Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 6412562, 11 pages
https://doi.org/10.1155/2020/6412562

Research Article

Automated Pavement Crack Damage Detection Using Deep Multiscale Convolutional Features

Weidong Song,[1] **Guohui Jia**,[1] **Hong Zhu**,[2] **Di Jia**,[3] **and Lin Gao** [1]

[1]*School of Geomatics, Liaoning Technical University, Fuxin 123000, China*
[2]*College of Ecology and Environment, Institute of Disaster Prevention, Beijing 101601, China*
[3]*School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China*

Correspondence should be addressed to Guohui Jia; jgh6080@163.com

Received 23 July 2019; Accepted 6 November 2019; Published 8 January 2020

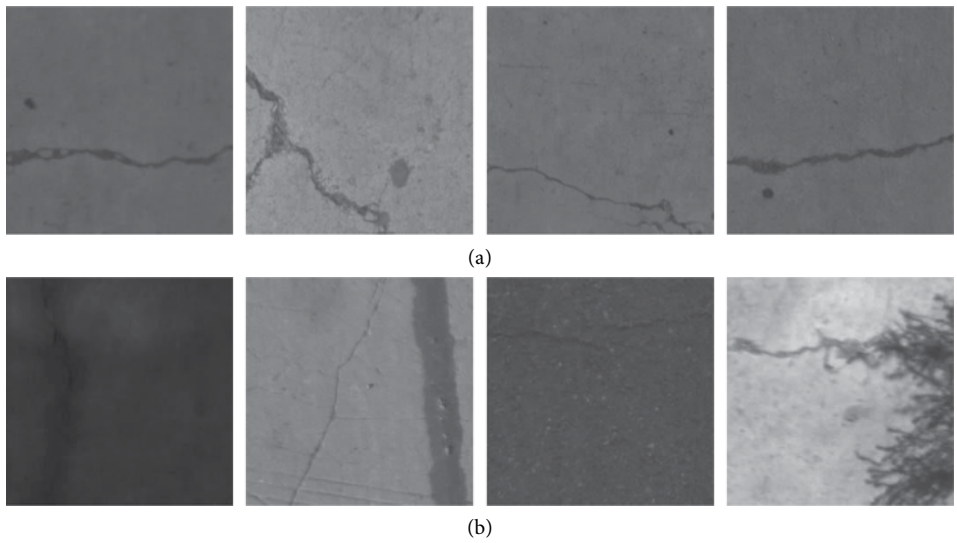
Academic Editor: Roberta Di Pace

Road pavement cracks automated detection is one of the key factors to evaluate the road distress quality, and it is a difficult issue for the construction of intelligent maintenance systems. However, pavement cracks automated detection has been a challenging task, including strong nonuniformity, complex topology, and strong noise-like problems in the crack images, and so on. To address these challenges, we propose the CrackSeg—an end-to-end trainable deep convolutional neural network for pavement crack detection, which is effective in achieving pixel-level, and automated detection via high-level features. In this work, we introduce a novel multiscale dilated convolutional module that can learn rich deep convolutional features, making the crack features acquired under a complex background more discriminant. Moreover, in the upsampling module process, the high spatial resolution features of the shallow network are fused to obtain more refined pixel-level pavement crack detection results. We train and evaluate the CrackSeg net on our CrackDataset, the experimental results prove that the CrackSeg achieves high performance with a precision of 98.00%, recall of 97.85%, F-score of 97.92%, and a mIoU of 73.53%. Compared with other state-of-the-art methods, the CrackSeg performs more efficiently, and robustly for automated pavement crack detection.

1. Introduction

Pavement crack detection plays an important role in the field of road distress evaluation [1]. Traditional crack detection methods depend mainly on manual work and are limited by the following: (i) they are time consuming and laborious; (ii) they rely entirely on human experience and judgment. Therefore, automatic crack detection is essential to detect and identify cracks on the road quickly and accurately [2]. This procedure is a key part of intelligent maintenance systems, to assist and evaluate the pavement distress quality where more continual road status surveys are required. Over the past decade, the development of high-speed mobile cameras and large-capacity hardware storage devices has made it easier to obtain large-scale road images. Through mobile surveying and mapping technology, integrated acquisition equipment is fixed to the rear of the vehicle roof frame to monitor both the road surface and the surrounding environment. The images can be acquired by processing and storing pavement surface images that are realized [3]. Currently, many methods utilize computer vision algorithms to process the collected pavement crack images and then obtain the final maintenance evaluation results [4].

Automatic crack detection is a very challenging image classification task with the goal of accurately marking crack areas. Figure 1 shows examples of data acquisition by a mobile pavement inspection vehicle. In a few cases, the cracks have good continuity and obvious contrast, as shown in Figure 1(a). However, in most cases, there is a considerable noise in cracks, which leads to poor continuity and low contrast, as shown in Figure 1(b). Therefore, automatic crack detection mainly includes the following three challenges. (i) In a poorly lit environment and complex background, the texture, and linearity of interference (weeds, stains, etc.) have similar features, resulting in greater intraclass differences. (ii) Boundary blurring occurs between small cracks and local noises. (iii) Blurred low-quality images from crack data collected at high speed are unavoidable. These three difficulties create considerable challenges in pavement crack detection.



(a)

(b)

Figure 1: Crack detection challenges. (a) The crack images in the ideal case. (b) Poor lighting conditions, stains, small cracks, and the occlusion of branches increase the difficulty of crack recognition.

The recent publications [5–7] assumed that the crack pixels are generally darker than their surroundings and then used the threshold method to extract the crack area. These methods lack the description of global information and are sensitive to noise. To improve the continuity of crack detection, researchers have attempted to detect cracks by introducing minimal path selection (MPS) [8–10], minimal spanning trees (MSTs) [11–13], and crack fundamental elements (CFEs) [14]. These methods can partially eliminate noise and improve crack detection continuity. However, only the low-level features can be roughly obtained, some complex high-level crack features may not be presented, and utilized correctly. A randomly structured forest-based method is presented in [15] to detect cracks automatically. This method can effectively suppress noise by manually selecting crack features and learning internal structures. Although it improves the recognition speed and accuracy but does not perform well when dealing with complex pavement crack situations. Therefore, traditional machine learning methods simulate cracks by manually setting color or texture features. In these methods, the features cover only some specific real-world situations. The set of crack features is simplified and idealized, which cannot achieve the robust detection requirements for pavement diseases.

In recent years, deep learning methods have been widely used to solve complex problems through hierarchical concepts. A deep convolutional neural network (DCNN) has shown great advantages in computer vision tasks, such as image classification [16–18], object detection [19], and semantic segmentation [20, 21]. The DCNN can acquire expressive features at different levels as it consists of several trainable layers [22]. The rich hierarchical features of DCNN have made great progress in pixel-level semantic segmentation tasks [23–24] and crack detection. In [25], the AlexNet is used to extract the crack characteristics, and then crack detection is performed based on probability maps. However, the detailed division of the crack could not be completed. In [26] and [27], 3D crack detection networks based on DCNN are proposed for automated pixel-level crack detection on 3D asphalt pavement. In [28], an effective detection model for concrete cracks is proposed through two modules of multi-view image feature detection and multitask crack detection. A robust algorithm by postprocessing the output feature mapping is proposed in [31] to detect cracks. The DeepCrack net is constructed based on the encoder-decoder architecture of the SegNet in [32], and the convolution features generated in the encoder network and decoder network are fused in pairs at the same scale to complete crack detection, but the width information of cracks may not be considered in the detection results.

Although most of the published methods have achieved ideal results, automated pavement crack detection in the complex backgrounds is still demanding. In this paper, we propose an end-to-end trainable deep convolutional neural network, called the CrackSeg, for pixel-wise crack detection from a complex scene. First, a multiscale dilated convolution module is proposed to obtain more abundant crack texture information. Additionally, to satisfy networks with a larger receptive field and spatial resolution, multiscale context information is captured by different dilated rates. Second, a pixel-level dense prediction mapping is generated by fusing the upsampling module of low-level features to recover the crack boundary details. Finally, the model is systematically evaluated in three crack data sets by quantitative evaluation methods, including comparing the results with manual marking. The results show that the proposed crack detection method can accurately extract cracks in different pavement types and complex backgrounds.

The contributions of this paper mainly include three aspects as follows:

(1) A novel trainable end-to-end crack segmentation network, the CrackSeg, is designed to detect road cracks at the pixel level. The network makes full use of the semantic information of hierarchical convolution features and is very effective for crack detection under a complex scene.

(2) A crack feature detection network based on a joint multiscale dilated convolution module is proposed. While the computational cost is controlled, the multiscale semantic information is fused to obtain more abundant features. In the upsampling stage, the high spatial resolution features of the shallow network are fused to obtain more refined crack segmentation results.

(3) A multisource pavement distress labelling dataset, the CrackDataset, is established that reflects the overall situation of road diseases in China.

The rest of this paper is organized as follows. Section 2 describes crack detection based on deep learning semantic segmentation. Section 3 demonstrates the effectiveness of the proposed scheme through comparative analyses of experiments. Section 4 discusses the detailed design of the two modules proposed in this paper. Finally, Section 5 concludes the paper.

2. Materials and Methods

In this section, we introduce a novel end-to-end trainable crack detection DCNN structure based on multiscale features, which is divided into three parts. In the first part, the overall structure of the crack detection network is introduced. In the second part, a multiscale dilated convolution module is introduced to obtain more abundant context information in the crack image, and feature mapping after crack detection fusion is preliminarily obtained. In the third part, we propose a new upsampling scheme based on different resolution feature maps.

2.1. The Structure of CrackSeg. The crack detection network is proposed based on a multiscale dilated convolution module and an upsampling module, as shown in Figure 2. The ResNet [31] pretraining model with a dilated network strategy is used to extract the crack characteristics. In the traditional CNN network structure, the use of a down-sampling layer can effectively increase the receptive field, and reduce the number of calculation parameters but also reduce the spatial resolution of learning features, making the final feature mapping size smaller. After final feature representation, multiscale crack semantic information is obtained by using a multiscale dilated convolution module, and global prior information is captured by fusing different levels of semantic information. Finally, the shallow and deep semantic information is fused by the upsampling module so that the network output feature mapping size is consistent with the input image size, and the probability that each pixel belongs to cracks or noncracks is calculated by the softmax function. The vector value $[0, 1]$ generated by the softmax function represents the probability distribution of a class, and the softmax function can be expressed as:

$$\hat{y}_j = \text{softmax}\left(X, W_j\right) = \frac{e^{X^T W_j}}{\sum_{k=1}^{K} e^{X^T W_j}}, \tag{1}$$

where W_j and Xrepresent the weights of the network and input data, respectively. In the task of pixel-wise prediction loss, different pixels are divided into different categories by cross-entropy loss [32] and can be expressed as:

$$L(\boldsymbol{W}) = -\frac{1}{N}\sum_{n=1}^{N} y_n \log \hat{y}(x_n, \boldsymbol{W}) + (1 - y_n)(1 - \log \hat{y}(x_n, \boldsymbol{W})), \tag{2}$$

where $x_n \in [0, 255]$ is the input pixel value, $y_n \in \{0, 1\}$ is the ground truth label, $\hat{y}_n \in \{0, 1\}$ is the prediction probability, $\boldsymbol{W}$ is the network weight matrix, and L is the loss function.

2.2. Multiscale Dilated Convolution. Using a top-down convolutional neural network can identify the target region with strong discrimination, but for the target region with weak discrimination, the classification performance is reduced [33]. In DCNNs, the size of the receptive field represents the amount of available information. Increasing the receptive field of convolution kernels can effectively mix the semantic information around the target, thus improving the classification ability of regions with weak discrimination [34]. Dilated convolution is a special form of standard convolution. Zero values are inserted between the pixels of the convolution kernel to increase the image resolution of intermediate feature maps, thus enabling dense feature extraction in DCNNs with an enlarged convolution kernel field:

$$RF_{K_l} = k + (k - 1) \times \left(D_k - 1\right), \tag{3}$$

where D_k represents the dilated rate of convolutional kernel K_l to specify the number of zeros placed between pixels. Because of the dilation, only $k \times k$ pixels are involved in convolution calculation, which increases the receptive field, and reduces the computational cost, thus increasing the receptive field without losing resolution.

Inspired by the mentioned findings, a novel classification network with multiple dilated convolution blocks is proposed to generate dense localization. To capture the multiscale semantic crack information, the features of three different scales are fused. A multiscale dilated convolution module is constructed by combining multibranching with different kernels and dilated convolution layers, which forms a merged feature representation for different locations. After fusing different sizes and levels of feature, the 1×1 kernel size convolution operation is carried out to reduce the dimension of semantic features to $1/N$. The network structure and parameter details of the multiscale dilated convolution module are shown in Figure 3.

In the multiscale dilated convolution module, two main convolution operations are used: (i) obtaining accurate location mapping through standard convolution kernels to highlight the target areas with strong discrimination; (ii) introducing multiple dilated rates to expand the convolution kernel receptive fields to improve the target areas with weak discrimination. Thus, discriminant features from adjacent salient regions are transformed into target-related regions that have not been found. We find that convolution blocks with a large dilation rate introduce some irrelevant regions, such as some true-negative regions, which would be

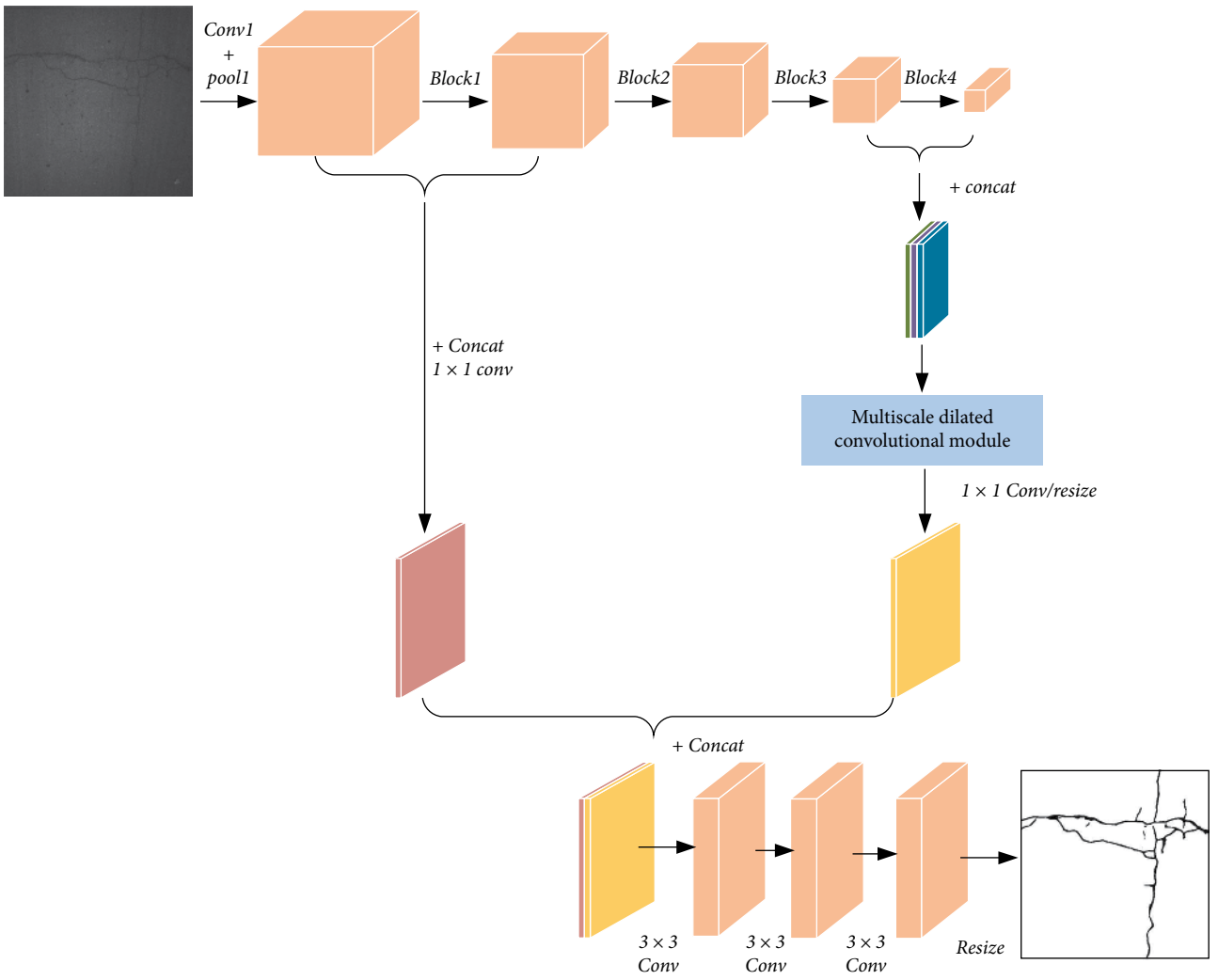


Figure 2: Illustration of the crack detection network, CrackSeg. The multiscale dilated convolution module is used to capture abundant crack features. After fusing with the lower level crack features in the network, three 3×3 convolution operations are used continuously to improve the feature expression ability. The output feature of the last convolution layer is the crack feature maps, which is the input into the binary classifier for crack pixel-wise prediction.

highlighted by adjacent discriminant objects. Therefore, a multiscale dilated convolution network with a small dilation rate is proposed.

2.3. Up-Sampling Module. The multiscale dilated convolution module in the encoding stage can transform the input image into rich semantic visual features. However, these features have a rough spatial resolution [35]. The purpose of upsampling is to restore these features to the input image resolution and then predict the crack spatial distribution.

The proposed upsampling module contains mainly two inputs: the low-resolution features with high-level semantic information and the high-resolution features on the bottom of the network that use features extracted at different scales to aggregate local and global context information. As shown in Figure 4, the features of the shallower encoding layer retain more spatial details, which helps to obtain sharper boundaries; the deeper features have stronger representation ability. The upsampling module first samples the output features of the multiscale dilated convolution four times and then fuses with the low-level features with the same spatial resolution in the network. To reduce the dimension of low-level features and convolute them by 1×1, three convolution operations with

Table 1: Details of upsampling module units.

Items	Kernel size	Numbers of feature maps	Feature size
Input	—	—	H, W
Low-level features	1×1	256	H/4, W/4
High-level features	1×1	512	H/16, W/16
Conv-1	3×3	256	H/4, W/4
Conv-2	3×3	256	H/4, W/4
Conv-3	3×3	256	H/4, W/4
Output	1×1	2	H, W

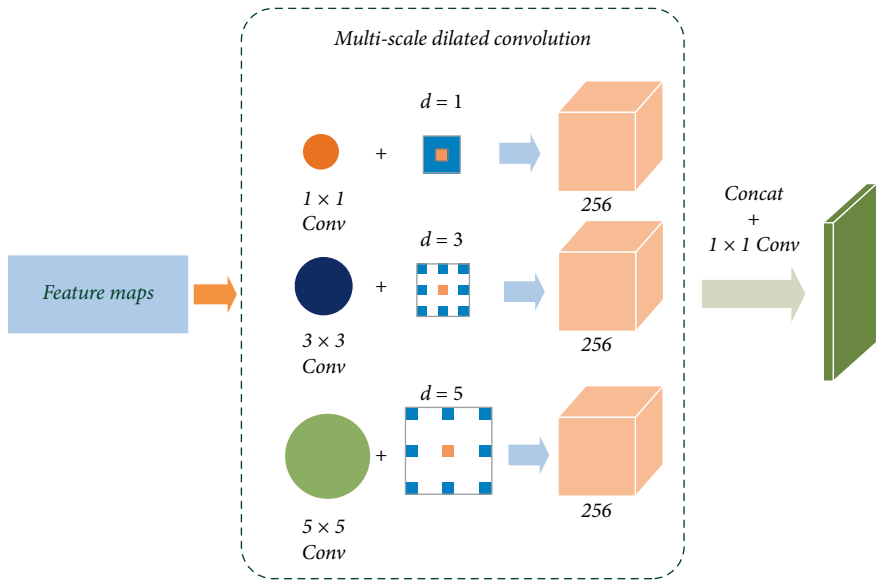


Figure 3: Multiscale dilated convolution module.

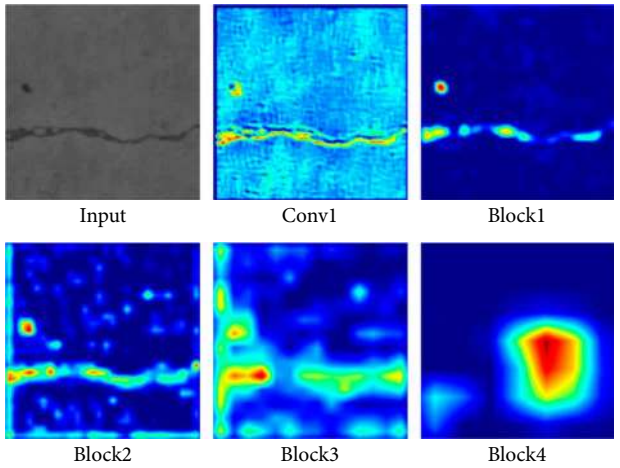


Figure 4: Visualization results of crack feature maps at different levels, in which the feature resolution of the shallow network is higher, retaining the crack characteristic details. The deep features are more abstract and have strong discriminant power.

3×3 kernel size are used to improve the feature expression ability after feature fusion. Because the upsampling module is learnable, it can recover the fine information lost in the bilinear upsampling (BU) operation. Details of the parameters of the upsampling module are described in Table 1, where H and W are the height and width dimensions of the input features, respectively.

3. Experiments and Analysis

To verify the effectiveness of our scheme, extensive experiments on pavement crack detection were conducted on various images. In this section, we depict the experimental setup and analyze our experimental results.

Table 2: The details of datasets.

Datasets	Training set	Validation set	Test set
Ours	4746	1036	2416
CFD [15]	—	—	118
AigleRN [10]	—	—	38

3.1. Experimental Setup

3.1.1. Dataset. Our CrackDataset consists of pavement detection images of 14 cities in the Liaoning Province, China. The data cover most of the pavement diseases in the whole road network. These images include collected images of different pavement, different illumination, and different sensors. The real values in the dataset provide two types of labels, cracks, and noncracks. The dataset is divided into three parts. The training set and the validation set are composed of 4736 and 1036 crack images, respectively. The test set contains 2416 images. In addition, two other crack datasets, CFD [15] and AigleRN [10], are used as test sets. The details of the datasets are shown in Table 2.

3.1.2. Implementation Details. We implement our CrackSeg using the TensorFlow, which is an open source platform for deep learning. Because of the large image size, training the CrackSeg network requires a large amount of memory, which results in overburdening the training process. Additionally, the crack areas occupy a small proportion of the whole image, and many background areas are meaningless for the training process. Therefore, the original road crack images are divided into several small blocks with a size of 256×256. To improve the robustness of the model, several transformations are made to the data, including random flip, color enhancement, and enlargement. We utilize the Adam [36] algorithm to converge the network. The network is trained with an initial learning rate of 0.0001. The momentum and weight decay are set to

Table 3: Comparisons of the proposed and other methods on the CrackDataset.

Method	OA	Precision	Recall	F-score	mIoU
CrackForest [15]	87.04	86.28	85.46	85.86	59.26
SegNet [38]	96.64	96.86	97.08	96.97	70.56
U-Net [21]	96.58	96.99	97.09	97.04	71.49
PSPNet [39]	96.25	96.90	96.88	96.89	69.63
DeepLabv3+ [40]	96.83	97.01	97.64	97.32	71.77
DeepCrack [31]	97.14	97.33	97.72	97.52	72.04
CrackSeg	98.79	98.00	97.85	97.92	73.53

0.9997 and 0.0005, respectively. All experiments in our work are performed using an NVIDIA GTX 1080 GPU and 8 GB of on-board memory.

3.1.3. Evaluation Metrics. In the evaluation of crack detection accuracy, crack and noncrack pixels are considered as two categories. The overall accuracy (OA), precision, recall, F-score, and mIoU are used as the metrics for the quantitative performance evaluation and comparison method in the experiment. These five indicators can be calculated as follows:

$$\text{Overall accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{5}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{6}$$

$$F_1\text{score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{7}$$

$$\text{mIoU} = \frac{\text{Intersection areas of detected and reference crack}}{\text{Union areas of detected and reference crack}}, \tag{8}$$

where TP represents the number of positive cases correctly divided, FP is the number of incorrectly classified positive pixels, FN is the number of incorrectly classified negative cases, OA, mIoU, and F-score are comprehensive indicators, and the larger the value is, the higher the accuracy.

3.2. Result and Analysis. To demonstrate the feasibility of the proposed scheme, we compare our CrackSeg with SegNet [38], U-Net [21], PSPNet [39], DeepCrack [31], and DeepLabv3+ [40]. In addition, to verify the advantages of the deep learning semantic segmentation model in crack detection, the nondeep learning method CrackForest is introduced to compare based on different comparative experiments.

The quantitative comparison testing results in our CrackDataset are shown in Table 3, which shows that the crack detection accuracy based on the deep learning method in a complex background is higher and has good advantages. Compared with other segmentation methods based on deep learning, the CrackSeg achieves the highest OA, recall, precision, mIoU, and F-score. The mIoU value of CrackSeg reached the highest 73.53%, followed by DeepCrack, and Deeplabv3+, with the mIoU of 72.04 and 71.77. The mIoU of CrackForest, SegNet, U-Net, and PSPNet are 14.27%, 2.97%, 2.04%, and 3.90% lower than the results of CrackSeg. The performance improvement is mainly due to the use of a multiscale dilated convolution module in the encoding stage, which captures a multiscale context for accurate semantic mining. On the basis of obtaining rich semantic information, the boundary information of the target is recovered by using low-level high-resolution features, and more accurate segmentation results are obtained by using continuous convolution operations.

Figure 5 describes the visual comparisons of the crack detection results using different methods. The first row is the original image containing cracks, some of which are accompanied by noise such as shadows, oil spots, and watermarks, which are the main factors affecting the detection of cracks. The experimental results show that the CrackForest method based on traditional machine learning features can extract cracks in a simple background, but it still retains more noise and cannot adapt to the automatic crack detection in complex scenes. For SegNet and U-Net, the detection results are acceptable, but these methods produce many false detections in complex backgrounds. The DeepCrack performs well in extracting the thin cracks in the complex backgrounds, however, some width information of cracks is lost in the detection results. The DeepLabV3+ has good performance in detecting light cracks, but nonexistent cracks occur because of its large dilation rate. Furthermore, its single convolution kernel size causes the loss of crack information. Our CrackSeg integrates low-level and high-level features in convolution stages at different scales and can further improve the accuracy of crack detection and robustness of background artifact suppression, effectively eliminate the influence of oil pollution, shadow, and complex backgrounds, and extract various complex topological crack relationships.

3.3. Network Robustness Analysis. To verify the stability of our proposed method, the other two datasets (CFD and AigleRN) are tested by CrackSeg. The visual crack detection results are shown in Figure 6. It is noteworthy that this method does not use the crack images in these two datasets in the training phase. The results show that the proposed method can extract most pavement cracks and that the model has strong robustness.

4. Discussion

In this section, to determine the optimal crack characteristics, we discuss the self-impact of the multiscale dilated convolution module. Then, the low-level feature selection and convolution operation structure in the upsampling module are discussed.

4.1. Principle for Choosing Multiscale Dilated Convolution. To compare the effect of the multiscale dilated convolution module on crack detection more clearly, the features are sampled 16 times in the upsampling stage by BU, and the

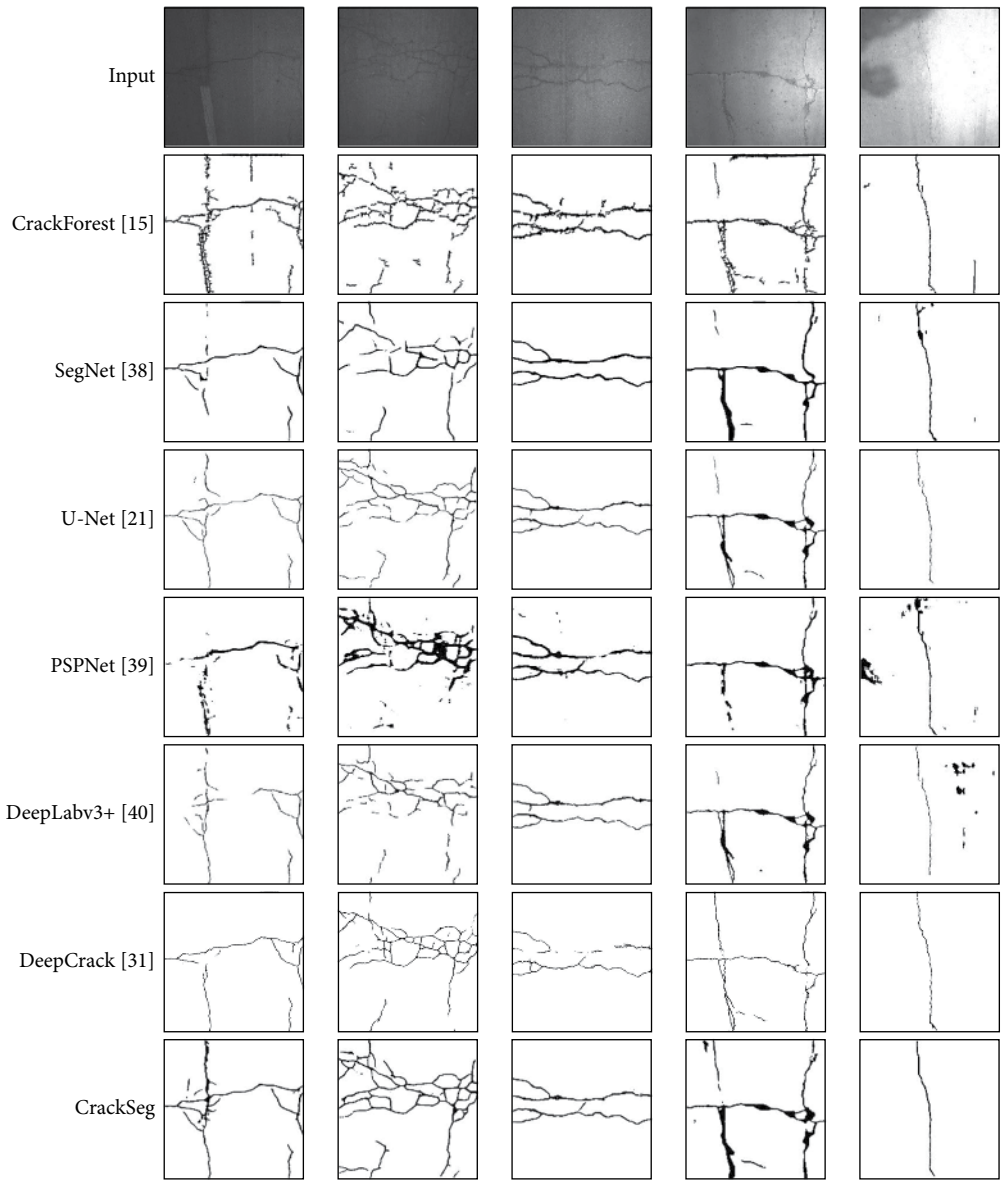


Figure 5: Comparison of results obtained by different methods on five sample images selected from our CrackDataset.

final prediction results are obtained. In the experiment, ResNet50 is used as the network backbone to validate the multiscale dilated convolution module. Figure 7 shows the change in mIoU of different dilated convolution modules after 20 epochs in the training stage. After 14 epochs, each method reaches a stable state. The BaseLine has the lowest performance, and the purple polyline (fusion-S-dilated) represents the highest mIoU score compared with the other methods. In summary, the multiscale dilated convolution module with fusion features achieves the best results.

As shown in Table 4, the experimental results are compared and analyzed for the selection of high-level features. The mIoU of the BaseLine model using ResNet50 as the feature detection network is only 65.07%. The performance of the ASPP [33] module is 1.25% higher than that of the BaseLine, which shows that dilated convolution can improve the performance of crack detection. To facilitate the effect of dilated convolution, different dilation rates were applied to the final Block4, and the receptive field size is increased. The experimental results show that the multiscale dilated convolution module with a large dilation rate and a small dilation rate increases by 1.47% and 2.05%, respectively. Although dilated convolution with a larger dilation rate has a larger receptive field, it introduces other unrelated regions while capturing crack characteristics, which affect the final crack identification. With the smaller dilation rate, better optimal convergence, and better detection effect can be obtained in model training. To explore the influence of different high-level features on the multiscale dilated convolution module, two high-level crack feature maps, Block3 and Block4, were fused in the experiment, and the network performance improved by 0.82%. The experimental results show that the high-level features fused at multiple levels have stronger representation ability, which helps locate crack pixels in the encoding process.

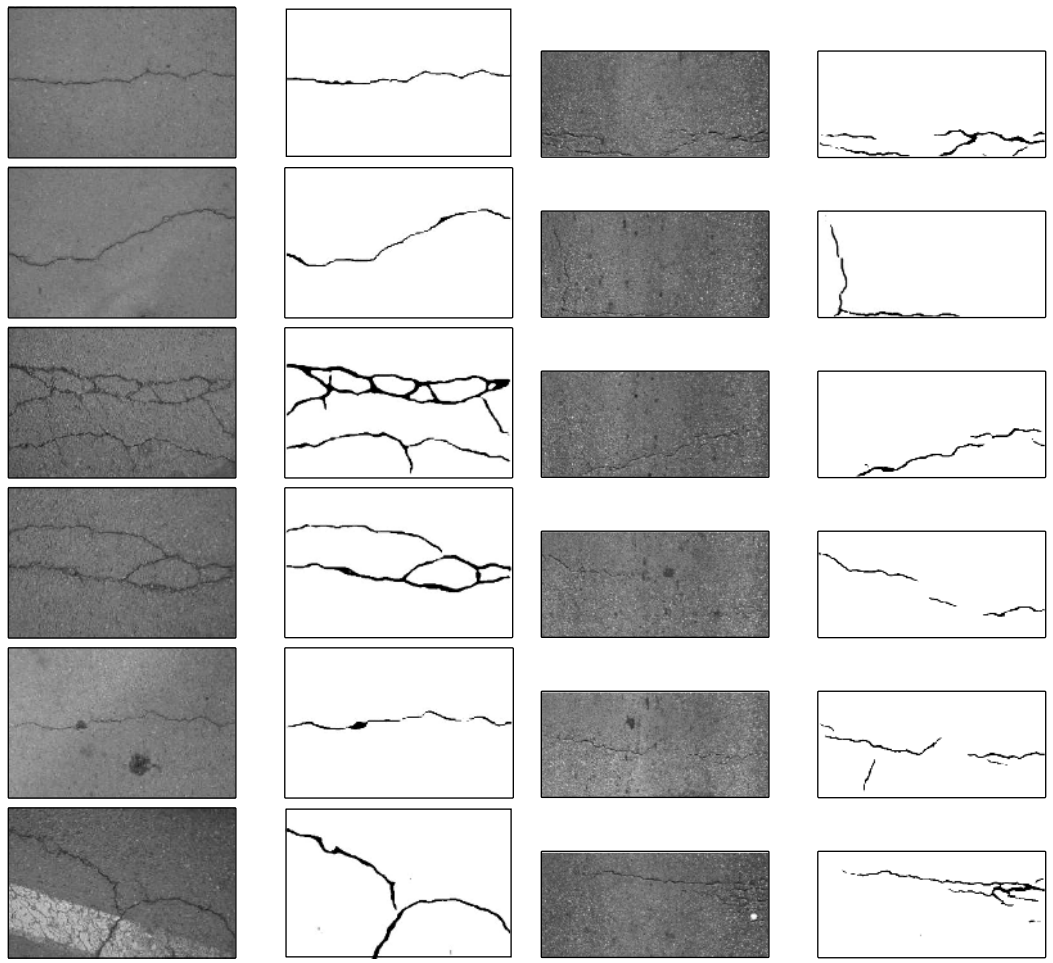


Figure 6: The results of crack detection on (a) CFD and (b) AigleRN datasets.

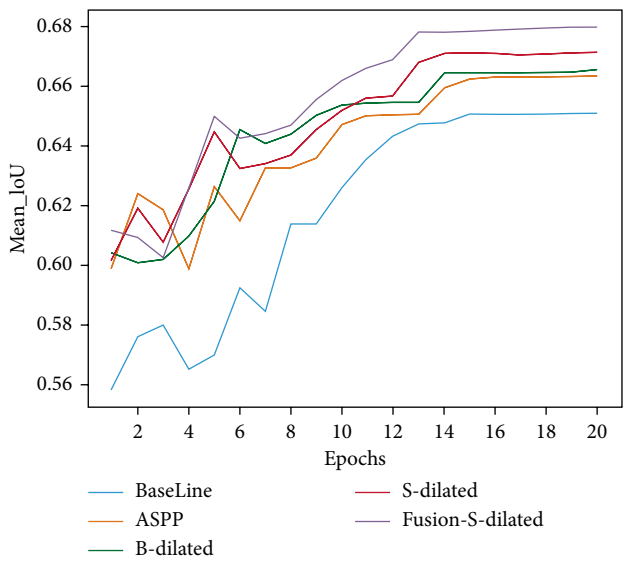


Figure 7: The curves of mIoU with different training stages.

4.2. Performance Improvement when Using Different Upsampling Strategies. As shown in Table 4, only 67.94% of the mIoU is obtained by using the simple BU method. To improve the crack detection accuracy, the features generated by the multiscale dilated convolution module are used as the high-level input feature of the upsampling module, which includes more discriminant semantic information. Low-level features in the network have a high spatial resolution, which retains the details of the crack boundary. After the fusion of low-level semantic information and discriminative high-level features, the convolution operations are used in the upsampling module to obtain sharper detection results. Table 5 shows the performance of the different upsampling features and structures. Through comparative analyses of experiments, the choice of the number of convolutions has a great impact

Table 4: Comparisons of the results with different feature maps and dilation rates in the multiscale dilated convolution module.

Method	Features		Dilation rate		OA	mIoU
	Block3	Block4	{1, 3, 5}	{6, 12, 18}		
BaseLine		✓			97.29	65.07
ASPP		✓		✓	97.29	66.32
B-dilated		✓		✓	97.31	66.54
S-dilated		✓	✓		97.11	67.12
Fusion-S-dilated	✓	✓	✓		97.32	67.94

Table 5: Comparisons of the results with different upsampling features and structures in the upsampling module.

Backbone	Low-level features		3×3 Conv structure	OA	mIoU
	Conv1	Block1			
ResNet50	✓		(3×3, 256)×1	97.33	70.43
	✓		(3×3, 256)×2	97.55	71.16
	✓		(3×3, 256)×3	97.56	71.28
	✓		(3×3,256)×4	97.60	70.86
		✓	(3×3, 256)×3	97.68	71.68
	✓	✓	(3×3, 256)×3	98.43	71.82
ResNet101	✓	✓	(3×3, 256)×3	98.79	73.53

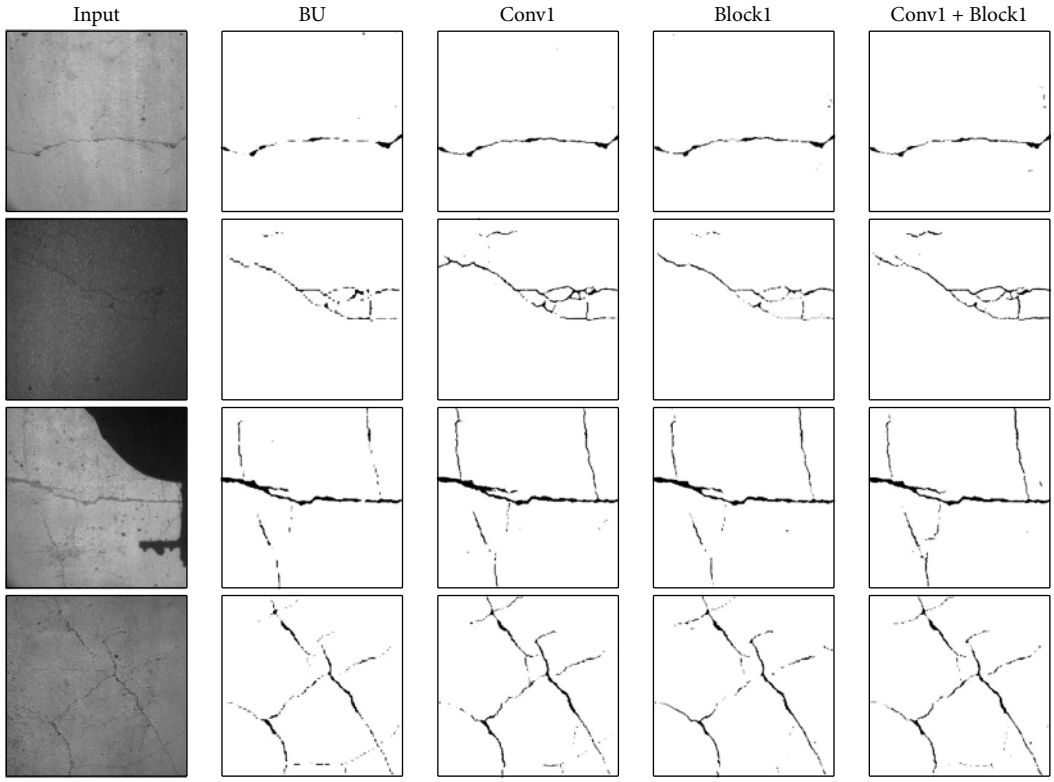


Figure 8: Comparisons of the crack detection results with different low-level features in upsampling modules.

on the final crack detection results of the model. After fixing the low-level features generated by Conv1, the best results are achieved by using three [3×3,256] convolution, compared with using one convolution operation and two convolution operations, and the mIoU values are increased by 0.85% and 0.12%, respectively. When convolution operations are used four times, the accuracy of crack detection begins to decline.

To evaluate the effectiveness of low-level features on boundary restoration, the low-level features generated by Conv1 in the upsampling module of the network are changed to Block1 and the combination of the two modules (Conv1 and Block1). As shown in Figure 8, the features generated by the combination of Conv1 and Block1 can restore the best crack detection edge. In Table 5, the OA and mIoU values are 98.43% and 71.82%, respectively. Compared with the single low-level feature Conv1 and Block1, the mIoU values increase by 0.96% and 0.14%, respectively. Compared with the simple BU method, the mIoU values increased by 3.88%. Using the ResNet101 network as the backbone for crack detection, the OA, and mIoU values reach 98.79% and 73.53%, respectively. Thus, the use of the upsampling module in the CrackSeg network combines shallow crack features with deep semantic information, which helps to aggregate multilevel features of cracks and improve the accuracy of crack detection.

5. Conclusions

In this paper, an end-to-end trainable pavement crack detection framework based on DCNN, CrackSeg, is proposed, which can automatically detect road cracks under complex backgrounds. First, a crack training dataset is established,

which covers a wide range of data sources and reflects the overall situation of pavement distress in the Liaoning Province, China. Second, through the fusion of high-level features in the backbone network, we propose the multiscale dilated convolution module. By capturing the features of context information at multiple scales, the crack detection network can learn rich semantic information in a complex background. Therefore, based on the dilated convolution theory, we design a novel network structure that can be inserted into the existing semantic segmentation system to improve the accuracy of crack feature detection. Finally, through the upsampling module, the low-level features, and continuous convolution features are fused to realize the crack pixel-level prediction. This feature aggregation, which combines different levels of feature information, can not only fully mine the crack features in the image but also restore and describe the details of the object boundary information. The experimental results of CrackSeg achieve high performance with a precision of 98.00%, recall of 97.85%, F-score of 97.92%, and a mIoU of 73.53%, which are higher than those of other networks. Furthermore, the model has strong stability and robustness to solve the noise interference caused by shadows, stains, and exposures in the process of data acquisition. The good performance of the CrackSeg network provides a possibility for large area automatic crack detection.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Public Welfare Research Fund in Liaoning Province (No. 20170003), the Key Natural Science Plan Fund of Liaoning Province (No. 20170520141), the National Natural Science Foundation of China (project Nos. 41871379 and 61601213).

Supplementary Materials

CrackDataset is an annotated road crack image database which can reflect road surface condition in general. (*Supplementary Materials*)

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," *Advanced Engineering Informatics*, vol. 29, no. 2, pp. 196–210, 2015.

[2] A. Cubero-Fernandez, F. J. Rodriguez-Lozano, R. Villatoro, J. Olivares, and J. M. Palomares, "Efficient pavement crack detection and classification," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, 2017.

[3] H. Zakeri, F. M. Nejad, and A. Fahimifar, "Image based techniques for crack detection, classification and quantification in asphalt pavement: a review," *Archives of Computational Methods in Engineering*, vol. 24, no. 4, pp. 935–977, 2017.

[4] Y.-C. Tsai, V. Kaul, and R. M. Mersereau, "Critical assessment of pavement distress segmentation methods," *Journal of Transportation Engineering*, vol. 136, no. 1, pp. 11–19, 2010.

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

[6] Q. Li and X. Liu, "Novel approach to pavement image segmentation based on neighboring difference histogram method," in *2008 Congress on Image and Signal Processing*, pp. 792–796, IEEE, Piscataway, NJ, 2008.

[7] F. Liu, G. Xu, Y. Yang, X. Niu, and Y. Pan, "Novel approach to pavement cracking automatic detection based on segment extending," in *2008 International Symposium on Knowledge Acquisition and Modeling*, pp. 610–614, IEEE, Piscataway, NJ, 2008.

[8] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart, "A new minimal path selection algorithm for automatic crack detection on pavement images," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 788–792, IEEE, Piscataway, NJ, 2014.

[9] Y. Xu, Y. Guizhen, W. Yunpeng, W. Xinkai, and M. Yalong, "Car detection from low-altitude UAV imagery with the faster R-CNN," *Journal of Advanced Transportation*, vol. 2017, Article ID 2823617, 10 pages, 2017.

[10] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart, "Automatic crack detection on two-dimensional pavement images: an algorithm based on minimal path selection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, pp. 2718–2729, 2016.

[11] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, "CrackTree: automatic crack detection from pavement images," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227–238, 2012.

[12] K. Fernandes and L. Ciobanu, "Pavement pathologies classification using graph-based features," in *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 793–797, IEEE, Piscataway, NJ, 2014.

[13] H. Chen, H. Zhao, D. Han, and K. Liu, "Accurate and robust crack detection using steerable evidence filtering in electroluminescence images of solar cells," *Optics and Lasers in Engineering*, vol. 118, pp. 22–33, 2019.

[14] Y.-C. Tsai, C. Jiang, and Y. Huang, "Multiscale crack fundamental element model for real-world pavement crack classification," *Journal of Computing in Civil Engineering*, vol. 28, no. 4, p. 04014012, 2014.

[15] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, pp. 1097–1105, 2012.

[17] H. K. Kim, J. H. Park, and H. Y. Jung, "An efficient color space for deep-learning based traffic light recognition," *Journal of*

Advanced Transportation, vol. 2018, Article ID 2365414, 12 pages, 2018.

[18] O. Sudakov, E. Burnaev, and D. Koroteev, "Driving digital rock towards machine learning: predicting permeability with gradient boosting and deep neural networks," *Computers & Geosciences*, vol. 127, pp. 91–98, 2019.

[19] R. Girshick, "Fast r-cnn," in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1440–1448, IEEE, Piscataway, NJ, 2015.

[20] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer International Publishing, Cham, 2015.

[22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[23] G. L. Oliveira, C. Bollen, W. Burgard, and T. Brox, "Efficient and robust deep networks for semantic segmentation," *The International Journal of Robotics Research*, vol. 37, pp. 472–491, 2018.

[24] S. Zheng, S. Jayasumana, B. Romera-Paredes et al., "Conditional random fields as recurrent neural networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1529–1537, IEEE, Piscataway, NJ, 2015.

[25] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, p. 3452, 2018.

[26] A. Zhang, K. C. P. Wang, B. Li et al., "Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network," *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 10, pp. 805–819, 2017.

[27] Y. Fei, K. C. P. Wang, A. Zhang et al., "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 273–284, 2020.

[28] B. Wang, W. Zhao, P. Gao, Y. Zhang, and Z. Wang, "Crack damage detection method via multiple visual features and efficient multi-task learning model," *Sensors*, vol. 18, p. 1796, 2018.

[29] Y. Li, H. Li, and H. Wang, "Pixel-wise crack detection using deep local pattern predictor for robot application," *Sensors*, vol. 18, p. 3042, 2018.

[30] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, "Deepcrack: learning hierarchical convolutional features for crack detection," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2019.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, IEEE, Piscataway, NJ, 2016.

[32] S. M. Azimi, P. Fischer, M. Körner, and P. Reinartz, "Aerial LaneNet: lane-marking semantic segmentation in aerial imagery using wavelet-enhanced cost-sensitive symmetric fully convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 2920–2938, 2018.

[33] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2921–2929, IEEE, Piscataway, NJ, 2016.

[34] S. Liu, D. Huang, and Y. Wang, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 404–419, Springer International Publishing, Cham, 2018.

[35] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[36] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd International Conference for Learning Representations*, ICLR, San Diego, CA, 2014.

[37] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: a deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[38] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6230–6239, IEEE, Piscataway, NJ, 2017.

[39] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 833–851, ECCV, Glasgow, UK, 2018.