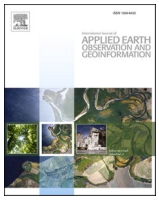
International Journal of Applied Earth Observations and Geoinformation 111 (2022) 102836

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag



SD-GCN: Saliency-based dilated graph convolution network for pavement crack extraction from 3D point clouds

Lingfei Ma [a,b,*], Jonathan Li [c,d,*]

[a] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China
[b] Engineering Research Center of State Financial Security, Ministry of Education, Central University of Finance and Economics, Beijing 102206, China
[c] Geospatial Sensing and Data Intelligence Laboratory, Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[d] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada

ARTICLE INFO

Keywords:
Point clouds
Pavement crack extraction
Feature saliency
Graph convolution
Dilated convolution

ABSTRACT

Accurate pavement crack extraction is significant for pavement routine maintenance and potential traffic disaster minimization. Due to unordered data formats, intensity distinctions, and crack shape variations from point clouds captured by mobile laser scanning (MLS) systems, many preceding rule-based approaches and learning-based approaches cannot achieve high extraction accuracy and efficiency. To tackle these problems, we develop a saliency-based dilated graph convolution network, named SD-GCN, for pavement crack extraction from MLS point clouds. This network mainly consists of four modules. First, Module I is designed to remove off-ground point clouds. Next, two feature saliency maps are constructed to leverage both height and intensity information in Module II. Then, in Module III, the inherent point features and high-level edge features in multiple local neighborhoods are further extracted using a cylinder-based dilated convolution strategy. Finally, an MLP-based net architecture is designed for crack extraction refinement in Module IV. Experimental results exhibit that the SD-GCN model delivers an average of precision, recall, and F_1-score of 79.5%, 77.1%, and 78.3%, respectively, which outperforms state-of-the-art methods in terms of extraction accuracy and computational efficiency.

1. Introduction

Accurate and automated road health condition extraction is significant to obtain fundamental road information for extensive intelligent transportation system (ITS)-related applications, including road construction (Shi et al., 2016), pavement maintenance and rehabilitation (Dung and Anh, 2019), and road object recognition (Pu et al., 2011). However, due to heavy-duty or overloaded trucks, weather and environment conditions, and late pavement repair, road surfaces in urban road environments usually suffer from a diversity of cracks with different sizes, distinct structures, and various levels (see Fig. 1). Such pavement cracks or road distresses have considerably negative influences on supporting smooth traffic flows, ensuring reasonable driving behaviors, and even reducing severe traffic threats (Cao et al., 2020). Therefore, accurate pavement crack extraction is essential for potential traffic disaster minimization and pavement security improvement, which greatly promotes the advancement of high-definition (HD) maps and digital twins, especially in complicated urban road environments (Li et al., 2020).

Initially, pavement crack extraction is mainly conducted based on regular in-situ field measurements. Such manual visual inspection method highly depends on the expertise of inspectors, which are very unsafe, labor-intensive, and time-consuming. Benefitting from the rapid advancement of high-quality optical sensors and computer vision techniques, many studies have been performed for pavement distress inspection based on the digital images collected by satellites, unmanned aerial vehicles (UAVs), and airborne or mobile mapping systems (MMS). Compared to manual inspection, the image processing methods are safer and more effective, particularly for express highways and complex road scenarios. However, due to the varying illumination conditions, diverse crack structures, inevitable shadows, and low contrast with surroundings, it is still a difficult task to precisely and robustly extract pavement cracks from optical images (Hsieh and Tsai, 2020). More recently, mobile laser scanning (MLS) systems consisting of LiDAR sensors are widely

* Corresponding authors at: School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 102206, China (L. Ma). Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (J. Li).
E-mail addresses: l53ma@cufe.edu.cn (L. Ma), junli@uwaterloo.ca (J. Li).

https://doi.org/10.1016/j.jag.2022.102836
Received 16 March 2022; Received in revised form 6 May 2022; Accepted 19 May 2022
Available online 28 May 2022

employed to provide highly precise and dense 3D point clouds with reliable geospatial information. The point clouds acquired by high-end laser scanners can reach over 10,000 pts/m^2 with a driving speed up to 100 km/h, while it handles the dilemmas for both terrestrial and airborne LiDAR systems to achieve such high survey flexibility and measurement accuracy (Ma et al., 2021).

Meanwhile, deep learning-based models have exhibited their superior performance for crack extraction depending on the enhanced feature encoding capabilities from 3D point clouds (Yu et al., 2021). Nevertheless, most preceding methods pose three major shortcomings. (1) Unlike 2D images with regular grid structures, 3D point clouds collected by LiDAR systems are in unstructured and unordered data formats. Thus, most studies typically focus on transforming 3D point clouds to 2D images before feeding them into deep learning frameworks. This data dimension reduction scheme, yet, results in unnecessary topological and spatial information loss. Notably, such data transformation methods neglect the height information of 3D point clouds, which plays a vital role in pavement crack extraction. (2) Because of the type uncertainties and distribution diversities of pavement cracks, it faces a considerable challenge to directly process point clouds and accurately extract pavement cracks, particularly for complex and large-scale road scenarios. Besides, the geospatial correlations and adjacency relationships among neighboring points are not thoroughly explored. (3) The 3D point-based deep learning methods normally bring in a large number of parameters, and the computational costs greatly increase in the more complex and deeper net architecture.

To solve these drawbacks, we explore the practicality of pavement crack extraction from MLS point clouds by designing a saliency-focused graph convolutional network with the assistance of dilated convolutions. Typically, the unstructured point clouds can be well represented by a graph structure, while the nodes denote the unevenly distributed point clouds and the edges indicate the spatial relationships between two adjacent points. Thus, graph convolutional networks (GCN) are effective to characterize the local features of different points, contributing to increasing receptive fields and enhancing feature representation capabilities. The significant contributions of this paper are emphasized as threefold aspects. (1) We designed two salient feature mapping spaces, i.e., height and intensity feature maps, which could efficiently and completely represent the point characteristics to amplify the intensity differences and height distinction between crack points and pavement points. (2) We proposed a novel cylinder-based dilated convolution architecture. To the best of our knowledge, it is the first study that integrates graph convolutions with dilated convolution strategy. This net architecture could not only construct strong adjacent relationships among neighboring points but increase the receptive field of graph convolutions with a reduced computation complexity. (3) We performed a comparative analysis to prove that our proposed methods could outperform state-of-the-art (SOTA) algorithms regarding accuracy and efficiency.

We design the rest parts as follows. Section 2 introduces the previous studies about pavement crack extraction. Section 3 presents the developed pavement crack extraction method. Section 4 indicates the datasets and evaluation metrics used in this work. We analyze the experimental results in Section 5 and summarize this study followed by future research trends in Section 6.

2. Related work

2.1. 2D Image-based methods

Different photogrammetric sensors and surveying platforms, such as satellite-mounted, airborne-mounted, drone-mounted, and vehicle-mounted imaging systems, provide various remotely sensed images with spatial, spectral, and texture information for pavement crack extraction. Pavement cracks on optical images commonly show irregular geometric topologies and low contrast compared with surroundings (Ragnoli et al., 2018). Accordingly, many image-based methods, including threshold-based segmentation (Kamaliardakani et al., 2016), edge detection (Li et al., 2018), and region search (Tan and Li, 2019), have been proposed to conduct crack extraction tasks.

Recently, benefiting from the inherent features derived from optical imagery, many deep learning-based methods have achieved impressive performance with improved efficiency and accuracy for pavement health inspection. Pan et al. (2018) implemented several classical machine learning methods, including support vector machine (SVM), artificial neural network, and random forest, to recognize pavement cracks from multispectral pavement images obtained from UAVs. A pixel-level convolutional neural network (CNN) architecture was proposed, called CrackNet (Zhang et al., 2017), for accurate concrete crack extraction without the assistance of pooling layers. CrackNet can directly learn the inherent features from input images and generate various feature maps without down-sampling operations. As the improved version, CrackNet-V (Fei et al., 2020) was accordingly proposed for automated pixel-wise crack extraction with fewer training parameters and deeper network architecture, contributing to enhanced extraction accuracy and reduced computational cost. A CNN-based deep learning framework was developed to detect pavement cracks without the conjugation of image processing techniques for feature extraction (Cha et al., 2017).

Moreover, according to the orthoframes collected by a mobile mapping system, a deep learning framework was developed towards the implementation of the detector for pavement distress detection (Riid et al., 2019). Zou et al. (2018) introduced the DeepCrack model for automatically extracting pavement cracks in an end-to-end way depending on the multi-scale convolutional features. Furthermore, a deep learning-based architecture was introduced, called feature pyramid and hierarchical boosting network (FPHBN), to effectively extract pavement cracks (Yang et al., 2019). This method combined context information with low-level features derived from input images in a pyramid pattern, while balancing the contributions of different input training samples in a hierarchical pattern. Additionally, Yu et al. (2020) developed a capsule-based neural network for crack extraction from pavement images by designing a feature pyramid architecture and fusing different levels of capsule features. Such image-based deep learning architectures have provided promising solutions in crack

Fig. 1. Pavement cracks with different health conditions.

extraction. Due to the poor continuity of cracks, inhomogeneous intensity, and high sensitivity to ambient brightness, however, it remains a quite difficult task to precisely extract pavement cracks from optical images, especially in complex urban road conditions.

2.2. 3D Point-based methods

With the tremendous development of laser scanning techniques and decreasing costs of commercial mobile LiDAR systems, many research efforts have been focusing on pavement crack extraction from 3D point clouds (Ma et al., 2018). Unlike 2D optical images, 3D point clouds present more precise spatial coordinates and intensity information that is independent of ambient brightness, conducing to more accurate pavement crack extraction results.

The pavement surfaces were firstly extracted based on the vehicle trajectory data collected by mobile LiDAR systems. Then, the anomalous seed points were identified using geometric analysis of individual scan lines, and local neighborhood analysis was conducted to determine whether such seed points belong to pavement cracks (Ravi et al., 2021). Additionally, to take advantage of advanced image processing techniques, Zhong et al. (2020) introduced a pavement crack extraction approach by converting MLS point clouds into regular grid structures. This study introduced a 2D index for each 3D point with respect to its acquisition time or incident angle. Then, crack points were detected by considering both intensity and height differences. Crack curves were finally generated depending on morphological filtering and thinning algorithms, followed by the Freeman code method. Xu and Yang (2019) proposed a unique strategy for efficient pavement crack extraction from terrestrial laser scanning (TLS) point clouds by optimizing signal-to-noise ratio gradient for Gaussian filtering. Li et al., (2019a) performed a random forest classification (RFC) method using LiDAR point clouds captured by UAVs. In this method, by analyzing both spatial and spectral features of pavement cracks, a total of 48 multi-scale and multi-dimensional features were extracted depending on intensity and height information of point clouds, and such features were subsequently used as inputs of the RFC method for crack extraction.

Furthermore, an iterative tensor voting (ITV)-based framework was proposed for automated crack extraction (Guan et al., 2014). Firstly, the pavement points collected by an MLS system were separated from non-road points depending on vehicle trajectory data. Then, such pavement points were transformed into georeferenced feature images by an improved inverse distance weighted (IDW) method. Finally, the crack candidates were detected based on the ITV-based crack extraction framework, followed by a morphological thinning algorithm. Meanwhile, the Otsu thresholding approach was first used to extract intensity differences from MLS point clouds for crack skeleton recognition, followed by a spatial density filter for noise removal. Then, a Euclidean distance clustering algorithm was performed to group crack points into different crack lines. Finally, crack skeletons were generated using an L_1-median line-shaped extraction method (Yu et al., 2014). These rule-based or thresholding-based approaches usually require a wealth of prior knowledge, leading to poor practical applicability in complex urban road conditions. Hence, it is considerably challenging for these methods to accurately and robustly extract pavement cracks with low connectivity and irregular geometric topologies.

3. Method

In this section, we present the technical and practical implementations of the proposed deep learning framework, named saliency-based dilated graph convolution network (SD-GCN), for accurate and efficient point-wise pavement crack extraction using MLS point clouds. This network consists of four modules: data preprocessing, feature mapping and saliency construction, cylinder dilated graph convolution, and multi-layer perceptron (MLP)-based prediction.

Fig. 2 details the workflow of the proposed SD-GCN framework. More specifically, Module I is designed to remove the off-road point clouds from raw MLS point clouds for computational efficiency enhancement using an improved curb-related pavement extraction approach. In Module II, two feature maps are first created to boost the height and intensity saliency of pavement crack points by a cotangent function. Next, a fixed-radius nearest neighbor progressive competition (FR-NNPC) approach (Zhou et al., 2019) is applied to search candidate points and then generate height saliency and intensity saliency matrices. In Module III, the inherent point features and edge features in multiple local regions are further derived by developing a cylinder-based dilated

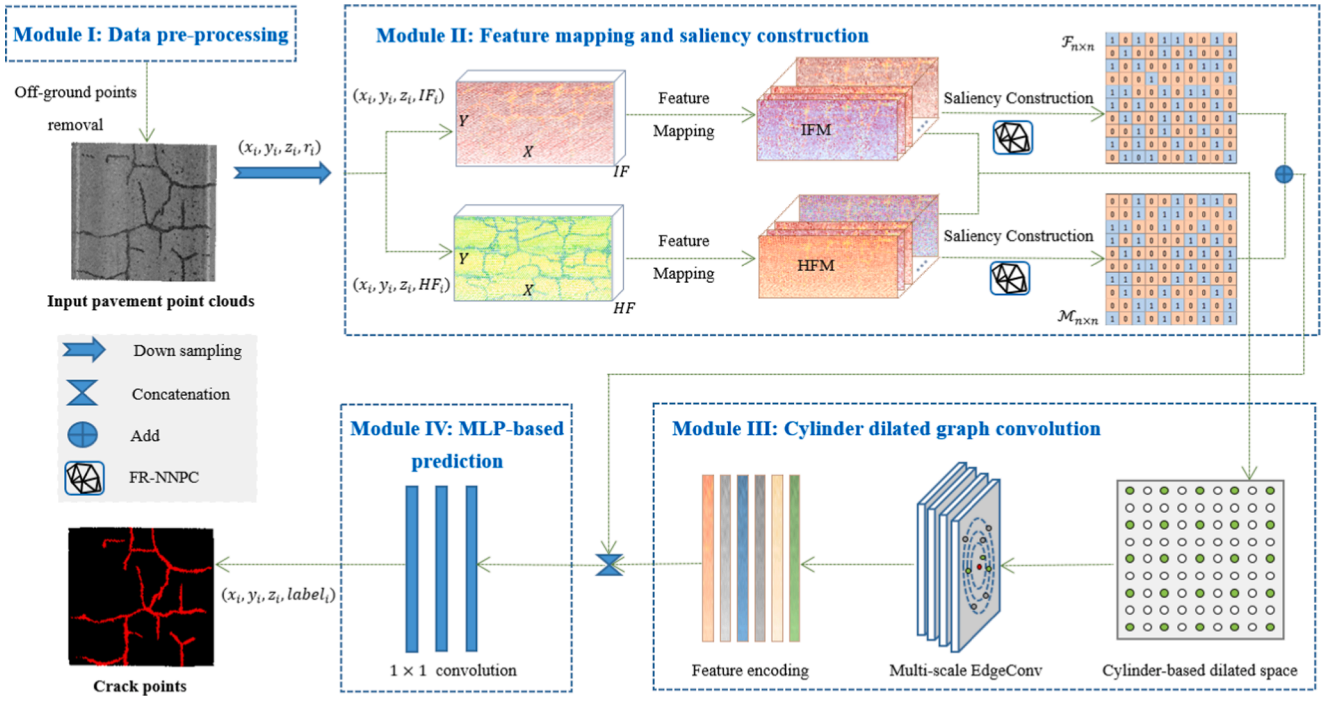


Fig. 2. Workflow of the SD-GCN model.

convolution strategy. Finally, in Module IV, an MLP-based net architecture is designed for binary classification result refinement, and pavement crack points are thus extracted.

3.1. Data pre-processing

Because we mainly focus on pavement cracks in this work, the off-road objects (e.g., poles, buildings, and cars) in raw point clouds are first removed using a revised curb-related pavement extraction method for memory cost reduction and computational efficiency improvement. The experimental details and parameter settings of this pavement extraction method were presented in Ma et al. (2019). Then, to achieve the lightweight end-to-end neural networks and boost the computational speed, we employ the farthest point sampling (FPS) down-sampling method (Qi et al., 2017a) to decrease the number of pavement point clouds. More specifically, only 1% of the total number of points in each point cloud segment is remained and then used as the inputs for feature mapping and saliency construction tasks.

3.2. Feature mapping and saliency construction

3.2.1. Feature mapping space construction

Compared to the surrounding pavement points, the crack points usually show lower height and intensity values. Moreover, these crack points have uneven data distributions and irregular geometrical shapes. In contrast, pavement points on flat terrains are normally located in a horizontal planar with similar heights. Because of delayed road maintenance and dusty road surfaces, both height and intensity differences between pavement points and crack points are quite small, which makes accurate road crack extraction challenging. Thus, two feature maps (i.e., height feature map and intensity feature map) are constructed to enhance the height and intensity saliency of crack points.

The point clouds obtained from MLS systems contain spatial information (i.e., *X, Y, Z*) and intensity (*R*) information. For each point $p_i = (x_i, y_i, z_i, r_i)$, $i = 1, 2, \cdots, n$, where n is the total number of input point clouds, the new height feature (HF) and intensity feature (IF) can be calculated as follows:

$$HF_i = cot(\frac{1}{1 + e^{-\lambda \times z_i}}) \tag{1}$$

$$IF_i = cot(\frac{1}{1 + e^{-\lambda \times r_i}}) \tag{2}$$

where HF_i and IF_i denote the new salient height features and intensity features of each point p_i, $cot(\bullet)$ indicates the cotangent function. λ is a predefined coefficient, which indicates the transformation degree between the original height and intensity features to the new ones. Specifically, $\lambda = 1$ is set in this paper.

In this paper, both HF and IF are designed to amplify intra-class compactness and alleviate inter-class imbalance. Consequently, we generate two new feature maps, i.e., the height feature map (HFM) and intensity feature map (IFM), respectively. For each point p_i, we apply both Eqs. (1) and (2) to convert original inputs p_i into two mapping outputs $m_i = (x_i, y_i, z_i, HF_i)$ and $n_i = (x_i, y_i, z_i, IF_i)$, respectively. Then, we can obtain two output datasets, i.e., $M = \{m_1, m_2, \cdots, m_n\}$ and $N = \{n_1, n_2, \cdots, n_n\}$. Such datasets can be further employed as inputs for feature encoding and saliency construction in the process of pavement crack extraction.

3.2.2. Saliency matrix

In both HFM and IFM feature mapping spaces, the feature saliency is further estimated by calculating the distances from the normal of each point to the principal normal of the input point clouds. For an input point cloud, the centroid of this point cloud is regarded as the principal normal vector. Given $\forall m_i, m_j \in M$, we compute the height feature distance (HFD) $\mathscr{D}_m$ using the following equation:

$$\mathscr{D}_m(m_i, m_j) = \|m_i - m_j\|_2 \tag{3}$$

Where $\| \bullet \|_2$ represents the Euclidean distance. Likewise, the intensity feature distance (IFD) $\mathscr{D}_n$ could be also calculated using the same operation.

In addition, the FR-NNPC method is employed to obtain the surrounding points of each point in the datasets *M* and *N*, respectively. In comparison with K-nearest neighbors (KNN), the FR-NNPC algorithm uses the fixed radius search and neighbor progressive competition rules to search candidate points and compute their scores, ensuring the sum of these scores of crack points is higher than pavement points. Because there is a problem of the imbalance of classes in the dataset created in this paper, resulting in the number of pavement points is greatly higher than the number of crack points. Accordingly, the FP-NNPC algorithm could capture the feature differences magnified for all points during the HFM and IFM space construction stages, addressing the inter-class imbalance dilemma with few manual parameter settings. The following equation is performed to determine the scores:

$$\begin{cases} S_{crack} = IR \times \dfrac{P_{all} - L_p}{P_{all} - 1 - \sqrt{\mathrm{IR}}} \\ S_{pavement} = -(IR^{(\frac{L_p}{P_{all}})}) \end{cases} \tag{4}$$

where P_{all} denotes the total number of input point clouds, *IR* represents the imbalance ratio. In a binary classification problem, the imbalance ratio is typically defined to measure the degree of imbalance of a specific dataset. The higher the imbalance ratio is, the higher scores will be assigned to few training samples, namely crack points in this study. In this paper, $IR = P_{maj}/P_{min}$ is applied to determine the imbalance ratio, where P_{maj} and P_{min} represent the number of point samples of the majority (i.e., pavement points) and minority (i.e., crack points) categories, respectively. L_p denotes the grade value of samples. If a point p_i belongs to cracks, it will be assigned by lower grade values but higher scores than pavement points, and vice versa.

Furthermore, the height saliency matrix $\mathscr{M}_{n\times n} = (\mathscr{M}_1, \mathscr{M}_2, \cdots, \mathscr{M}_n)$ of the dataset *M* and the intensity saliency matrix $\mathscr{T}_{n\times n} = (\mathscr{T}_1, \mathscr{T}_2, \cdots, \mathscr{T}_n)$ of the dataset *N* are generated, respectively. The height saliency vector of m_i is constructed as $\mathscr{M}_i = (m_{i1}, m_{i2}, \cdots, m_{in})$, $m_{ij}(j \in [1, n])$ is computed by the following equation:

$$m_{ij} = \begin{cases} 1, if\ \mathscr{D}_m(m_i, m_j) \le R \\ 0, else \end{cases} \tag{5}$$

where *R* indicates a fixed radius (or cutoff distance), it is obtained from the mean square Euclidean distance between neighboring points. Similarly, the intensity saliency vector of n_i is constructed as $\mathscr{T}_i = (f_{i1}, f_{i2}, \cdots, f_{in})$, $n_{ij}(j \in [1, n])$ can be also calculated by repeating the above operations.

3.3. Cylinder dilated graph convolutions

Although CNNs have been demonstrated to achieve dominant performance in various fields, it is still challenging to solve problems with non-Euclidean data, particularly for unordered 3D point clouds (Li and Baciu, 2021). Accordingly, graph convolutions are proposed to obtain local geometrical features, as well as edge features between a discrete point and its neighboring points. These discrete points are considered as nodes, while the feature connections between two adjacent points are considered as edges in a graph structure. Moreover, in the computer vision and image processing domains, dilated convolutions are developed to decrease spatial information loss caused by pooling operations, which expands the convolution kernel and amplifies the receptive field with a relatively small number of parameters. (Li et al., 2019b). For instance, Zhou et al. (2018) employed dilated convolutions in both cascade and parallel modes to preserve detailed spatial and texture information, and thus significantly enhanced the road segmentation

accuracy from satellite images.

In this paper, we employ dilated aggregation operations in a graph-based convolutional net structure for receptive field adjustment without downsizing the resolutions of feature maps. To this end, we first propose a cylinder-based dilated convolution strategy to partition the input feature points obtained from the HFM feature space. Then, the EdgeConv operation in the DGCNN network (Wang et al. 2019) is used to capture fine-grained geometric features and global shape properties of road cracks within each cylinder point cloud space.

Specifically, an input point cloud named *PC* is partitioned into many cylinder point cloud subsets with a size of *R* and *H*, where *R* is the radius of the base of the cylinder in the *XY* plane and *H* is the height of the cylinder (see Fig. 3). In particular, the value of *H* is ascertained by the maximum height value of all point clouds. D is the dilation rate, referring to the number of cylinder intervals in this paper. Consequently, the cylinder dilated convolutions return interval cylinder point cloud sections in the HFM feature space by skipping every *D* cylinders. As illustrated in Fig. 3, the white cylinder point cloud spaces are discarded, while the green ones are kept for computational cost reduction.

Inspired by the EdgeConv proposed in the DGCNN network, we advocate a multi-scale EdgeConv as the classifier to accurately extract pavement cracks within each cylinder point cloud section. DGCNN applies the EdgeConv operation to capture geometrical features and edge structures in local areas. Unlike conventional graph convolutions introduced in DGCNN, the graphs constructed in this paper are in multiple scales instead of a fixed scale. To this end, within each cylinder-based HFM feature space, the K-nearest neighbors of a point are first ascertained in multiple scales, i.e., $k = 16, 32$ and 64, respectively. Then, all features captured in multiple scales are combined together as the new inputs to feed into the following multi-scale EdgeConv architectures. Thus, by performing multi-scale EdgeConv operations, not only point-wise geometrical features but edge descriptive features among different points could contribute to encoding more inherent features in multiple local neighborhoods.

Furthermore, this multi-scale EdgeConv descriptor is constructed depending on a channel-based symmetric aggregation processing on the edge features from all points, and a max operation is used to reduce the computational consumption. In each multi-scale EdgeConv module, a symmetric edge function is adopted to catch both the local neighbor structures and global characteristics in an effective way using the following equation:

$$g_\psi(x_i, y_i) = g_\psi(x_i, x_j - x_i)/2 \tag{6}$$

where g_ψ indicates a nonlinear parameterized function, x_i and x_j denote the *X* coordinate of two adjacent points, respectively. Likewise, in the IFM feature space, the same cylinder-based dilated partition strategy is conducted to divide the input points into many point cloud sections, followed by the multi-scale EdgeConv operation for learning more high-level features.

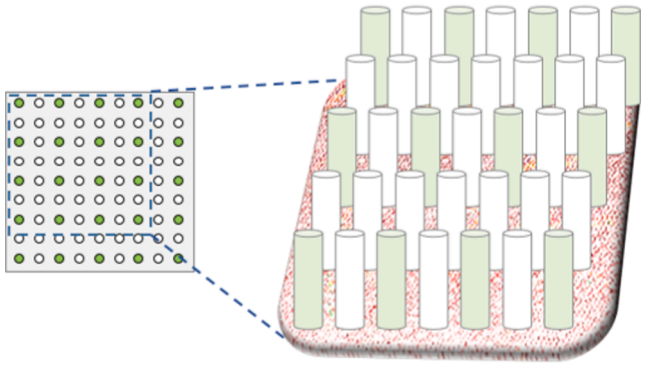


Fig. 3. Illustration of the proposed cylinder-based dilated convolution in 3D space. Left: The 3D cylinder-based space partition scheme in the top view. Right: The 3D cylinder-based space partition with a kernel size of 7×7 and dilation rate of 2 in a zoom-in view.

3.4. MLP-based prediction Module

Finally, the fused features learned from both HFM and IFM feature spaces through several multi-scale EdgeConv operations are fed into an MLP-based prediction module, as depicted in Fig. 2. This prediction module uses three MLP layers with 1×1 convolutions on the fused features of each point to output the point-wise binary classification results, i.e., pavement points and crack points.

3.5. Implementation details

The developed SD-GCN model was developed and evaluated using Python 3.6.9, Tensorflow 2.3.0, Nvidia® RTX 3090 with 24 GB memory, Intel® I7-11700 8-core CPU @2.5 GHz, and 64 GB RAM on the Ubuntu 18.04 operating system. Batch normalization and ReLU activation function were used after each convolutional operation. Moreover, the built-in Adam optimizer in Tensorflow was employed for model optimization. Several hyperparameters, including the initial learning rate and batch size, were also fine-tuned at the training stage to ascertain the optimal settings. According to multiple experiments and prior knowledge, the batch size and dropout rate were predefined as 32 and 0.5, respectively. The initial learning rate was 0.0001 with a decreasing rate of 20% after every 20 epochs, while a total of 200 iterations was set to train the developed model.

4. Dataset and evaluation matrix

4.1. Experimental data

In this work, the MLS point clouds were collected by a RIEGL-VMX 450 system in Qinghai-Tibet (QT) Freeway, China, which has a total length of 1,937 km and reaches its highest evaluation about 5,230 m at Tanggula Mountains Pass. Because both urban roads and plain highways are frequently maintained, pavement cracks on such road sections are repaired in time. Due to the serious weather conditions, thawing permafrost, and variable topography of the QT Freeway, however, it is considerably challenging to regularly maintain and recover pavement cracks. Fig. 4(a) shows the pavement cracks captured by the RIEGL-VMX 450 MLS system.

The RIEGL-VMX 450 MLS system, consisting of two high-end RIEGL VQ-450 laser heads, can reach a 400 lines/sec scanning speed and an 800 m effective measurement range in an "X-shape" configuration pattern. The maximum absolute positioning accuracy could achieve 8 mm with an average driving speed of 80 km/h on QT Freeway. Table 1 indicates the specifications of the RIEGL VQ-450 laser scanner. Accordingly, a total number of 115 pavement point cloud segments were generated from the collected QT Freeway point cloud dataset to train and test the developed neural networks.

The 3D crack points were manually labeled in each pavement point cloud segment. The proposed methods in this work were trained and evaluated using the created pavement crack benchmark dataset. To this end, all 3D points were annotated point-by-point as either "crack" or "pavement" on such point cloud segments according to manual inspection. Each road segment usually has an average of 220,000 points in a $10 \times 10\ m^2$ area, while the ratio between crack points and pavement points is about 20%:80%. The annotated ground truth data is shown in Fig. 4 (b). Moreover, the whole labeled points were split into 70%, 10%, and 20% subsets for training, validating, and testing purposes.

4.2. Evaluation metrics

To estimate the performance of the developed SD-GCN model, a quantitative accuracy assessment is performed from pavement point

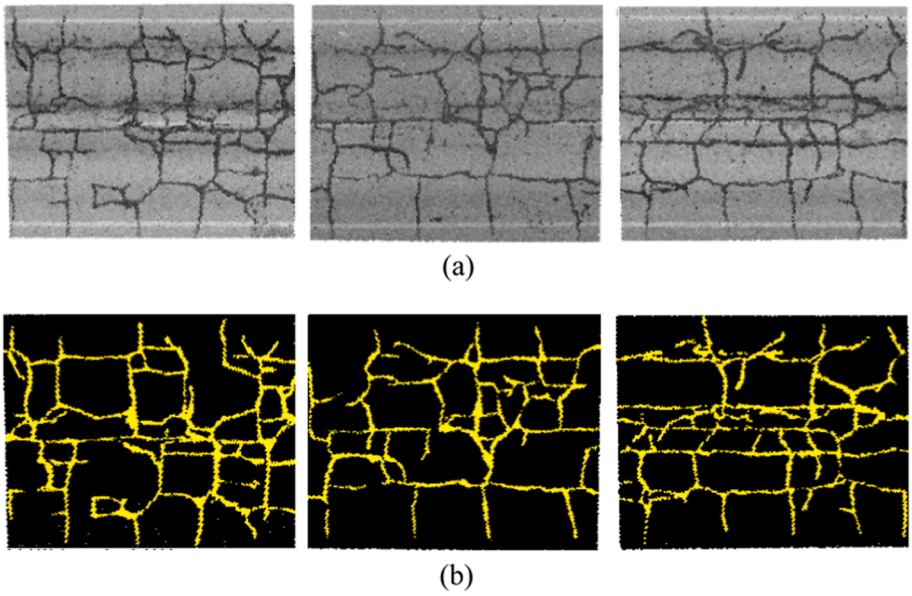


Fig. 4. RIEGL VMX-450 MLS point cloud samples. (a) Road crack samples with intensity information. (b) Manually annotated ground truth data, the black color represents pavement points and the yellow color denotes crack points.

Table 1
Specifications of the RIEGL VQ-450.

Specification	Value
Scan frequency	200 lines/sec
Measurement range	1.5–800 *m*
Measurement precision	5 mm
Absolute accuracy	8 mm
Angle measurement resolution	0.001
Scanner FOV	Full circle
Laser wavelength	Near infrared

cloud segments. Three evaluation metrics, including Precision (Pr), Recall (Re), and F_1-score (F_1) are accordingly employed, which are formulated as follows:

$$Pr = \frac{P_{tp}}{P_{tp} + P_{fp}} \tag{7}$$

$$Re = \frac{P_{tp}}{P_{tp} + P_{fn}} \tag{8}$$

$$F_1 = \frac{2*Pr*Re}{Pr + Re} \tag{9}$$

where P_{tp}, P_{fp}, and P_{fn} denote the number of true-positive points, false-positive points, and false-negative points, respectively. It is noteworthy that the precision shows the valid percentage of the extracted pavement cracks, while the recall addresses the completeness of the extracted cracks. In addition, F_1-score indicates an overall evaluation index by assessing both precision and recall.

5. Result and discussion

This section introduces the optimal hyperparameter settings, followed by experimental results. Then, an assessment evaluation is implemented to quantitatively and qualitatively analyze the model performance. Finally, we perform a comparative study to compare the developed SD-GCN model with SOTA methods.

5.1. Hyperparameter optimization

The developed SD-GCN model has two essential hyperparameters, i.e., R, the radius of the cylinder base in the XY plane; and D, the dilation rate during the process of cylinder-based dilated convolutions. Multiple experiments were performed to ascertain the optimal hyperparameter combinations. Based on the variable-controlling strategy, we evaluated the model performance of SD-GCN using 15 different combinations (i.e., 5 alternatives of R and 3 alternatives of D).

Size of *R*: The size of radius R plays a crucial role in implementing the SD-GCN model. A suitable value of R not only enables the model to capture more inherent local features but greatly reduces the computational costs. Typically, the larger R is, the more features the model could learn, yet the more computational burden involves. Since each road segment has a 10×10 m^2 area, we tested the value of R changing in the range of [0, 0.5 m]. More specifically, different values, i.e., 0.1 m, 0.2 m, 0.3 m, 0.4 m, and 0.5 m, respectively, were determined to calculate their F_1-score and seek an optimal tradeoff between computational burden (measured as the forward propagation time) and model performance, while remaining $D = 2$ during the testing phase.

Fig. 5(a) illustrates different F_1-score and forward propagation time obtained by varying R values. Intuitively, the performance curves locate in the top-left area of two diagrams, which means the SD-GCN could achieve high F_1-score with low computational complexity. More specifically, the forward pass time of SD-GCN increases with the decrease of the radius size R. The reason is that at the stage of cylinder dilated graph convolutions, more cylinder data spaces will be built with the decreasing sizes of R. Note that, crack extraction performance increases by 1.7% when varying the radius sizes $R = 0.5$m (F_1-score = 77.6%) to $R = 0.4$m (F_1-score = 78.3%), that is because the proposed SD-GCN could learn more high-level features and avoid the overfitting issue by using less training data in smaller cylinder data spaces. Therefore, $R = 0.4$m was defined as the fine-tuned hyperparameter setting so that the SD-GCN could achieve high extraction accuracy and low computational costs.

Size of D: At the stage of cylinder-based dilated convolutions, an appropriate value of the dilation rate D could deliver both effective and robust pavement crack extraction results in an efficient manner. Herein, considering various point densities within different input point cloud segments, we tested the SD-GCN model performance by setting D = 1, 2, and 3 through trial and error, respectively, on the testing dataset while keeping $R = 0.4$ m all the time. Fig. 5(b) shows the model

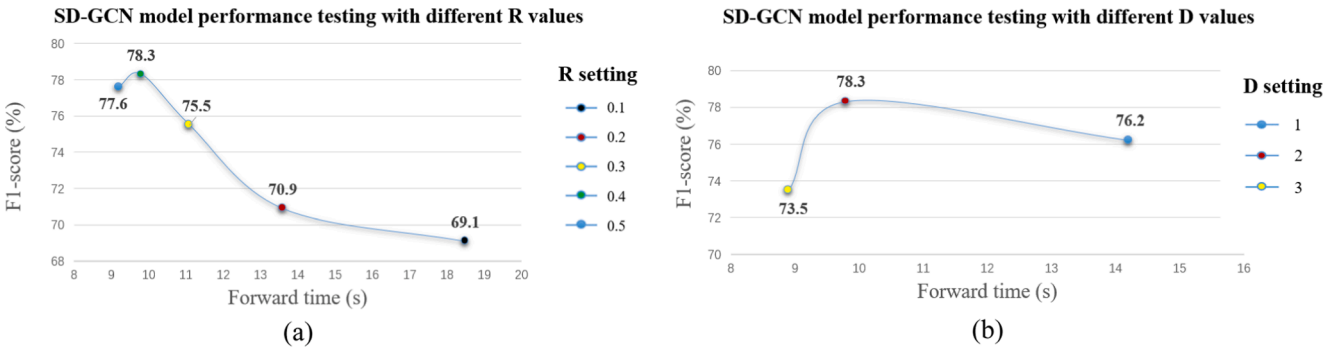


Fig. 5. SD-GCN model performance evaluation based on F_1-score and forward propagation time. (a) Testing with different R values. (b) Testing with different D values.

performance by setting different *D* values. As can be perceived, to make the best balance between representative feature encodings and relatively low computational burden, it can deliver the best performance for pavement crack extraction by setting D = 2. Note that, for the point-wise crack extraction task, the F_1-score increases by 4.8% when changing D = 3 to D = 2. On the whole, the smaller the dilation rate is, the more local and geometrical features the model can learn, but the lower computational cost. Therefore, in the SD-GCN, we determined D = 2 as the optimal hyperparameter setting.

5.2. Crack extraction results

According to various experiments by setting different hyper-parameter combinations, we predefine the optimal settings as $R = 0.4$ m and D = 2 in the training process. Meanwhile, the dropout rate, batch size, initial learning rate, and momentum of Adam are 0.5, 32, 0.0001, and 0.9, respectively, which could achieve the best extraction results. Fig. 6 presents the pavement crack extraction results from MLS point clouds, which demonstrates that the designed SD-GCN model could deliver a promising solution for the crack extraction task in complex road environments. Despite pavement cracks obtained from highway environments in different structures and sizes, the experimental results indicate a variety of pavement cracks were completely extracted.

SD-GCN network has powerful feature saliency construction and graph representation capabilities in local regions, contributing to accurate point-wise crack extraction results. However, complex structures and low connectivity of pavement cracks in varying shapes have profound impacts on the feature encoding performance. As shown in Fig. 6, some crack points were misclassified as pavement points. The crack occlusion, decay, and moving road users (e.g., vehicles) in the QT Freeway can conduce to the false point-wise classification.

5.3. Comparative study

To demonstrate the model performance and practical feasibility, we compared the developed SD-GCN model with the existing methods, in terms of pixel-wise methods, i.e., U-Net (Ronneberger et al., 2015) and AU-Net (Oktay et al., 2018), and point-wise methods, i.e., 3D-Skeleton (Yu et al., 2014), PointNet (Qi et al., 2017b), DGCNN (Wang et al., 2019), and Feng's method (Feng et al., 2021). We adopted the same testing protocols used in Feng's method (Feng et al., 2021). Fig. 7 shows the extracted pavement cracks using five different methods.

More specifically, for pixel-wise comparison methods, the point cloud samples were first transformed into 2D intensity images. Then, such images were used to train and test both U-Net and AU-Net networks, resulting in average pixel-wise evaluation metrics. Meanwhile, all hyperparameters evolving in the process of training are predefined as default. Nevertheless, such methods normally cause information loss during the dimensionality reduction process.

For point-wise comparison methods, 3D-Skeleton is a traditional threshold-based method that consists of Otsu thresholding, density filtering, Euclidean distance-based clustering, and L_1-median skeleton extraction algorithms. We used the same hyperparameter settings as in Feng et al. (2021) by keeping the density threshold $ds = 1.2$ and the local radius $rd = 0.2$ m for crack extraction and outlier removal purposes. Nevertheless, it is remarkably challenging to ascertain the fine-tuned threshold settings in different testing scenarios. PointNet is the pioneer of point-wise classification methods, which comprises a symmetric function with respect to permutation invariance. But PointNet cannot effectively learn local features, which makes fine-grained pattern recognition difficult, particularly for complex road scenes. DGCNN introduced a dynamic edge convolution for 3D point cloud segmentation. Furthermore, Feng's method developed a graph-widen deep

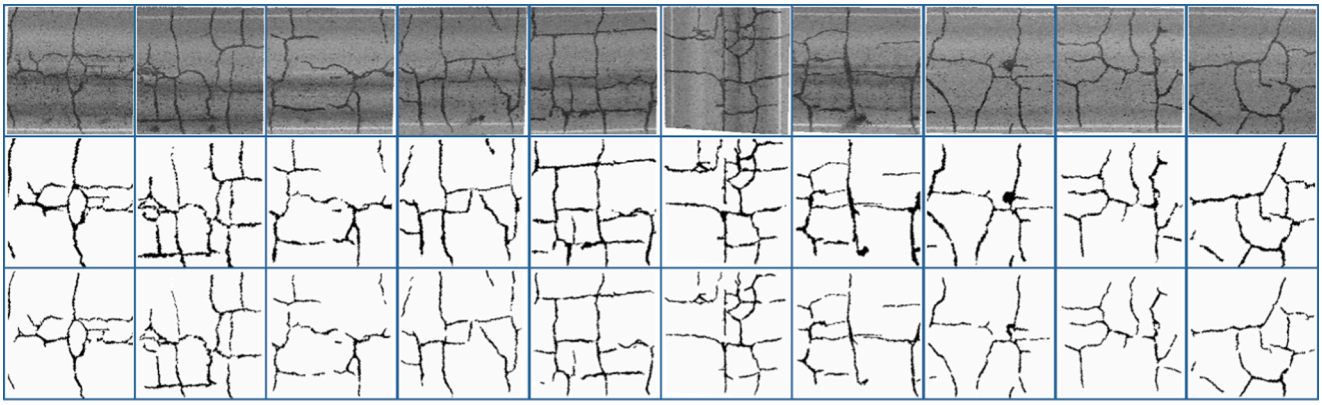


Fig. 6. Pavement crack extraction results using the proposed SD-GCN model. Top: Original input point cloud samples with intensity information. Middle: Ground truth data. Bottom: Pavement crack extraction results. The black parts are crack points and the white ones are pavement points.

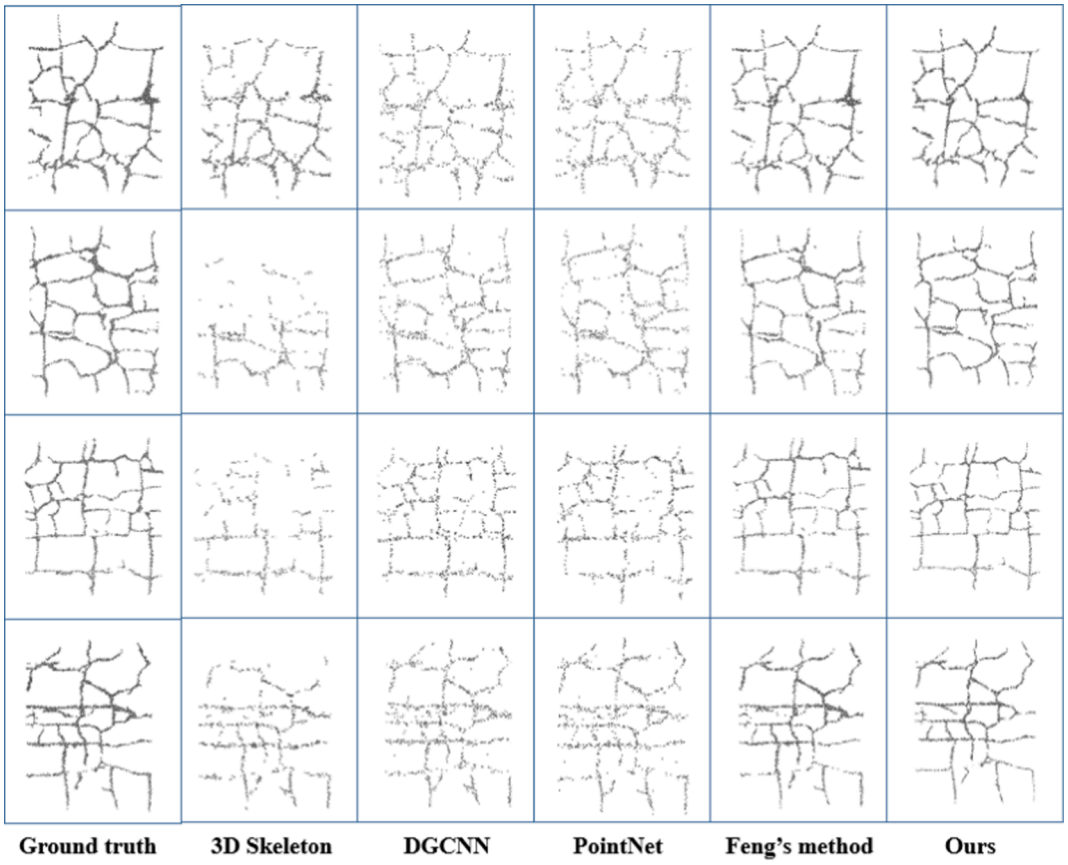


Fig. 7. Comparison results of pavement crack extraction using different methods.

learning framework to extract pavement cracks from MLS point clouds by constructing descriptive graph structures and salient feature representations. However, these graph-based methods cannot tackle the sample imbalance problem when using KNN to explore the surrounding points in large-scale point cloud spaces. Likewise, all hyperparameters used in the training phase are set as default.

Table 2 indicates the model performance comparison results by calculating precision, recall, and F_1-score evaluation metrics. For pixel-wise comparison methods, U-Net and AU-Net achieve an average of precision, recall, and F_1-score of [80.1%, 72.9%, 76.3%] and [75.8%, 77.2%, 76.5%], respectively. For point-wise comparison methods, 3D-Skeleton, DGCNN, PointNet, and Feng's method achieve an average of precision, recall, and F_1-score of [33.3%, 71.8%, 45.5%], [73.4%, 67.2%, 70.2%], [69.7%, 64.5%, 67.0%], and [70.2%, 73.5%, 71.8%], respectively. In contrast, the proposed SD-GCN model in this work delivers an average of precision, recall, and F_1-score of [79.5%, 77.1%, 78.3%], which is 32.8% higher than 3D-Skeleton and 11.3% higher than PointNet in term of F_1-score. The experimental results demonstrate that the SD-GCN outperforms most traditional rule-based and supervised learning-based algorithms, mainly because of the descriptive feature encoding capability reinforced by feature saliency space construction and dilated graph convolutions in multiple scales. As can be seen in Fig. 7, the pavement cracks are incompletely extracted using four comparative methods, while the pavement cracks extracted by the SD-GCN network could achieve higher completeness and fewer outliers by comparison. On the whole, the SD-GCN model proposes a potential solution for pavement crack extraction from unorganized and massive MLS point clouds.

Table 2
Pavement crack extraction results obtained by varying methods.

Method	Type	Precision (%)	Recall (%)	F_1-score (%)
U-Net (Ronneberger et al., 2015)	Pixel-wise	**80.1**	72.9	76.3
AU-Net (Oktay et al., 2018)	Pixel-wise	75.8	**77.2**	76.5
3D-Skeleton (Yu et al., 2014)	Point-wise	33.3	71.8	45.5
DGCNN (Wang et al., 2019)	Point-wise	73.4	67.2	70.2
PointNet (Qi et al., 2017b)	Point-wise	69.7	64.5	67.0
Feng's Method (Feng et al., 2021)	Point-wise	70.2	73.5	71.8
Ours	Point-wise	79.5	77.1	**78.3**

5.4. Efficiency evaluation

Moreover, to highlight the performance efficiency of the SD-GCN model, we carried out a model complexity comparison by taking model size, forward propagation time, and F_1-score into account using the same training and testing datasets created in this study. Table 3 shows the comparison results. Note that, the model size of our method is 66% smaller than PointNet and 58% smaller than DGCNN, respectively. Besides, compared to PointNet and DGCNN, our method only consumes

Table 3
Model complexity, forward time, and performance of different methods.

Method	Model size (MB)	Forward time (s)	F_1-score (%)
PointNet (Qi et al., 2017b)	23.8	27.5	67.0
DGCNN (Wang et al., 2019)	19.3	42.2	70.2
Ours	8.1	7.7	78.3

28% and 18% forward pass time but achieves 11.3% and 8.1% F_1-score improvement, respectively. The reason is that the SD-GCN model is designed in a semi-supervised manner, which consumes less training data due to downsampling operations and dilated graph convolutions. In conclusion, the developed SD-GCN model in this paper not only provides more accurate pavement crack extraction results but also consumes less processing time than most existing methods, including PointNet and DGCNN. Accordingly, multiple GPU parallel computing techniques could be further employed to dramatically accelerate the SD-GCN framework for real-time applications, e.g., autonomous driving.

6. Conclusion

In this work, we have developed a novel deep learning framework, named SD-GCN, by exploring feature saliency space construction and cylinder-based dilated graph convolutions for pavement crack extraction from MLS point clouds. Experimental results demonstrate that the SD-GCN model could achieve high extraction accuracy and low computational costs, benefitting from effective feature mapping and saliency construction, powerful feature encoding capability, and compact-designed neural network architecture. Consequently, the SD-GCN model delivers an average of precision, recall, and F_1-score of 79.5%, 77.1%, and 78.3%, respectively. Moreover, comparison results indicate that our method outperforms SOTA point-wise segmentation approaches regarding extraction accuracy and computational efficiency. To summarize, we have developed a promising solution, which significantly reinforces the pavement crack extraction performance from MLS point clouds. For further research, we are dedicated to exploiting more effective loss functions to tackle sample imbalance problems and designing more compact network structures in an unsupervised pattern to capture more geospatial correlations and adjacency relationships among neighboring points.

CRediT authorship contribution statement

Lingfei Ma: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Investigation, Validation, Funding acquisition. **Jonathan Li:** Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was partially funded by the National Natural Science Foundation of China under Grant 42101451. We acknowledge Xiamen University for providing us with raw point clouds and ground truth data used in this paper.

References

Cao, W., Liu, Q., He, Z., 2020. Review of pavement defect detection methods. IEEE Access 8, 14531–14544.
Cha, Y.J., Choi, W., Büyüköztürk, O., 2017. Deep learning-based crack damage detection using convolutional neural networks. Comput.-Aided Civ. Infrastruct. Eng. 32 (5), 361–378.
Dung, C.V., Anh, L.D., 2019. Autonomous concrete crack detection using deep fully convolutional neural network. Autom. Constr. 99, 52–58.
Fei, Y., Wang, K.C.P., Zhang, A., Chen, C., Li, J.Q., Liu, Y., Yang, G., Li, B., 2020. Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V. IEEE Trans. Intell. Transport. Syst. 21 (1), 273–284.
Feng, H., Li, W., Luo, Z., Chen, Y., Fatholahi, S. N., Cheng, M., & Li, J. (2021). GCN-Based Pavement Crack Detection Using Mobile LiDAR Point Clouds. *IEEE Trans. Intell. Transport. Syst.*, 10.1109/TITS.2021.3099023.
Guan, H., Li, J., Yu, Y., Chapman, M., Wang, H., Wang, C., Zhai, R., 2014. Iterative tensor voting for pavement crack extraction using mobile laser scanning data. IEEE Trans. Geosci. Remote Sens. 53 (3), 1527–1537.
Hsieh, Y.A., Tsai, Y.J., 2020. Machine learning for crack detection: Review and model performance comparison. J. Comput. Civ. Eng. 34 (5), 04020038.
Kamaliardakani, M., Sun, L., Ardakani, M.K., 2016. Sealed-crack detection algorithm using heuristic thresholding approach. J. Comput. Civ. Eng. 30 (1), 04014110.
Li, Z., Cheng, C., Kwan, M.P., Tong, X., Tian, S., 2019a. Identifying asphalt pavement distress using UAV LiDAR point cloud data and random forest classification. ISPRS Int. J. Geoinf. 8 (1), 39.
Li, G., Muller, M., Thabet, A., Ghanem, B., 2019b. Deepgcns: Can GCNs go as deep as CNNs?. In *Proc. IEEE Conf. Comput. Vis.* Pattern Recognit. CVPR, 9267–9276.
Li, H., Song, D., Liu, Y., Li, B., 2018. Automatic pavement crack detection by multi-scale image fusion. IEEE Trans. Intell. Transport. Syst. 20 (6), 2025–2036.
Li, Y., Ma, L., Zhong, Z., Liu, F., Chapman, M.A., Cao, D., Li, J., 2020. Deep learning for lidar point clouds in autonomous driving: A review. IEEE Trans. Neural Netw. Learn. Syst. 32 (8), 3412–3432.
Li, Y., Baciu, G., 2021. HSGAN: Hierarchical Graph Learning for Point Cloud Generation. IEEE Trans. Image Process. 30, 4540–4554.
Ma, L., Li, Y., Li, J., Wang, C., Wang, R., Chapman, M.A., 2018. Mobile laser scanned point-clouds for road object detection and extraction: A review. Remote Sens. 10 (10), 1531–1563.
Ma, L., Li, Y., Li, J., Zhong, Z., Chapman, M.A., 2019. Generation of horizontally curved driving lines in HD maps using mobile laser scanning point clouds. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12 (5), 1572–1586.
Ma, L., Li, Y., Li, J., Junior, J. M., Gonçalves, W. N., & Chapman, M. A. (2021). BoundaryNet: Extraction and Completion of Road Boundaries with Deep Learning Using Mobile Laser Scanning Point Clouds and Satellite Imagery. *IEEE Trans. Intell. Transport. Syst.*, 10.1109/TITS.2021.3055366.
Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Rueckert, D., 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv: 1804.03999*.
Pan, Y., Zhang, X., Cervone, G., Yang, L., 2018. Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (10), 3701–3712.
Pu, S., Rutzinger, M., Vosselman, G., Elberink, S.O., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. ISPRS J. Photogramm. Remote Sens. 66 (6), S28–S39.
Qi, C.R., Yi, L., Su, H., Guibas, L.J., 2017a. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Adv. Neural Inf. Process. Syst., 30.
Qi, C.R., Su, H., Mo, K., Guibas, L.J., 2017b. Pointnet: Deep learning on point sets for 3d classification and segmentation. Proc. IEEE Conf. Comput. Vis Pattern Recognit. (CVPR) 652–660.
Ragnoli, A., De Blasiis, M.R., Di Benedetto, A., 2018. Pavement distress detection methods: A review. Infrastructures 3 (4), 58.
Ravi, R., Bullock, D., Habib, A., 2021. Pavement Distress and Debris Detection using a Mobile Mapping System with 2D Profiler LiDAR. Transp. Res. Rec. 2675 (9), 428–438.
Riid, A., Louk, R., Pihlak, R., Tepljakov, A., Vassiljeva, K., 2019. Pavement distress detection with deep learning using the orthoframes acquired by a mobile mapping system. Appl. Sci. 9 (22), 4829–4850.
Ronneberger, O., Fischer, P., Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In: Int. Conf. Med. Image Comput. Comput.-Assist Interv., Springer, Cham, pp. 234-241.
Shi, Y., Cui, L., Qi, Z., Meng, F., Chen, Z., 2016. Automatic road crack detection using random structured forests. IEEE Trans. Intell. Transport. Syst. 17 (12), 3434–3445.
Tan, Y., Li, Y., 2019. UAV photogrammetry-based 3D road distress detection. ISPRS Int. J. Geoinf. 8 (9), 409.
Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M., 2019. Dynamic graph CNN for learning on point clouds. ACM Trans. Graph. 38 (5), 1–12.
Xu, X., Yang, H., 2019. Intelligent crack extraction and analysis for tunnel structures with terrestrial laser scanning measurement. Adv. Mech. Eng., 11(9), 1687814019872650.
Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X., Ling, H., 2019. Feature pyramid and hierarchical boosting network for pavement crack detection. IEEE Trans. Intell. Transport. Syst. 21 (4), 1525–1535.
Yu, A., Mei, W., Han, M., 2021. Deep learning based method of longitudinal dislocation detection for metro shield tunnel segment. Tunn. Undergr. Space Technol., 113, 103949.
Yu, Y., Li, J., Guan, H., Wang, C., 2014, July. 3D crack skeleton extraction from mobile LiDAR point clouds. In: 2014 IEEE Geosci. Remote. Sens. Symp. (IGRSS). IEEE, pp. 914-917.
Yu, Y., Guan, H., Li, D., Zhang, Y., Jin, S., Yu, C., 2020. CCapFPN: A context-augmented capsule feature pyramid network for pavement crack detection. IEEE Trans. Intell. Transport. Syst., 10.1109/TITS.2020.3035663.
Zhang, A., Wang, K.C.P., Li, B., Yang, E., Dai, X., Peng, Y.i., Fei, Y., Liu, Y., Li, J.Q., Chen, C., 2017. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. Comput.-Aided Civ. Infrastruct. Eng. 32 (10), 805–819.
Zhong, M., Sui, L., Wang, Z., Hu, D., 2020. Pavement crack detection from mobile laser scanning point clouds using a time grid. Sensors 20 (15), 4198.

Zhou, L., Zhang, C., Wu, M., 2018. D-LinkNet: LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW) 182–186.

Zhou, P., Yi, J., Zhu, Z., Liu, P., 2019. Fixed-radius nearest neighbor progressive competition algorithm for imbalanced classification. J. Shandong Univ. (Nat. Sci.) 54 (3), 102–109. https://doi.org/10.6040/j.issn.1671-9352.1.2018.107.

Zou, Q., Zhang, Z., Li, Q., Qi, X., Wang, Q., Wang, S., 2018. Deepcrack: Learning hierarchical convolutional features for crack detection. IEEE Trans. Image Process. 28 (3), 1498–1512.