IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 4, 2019, accepted January 12, 2020, date of publication January 15, 2020, date of current version January 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2966881

Review of Pavement Defect Detection Methods

WENMING CAO[1,2,3], QIFAN LIU[1,2], AND ZHIQUAN HE[1,2]

[1]Shenzhen Key Laboratory of Media Security, Shenzhen University, Shenzhen 518060, China
[2]Guangdong Multimedia Information Service Engineering Technology Research Center, Shenzhen 518060, China
[3]Video Processing and Communication Laboratory, Department of Electrical and Computer Engineering, University of Missouri, Columbia, MO 65211, USA

Corresponding author: Zhiquan He (zhiquan@szu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61971290, Grant 61771322, and Grant 61871186.

ABSTRACT Road pavement cracks detection has been a hot research topic for quite a long time due to the practical importance of crack detection for road maintenance and traffic safety. Many methods have been proposed to solve this problem. This paper reviews the three major types of methods used in road cracks detection: image processing, machine learning and 3D imaging based methods. Image processing algorithms mainly include threshold segmentation, edge detection and region growing methods, which are used to process images and identify crack features. Crack detection based traditional machine learning methods such as neural network and support vector machine still relies on hand-crafted features using image processing techniques. Deep learning methods have fundamentally changed the way of crack detection and greatly improved the detection performance. In this work, we review and compare the deep learning neural networks proposed in crack detection in three ways, classification based, object detection based and segmentation based. We also cover the performance evaluation metrics and the performance of these methods on commonly-used benchmark datasets. With the maturity of 3D technology, crack detection using 3D data is a new line of research and application. We compare the three types of 3D data representations and study the corresponding performance of the deep neural networks for 3D object detection. Traditional and deep learning based crack detection methods using 3D data are also reviewed in detail.

INDEX TERMS Crack detection, image processing, deep learning, 3D imaging.

I. INTRODUCTION

With the rapid development of road traffic, people have paid more and more attention to the importance of pavement maintenance as road surface cracks not only affect the transportation efficiency but also pose a potential threat to vehicle safety. Many studies have been conducted to detect the cracks of pavement surfaces. In early pavement crack detection system, people analyzed the road images collected by line scan or area scan cameras to examine the road conditions. Such systems include the GERPHO [1] system used in France, the DHDV [2] detection system of American expressway, and the PAVUE [3] system of IMS in Sweden and so on. The development of hardware technologies such as the appearance of CCD [4] digital photography has greatly advanced the effect of pavement crack detection.

Defect detection is to distinguish the part with defect features from other defect free parts in the image, which has both links and differences with image segmentation. From the Wikipedia definition [5], image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics. The idea of image segmentation can be used to segment the defect and the rest part of the image. The defects appearing on the surface may have various shapes and types. Fig.1 shows a few examples. Therefore, defect detection usually contains two subtasks, i.e. locate the defect pixels and classify the type of the defects.

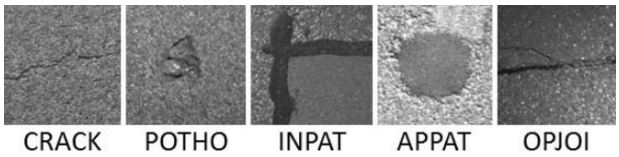

FIGURE 1. Sample surface defect types: CRACK-Crack, POTHO-Pothole, INPAT-Inlaid patch, APPAT-Applied patch, OPJOI-Open joint.

Researchers have conducted in-depth researches on road crack detection and proposed many methods to crack the problem, from image processing to machine learning methods, including deep learning methods which have been widely used nowadays. Image processing methods mainly

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

include three categories [6], threshold segmentation, edge detection and region growing methods. The threshold segmentation method divides the image pixels into several categories by setting a proper pixel intensity threshold, so as to separate the target crack from the background. The edge detection method detects the edges of the road crack through edge detection operators such as Sobel operator [7], Prewitt operator [8], and Canny operator [9]. The region growing method depicts the specific information inside the crack by assembling the pixels with similar characteristics to form a region.

The emergence of machine learning makes road crack detection rise to a new level. Image processing techniques can only be able to analyze some superficial defect features, while machine learning can learn some deep features. Machine learning takes advantage of the similarity between data through the design of algorithms, so that the computer can master the learning rules and predict from the unknown data by itself. Especially, deep learning methods have greatly advanced the accuracy of pavement crack detection.

Unlike other types of surface defects, pavement cracks are usually deep and have large size, such as block cracks and alligator cracks [10]. It is practically meaningful to measure and detect the depth of the cracks. The detection of crack depth can predict the future trend of the crack, which is helpful to repair the pavement in time and reduce potential safety risks [11]. In recent years, 3D imaging technology has achieved great progress, making cracks detection in 3D images has become a new research direction for scholars. Owing to the extra depth dimension, the 3D structure of road cracks can be constructed from the 3D images. Besides this, 3D images can reduce the effect of shadow and other noise [12].

In recent years, there have been several reviews available from the literature. Sylvie *et al*. summarized the application of image processing technologies in road detection, and proposed a new automatic road cracks detection and evaluation comparison protocol [13]. In the work of [14], Gopalakrishnan compared some deep learning frameworks, networks and hyper-parameters used in pavement crack detection, and classified the previous papers, which provided a good reference for developing pavement crack detection models. Tom *et al*. listed different kinds of pavement defects, discussed different defect detection methods and assessed different defect data acquisition devices [15]. In [16] Mathavan *et al*. discussed the detection of road surface lesions from the perspective of 3D image defect detection, summarized the application of 3D imaging technologies in road surface monitoring, analyzed the imaging principle of different devices and compared the advantages and disadvantages of different pavement detection technologies. These reviews address different emphasis or aspect on road surface detection. In this review, we provide a comprehensive review of pavement crack detection methods, especially the in-depth analysis of deep learning and 3D image based methods.

The rest of the paper is organized as follows. Section II briefly reviews the crack detection methods mainly based on image processing techniques. Crack detection based on machine learning methods, including unsupervised learning, traditional supervised learning and deep learning, are reviewed in Section III. Section IV talks about the 3D imaging technologies and corresponding methods for pavement defect detection. Discussions about the existing problems and the prospect of crack detection is presented in Section V. Section VI concludes this work.

II. CRACK DETECTION BASED ON IMAGE PROCESSING

Pavement is exposed to the natural environment for long time, often affected by rain, shadow, stains and other factors. Therefore, the images captured by imaging sensors usually contain a lot of noises, textures and interferences. Cracks on images appear as thin, irregular, dark curves, surrounded by strong textured noise. Researchers have proposed various image processing methods to reduce the influence of the noise on detection. These methods mainly include three categories: threshold segmentation, edge detection and region growing.

A. THRESHOLD SEGMENTATION METHODS

Threshold segmentation [17] is a classical method in image segmentation. For each pixel in the image, we can judge whether its characteristic attributes meet a threshold requirement to determine the pixel belongs to the target area or the background. This way, we can convert a gray image into a binary image. Let $f(x, y)$ be the original image and T be the threshold value, image segmentation can be written as

$$g(\mathrm{x}, \mathrm{y}) = \begin{cases} 1, & f(x, y) \geq T \\ 0, & f(x, y) < T \end{cases}$$

Obtaining reasonable threshold value is the key of this method. Dynamic threshold method and local threshold method have achieved good results in pavement defect detection. Oliveira and Correia [18] recognized the potential cracks by identifying dark pixels in images with dynamic threshold. In their work, thresholded images are divided into non-overlapping blocks by entropy computation, and secondary dynamic threshold of the generated Entropy Block Matrix is used as the basis for identifying image blocks containing crack pixels. Peng *et al*. proposed a twice-threshold segmentation [19]. Firstly, the improved Otsu threshold segmentation algorithm was used to remove the road markers in the runway image. Then, the improved adaptive iterative threshold segmentation algorithm was used to segment images which removed the markers. Finally, the outline of the crack can be obtained through morphological denoising. In [20], a new multi-scale local optimal threshold segmentation algorithm was proposed to segment pavement cracks through crack density distribution. Compared with the global threshold method and the optimal threshold method, this method achieved a better segmentation effect.

FIGURE 2. Detection effect of different edge operators.

B. EDGE DETECTION METHODS

Edge detection methods can also be used in crack detection. Common edge detection operators include Sobel operator, Roberts operator, Prewitt operator and Canny operator. Different operators have different detection effects on edge of the same type. Fig. 2 shows an example. Simply using a single operator can hardly reach the expected effect. Many scholars have improved the edge detection operators. Zhao *et al*. proposed an improved Canny edge detection method for road edge detection [21]. Mallat wavelet transform was used to enhance the blurred edge, and a better adaptive threshold Canny algorithm is obtained by using genetic algorithm [22]. Ayenu-Prah and Attoh-Okine [23] studied the road crack detection method which combines bi-dimensional empirical mode decomposition (BEMD) and Sobel edge detection. BEMD is an extension of EMD [24], which removes noise from the signal without the need for complex convolution processes.

C. REGION GROWING METHODS

The edge detection algorithm can get the edge distribution of crack defects and outline the crack contour, but it can not describe the information of internal pixels of cracks concretely. The recognition method based on region growing provides another idea for pavement crack detection. The basic idea of region growing is to gather similar pixels to form a region. The selection of seeds is very important, which greatly affects the accuracy of image segmentation. In the work of [25], after the road surface image was preprocessed, the lane was marked and the uneven background part was also processed. Then, the crack seeds were selected by grid cell analysis and connected by Euclidean minimum spanning tree structure. In this way, cracks can be detected quickly and effectively. Li *et al*. proposed an automatic cracks detection method based on FoSA-F* seed growth for better detection of blurred and discontinuous cracks [26]. It exploited seed-growing strategy to eliminate the requirement that start and end points should be surrounded in advance. The global search space is reduced to the interested local space to improve the search efficiency.

III. CRACK DETECTION BASED ON MACHINE LEARNING

Machine learning has become a hot research topic and widely used in various areas. It can give predictions by learning the rules embedded in the data. Supervised learning and unsupervised learning are commonly used for cracks detection and analysis.

A. UNSUPERVISED LEARNING METHODS

The biggest difference between unsupervised learning and supervised learning is absence of data labels in training. Training samples for unsupervised learning have no labels and no definite results for output, the computer needs to learn the similarity between samples by itself and classify the samples. The advantage of unsupervised learning is that there is no need to label, reducing the influence of human subjective factors on the results.

Akagic *et al*. proposed a new unsupervised road crack detection method based on gray histogram and Otsu method, and a better results were obtained under the condition of low signal-to-noise ratio [27]. In [28], Amhaz *et al*. introduced an improved unsupervised learning algorithm based on minimum path selection, which reduced the loop and peak artifacts in crack detection by estimating the crack width. In [29], Li *et al*. used a method based on the minimum intensity path of the window to extract candidate cracks at each scale in the image, compared the corresponding relations of different scale cracks, established a crack evaluation model based on multivariate statistical hypothesis.

B. SUPERVISED LEARNING METHODS

Supervised learning needs the labels of the training data. Common supervised learning algorithms include logistic regression [30], Naive Bayesian [31], Support Vector Machine [32], artificial neural network [33] and random forest [34]. Xu *et al*. used the self-learning characteristic of neural network to transform cracks recognition into crack probability judgment of each sub-block image in the work of [35]. They first divide the binary image of cracks into sub-images and extract the parameters representing the features of crack from each sub-image, then select representative images to train back propagation neural network. In [36], Crack Forest, a road crack detection framework based on random structure forest, was proposed to effectively solve the problems of uneven edge cracks and cracks with complex topological structures. The authors extracted crack features from multiple levels and directions to train the random forest model. In [37], an automatic pavement crack detection scheme is proposed. Firstly, the crack image is preprocessed to smooth its texture and enhance any existing cracks. Then the image is divided into several non-overlapping blocks, each block produces a feature vector, and the supervised learning algorithm support vector machine is used to detect the cracks. These methods heavily rely on the high-quality features extracted from the images, which needs careful design of the algorithms.

1) DEEP LEARNING METHODS

In recent years, deep learning technologies have achieved tremendous success in various computer vision tasks such as image classification, object detection and image segmentation [38]–[42]. Many deep learning based methods, especially deep convolution neural networks, have been proposed for

road crack detection. According to the way of handling the crack detection problem, these methods can roughly divided into three categories, pure image classification methods, object detection based methods and pixel-level segmentation methods.

a: CRACK DETECTION BASED ON CLASSIFICATION

Basically, this category of methods divide the input image into overlapping blocks, and then classify the block image into classes. If the block contains a certain number of defect pixels or more, the block is labeled as defective block.

Crack Detection Based on Binary Classification: This kind of methods divide the input images into overlapping blocks and then use a deep convolution network to decide if the block contains crack or not. For example, Lei *et al*. divided the road image of 3264 $\times$ 2248 into small patches of size 99 $\times$ 99 $\times$ 3, and used their convolution neural network to classify these small patches [43]. The output is the probability that the small patch is crack or not. In the work of [44], Li *et al*. modified GoogLeNet [45] to classify image blocks and realized crack detection on real pavement using smartphone. In [46], Cha *et al*. used MatConvNet [47] to classify the input pavement 256 $\times$ 256 images. Similarly, in [43], the authors generated image patches of 99 $\times$ 99 from original pavement images, where the patch is defective if its center pixel is within 5 pixels of the crack center. The CNN model was compared to the performance of SVM and boosting methods. Leo *et al*. studied the relationship between network depth and network accuracy using a self-designed CNN model [48]. Unlike the work mentioned above, Chen *et al*. processed pavement videos in [49]. In this work, a CNN model was designed to classify the image patches of size 120 $\times$ 120 sampled from video frame and then adopted a naive Bayes data fusion scheme to aggregate the information obtained from each video frame to enhance the overall performance and robustness of the system.

Crack Detection Based on Multi-Class Classification: Crack detection based binary classification is not suitable for the case when it is required to decide the defect types. In [50], Fan *et al*. used one CNN model to learn the structure of the pavement cracks as a multi-label classification problem. Small crack image patches of 27 $\times$ 27 were used as the input and the output layer had $s \times s$ nodes, representing the intensity states of square block centered at the crack pixel. For example, if $s = 5$, the model predicts 25 pixel state of the block image of 5 $\times$ 5. During training, the input 27 $\times$ 27 was resized to 5 $\times$ 5 as the ground truth. In [51], Li *et al*. proposed a deep CNNs for pavement crack classification based on 3D pavement images, and classify pavement patches cut from 3D images into five categories including the normal category. They trained four supervised CNNs classification models with different sizes of receptive field, and find that different size of receptive field have a slight effect on the classification accuracy. The method proposed by Wang and Hu [52] is quite different from above methods. In this work, the input pavement images are segmented into non-overlapping grids of size 32 $\times$ 32 or 64 $\times$ 64, then a simple CNN is used to classify the grid image to decide if it contains crack. After this, crack skeleton can be represented by the grid cells containing cracks. PCA (principal component analysis) is used to process the coordinate vector of the crack grid cells to decide the crack type to be longitudinal, transverse or alligator crack.

b: CRACK DETECTION BASED ON PIXEL SEGMENTATION

Pixel segmentation is to assign a label or a score to each pixel in the image. In [50] Fan *et al*. proposed a network structure with 4 convolutional layers with 2 max-pooling layers and 3 Fully Connected layers to directly segment the original images. The output can have different resolution, from 1 $\times$ 1 to 5 $\times$ 5. In [53] Jenkins *et al*. proposed a semantic segmentation algorithm for road cracks based on U-Net, where the U-Net is basically encoder-decoder structure [54]. This network can be divided into encoder layer and decoder layer. The encoder layer mainly realizes feature mapping of images, and the decoder layer is mainly used to promote feature vectors during segmentation and generate probability distribution of each pixel. Similarly, Zou *et al*. [55] proposed DeepCrack which uses encoder-decoder architecture to segment pavement image pixels into crack and background. And in [56], the propose network structure used 4 convolution layers and max poolings as the encoder to extract features and 4 subsequent modules as the decoder. The work of [57] employed residue connections inside each encoder and decoder block and attention gating block before the decoder to retain only spatially relevant features of the feature map in the shortcut connection. Fully convolutional network is also often used for segmentation purpose, such as [58], [59].

c: CRACK DETECTION BASED ON OBJECT DETECTION

Object detection is an important task in computer vision. Its goal is to locate the object with a bounding box in the image and decide the object type. Many deep CNN models have been proposed to improve the accuracy and efficiency, such as faster R-CNN [60], SSD [61], YOLO [62] etc. Object detection methods are also popular in road crack detection.

Faster R-CNN is widely used in object detection, which has three major steps, 1) extract image features using CNN structure like VGG, 2) propose candidate regions for objects (RPN), 3) classification of object types and bounding box coordinates regression. The CNN structure in step 1 is shared by step 2 and 3. In [63], Suh and Cha used faster R-CNN to detect the damages in civil infrastructure. Cha *et al*. modified the faster R-CNN by using a ZF-net to speedup the feature extraction in step 1 [64]. ZF-net [65] is slightly modified from AlexNet [66] which is relatively simple and fast. In [67] Li *et al*. used the faster R-CNN to detect six kinds of road defects. The model can automatically identify and locate defects under different lighting conditions with high accuracy and stability.

SSD [61] combines predictions from multiple feature maps with different resolutions to naturally handle objects of various sizes and completely eliminates proposal generation

and encapsulates the region classification and coordinates regression in a single network. This makes SSD much faster than faster R-CNN. And MobileNet [68] is a well known light weight deep neural networks for mobile applications. To test the crack detection on devices with limited resources, Hiroya *et al*. compared SSD using MobileNet, SSD using Inception v2 [69] for object detection on smart phones and found that SSD using Inception v2 is two times slower than SSD-MobileNet [70]. This conclusion is not surprising as MobileNet is designed for acceleration purpose.

Unlike above methods, Crack-pot method in [71] combined traditional image processing techniques and deep learning methods to detect the potholes and cracks in the road. In these method, edge detection, dilation, contour detection were applied to generate candidate bounding boxes for suspected potholes and cracks. Then these regions were feed into a classification model which is modified from SqueezeNet [72] by replacing the last pooling layer with a learned dictionary [73].

Methods based on object detection like SSD and faster R-CNN propose multiple candidate regions and perform the location regression using the image features extracted from CNN structure is a systematic way for object detection. For defects with compact shapes, these methods may work well. However, for defects like long curves or scratches on the surface, the methods may fail to detect due to the overly large bounding box proposed by the Region Proposal Network (RPN).

2) METRICS TO EVALUATE MODEL PERFORMANCE

a: PRECISION, RECALL AND F1

The three most commonly used parameters for evaluating crack detection performance are precision, recall, and F1. Precision is the ratio of the correct detected results to all the actual detected results, recall is the ratio of the correct detected results to all the results that should be detected. The F1 is the harmonic mean of the precision and the recall. $Precision = \frac{TP}{TP+FP}$, $Recall = \frac{TP}{TP+FN}$ and $F_1 = \frac{2*TP}{2*TP+FP+FN}$. The detection accuracy is defined as $Acc = \frac{TP+TN}{TP+TN+FP+FN}$. Table 1 shows the definition of FN (False Negative), FP (False Positive), TN (True Negative) and TP (True Positive).

TABLE 1. Definition of FN, FP, TN and TP.

		Prediction	
		Positive	Negative
Ground Truth	Positive	TP	FN
	Negative	FP	TN

b: ROC, AUC, and IOU

ROC (Receiver Operating Characteristic) [74] curve and AUC (Area Under Curve) [75] can also be used to measure the detection performance. ROC curve describes the relationship between TP rate and FP rate. Fig. 3 shows two ROC curves. If the ROC curve is closer to the upper left corner, that's mean, FP is low, TP is high, and the better the model works. Therefore, the Area under the ROC curve, namely AUC is used to compare two ROC curves.

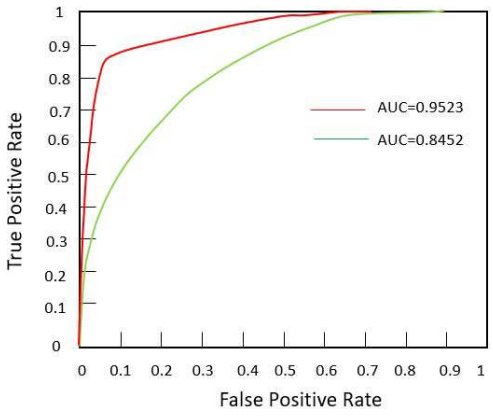

FIGURE 3. Two ROC curves and AUC.

In object detection using models such as SSD, IOU (Intersection over Union) is often used to decide if the object is correctly detected. The IOU means the overlap rate between the bounding box given by the model and the ground truth bounding box. If the IOU is larger than a predefined threshold, which is usually 0.5, the object detection is considered successful.

$$IOU = \frac{\text{Detection Result} \cap \text{Ground Truth}}{\text{Detection Result} \cup \text{Ground Truth}}$$

c: AIU, ODS and OIS

In [76], the authors proposed three new evaluation metrics, AIU, ODS and OIS. AIU is the average intersection over union between the predicted area and ground truth area. ODS represents the best F1 score on the dataset with fixed scale, and OIS represents the aggregated F1 score on the dataset with the best proportion of each image. ODS and OIS are defined as follows:

$$\text{ODS} = \max\left\{2\frac{P_t \times R_t}{P_t + R_t} : \text{t} = 0.01, 0.02, \ldots, 0.99\right\}$$

$$\text{OIS} = \frac{1}{N_{img}}\sum_{i}^{N_{img}} \max\left\{2\frac{P_t^i \times R_t}{P_t + R_t} : \text{t} = 0.01, 0.02, \ldots, 0.99\right\}$$

where t represents the threshold value, i is the index of image, N_{img} is the total number of images, P_t and R_t are precision and recall at threshold t on the dataset. P_t^i and R_t^i represent the accuracy rate and recall rate on image I respectively.

3) PUBLIC DATASETS FOR ROAD CRACK DETECTION

Road crack detection has been research topic for years. There are many public datasets to help us do better research.

a: CRACKFOREST DATASET (CFD)

The CrackForest dataset consists of 118 images of cracks on urban road surface in Beijing taken by iphone5. Each image is resized to 480×320 pixels and has been labeled. It is available at https://github.com/cuilimeng/CrackForest-dataset.

b: AIGLERN DATASET

AigleRN dataset contains 38 pre-processed gray-scale images on French pavement. Half of them are 991 × 462 and half of them are 311 × 462. The dataset is available at http://telerobot.cs.tamu.edu/bridge/Datasets.html.

c: CRACK500

500 pictures of pavement cracks with the size of 2000 × 1500 were taken by smartphone. Each crack image has a binary mask image for annotation. The dataset is divided into three parts, 250 images for training, 50 for validation, and 200 for test. It is available at https://github.com/fyangneil/pavement-crack-detection.

d: GAPs DATASET

German asphalt pavement disease (Gaps) dataset, including 1969 gray-scale pavement images, is partitioned into 1418 training images, 51 validation images, and 500 test images. The image resolution is 1920 × 1080 pixels. It is available at http://www.tu-ilmenau.de/neurob/data-sets-code/gaps/.

e: RESULTS ON BENCHMARK DATASETS

The following tables list the results comparison on different benchmark datasets. In Table 2 and Table 3, the tolerance margin is the number of pixels of the predicted pixel away from the ground truth pixel when we count the true negatives. For example, if the tolerance margin is 2, a ground truth pixel is hit if there is a predicted pixel within its 2-pixel neighborhood. AIU, ODS, OIS are used to compare the performance of different methods on CRACK500 dataset in Table 4.

TABLE 2. Test results on CFD dataset.

Method	Tolerance margin (pixel)	Precision	Recall	F1
FPCNet [56]	2	0.9748	0.9639	0.9693
FCN [58]	2	0.9729	0.9456	0.9590
Fan *et al.* [50]	2	0.9119	0.9481	0.9244
U-Net-A [59]	5	0.9693	0.9345	0.95
U-Net-B [59]	5	0.9731	0.9428	0.9575

TABLE 3. Test results on AigleRN dataset.

Method	Tolerance margin (pixel)	Precision	Recall	F1
König *et al.* [57]	2	0.8690	0.9304	0.8986
Fan *et al.* [50]	2	0.9178	0.8812	0.8954
CrackForest [36]	5	0.9028	0.8658	0.8839
U-Net [77]	5	0.9202	0.9321	0.9261

Reference [80] presented GAPs dataset to test pavement defect type classification. On this dataset, the authors compared four methods, shown in Table 5, where the RCD net [43] is just a simple and small CNN with four blocks of alternating convolutional and max-pooling layers, and the ASINVOS net [80] is modified from RCD net by adding more blocks, the ASINVOS-mod [80] is a further version of ASINVOS net by replacing large convolutional filters by multiple smaller filters.

TABLE 4. Results comparison on CRACK500 dataset.

Method	AIU	ODS	OIS
HED [78]	0.481	0.575	0.625
RCF [79]	0.403	0.490	0.586
FCN [58]	0.379	0.513	0.577
CrackForest [36]	N/A	0.199	0.199
FPHBN [76]	0.489	0.604	0.635

TABLE 5. Test results on Gaps dataset.

Method	Acc	F1
Crack-pot [71]	0.9893	0.7314
ASINVOS net [80]	0.9772	0.7246
ASINVOS-mod [80]	0.9723	0.6707
RCD net [43]	0.9732	0.6642

4) DATA AUGMENTATION

The training of deep neural network model requires a large amount of data. However, it is costly to acquire and label this amount of data. Data augmentation is an effective technique to relieve the problem. Common data augmentation methods include image rotation, flipping, mirroring, adding noise, changing the illumination etc. These techniques are usually combined to get more data. Table 6 shows the data augmentation techniques used in road crack detection.

IV. CRACK DETECTION BASED ON 3D DATA

Most of existing crack detection methods are based on 2D images. With the development of stereo camera and range-based sensors, stereovision is becoming a promising approach in crack detection as it can provide accurate and robust data for the depth information.

A. REPRESENTATION OF 3D DATA

Basically, there are three kinds of 3D data representations, namely, multi-view, point cloud and voxel data.

Earlier representations of 3D images were made through multi-view. Multi-view represents a collection of 2D images of a rendered polygon grid captured from different viewpoints to convey 3D geometry in a simple manner, as shown in Fig.4(a). This method is easy to understand, but difficult to express the spatial structure of 3D data. On the other hand, since multi-view projections can only represent 2D contours of 3D objects, some detailed geometrical information is inevitably lost during the projection process [81].

Point cloud is a set of points in the 3D space, where each point is specified by the 3D coordinates (x, y, z) and other information such as RGB value of color. These huge amount of points are used to interpolate the geometric shape

TABLE 6. Data augmentation.

Reference	Dataset type	Methods	Result
[43]	500 images (private)	1.crop 2.rotate	640,000 samples for train 160,000 samples for validation 200,000 samoles for test
[44]	1,250 images (private)	crop	60,000 patches
[53]	CFD (public)	flip	-
[49]	Video frame data (private)	1.crop 2.rotate 3.flip and rotate 4.Add noise	147,344 crack patches 149,460 noncrack patches
[55]	CrackTree260 (public)	1.rotate 2.flip 3.crop	35,100 samples for train
[64]	297 images (private)	flip	2,366
[46]	282 images (private)	crop	40,000 patches

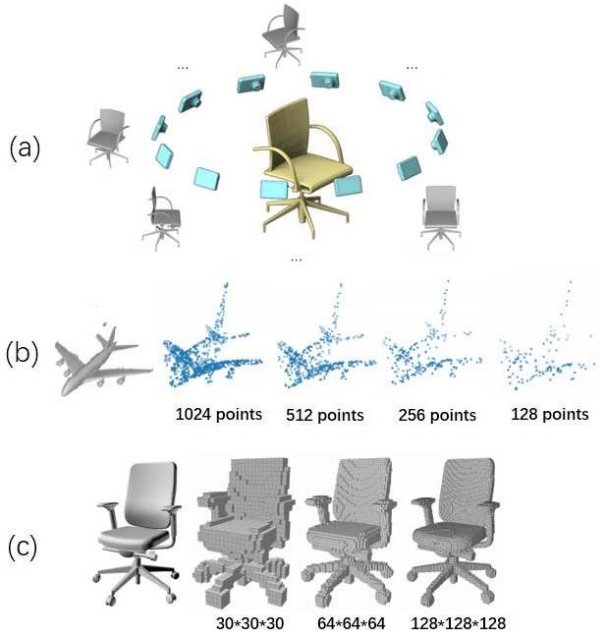

FIGURE 4. Three expressions of 3D data, (a) Multi-view, (b) point cloud and (c) voxels.

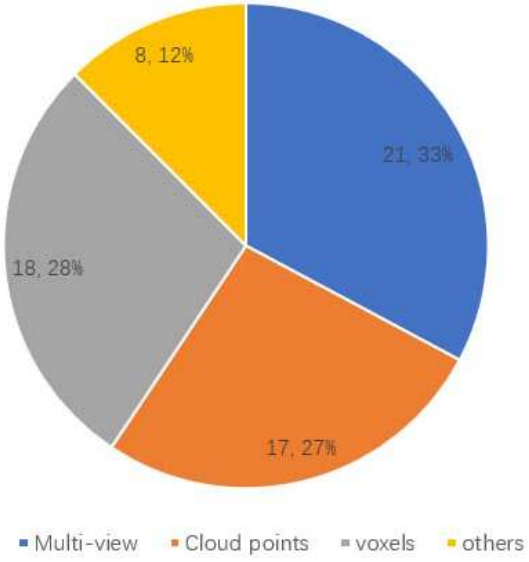

FIGURE 5. Distribution of 3D object classification methods on data representations.

of object surface, the more dense point clouds are, the more accurate models can be created, this process is called 3D reconstruction, as shown in Fig.4(b). 3D scanners and LiDAR devices can be used to generate point cloud data [82].

Point cloud data can convert to structured 3D regular grids [83], namely, voxel. Voxel is the smallest unit of digital data in 3D space segmentation, each unit can be viewed as a grid with fixed coordinates. Similar to 2D image, it also has a resolution, the finer the 3D space is divided, the smaller each grid is, and the greater the resolution is. Fig.4(c) shows 3D occupancy grids in different resolution. For easy reference, we compared these three kinds of representation in Table 7.

B. COMPARISON OF DIFFERENT 3D REPRESENTATIONS

Different 3D data representation will affect the effectiveness of the methods. We compared different methods in terms of object classification performance on benchmark Modelnet40 [86]. Modelnet40 contains 40 categories of CAD 3D models and is a standard dataset for evaluating semantic segmentation and classification of 3D deep learning models [87]. For 3D object classification, we studied the 60 methods submitted to the web site, Fig. 5 shows the distribution of these methods on different data types. We can see that 21.33% of these methods were based on multi-view, 17.27% were based on point cloud data, 18.29% were based on voxel, and 7.11% were based on other methods. The highest classification accuracy (97.37%) was achieved by RotationNet [88], which jointly estimates the object categories and viewpoints for each single-view image and aggregates object class predictions from partial multi-view image sets. As just mentioned, different data representation may affect the classification performance. We analyzed three different 3D data representation methods in terms of classification performance. The average accuracy based on multi-view is 92.31%, based on point cloud data is 90.43%, and based on

TABLE 7. 3D data representation.

3D Data Representation	Introduction	Pros and cons
Multi-view [84]	Multi-view is a set of 2D rendered images obtained from different perspectives of complete 3D model, which is a simple process of converting 3D images into 2D images.	Pros: Mature technology Cons: Cannot represent a 3D structure completely
Point cloud [82]	Point cloud data is generated by laser light hitting the surface of an object, and each point records the 3D coordinates x, y, and z, color information, gray value, and depth information, it is the collection of all sampling points on the object surface.	Pros: 1. Can express the space outline and specific position of the object 2. The point cloud is independent of the angle of view and can rotate at will Cons: 1. Point clouds are disordered 2. Point clouds are sparse 3. Point cloud data expresses limited data information
Voxel grid [85]	Voxel is similar to pixels in a 2D image, which is equivalent to pixel points in a 3D image. The voxel grid is regarded as a point cloud with fixed size and quantization, and each voxel has its own coordinates.	Pros: 1. Voxel is the closest 3D representation of an image 2. The voxel data can be used to deal with the data which is not fully sampled Cons: 1. No rotational invariance 2. The general voxel representation use a lot of memory

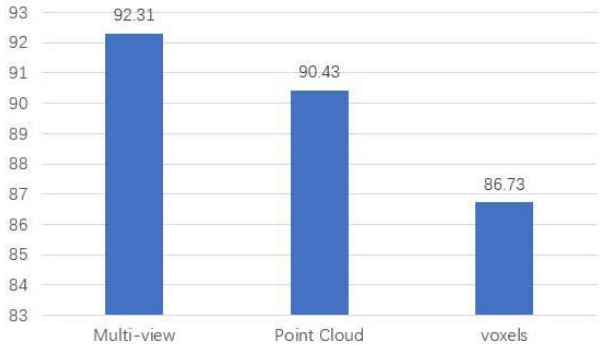

FIGURE 6. Average accuracy of different classification methods.

voxel is 86.73%, as shown in Fig. 6. It can be found that in the classification task, the method based on multiple views and point cloud are more accurate than that based on voxel.

C. DEEP NETWORKS FOR 3D OBJECT CLASSIFICATION

In the work of [84], the authors presented a CNN architecture that combines information from multiple views of a 3D shape into a single and compact shape descriptor offering even better recognition performance. In this method, images from each view were passed through a separate CNN to extract view-based features. Then, an additional CNN is used to combine these features for final classification.

Following the first volumetric CNN is 3D ShapeNets [86], Maturana *et al*. proposed VoxNet in [85] to process volumetric data with grid resolution of 32×32, where the model consists of 4D convolution filters to hold 3D spatial features. Rahul Dev also proposed CNN models to classify 3D object based on volumetric data [89]. LightNet [81] is a faster version of VoxNet to address heavy computation problem for real time 3D object recognition.

Point cloud is an unordered set of points scanned from the 3D object. The critical problem to solve is to make the model invariant to the permutation of the data points. PointNet [90] is the first CNN model to directly work on the raw point cloud. The method operates on each point separately and accumulate features from all the points by a symmetric function, which is a max pooling layer. Pointnet++ introduces a hierarchical neural network that applies PointNet recursively on a nested partitioning of the input point set. By exploiting metric space distances, the method is able to learn local features with increasing contextual scales [91]. To further address the problem, DGCNN was proposed in [92]. Instead of working on individual points like PointNet, this method constructs a neighborhood graph to capture the local geometric information and proposes EdgeConv operation to apply convolution-like operations on the edges.

These methods were all tested on modelnet40 dataset. We compared them in terms of the number of model parameters, input type, forward time, accuracy and the deep learning framework in Table 8. We can see that, the multi-view model is much larger than the other two methods in terms of model parameters. In terms of classification accuracy, data representation based on multi-view and point cloud is slightly higher than based on voxel. This is caused by the resolution of voxel, the higher the resolution of voxel, the larger calculation amount and the more complex the model is. Generally, only $32 \times 32 \times 32$ or $64 \times 64 \times 64$ resolutions are selected for training.

For multi-view, the performance of the model will get better as the number of images from different perspectives increases. The same is true to point cloud data. The more points used to describe an object, the more comprehensive the 3D information of the object will be, and the classification accuracy will be improved. Similarly, the higher the resolution of voxel data, the better the performance of the model.

D. FEATURE EXTRACTION USING 3D DATA

Feature extraction is a very important step in crack detection. 3D data can provide richer features than 2D images.

TABLE 8. Comparison of different methods on modelnet40 dataset.

Method	Parameters size (M)	Input	Forward time	Accuracy(%)	Deep learning framework
MVCNN [84]	99	Multi-view : 80 views	-	90.2	TensorFlow
	99	Multi-view : 12 views	-	89.5	TensorFlow
PointNet [90]	3.5	point cloud : 1024 points	25.3	89.2	Caffe
	3.48	point cloud : 1024 points	14.7	89.2	TensorFlow
	3.5	point cloud : 1024 points	3.1	89.1	Pytorch
pointnet++ [91]	1.99	point cloud : 1024 points	32	90.7	TensorFlow
DGCNN [92]	1.84	point cloud : 1024 points	27.2	92.9	TensorFlow
	-	point cloud : 2048 points	-	93.5	TensorFlow
VoxNet [85]	1.4	voxels : $32 \times 32 \times 32$	1.3	85.6	Pytorch
	3.4	voxels : $32 \times 32 \times 32$	-	83	Theano
Lightnet [81]	1.1	voxels : $32 \times 32 \times 32$	-	86.9	Theano
Rahul Dev [89]	-	voxels : $32 \times 32 \times 32$	-	85.9	-
	-	voxels : $64 \times 64 \times 64$	-	88.5	-
	-	voxels : $128 \times 128 \times 128$	-	91.4	-

Several methods explicitly extract features from 3D data to feed to traditional machine learning models. For example, in the work of [93], the authors combined the extracted features from 2D and 3D to train classifiers, and in [94], spatiotemporal features were extracted from videos using 3D ConvNets. These features followed by a linear classifier achieved state-of-the-art results at the publication time.

1) SPATIOTEMPORAL FEATURES

In [94] Tran *et al*. proposed a simple and efficient method to learn spatial feature of 3D data by using 3D convolutional neural network to learning spatiotemporal features for videos. They found that $3\times3\times3$ convolutional kernels in all layers are among the best performing architectures for 3D ConvNets. In [95] Owoyemi and Hashimoto proposed an end-to-end spatiotemporal gesture learning method for 3D point cloud data, mapping the point cloud data into a dense occupancy grid and learning the spatiotemporal characteristics of the data. In this work, 3D ROI jittering method is used in training to expand 3D data.

2) GEOMETRIC FEATURES

In [96] Furuya and Ohbuchi proposed a deep local feature aggregation network (DLAN) for 3D model retrieval. It combines the extraction of rotation invariant 3D local features with their aggregation in a single depth architecture. DLAN describes the local 3D region of a 3D model by using a set of 3D geometric features that are not affected by local rotation. Zheng *et al*. proposed a data-driven model, 3DMatch [97], which learns a local volumetric patch descriptor to establish corresponding relationships between local 3D data and can match local geometric features well in real depth images. Deng *et al*. proposed PPFNet [98], a 3D local feature descriptor for in-depth learning of global information, which can be matched to corresponding parts in disordered point cloud data. PPFNet uses a new n-tuple loss and architecture to naturally inject global information into local descriptors and enhance the representation of local features.

E. 3D PAVEMENT DEFECT DETECTION

With 3D data acquisition is becoming easier, the application of 3D technology to pavement defect detection is more and more common. 3D data can well represent the spatial information (length, width and depth) of road defects, and conduct multi-directional analysis on the area, volume and other aspects of defects.

Xu *et al*. [99] used 3D mobile LiDAR to collect road point cloud data and studied the automatic extraction of road curbs, in order to improve the robustness and accuracy of the model, they designed a new energy function to extract the constrained candidate points and refined the candidate points with the least cost path model. They sampled the point cloud data at a rate of 100%, 50%, 10% and 1% respectively. Even if the point cloud drops to 1%, the method proposed in this paper can still extract the road curbs.

1) TRADITIONAL METHODS FOR 3D CRACK DETECTION

Zhang *et al*. utilized the Microsoft Kinect to reconstruct pavement surfaces and capture geometric features of pavement cracking, including crack width, length, and depth to identify the distress severities of three major types of pavement cracks, namely, alligator cracking, traverse cracking, longitudinal cracking [100]. In the work of [101], Li *et al*. employed laser-imaging techniques to model the pavement surface with dense 3D points and used an algorithm based on frequency analysis (Fourier transformation) separate potential cracks from the control profile and material texture of the pavement assuming that the road pavement in the absence of pavement distresses commonly holds a relatively uniform control profile. Tsai and Li proposed a dynamic-optimization-based crack segmentation method to

TABLE 9. Network performance comparison.

Network	Input size	Hardware devices	Number of train images	Number of test images	Forword time(s)	Precision(%)	Recall(%)	F1(%)
CrackNet	1024 × 512	GPU: two GeForce GTX TITAN	1800	200	5.37	90.13	87.63	88.86
CrackNet II	1024 × 512	CPU: six CPU cores/12 logical processors	2500	200	1.26	90.20	89.06	89.62
CrackNet	512 × 256	GPU:GeForce GTX 1080Ti	2568	500	1.21	90.86	80.96	85.62
CrackNet-V	512 × 256	GPU:GeForce GTX 1080Ti	2568	500	0.33	84.31	90.12	87.12
CrackNet	1024 × 512	GPU:Two NVidia GeForce GTX 1080 Ti	3000	500	2.894	83.89	89.41	86.57
CrackNet-R	1024 × 512	GPU:Two NVidia GeForce GTX 1080 Ti	3000	500	0.713	88.89	95.00	91.84

test 1 to 5 mm wide cracks collected by 3D laser at different depths and lighting conditions [102]. To detect similar cracks in masonry, the work [103] presented mathematics to determine the minimum crack width detectable with a terrestrial laser scanner, in which the main features used include orthogonal offset, interval scan angle, crack orientation, and crack depth. In [93], the whole image is divided into sub images of 128 × 128 pixels and filtered by a set of Gabor filters. The maximum value of the magnitude of every filtered image is the feature used to train weak classifiers. To detect crack in pavement images, binary segmentation is a straightforward way. Unlike most 2D thresholding techniques based on the assumptions that the distress pixels are darker than their surroundings, [104] proposed a probabilistic relaxation labeling technique to enhance the accuracy of the distress detection, which take account of the non-uniform illumination and complicated contents on the pavement surface areas. The work of [105] proposed a unique method which uses Dempster-Shafer (D-S) theory to combine the 2D gray-scale image and 3D laser scanning data as a mass function, and the corresponding detection results are fused at the decision-making level.

2) DEEP NETWORK FOR 3D CRACK DETECTION

Applying deep learning neural network in 3D crack detection is currently a new and hot research direction. In 2017, Zhang *et al*. proposed CrackNet network to implement pixel-level detection of pavement cracks and defects [106]. The model consists of five layers with two fully connected layers, two convolution layers and one output layer. The feature extractor utilizes line filters oriented at various directions and with varied lengths as well as widths to enhance the contrast between cracks and the background. The model was trained with 1,800 3D pavement images collected from DHDV [2].

Later on, in the work of [107], the authors proposed an improved architecture of CrackNet called CrackNet II for enhanced learning capability and faster performance. CrackNet II has a deeper architecture with more hidden layers but fewer parameters. Such an architecture yields five times faster performance compared with the original CrackNet. Similar to the original CrackNet, CrackNet II still uses invariant image width and height through all layers to place explicit requirements on pixel-perfect accuracy. In addition, they deepened the network and the combination of repeated convolution and 1 × 1 convolution is used to learn the local features with different local receptive fields. Recently, Zhang's team put forward the CrackNet V [108], which includes a pre-processing layer, eight convolutional layers and an output layer. They used a 3 × 3 filter for the first six convolutions, and stack multiple 3 × 3 convolutions together for depth extraction, which reduced the number of parameters and improves the efficiency of feature extraction. In addition, they designed a new activation function to improve the detection accuracy of shallow cracks.

In order to improve the recall rate, they put forward CrackNet-R [109] based on recurrent neural network. As a recursive unit, gated recurrent multi-layer perceptron (GRMLP) is designed to update the internal memory of CrackNet-R recursively. GRMLP aims to abstract the features of input and hidden state more deeply by multi-layer nonlinear transformation at gate unit. The resultant model achieved about four times faster and introduces tangible improvements in detection accuracy, when compared to CrackNet. The performance comparison of the networks shown in Table 9.

3) FACTORS AFFECTING 3D PAVEMENT DEFECT DETECTION

There are many factors that can influence the detection of pavement defects. Yi *et al*. [102] proposed a dynamic-optimization-based crack segmentation method to test 1 to 5 mm wide cracks collected by 3D laser at different depths and lighting conditions. Experiments show that cracks with width equal to or greater than 2 mm can be effectively separated from the pavement background, while cracks with width of 1 mm can only be partially separated. In addition, it was found that the light intensity had little effect on the test results.

Li *et al*. [101] used laser imaging technology to model 3D dense point road surface and proposed a 3D point cloud crack detection method based on sparse point grouping, which can reduce the influence of light variation and shadow on crack detection. They tested the effect of the data acquisition vehicle on the performance of the proposed method at different speeds(10km/h to 80km/h). The experimental results show

that at different speeds, the crack test effect is roughly the same, but the slower the speed, the more detailed the crack contour description.

Debra *et al*. [103] found through the experiment that crack depth depends on three factors: scanning distance, scanning angle and crack width.The scanning distance is the distance between the crack and the laser scanner, and the scanning angle is the offset angle between the crack and the laser scanner. Cracks with a width of 1 to 7 mm were scanned at distances of 5m and 7.5m and angles of 0°, 15° and 30°. The results show that the crack depth cannot be detected when the crack width is less than 1 mm, because the smaller the crack width is, the more difficult to obtain the depth information of crack. As the crack width increases, the detection of the crack depth becomes more accurate. With the increase of scanning angle, the error of crack depth detection will also increase. The closer the scanning distance is, the higher the detection accuracy will be.

Khurram *et al*. [110] used Kinect to predict and analyze the depth and volume of pothole, the mean percentage error are 2.58% and 5.47%, respectively. In addition, the test performance of pothole with water, dust and oil is also discussed. Experimental results show that the error of test results will increase with the increase of water, dust and oil content, and the error is also related to the types of these media.

V. EXISTING PROBLEMS AND RESEARCH PROSPECTS

After years of development, many achievements have been made in pavement defects detection, which has made great contributions to the maintenance of pavement and the safety of vehicles. However, there are still some problems in the practical application:

1) Due to the complex and dynamic environmental factors, there may be some errors in the detection of road cracks under the condition of poor light in rainy days or when there is water on the road.
2) Different algorithms are needed to test on different road surface conditions, and the algorithm transplantation performance is poor.
3) The process of defects detection is always offline, so the performance of real-time is not good in reality.

Therefore, we need to further enhance the detection accuracy and real-time performance of the algorithm to ensure the optimal detection results in real applications. The generalization and robustness of the methods is also very important as the factors such as road and weather conditions greatly affect the detection. As for 3D cracks detection, the depth information of cracks is added to make the cracks have spatial structure. Although the overall information of cracks is more complete, it undoubtedly increases the complexity of the algorithm and greatly increases the computational cost. The algorithm can be improved and the computing cost can be reduced by referring to some progress in deep convolutional neural networks for 2D images such as network architecture and model compression techniques. On the other hand, there are few public 3D cracks datasets, researchers collect pavement crack data for training and testing by themselves, and it is impossible to conduct performance analysis on the same dataset. Collecting 3D crack benchmark datasets will greatly benefit future study of the 3D crack detection.

VI. CONCLUSION

The automatic detection of pavement crack has been studied extensively due to its practical significance. From traditional image processing methods to machine learning methods to deep learning algorithms that have become popular in recent years. In this work, we review these methods, and we focus on the detailed comparison and analysis on deep learning methods and 3D image based methods. Particularly, deep learning methods are grouped and reviewed in three categories, image classification, object detection and pixel-level segmentation. For 3D crack detection methods, we compare the different data representations and study the corresponding performance of the deep neural networks for 3D object classification. Traditional and deep learning based crack detection methods using 3D data are also reviewed.

REFERENCES

[1] G. Caroff, P. Joubert, F. Prudhomme, and G. Soussain, "Classification of pavement distresses by image processing (MACADAM SYSTEM)," in *Proc. ASCE*, 1989, pp. 46–51.
[2] K. C. Wang, Z. Hou, and W. Gong, "Automation techniques for digital highway data vehicle (DHDV)," in *Proc. 7th Int. Conf. Manag. Pavement Assets*. Citeseer, 2008. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download
[3] L. Sjogren and P. Offrell, "Automatic crack measurement in Sweden," in *Proc. 4th Int. Symp. Pavement Surface Characteristics Roads Airfields World Road Assoc. (PIARC)* 2000.
[4] L. Jin-Hui, L. Wei, and J. Shou-Shan, "A study on road surface defects detecting technology with CCD camera," *J. Xi'an Inst. Technol.*, vol. 2, 2002.
[5] K. K. Singh and A. Singh, "A study of image segmentation algorithms for different types of images," *Int. J. Comput. Sci. Issues)*, vol. 7, no. 5, p. 414, 2010.
[6] S. Kamdi and R. Krishna, "Image segmentation and region growing algorithm," *Int. J. Comput. Technol. Electron. Eng.*, vol. 2, no. 1, 2012.
[7] N. Kanopoulos, N. Vasanthavada, and R. Baker, "Design of an image edge detection filter using the Sobel operator," *IEEE J. Solid-State Circuits*, vol. SSC-23, no. 2, pp. 358–367, Apr. 1988.
[8] W. Dong and Z. Shisheng, "Color image recognition method based on the Prewitt operator," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, vol. 6, 2008, pp. 170–173.
[9] L. Er-Sen, Z. Shu-Long, Z. Bao-Shan, Z. Yong, X. Chao-Gui, and S. Li-Hua, "An adaptive edge-detection method based on the canny operator," in *Proc. Int. Conf. Environ. Sci. Inf. Appl. Technol.*, vol. 1, Jul. 2009, pp. 465–469.
[10] B. J. Lee and H. D. Lee, "Position-invariant neural network for digital pavement crack analysis," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 19, no. 2, pp. 105–118, Mar. 2004.
[11] J.-Y. Jung, H.-J. Yoon, and H.-W. Cho, "A study on crack depth measurement in steel structures using image-based intensity differences," *Adv. Civil Eng.*, vol. 2018, pp. 1–10, 2018.
[12] F. Blais, M. Rioux, and J.-A. Beraldin, "Practical considerations for a design of a high precision 3-D laser scanner system," in *Proc. Optomech. Electro-Opt. Design Ind. Syst.*, Nov. 1988, pp. 225–246.
[13] S. Chambon and J.-M. Moliard, "Automatic road pavement assessment with image processing: Review and comparison," *Int. J. Geophys.*, vol. 2011, pp. 1–20, 2011.
[14] K. Gopalakrishnan, "Deep learning in data-driven pavement image analysis and automated distress detection: A review," *Data*, vol. 3, no. 3, p. 28, Jul. 2018.
[15] T. B. Coenen and A. Golroo, "A review on automated pavement distress detection methods," *Cogent Eng.*, vol. 4, no. 1, p. 1374822, 2017.

[16] S. Mathavan, K. Kamal, and M. Rahman, "A review of three-dimensional imaging technologies for pavement distress detection and measurements," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2353–2362, Oct. 2015.
[17] S. Zhu, X. Xia, Q. Zhang, and K. Belloulata, "An image segmentation algorithm in image processing based on threshold segmentation," in *Proc. 3rd Int. IEEE Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 2007, pp. 673–678.
[18] H. Oliveira and P. L. Correia, "Automatic road crack segmentation using entropy and image dynamic thresholding," in *Proc. IEEE 17th Eur. Signal Process. Conf.*, 2009, pp. 622–626.
[19] L. Peng, W. Chao, L. Shuangmiao, and F. Baocai, "Research on crack detection method of airport runway based on twice-threshold segmentation," in *Proc. 5th Int. Conf. Instrum. Meas., Comput., Commun. Control (IMCCC)*, Sep. 2015, pp. 1716–1720.
[20] S. Wang and W. Tang, "Pavement crack segmentation algorithm based on local optimal threshold of cracks density distribution," in *Proc. Int. Conf. Intell. Comput.* Springer, 2011, pp. 298–302.
[21] H. Zhao, G. Qin, and X. Wang, "Improvement of canny algorithm based on pavement edge detection," in *Proc. 3rd Int. Congr. Image Signal Process.*, Oct. 2010, pp. 964–967.
[22] C.-C. Zhou, G.-F. Yin, and X.-B. Hu, "Multi-objective optimization of material selection for sustainable products: Artificial neural networks and genetic algorithm approach," *Mater. Des.*, vol. 30, no. 4, pp. 1209–1215, Apr. 2009.
[23] A. Ayenu-Prah and N. Attoh-Okine, "Evaluating pavement cracks with bidimensional empirical mode decomposition," *EURASIP J. Adv. Signal Process.*, vol. 2008, no. 1, Art. no. 861701, 2008.
[24] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 460, no. 2046, pp. 1597–1611, Jun. 2004.
[25] Y. Zhou, F. Wang, N. Meghanathan, and Y. Huang, "Seed-based approach for automated crack detection from pavement images," *Transp. Res. Rec.*, vol. 2589, no. 1, pp. 162–171, Jan. 2016.
[26] Q. Li, Q. Zou, D. Zhang, and Q. Mao, "FoSA: F* Seed-growing approach for crack-line detection from pavement images," *Image Vis. Comput.*, vol. 29, no. 12, pp. 861–872, Nov. 2011.
[27] A. Akagic, E. Buza, S. Omanovic, and A. Karabegovic, "Pavement crack detection using Otsu thresholding for image segmentation," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, May 2018, pp. 1092–1097.
[28] 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 Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2718–2729, Oct. 2016.
[29] H. Li, D. Song, Y. Liu, and B. Li, "Automatic pavement crack detection by multi-scale image fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2025–2036, Jun. 2019.
[30] R. E. Wright. Logistic regression. American Psychological Association. Accessed: 1995. [Online]. Available: https://psycnet.apa.org/record/1995-97110-007
[31] K. M. Leung. *Naive Bayesian Classifier*. Accessed: 2007. [Online]. Available: http://cis.poly.edu/~mleung/FRE7851/f07/naiveBayesianClassifier.pdf
[32] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining Knowl. Discovery*, vol. 2, no. 2, pp. 121–167, 1998.
[33] A. Jain, J. Mao, and K. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, Mar. 1996.
[34] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
[35] G. Xu, J. Ma, F. Liu, and X. Niu, "Automatic recognition of pavement surface crack based on BP neural network," in *Proc. Int. Conf. Comput. Elect. Eng.*, Dec. 2008, pp. 19–22.
[36] 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.
[37] A. Marques and P. L. Correia, "Automatic road pavement crack detection using SVM," Ph.D. dissertation, Elect. Comput. Eng., Instituto Superior Técnico, Lisbon, Portugal, 2012.
[38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[39] W. Cao, Q. Lin, Z. He, and Z. He, "Hybrid representation learning for cross-modal retrieval," *Neurocomputing*, vol. 345, pp. 45–57, Jun. 2019.
[40] W. Cao, J. Yuan, Z. He, Z. Zhang, and Z. He, "Fast deep neural networks with knowledge guided training and predicted regions of interests for real-time video object detection," *IEEE Access*, vol. 6, pp. 8990–8999, 2018.
[41] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang, and B. Hu, "Liver fibrosis classification based on transfer learning and FCNet for ultrasound images," *IEEE Access*, vol. 5, pp. 5804–5810, 2017.
[42] D. Meng, G. Cao, Y. Duan, M. Zhu, L. Tu, D. Xu, and J. Xu, "Tongue images classification based on constrained high dispersal network," *Evidence-Based Complementray Alternative Med.*, vol. 2017, no. 4, pp. 1–12, 2017.
[43] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, "Road crack detection using deep convolutional neural network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3708–3712.
[44] S. Li and X. Zhao, "Convolutional neural networks-based crack detection for real concrete surface," *Proc. SPIE*, vol. 10598, Mar. 2018, Art. no. 105983V.
[45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
[46] 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.
[47] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. 23rd ACM Int. Conf. Multimedia (MM)*, 2015.
[48] L. Pauly, H. Peel, S. Luo, D. Hogg, and R. Fuentes, "Deeper networks for pavement crack detection," in *Proc. 34th Int. Symp. Autom. Robot. Construct. (ISARC)*, Jul. 2017, pp. 479–485.
[49] F.-C. Chen and 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.
[50] Z. Fan, Y. Wu, J. Lu, and W. Li, "Automatic pavement crack detection based on structured prediction with the convolutional neural network," 2018, *arXiv:1802.02208*. [Online]. Available: https://arxiv.org/abs/1802.02208
[51] B. Li, K. C. Wang, A. Zhang, E. Yang, and G. Wang, "Automatic classification of pavement crack using deep convolutional neural network," *Int. J. Pavement Eng.*, pp. 1–7, Jun. 2018.
[52] 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.
[53] M. D. Jenkins, T. A. Carr, M. I. Iglesias, T. Buggy, and G. Morison, "A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks," in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 2120–2124.
[54] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[55] 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.
[56] W. Liu, Y. Huang, Y. Li, and Q. Chen, "FPCNet: Fast pavement crack detection network based on encoder-decoder architecture," 2019, *arXiv:1907.02248*. [Online]. Available: https://arxiv.org/abs/1907.02248
[57] J. Konig, M. D. Jenkins, P. Barrie, M. Mannion, and G. Morison, "A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 1460–1464.
[58] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[59] U. Escalona, F. Arce, E. Zamora, and J. H. S. Azuela, "Fully convolutional networks for automatic pavement crack segmentation," *Comput. Sist.*, vol. 23, no. 2, pp. 451–460, 2019.
[60] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.
[61] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 21–37.

[62] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[63] G. Suh and Y.-J. Cha, "Deep faster R-CNN-based automated detection and localization of multiple types of damage," *Proc. SPIE*, vol. 10598, Mar. 2018, Art. no. 105980T.

[64] Y.-J. Cha, W. Choi, G. Suh, S. Mahmoudkhani, and O. Büyüköztürk, "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 731–747, Sep. 2018.

[65] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, 2013, pp. 818–833.

[66] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[67] J. Li, X. Zhao, and H. Li, "Method for detecting road pavement damage based on deep learning," *Proc. SPIE*, vol. 10972, Apr. 2019, Oct. 109722D.

[68] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[69] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2818–2826.

[70] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection using deep neural networks with images captured through a smartphone," 2018, *arXiv:1801.09454*. [Online]. Available: https://arxiv.org/abs/1801.09454

[71] S. Anand, S. Gupta, V. Darbari, and S. Kohli, "Crack-pot: Autonomous road crack and pothole detection," in *Proc. Digit. Image Comput., Techn. Appl.*, 2018, pp. 1–6.

[72] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," 2016, *arXiv:1602.07360*. [Online]. Available: https://arxiv.org/abs/1602.07360

[73] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach, "Supervised dictionary learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 1033–1040.

[74] D. K. Mcclish, "Analyzing a portion of the ROC curve," *Med. Decis. Making*, vol. 9, no. 3, pp. 190–195, Aug. 1989.

[75] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognit.*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997.

[76] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, and H. Ling, "Feature pyramid and hierarchical boosting network for pavement crack detection," 2019, *arXiv:1901.06340*. [Online]. Available: https://arxiv.org/abs/1901.06340

[77] J. Cheng, W. Xiong, W. Chen, Y. Gu, and Y. Li, "Pixel-level crack detection using U-net," in *Proc. IEEE Region 10 Conf. TENCON*, Oct. 2018, pp. 462–466.

[78] 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.

[79] Y. Liu, M.-M. Cheng, X. Hu, J.-W. Bian, L. Zhang, X. Bai, and J. Tang, "Richer convolutional features for edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 8, pp. 1939–1946, Aug. 2019.

[80] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, and H.-M. Gross, "How to get pavement distress detection ready for deep learning? A systematic approach," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2039–2047.

[81] S. Zhi, Y. Liu, X. Li, and Y. Guo, "Toward real-time 3D object recognition: A lightweight volumetric CNN framework using multitask learning," *Comput. Graph.*, vol. 71, pp. 199–207, Apr. 2018.

[82] F. Chazal, L. J. Guibas, S. Y. Oudot, and P. Skraba, "Analysis of scalar fields over point cloud data," in *Proc. 20th Annu. ACM-SIAM Symp. Discrete Algorithms*, Jan. 2009, pp. 1021–1030.

[83] M. J. Lee, "Method and apparatus for transforming point cloud data to volumetric data," U.S. Patent 7 317 456, Jan. 8, 2008.

[84] H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 945–953.

[85] D. Maturana and S. Scherer, "VoxNet: A 3D convolutional neural network for real-time object recognition," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 922–928.

[86] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1912–1920.

[87] *ModelNet*. Accessed: Oct. 2019. [Online]. Available: https://modelnet.cs.princeton.edu/

[88] A. Kanezaki, Y. Matsushita, and Y. Nishida, "RotationNet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5010–5019.

[89] R. D. Singh, A. Mittal, and R. K. Bhatia. *3D Convolutional Neural Network for Object Recognition*. Accessed: 2017. [Online]. Available: https://pdfs.semanticscholar.org/218b/b5f163046166a5d13f7832d10f0de2ab8286.pdf

[90] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 652–660.

[91] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5099–5108.

[92] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *ACM Trans. Graph.*, vol. 38, no. 5, pp. 1–12, Oct. 2019.

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

[94] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 4489–4497.

[95] J. Owoyemi and K. Hashimoto, "Spatiotemporal learning of dynamic gestures from 3D point cloud data," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018.

[96] T. Furuya and R. Ohbuchi, "Deep aggregation of local 3D geometric features for 3D model retrieval," in *Proc. Brit. Mach. Vis. Conf.*, 2016, pp. 1–121.

[97] A. Zeng, S. Song, M. Niebner, M. Fisher, J. Xiao, and T. Funkhouser, "3DMatch: Learning local geometric descriptors from RGB-D reconstructions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017.

[98] H. Deng, T. Birdal, and S. Ilic, "PPFNet: Global context aware local features for robust 3D point matching," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 195–205.

[99] S. Xu, R. Wang, and H. Zheng, "Road curb extraction from mobile LiDAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 996–1009, Feb. 2017.

[100] Y. Zhang, C. Chen, Q. Wu, Q. Lu, S. Zhang, G. Zhang, and Y. Yang, "A Kinect-based approach for 3D pavement surface reconstruction and cracking recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 3935–3946, Dec. 2018.

[101] Q. Li, D. Zhang, Q. Zou, and H. Lin, "3D laser imaging and sparse points grouping for pavement crack detection," in *Proc. 25th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017.

[102] Y.-C.-J. Tsai and F. Li, "Critical assessment of detecting asphalt pavement cracks under different lighting and low intensity contrast conditions using emerging 3D laser technology," *J. Transp. Eng.*, vol. 138, no. 5, pp. 649–656, May 2012.

[103] D. F. Laefer, L. Truong-Hong, H. Carr, and M. Singh, "Crack detection limits in unit based masonry with terrestrial laser scanning," *NDT E Int.*, vol. 62, pp. 66–76, Mar. 2014.

[104] E. Salari and G. Bao, "Automated pavement distress inspection based on 2D and 3D information," in *Proc. IEEE Int. Conf. ELECTRO/INFORMATION Technol.*, May 2011, pp. 1–4.

[105] J. Huang, W. Liu, and X. Sun, "A Pavement crack detection method combining 2d with 3d information based on dempster-Shafer theory," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 29, no. 4, pp. 299–313, Apr. 2014.

[106] 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," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 10, pp. 805–819, Oct. 2017.

[107] A. Zhang, K. C. P. Wang, Y. Fei, Y. Liu, S. Tao, C. Chen, J. Q. Li, and B. Li, "Deep learning-based fully automated pavement crack detection on 3D asphalt surfaces with an improved CrackNet," *J. Comput. Civ. Eng.*, vol. 32, no. 5, Sep. 2018, Art. no. 04018041.
[108] 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.*, vol. 21, no. 1, pp. 273–284, Jan. 2020.
[109] A. Zhang, K. C. P. Wang, Y. Fei, Y. Liu, C. Chen, G. Yang, J. Q. Li, E. Yang, and S. Qiu, "Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 34, no. 3, pp. 213–229, Mar. 2019.
[110] K. Kamal, S. Mathavan, T. Zafar, I. Moazzam, A. Ali, S. U. Ahmad, and M. Rahman, "Performance assessment of Kinect as a sensor for pothole imaging and metrology," *Int. J. Pavement Eng.*, vol. 19, no. 7, pp. 565–576, Jul. 2018.

WENMING CAO received the M.S. degree from the System Science Institute, Chinese Academy of Sciences, Beijing, China, in 1991, and the Ph.D. degree from the School of Automation, Southeast University, Nanjing, China, in 2003. From 2005 to 2007, he was a Postdoctoral Researcher with the Institute of Semiconductors, Chinese Academy of Sciences. He is currently a Professor with Shenzhen University, Shenzhen, China. His research interests include pattern recognition, image processing, and visual tracking.

QIFAN LIU is currently pursuing the M.Eng. degree in communication and information engineering with Shenzhen University, Shenzhen, China. His research interests include image processing and machine learning.

ZHIQUAN HE received the M.S. degree from the Institute of Electronics, Chinese Academy of Sciences, in 2001, and the Ph.D. degree from the Department of Computer Science, University of Missouri-Columbia, in 2014. He is currently an Assistant Professor with the College of Information Engineering, Shenzhen University, China. His research interests include image processing, computer vision, and machine learning.