria are further clarified.

V. CONCLUSION

In this paper, a novel trainable convolutional network was proposed for automatic detection of cracks in complex environments. In consideration of the different characteristics of different level features, we designed an MDA feature extraction module containing different dilated convolutions at multiple scales and a channel-wise attention module to capture the semantic high-level features. Then, crack pixel-level prediction is achieved by an FFU module that is combined with low-level features and continuous convolution. The experimental results show that both the MDA module and the FFU module contribute to the improvement of crack detection performance. Compared to other segmentation networks, our proposed crack detection network achieves state-of-the-art performance with Precision 98.74%, Recall 98.05%, F-score 98.39% and mIoU 74.81%. The experimental results indicate that our network is insensitive to noise crack marking and can effectively distinguish the low contrast caused by shadows, stains, and exposures during data acquisition.

Cracks were labeled according to the types defined in the Chinese distress category, with each different crack present in a given image receiving the appropriate label. Moreover, a novel methodology for the assignment of crack severity levels was introduced. The conclusion can be drawn from the experiment that our method is a universal and robust automatic method to simultaneously determine the type of crack and severity levels—both of which are crucial for roadway agencies to assess pavement quality. For future developments in research, we will continue to investigate the influence of the attention mechanism on crack feature extraction. Likewise, the crack classification algorithm will be optimized, especially for classification of the block and alligator cracks. Additionally, other types of distress, such as potholes and crack sealings will be taken into account to improve the procedure of automatic crack detection.

REFERENCES

[1] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Adv. Eng. Inform.*, vol. 29, no. 2, pp. 196–210, Apr. 2015.

[2] H.-S. Yoo and Y.-S. Kim, "Development of a crack recognition algorithm from non-routed pavement images using artificial neural network and binary logistic regression," *KSCE J. Civil Eng.*, vol. 20, no. 4, pp. 1151–1162, May 2016.

[3] A. Mohan and S. Poobal, "Crack detection using image processing: A critical review and analysis," *Alexandria Eng. J.*, vol. 57, no. 2, pp. 787–798, Jun. 2018.

[4] A. Ahmadi, S. Khalesi, and M. Bagheri, "Automatic road crack detection and classification using image processing techniques, machine learning and integrated models in urban areas: A novel image binarization technique," *J. Ind. Syst. Eng.*, vol. 11, pp. 85–97, Sep. 2018.

[5] Q. Li, Q. Zou, D. Zhang, and Q. Mao, "FoSA: F* Seed-growing approach for crack-line detection from pavement images," *Imag. Vis. Comput.*, vol. 29, no. 12, pp. 861–872, Nov. 2011.

[6] Q. Li and X. Liu, "Novel approach to pavement image segmentation based on neighboring difference histogram method," in *Proc. 1st Congr. Image Signal Process. (CISP)*, vol. 2, May 2008, pp. 792–796.

[7] F. Liu, G. Xu, Y. Yang, X. Niu, and Y. Pan, "Novel approach to pavement cracking automatic detection based on segment extending," in *Proc. Int. Symp. Knowl. Acquisition Modeling (KAM)*, Wuhan, China, Dec. 2008, pp. 610–614.

[8] 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.*, Oct. 2014, pp. 778–782.

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

[10] Z. Qu, L. Bai, S.-Q. An, F.-R. Ju, and L. Liu, "Lining seam elimination algorithm and surface crack detection in concrete tunnel lining," *Proc. SPIE*, vol. 25, no. 6, Nov. 2016, Art. no. 063004.

[11] D. H. Ai, G. Y. Jiang, S.-K. Lam, and C. W. Li, "Automatic pixel-level pavement crack detection using information of multi-scale neighborhoods," *IEEE Access*, vol. 6, pp. 24452–24463, 2018.

[12] 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. 2014.

[13] 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.

[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[15] 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.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, Munich, Germany, 2015, pp. 234–241.

[17] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[18] F.-C. Chen and R. M. R. Jahanshahi, "NB-CNN: Deep learning-based crack detection using convolutional neural network and Naïve Bayes data fusion," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4392–4400, May 2018.

[19] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 12, pp. 1127–1141, Jun. 2018.

[20] 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.

[21] 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.

[22] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, no. 10, p. 3452, Oct. 2018.

[23] 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.

[24] A. Zhang, K. C. P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J. Q. Li, and C. Chen, "Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network," *J. Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 10, pp. 805–819, Oct. 2017.
[25] Y. Fei, K. C. P. Wang, A. Zhang, C. Chen, J. Q. Li, Y. Liu, G. Yang, and B. Li, "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V," *IEEE Trans. Intell. Transp. Syst.*, to be published.
[26] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "DeepCrack: Learning hierarchical convolutional features for crack detection," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1498–1512, Mar. 2019.
[27] Q. Song, Y. Wu, X. Xin, L. Yang, M. Yang, H. Chen, C. Liu, M. Hu, X. Chai, and J. Li, "Real-time tunnel crack analysis system via deep learning," *IEEE Access*, vol. 7, pp. 64186–64197, 2019.
[28] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," Dec. 2017, *arXiv:1706.05587*. [Online]. Available: https://arxiv.org/abs/1706.05587
[29] 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.
[30] A. Ouyang, Q. Dong, Y. Wang, and Y. Liu, "The classification of pavement crack image based on beamlet algorithm," in *Proc. Int. Conf. Comput. Comput. Technol. Agricult. (CCTA)*, 2013, pp. 129–137.
[31] H. Oliveira and P. L. Correia, "Automatic road crack detection and characterization," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 155–168, Mar. 2013.
[32] W. Wang, A. Zhang, K. C. P. Wang, A. F. Braham, and S. Qiu, "Pavement crack width measurement based on Laplace's equation for continuity and unambiguity," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 2, pp. 110–123, Feb. 2018.
[33] A. Cubero-Fernandez, F. J. Rodriguez-Lozano, R. Villatoro, J. Olivares, and J. M. Palomares, "Efficient pavement crack detection and classification," *EURASIP J. Image Video Process.*, vol. 2017, no. 1, Jun. 2017, Art. no. 39.
[34] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jun. 2017, pp. 2881–2890.
[35] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters–Improve semantic segmentation by global convolutional network," in *Proc. IEEE CVPR*, Honolulu, HI, USA, Jun. 2017, pp. 4353–4361.
[36] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," Apr. 2017, *arXiv:1704.06904*. [Online]. Available: https://arxiv.org/abs/1704.06904
[37] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," Apr. 2018, *arXiv:1804.09337*. [Online]. Available: https://arxiv.org/abs/1804.09337
[38] Z. Yan, W. Liu, S. Wen, and Y. Yang, "Multi-label image classification by feature attention network," *IEEE Access*, vol. 7, pp. 98005–98013, 2019.
[39] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, Munich, Germany, 2018, pp. 801–818.
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.
[41] L. Ying and E. Salari, "Beamlet transform-based technique for pavement crack detection and classification," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 25, no. 8, pp. 572–580, Jun. 2010.

WEIDONG SONG received the Ph.D. degree from the Liaoning Technical University, Liaoning, China. He is currently a Professor with the Liaoning Technical University of China. His main research interests include remote sensing image processing, 3D reconstruction, computer vision, and deep learning.

GUOHUI JIA received the B.S. degree from the University of Science and Technology Liaoning, China. He is currently pursuing the Ph.D. degree with the Liaoning Technical University, China. His research interests include semantic segmentation, computer vision, and deep learning.

DI JIA received the Ph.D. degree in computer application from Northeastern University. He is currently an Associate Professor with the Liaoning Technical University of China. His main research interests include stereo matching and 3D reconstruction, and computer vision positioning system.

HONG ZHU received the Ph.D. degrees from the Liaoning Technical University, Liaoning, China. She has completed her postdoctoral experience at the Land Satellite Remote Sensing Application Center, Beijing, China. She has authored some articles in the field of remote sensing image processing and has coauthored one book. Her research interests include remote sensing image processing, satellite attitude, and deep learning.