Construction and Building Materials 398 (2023) 132499

Contents lists available at ScienceDirect

Construction and Building Materials

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

3D reconstruction and segmentation system for pavement potholes based on improved structure-from-motion (SFM) and deep learning

Niannian Wang [a,b,c,d], Jiaxiu Dong [a,b,c,d,*], Hongyuan Fang [a,b,c,d], Bin Li [a,b,c,d], Kejie Zhai [a,b,c,d], Duo Ma [a,b,c,d], Yibo Shen [a,b,c,d], Haobang Hu [a,b,c,d]

[a] Yellow River Laboratory, Zhengzhou University, Zhengzhou, Henan 450001, China
[b] School of Water Conservancy Engineering, Zhengzhou University, Zhengzhou 450001, China
[c] National Local Joint Engineering Laboratory of Major Infrastructure Testing and Rehabilitation Technology, Zhengzhou 450001, China
[d] Collaborative Innovation Center of Water Conservancy and Transportation Infrastructure Safety, Zhengzhou, Henan 450001, China

ARTICLE INFO

Keywords:
Pavement pothole
SFM
Deep learning
3D reconstruction
Intelligent segmentation

ABSTRACT

Traditional pothole detection based on two-dimensional images lacks three-dimensional (3D) quantitative information such as depth and volume, although the high accuracy. In addition, existing 3D collection and reconstruction equipment based on depth cameras and lasers are expensive and difficult to operate. To solve the above problems, a low cost and automatic 3D reconstruction and segmentation system for potholes is proposed in this study. The system consists of a pavement pothole Structure-from-motion (PP-SFM) and a 3D point cloud segmentation network for potholes. Firstly, a point cloud reconstruction method called PP-SFM is proposed to reconstruct the easily obtained multi-view 2D potholes images. In addition, to a certain extent, the point cloud sparse problem is solved by the proposed PP-SFM, and the stereo display of potholes is realized. Next, Trans-3D-Seg is developed based on an improved 3D segmentation network modified by the transformer module to ealize effective segmentation of 3D point cloud data of potholes. The accuracy of the improved system is 93.44%, and the F1-score is 92.58%. In addition, the precision-recall (P-R) curve is near the upper right. Comparative experiments show that the proposed system has better segmentation performance. Compared with PointNet++, PointRCNN and PointCNN, the segmentation accuracy are increased by 4.13%, 2.96% and 3.17%, respectively. The F1-score are increased by 2.41%, 2.43% and 2.93%, respectively. The proposed system requires only a few multi-view images taken from ordinary high-definition cameras, which is a high accuracy and low cost method for 3D reconstruction and segmentation of pavement potholes.

1. Introduction

With the increase in the number of vehicles and traffic accidents, road safety has become a major concern worldwide [1]. Pavements suffer from damage such as potholes, cracks and ruts due to severe weather conditions, changing loads and aging. It has been investigated that the repair cost of damaged pavements can increase seven times in a period of five years. In addition, potholes are the most important cause of vehicle suspension damage and the biggest contributor to poor ride quality and accidents [2,3]. Frequent inspection and repair of pavement potholes is a crucial road maintenance task [4]. Therefore, continuous monitoring and maintenance of pavement is essential to provide safe driving conditions [5]. In addition, well-designed and well-maintained roads can reduce the probability and severity of road traffic accidents. As a result, there is an increasing need for automated road pothole detection systems, especially those developed based on state-of-the-art (SoTA) computer vision and machine learning technologies.

Manual inspection of pavement potholes is time-consuming, laborious and subjective [6,7]. Therefore, it is crucial that automated detection and evaluation systems are developed for efficient, objective and accurate automatic detection of pavement potholes. According to different sensors, pothole detection algorithms can be divided into three categories: vibration based, camera based and 3D laser [8]. The vibration-based approach uses accelerometers and GPS mounted on vehicles to record the impact of road damage on vehicle dynamics [9,10]. Although the method based on vibration has the advantages of low cost and fast processing speed, it will be affected by noise. Image-based pavement pothole identification and detection techniques can

* Corresponding author at: Yellow River Laboratory, Zhengzhou University, Zhengzhou, Henan 450001, China.
E-mail address: dongjiaxiu@gs.zzu.edu.cn (J. Dong).

https://doi.org/10.1016/j.conbuildmat.2023.132499
Received 12 March 2023; Accepted 11 July 2023
Available online 17 July 2023

effectively determine the location and area of damage [11–13]. The above methods collect only two-dimensional information about pavement potholes, and measurements of spatial attributes and the amount of damage to assess the severity of potholes are provided. However, three-dimensional information about potholes is lacking.

To improve the quality of pavement potholes detection, 3D laser point cloud detection methods have been developed to extract depth information of potholes. The 3D data for pavement pothole detection is generally provided by laser scanners, Microsoft Kinect sensors, or passive sensors [14,15]. However, laser scanning devices and their long-term maintenance are very expensive [3]. In addition, the Microsoft Kinect sensor is subject to significant infrared rays (IR) saturation in direct sunlight. Therefore, passive sensors, such as a single moving camera or multiple synchronized cameras, are more suitable for acquiring 3D road data and pothole detection [16]. With the advent of inexpensive, high-resolution cameras, many methods of obtaining 2D images for 3D reconstruction using a single HD camera have become a reality. The reconstructed 3D results are obtained, and effective extraction of the pothole area is achieved by 3D point cloud data detection and segmentation models.

1.1. Related work

1.1.1. 3D reconstruction method

Three-dimensional information of an object is essential for feature description. Therefore, 3D reconstruction of objects is an important element in computer graphics. Currently, the main methods used for 3D reconstruction are 3D laser scanning [17], and monocular cameras [18].

3D laser scanners has a high sampling rate and can obtain high-density point cloud data of the target surface [19,20]. In addition, 3D laser scanners can capture the geometric information of an object from the light reflected from its surface. J. Liang et al. proposed the well-known least-squares fitting algorithm to realize the three-dimensional reconstruction of power lines in LiDAR point clouds [21]. S. Dong et al. extracted the macrotexture and microtexture of pavement point cloud data based on spectral analysis techniques [22]. In addition, pavement texture information was measured. A. Yu et al. used a laser scanner to obtain the point cloud of an underpass shield tunnel [23]. Then, a method based on deep convolutional neural network was proposed to detect the longitudinal misalignment in the point cloud data. However, for existing laser devices, data noise is inevitable, resulting in incorrect texture features [24]. Furthermore, 3D laser scanners are too expensive and not easy to popularize [25].

The use of monocular images for 3D reconstruction of targets is a useful attempt. Monocular image reconstruction can be translated into a monocular depth estimation problem [25]. Monocular depth estimation is to estimate the depth of a target from a two-dimensional image [26]. This approach is relatively inexpensive and tractable. B. Li et al. proposed a model that fuses deep convolutional neural networks with conditional random fields for depth estimation of monocular images [27]. V. Guizilini et al. proposed a new deep network, PackNet, for self-supervised monocular depth estimation [28]. These monocular view-based methods rely mainly on shape features or color features and are sensitive to the scene. Furthermore, the robustness and generalization ability of the models need to be improved.

Therefore, the model called PP-SFM is proposed to estimate the 3D structure in a series of 2D images containing visual motion information.

1.1.2. 3D point cloud segmentation method

The identification of 3D point cloud target regions is also a crucial task after achieving 3D reconstruction [29]. At present, the methods used for 3D object extraction mainly include conventional processing method and deep learning-based segmentation method. F. Bosche et al. proposed a 3D CAD recognition method in the architectural field and verified the method through laboratory experiments [30]. J. Lam et al. proposed a segmentation method presented is presented which is capable of segmenting 3D images of free-form objects using piece-wise boundary curves and regions reconstructed from extracted interest points [31]. This method solves the problem of local signal fading or occlusion in chaotic scenes. F. Li et al. proposed a probabilistic graphical model to extract the road furniture in moving laser scan data [32]. J. Valença et al. proposed a detection method based on MCrack-TLS [33]. This method combines image processing and ground laser scanning to detect cracks on concrete surface.

Deep learning provides an important approach for 3D point cloud target recognition [34]. Y. Ben-Shabat et al. proposed a classification model based on 3D convolutional neural network to achieve classification of 3D point cloud data [35]. Y. Kim et al. proposed an automatic primitive classification recognition method based on curvature information and CNN to realize intelligent classification and recognition of pipes and elbows in three-dimensional point cloud data [36]. D. Bobkov et al. proposed a four-dimensional convolutional neural network to achieve target retrieval and classification of point cloud data [37]. H. Kim et al. proposed an automatic recognition method based on deep learning to intelligently extract bridge components [38]. This method greatly reduces the time of point cloud preprocessing.

1.2. Contribution of this study

To remedy the problem of information loss, high cost, unintuitive display and low segmentation accuracy, an innovative and low-cost PP-SFM and Trans-3D-Seg pavement pothole 3D point cloud data reconstruction and segmentation system is proposeda. The main contributions of this study are as follows:

First, in order to solve the problems of pavement pothole 3D point cloud data scarcity and high cost of 3D reconstruction, an improved 3D data generation model based on SFM called PP-SFM was established to generate high-quality 3D dense point cloud data of potholes.

Next, to solve the problem of effective segmentation of reconstructed point clouds, a new Trans-3D-Seg is developed based on an improved encoder and decoder modified by an transformer module to efficient and effective segmentation of reconstructed three-dimensional point cloud data of potholes. Transformer can sense pixels at a distance to learn a more comprehensive feature representation. Furthermore, the low-order features are fused with high-order features to enhance the correlation between features.Therefore, the accuracy of three-dimensional point cloud segmentation of potholes is effectively improved. In addition, the influence of asphalt aggregate gap on target extraction is solved.

The rest of this paper is organized as follows. In Section 2, the methodology used in this study is described. In Section 3, the process of PP-SFM to achieve 3D reconstruction of pavement potholes is elaborated. In Section 4, the training, validation and testing process based on Trans-3D-Seg pavement pothole 3D point cloud segmentation is described in detail. Finally, the conclusions are given in Section 5.

2. Methodology

2.1. 3D reconstruction method of pavement potholes based on PP-SFM

The SFM algorithm is an offline algorithm for 3D reconstruction based on various collected unordered images. That is, 3D information is extrapolated from the 2D images of the time series. The algorithm observes the same scene from two or more viewpoints to obtain multiple perceptual images of the scene from multiple different perspectives. The basic principle of triangulation is applied to calculate the deviation of position between image pixels. Thus, the 3D depth information of the scene is obtained. In addition, this method only requires a normal RGB camera. Therefore, the cost is low and the environment is less constrained.

In this study, the PP-SFM algorithm is proposed, and the structure of this algorithm is shown in Fig. 1. To solve the problem of 3D reconstruction of potholes under weak texture and dark light conditions, the

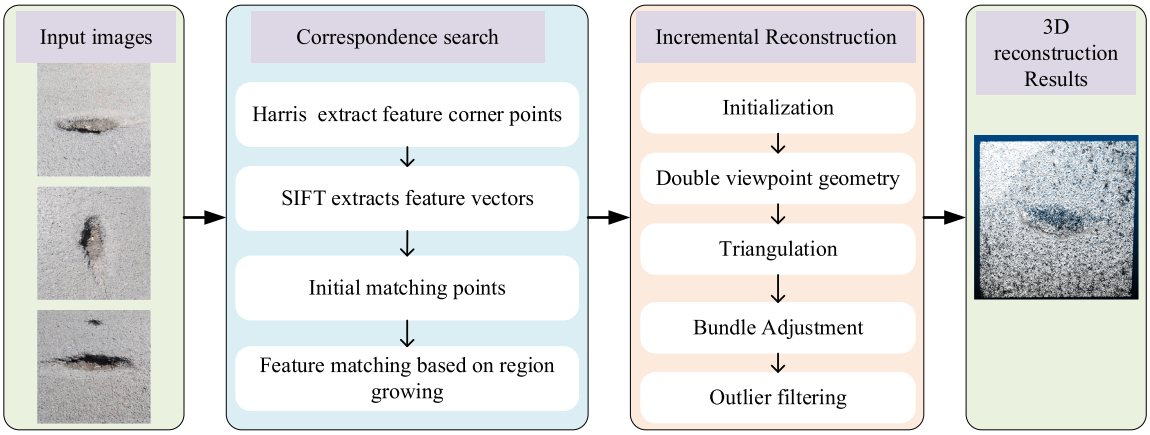


Fig. 1. PP-SFM algorithm structure.

feature extraction algorithm fusing Harris and SIFT is constructed. In order to deal with the sparse reconstruction results of the original SFM algorithm, the region growth matching technique is introduced into the algorithm.

2.1.1. Improved feature extraction and matching algorithm

The same pothole image is captured from multiple angles using a high-definition camera. During the 2D image acquisition, there may be problems such as noise, uneven image grayscale distribution and lighting variations. The Harris corner point detection algorithm extracts more comprehensive feature corner points with translation and rotation invariance [39]. Therefore, it has good robustness to the above mentioned difficulties. However, it is sensitive to image scale changes and does not have scale invariant property. The Scale-invariant Feature Transform (SIFT) [40] algorithm is invariant to rotation, scale, luminance and other changes. However, the accuracy of the obtained feature points is not very high, and the feature points cannot be extracted accurately for the targets with smooth edges. In addition, the computational effort is large and the efficiency is low. Therefore, Harris and SIFT algorithms are combined in this study. In order to obtain the dense point cloud data of potholes, the region growth matching technique is introduced in the process. The flow of the process is shown in Fig. 2.

(1) Acquisition of feature corner points

First, feature corner points in 2D images of potholes are extracted by Harris algorithm. The Harrs corner points detection algorithm uses the moving local window to calculate the gray change value in the image. If there is a large grayscale change when moving in any direction, it is considered that there are corner points in the window. In order to detect as many feature points as possible in the weakly textured regions of the 2D image, the image pixel-by-pixel grayscale changes are calculated. The corner points extracted from the 2D image can be expressed by Eq. (1).

$$R(x,y) = \det[H(x,y)] - \lambda\{tr[H(x,y)]\}^2 \quad (1)$$

In the above equation, *R(x,y)* is the corner point response function, *det(H)* is the determinant of the matrix *H*, and *tr(H)* is the trace of the matrix H. Theλ is a constant. The Hessian matrix *H* is calculated as shown in Eq. (2).

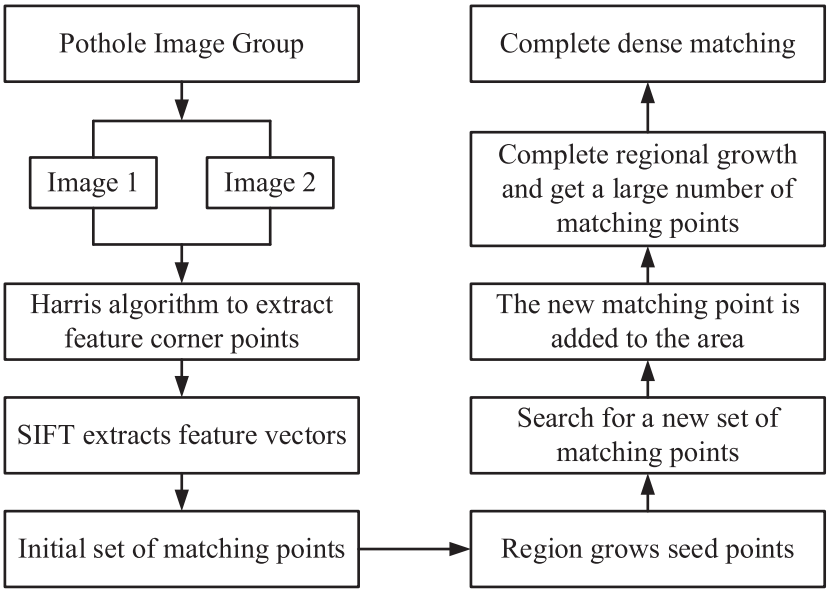


Fig. 2. Process of feature extraction and matching.

$$H(x,y) = \begin{bmatrix} I_x^2, I_x I_y \\ I_x I_y, I_y^2 \end{bmatrix} \quad (2)$$

In the above equation, I_x, I_y are the gradient values of the pixel points in the *x*-direction and *y*-direction, respectively.

(2) Extraction of feature vectors

Pavement pothole feature corner points are extracted by Harris algorithm. In addition, the SIFT algorithm is employed to describe and build the feature vectors. The SIFT uses the gradient direction distribution properties of the pixels in the neighborhood of the feature corner points to specify direction parameters for each key point, which gives the operator rotational invariance. The SIFT algorithm is widely used in image processing and its related fields such as image stitching and alignment [41].

After the precise feature corner points are obtained, the gradient information of the pixel points within a certain neighborhood of the key point is counted. The gradient information includes size and direction. Pixel point gradient size and direction are calculated according to Eqs. (3)-(4).

$$m(x,y) = \sqrt{[L(x+1,y) - L(x-1,y)]^2 + [L(x,y+1) - L(x,y-1)]^2} \quad (3)$$

$$\theta(x,y) = \tan^{-1}\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)} \quad (4)$$

In the above equation, *m(x, y)* is the magnitude of the gradient at *(x, y)*, *θ(x, y)* is the direction of the gradient at *(x, y)*, and *L(x, y)* is the size or scale of all key points.

After the feature corner obtained the scale and main direction information, descriptors with particularity and robustness were established for each key point. By this processing, the matching accuracy is improved. A feature vector is determined from the gradient information of each pixel in the neighborhood of the key point.

(3) Precise feature matching of datas

In order to reconstruct the complete 3D structure and rich data information, then more spatial points with accurate locations are needed. Therefore, certain strategies are needed to obtain a larger order of magnitude of image matching points. Thus, the final 3D point cloud obtained is more dense. The region growth matching technique is used in this study to achieve dense matching.

Region growth takes a known seed point as a starting point. Based on the pre-defined growth criteria, pixels similar to the seed point are continuously merged to finally generate a region. The pixel points within the region and their corresponding points can all be considered as matching points. The region growth is grown using a known seed point as a reference. Each seed point will check whether each pixel point can be classified into that matching region within a certain area. When a certain number of seed points are identified, it is possible to determine whether a pixel can be used as a new matching point according to certain similarity discrimination criteria. Finally the region growth ends and a large number of matched point pairs are obtained.

2.1.2. Reconstruction of 3D point cloud

After obtaining a large number of matched points on the 2D images of pavement potholes, the correctly matched feature points are recovered into spatial 3D points by triangulation method, and then a 3D point cloud is obtained. As shown in Fig. 3, I1 and I2 are a set of 2D images taken by two cameras from different locations, P_1 and P_2 are the matched feature points on the two images, and the camera centers are set to O_1 and O_2, respectively. The plane of observation called O_1O_2P is the epipolar plane. The lines of intersection between the epipolar plane and the image plane, called p_1e_1 and p_2e_2, are the epipolar lines。

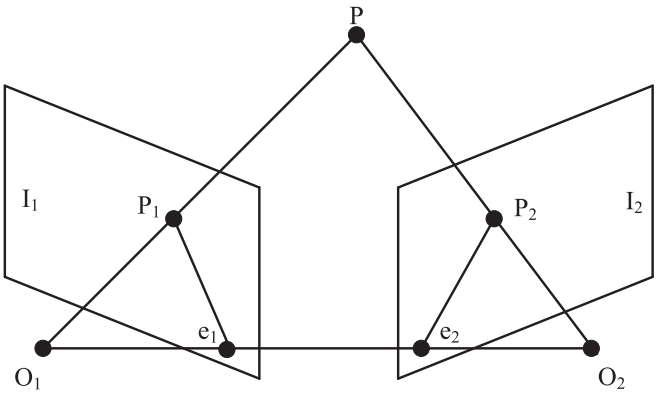


Fig. 3. Double view point geometry.

Extend the line between the camera center of each image and its feature point in space at point P, then point P is a three-dimensional point in the world coordinate system restored by the point on the image. After the 3D points are obtained, the 3D point cloud of the object is estimated by BA adjustment and filtering outliers.

2.2. Trans-3D-Seg based 3D point cloud segmentation model

The traditional deep learning-based pavement damage image detection system can collect two-dimensional information of the damage and be used to assess the severity of the damage. However, the above method cannot extract three-dimensional information of pavement damage. Therefore, the measurement of pothole damage depth and volume becomes difficult. In this study, a Trans-3D-Seg-based 3D point cloud segmentation model is developed to accurate extraction of pavement pothole 3D information. Damage segmentation uses features learned from similar point cloud data to find target regions from individual point cloud data.

As shown in Fig. 4, the Trans-3D-Seg network consists of a modified encoder and a modified decoder. The output size of the feature map for each layer is shown under the layer name, and the details of the network are shown in Table 1. The Encoder part is used to analyze the whole point cloud and perform feature extraction, while the corresponding Decoder part is used to generate the segmented point cloud data.

First, a 132 × 132 × 116 voxel block with 3 channels is input to the modified encoder. Secondly, two regular convolution operations are performed by using 3 × 3 × 3 three-dimensional convolution, which is used to extract local features of the pothole point cloud. For better convergence of the network, Batch Normalization is employed and output through the activation function ReLU. In addition, transformer is used for extracting the global features of the point cloud data. Finally, max pooling with 2 × 2 × 2 and step stride = 2 is operated. The number of the network parameters is reduced and the training efficiency of the network is improved. The above operation is repeated four times, and the output of the transformer-based encoder is obtained with a size of 9 × 9 × 7.

The output of the above encoder is fed into a modified decoder network. First, the upconvolution of 2 × 2 × 2 with step stride = 2 is used to reduce the feature map obtained in encoder to pixel space. That is, the size of the feature map output by encoder is expanded twice. Then, the feature map obtained by upconvolution is fused with the symmetric feature map in encoder to improve the accuracy of feature extraction. Next, the 3 × 3 × 3 3D convolution operation is performed to extract the local features of the fused feature map. This operation still includes the Batch Normalization operation. The output needs to be obtained through the activation function ReLU. In addition, the transformer operation is used to extract the global features of the fused feature map. In this study, the above operation was repeated three times to obtain the final output of the segmentation network. The size of the

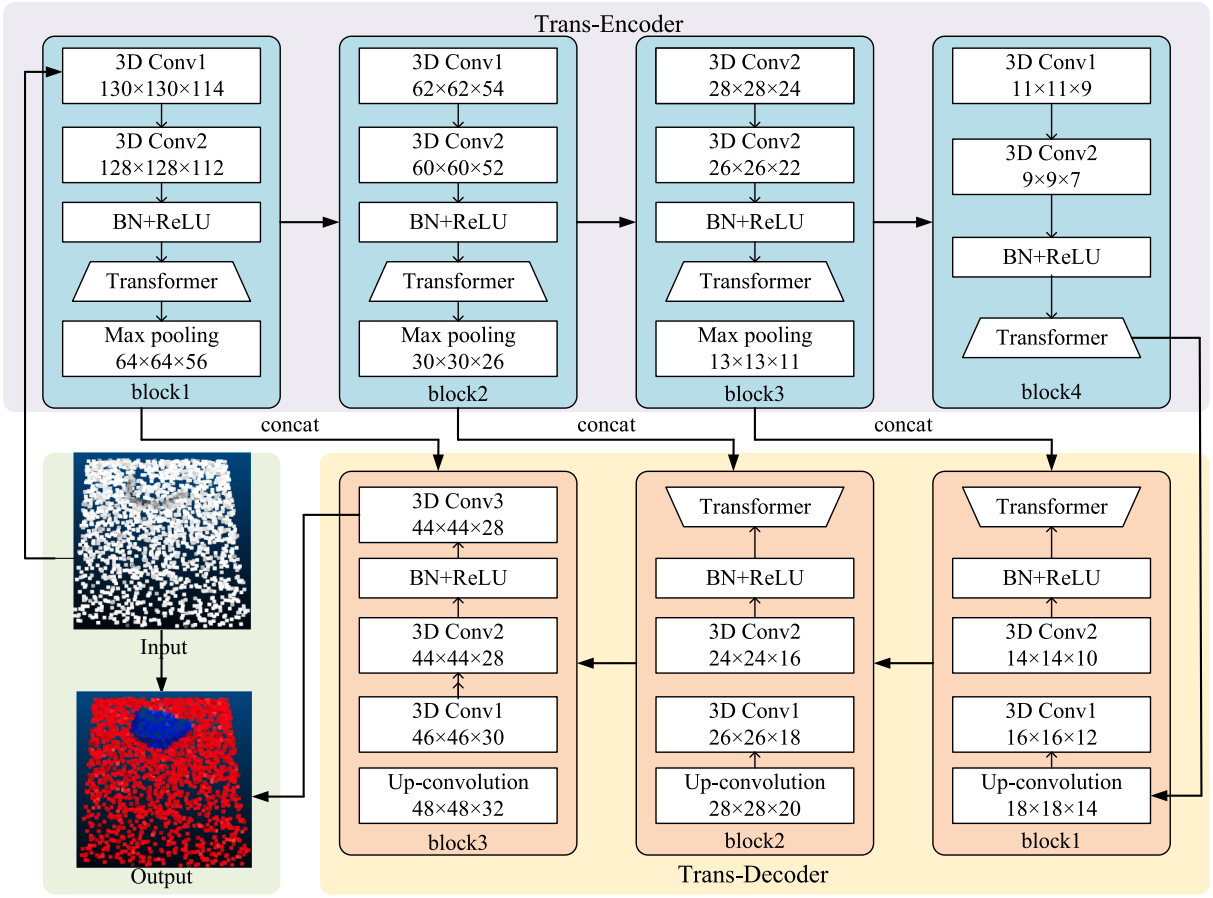


Fig. 4. Trans-3D-Seg network structure.

output is 44 × 44 × 28 voxels. This voxel is the segmentation result of the 3D point cloud of pavement potholes.

3. 3D reconstruction of pavement potholes

3.1. Multi-view 2D image acquisition of potholes

Machine vision-based evaluation systems have received attention for automated quality inspection of roads. Image recognition systems basically collect two-dimensional information to provide spatial attributes and measurements of the damage amount used to assess the severity of the damage. However, three-dimensional information about the damage is lacking [42]. In order to improve the quality of pavement damage evaluation, point cloud data containing 3D information is needed. The 3D scanner can be used to generate 3D profiles of the pavement potholes. However, laser scanners are very expensive. With the availability of inexpensive high-resolution camera devices (e.g., HD cameras, smartphones), 3D reconstruction using a single camera has become a reality.

Therefore, smartphones are used to acquire multi-view 2D images of pavement potholes on Changchun Road and Science Avenue in Zhengzhou High-tech Zone. The multi-view angle includes the acquisition of 2D images from three angles: front, rear, and left of the pavement pothole area. The multi-view 2D image set of potholes is composed of strict selection of the original image with high definition. The pavement pothole image set has a total of 2858 groups.

3.2. 3D reconstruction of pavement potholes based on PP-SFM

Similar to the traditional deep learning-based target detection methods, the Trans-3D-Seg-based 3D point cloud data segmentation model also requires the support of big data. However, the multi-view 2D image set of pavement potholes acquired in Section 3.1 cannot be directly used in the proposed Trans-3D-Seg. Therefore, a 3D reconstruction method based on PP-SFM is established in this study to achieve 3D reconstruction of potholes. For the development and application of the proposed algorithm, Ubuntu 18.04 server is used. A 64 GB RAM I7 8700 K CPU is employed in this server. The proposed PP-SFM algorithm is implemented through OpenCV3.0 module and python3.6. In addition, an 11 GB GeForce RTX 2080Ti GPU is used to save the algorithm computation time.

In the process of pavement pothole 3D reconstruction, the 2858 sets of pavement pothole image sets in Section 3.1 are fed into the PP-SFM network. The final 3D reconstruction of pavement potholes is achieved. The average reconstruction efficiency of CPU and GPU is calculated respectively. The average reconstruction efficiency is 23 fps under CPU and 40 fps under GPU acceleration. It can be seen that the 11 GB GeForce RTX 2080Ti GPU can effectively improve the 3D reconstruction efficiency of pavement potholes. An example of pavement pothole reconstruction results is shown in Fig. 7. As can be seen from Fig. 5, the proposed PP-SFM algorithm can effectively realize the three-dimensional reconstruction of potholes.

3.3. 3D point cloud dataset of pavement potholes

A 3D point cloud dataset containing 2858 pavement potholes is constructed by the PP-SFM algorithm. In this study, the open source CloudCompare software is used to implement the data annotation. The pothole area is noted as 1 and the background area is noted as 0. In this study, the annotated point cloud dataset was randomly divided into a

Table 1
Detailed structure parameters of Trans-3D-Seg.

Structure	Module	Layer name	Convolution kernel size	Stride	Output Size
Input					132 × 132 × 116
Trans-Encoder	Block1	3D Conv 1	3 × 3 × 3	1	130 × 130 × 114
		3D Conv 2	3 × 3 × 3	1	128 × 128 × 112
		Transformer			128 × 128 × 112
		Max pooling	2 × 2 × 2	2	64 × 64 × 56
	Block2	3D Conv 1	3 × 3 × 3	1	62 × 62 × 54
		3D Conv 2	3 × 3 × 3	1	60 × 60 × 52
		Transformer			60 × 60 × 52
		Max pooling	2 × 2 × 2	2	30 × 30 × 26
	Block3	3D Conv 1	3 × 3 × 3	1	28 × 28 × 24
		3D Conv 2	3 × 3 × 3	1	26 × 26 × 22
		Transformer			26 × 26 × 22
		Max pooling	2 × 2 × 2	2	13 × 13 × 11
	Block4	3D Conv 1	3 × 3 × 3	1	11 × 11 × 9
		3D Conv 2	3 × 3 × 3	1	9 × 9 × 7
		Transformer			9 × 9 × 7
Trans-Decoder	Block1	upconvolution	2 × 2 × 2	2	18 × 18 × 14
		3D Conv 1	3 × 3 × 3	1	16 × 16 × 12
		3D Conv 2	3 × 3 × 3	1	14 × 14 × 10
		Transformer			14 × 14 × 10
	Block2	upconvolution	2 × 2 × 2	2	28 × 28 × 20
		3D Conv 1	3 × 3 × 3	1	26 × 26 × 18
		3D Conv 2	3 × 3 × 3	1	24 × 24 × 16
		Transformer			24 × 24 × 16
	Block3	upconvolution	2 × 2 × 2	2	48 × 48 × 32
		3D Conv 1	3 × 3 × 3	1	46 × 46 × 30
		3D Conv 2	3 × 3 × 3	1	44 × 44 × 28
		3D Conv 3	1 × 1 × 1	1	44 × 44 × 28

training set, a validation set and a test set in the ratio of 3:1:1. The training set contains the data samples for model fitting. Validation sets are used to adjust the hyperparameters of the model. In addition, the segmentation ability of the model is preliminarily evaluated. The test set is used to evaluate the final generalization ability of the model.

4. Segmentation of pavement potholes based on Trans-3D-Seg

It is well known that the training of deep learning algorithms requires the support of high-performance devices. In this study, a computer with an Intel i7-8700x CPU, 11 GB GeForce RTX 2080Ti is used to train the model. The network is built in Ubuntu 18.04. The source code of Trans-3D-Seg is compiled from Python 3.6, Torch's Python library, and TensorFolw. CUDA 10.2 and CUDNN 7.6.5 acceleration libraries are used to improve the computational performance of the GPU.

4.1. Training and validation of model

The parameters of the proposed Trans-3D-Seg based segmentation algorithm include weights, biases and hyperparameters. The weights and biases are the core parameters of each layer of the Trans-3D-Seg network. In general, the optimal weights and biases can be obtained by continuous training of the model and BP back propagation algorithm. In addition, the hyperparameters are set before the model starts training. Therefore, manual optimization of hyperparameters is used. A set of optimal hyperparameters is selected to improve the training performance of the model. For the proposed Trans-3D-Seg algorithm, the hyperparameters that need to be manually optimized mainly include the learning rate, Mini-batch, Momentum and Weight decay. As shown in Table 2, eight different hyperparameter combinations are considered, and the optimal hyperparameter combination is case 4, and bold is used. The learning rate, Mini-batch, Momentum and Weight decay are 10^{-4}, 2, 0.98 and 5×10^{-4}, respectively.

For deep learning algorithms, epoch has a large impact on the model. With the increase of epoch, the loss of the model is gradually reduced and the accuracy is gradually improved. If the training process is stopped early, the model cannot learn effective features from the training set. Conversely, if the training process is stopped later, overfitting may occur. The features learned by the model are close to those of the training data, resulting in higher accuracy on the training set.

Table 2
Hyperparameter optimization results.

Case	Learning rate	Mini-batch	Momentum	Weight decay	Accuracy
1	10^{-5}	2	0.90	10^{-4}	92.37%
2	10^{-5}	1	0.90	5×10^{-4}	90.63%
3	10^{-5}	2	0.95	10^{-5}	91.34%
4	10^{-4}	1	0.98	5×10^{-4}	93.41%
5	$\mathbf{10^{-4}}$	**2**	**0.98**	$\mathbf{5 \times 10^{-4}}$	**94.55%**
6	10^{-4}	5	0.98	5×10^{-4}	90.28%
7	5×10^{-4}	2	0.85	5×10^{-4}	90.75%
8	5×10^{-4}	2	0.98	10^{-4}	92.16%

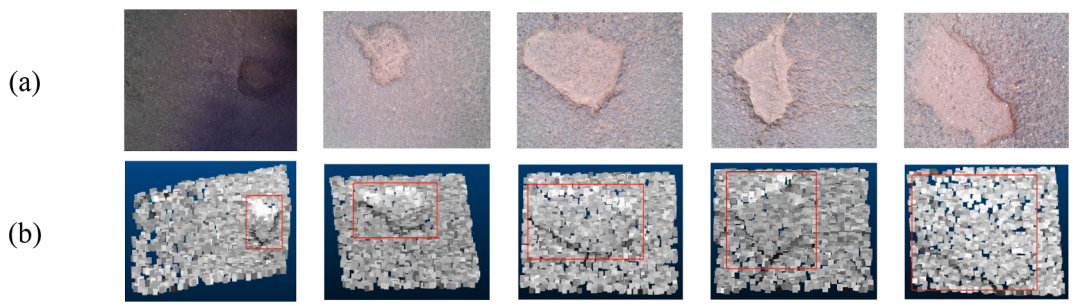


Fig. 5. 3D reconstruction results: (a) 2D images, (b) 3D reconstruction.

However, the performance is low on the new test set. As shown in Fig. 6, the training loss variation curve for 100 epochs is demonstrated. In addition, the model saved by each epoch is verified in the verification set, and the change curve of accuracy is shown in Fig. 7. As can be seen from Fig. 6, after training with 75 epochs, the loss curve of the model reaches convergence. Therefore, the model after 75 epoch training is saved for evaluation and detection.

4.2. Model testing

In this study, the test set is used to check the performance and generalization of the model. The pavement pothole 3D point cloud data of Trans-3D-Seg is tested on the test set. The test results are shown in Fig. 8, where the red area is the pavement background and the blue area is the pothole area. It can be seen that the pothole area can be effectively extracted by Trans-3D-Seg. In addition, the pothole segmentation accuracy is 93.44% and the F1-score is 92.58%.

4.3. Model robustness analysis

4.3.1. Segmentation effect of Trans-3D-Seg under different acquisition heights

To further test the robustness and generalization ability of the Trans-3D-Seg model, point cloud segmentation experiments are conducted at three shooting heights of 40 cm, 80 cm, and 120 cm. In this study, the Trans-3D-Seg model is tested through 165 pavement pothole 3D point cloud data. The accuracy variation curves verified on the above point cloud data are shown in Fig. 9. An example of the model testing results is shown in Fig. 10.

To further verify the segmentation performance in this case, the segmentation accuracy and F1-score are calculated as shown in Table 3. The average segmentation accuracy of 165 pavement pothole 3D point clouds reached 93.17% and the F1-score was 92.31%. It proves that the proposed model can still achieve effective segmentation for different shooting heights. From Fig. 10 and Table 3, it can be seen that the reconstructed point cloud data have higher resolution when the shooting height is 40 cm. The asphalt aggregate gap may be mistakenly detected as a pothole area. When the shooting height is 120 cm, the recall rate may be reduced. In this case, the pothole features in the point cloud data may not be extracted effectively. However, the accuracy rates of the above two cases can reach 91.25% and 90.36%, respectively. In addition, the F1-score can reach 90.75% and 89.87%, respectively. It shows that the proposed model can still meet the requirement of extracting the pothole area of 3D point cloud data. It is proved that the trained Trans-3D-Seg has good robustness for data under different acquisition heights.

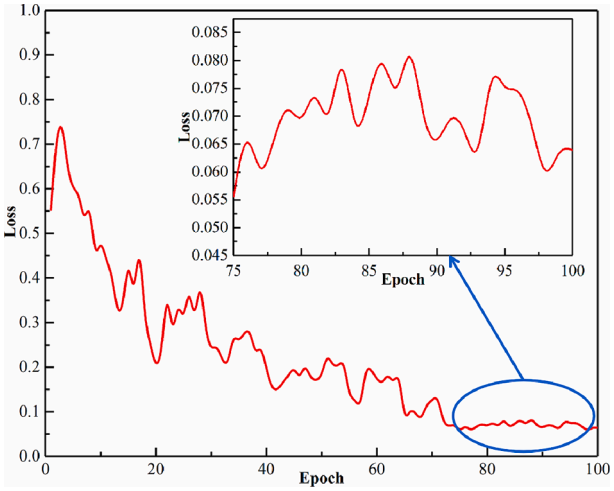


Fig. 6. Model training loss rate change curve.

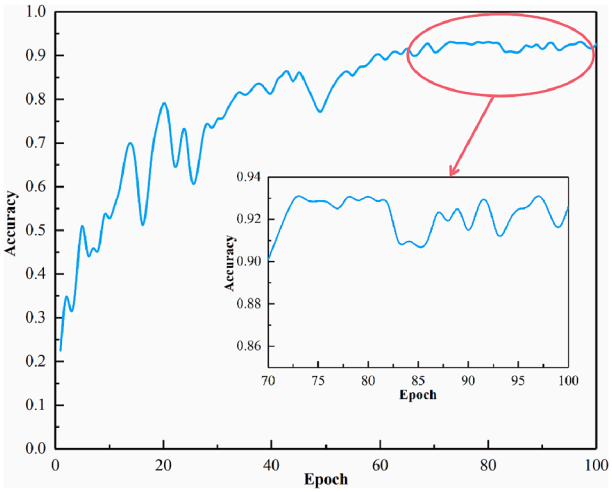


Fig. 7. Model training accuracy rate change curve.

4.3.2. Segmentation effect of Trans-3D-Seg under dark light

The proposed model has an effective segmentation effect under normal light. However, in the process of two-dimensional data acquisition, there is inevitable interference of dark light. Dark light represents dimly lit conditions. To verify the robustness of the proposed model under dark light, 157 tests of pavement pothole point clouds under dark light are performed. The test results of the proposed model are shown in Fig. 11, from left to right, indicating the model input, ground truth results and Trans-3D-Seg prediction results, respectively. From Fig. 11, it can be seen that the pothole area can still be extracted effectively under dark light conditions. In the proposed model, transformer technology has been adopted to optimize the model. Therefore, the proposed Trans-3D-Seg can simultaneously extract local and global features of point cloud data. In addition, the fusion of low-order features with high-order features is performed to enhance the correlation between features. Thus, the dark light point cloud recognition accuracy is improved. The measured accuracy of pothole segmentation under 57 dark light conditions is 91.86%, and the F1-score is 92.13%. It proves that the proposed method has good robustness for dark light data.

4.4. Trans-3D-Seg discussion

4.4.1. Effect of with and without transformer on the model

Transformer consists of a self-attentive mechanism for solving machine translation problems. Transformer can sense pixels at a distance to learn a more comprehensive feature representation. In addition, transformer technology has been applied to image processing tasks, such as target detection, semantic segmentation, and so on. In this study, transformer technology is innovatively integrated into the 3D encoder-decoder network, called Trans-3D-Seg, to improve the segmentation effect of the original model. To discuss the effectiveness of the transformer technique for model improvement, the proposed pavement pothole segmentation experiments with and without the transformer model are compared. The same dataset is used for training, validation and testing of both models mentioned above. On the validation set, the validation is performed. With the increase of epoch, the accuracy change curve is shown in Fig. 12. As can be seen from Fig. 12, the average accurate value of the proposed model can reach 94.55% when the iteration is 75 times. Compared without Transformer model, the segmentation accuracy is improved by 6.18%.

In order to further compare the two models, performance indicators are tested, as shown in Table 4 and Table 5. Table 4 and Table 5 show that the proposed model with Transformer has better accuracy and F1-

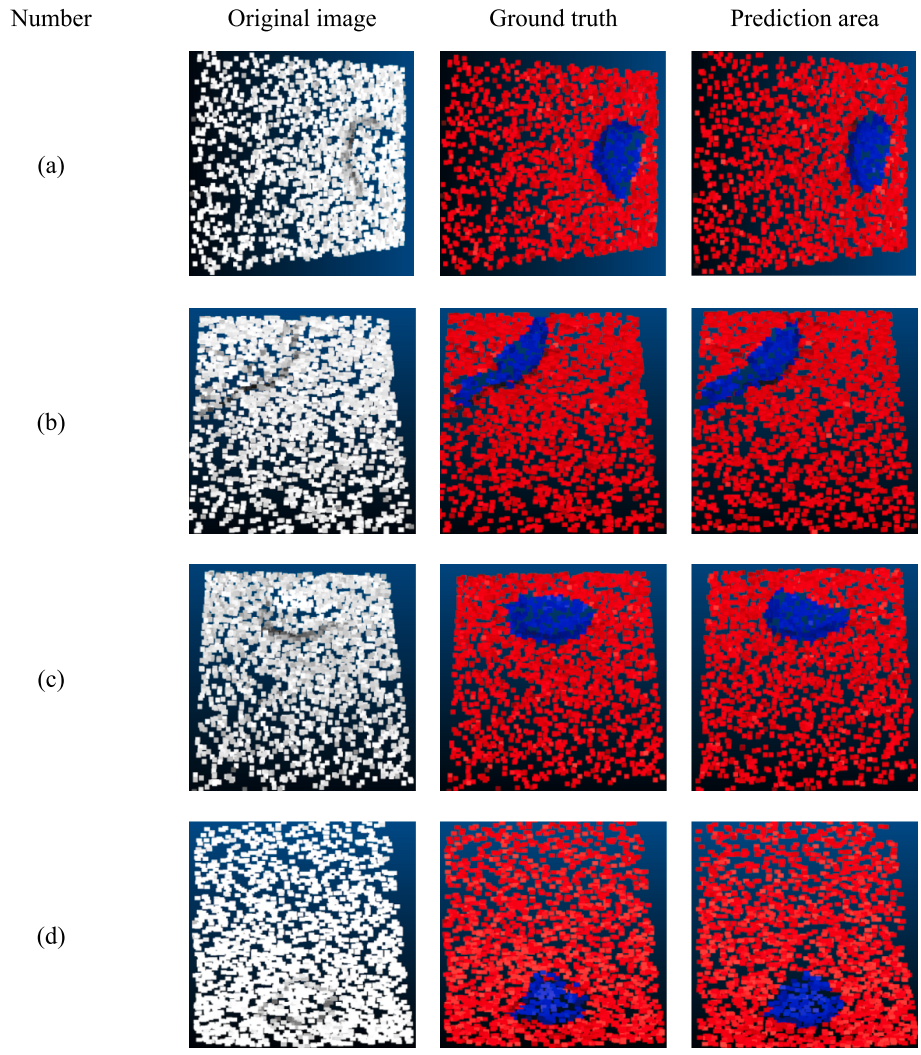
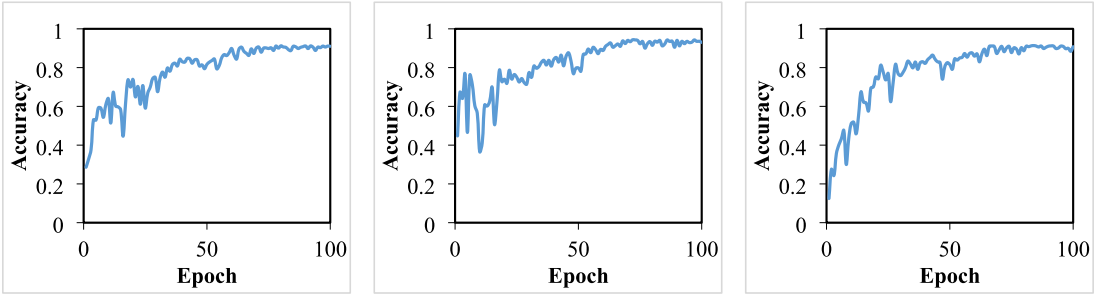


Fig. 8. Trans-3D-Seg test results.



The collection height is 40cm The collection height is 80cm The collection height is 120cm

Fig. 9. Variation curve of accuracy rate under different heights.

score, which can reach 93.44% and 92.58% respectively. Compared with the segmentation model without transformer, they are 5.46% and 4.05% higher respectively. Therefore, transformer is used to optimize the 3D encoder-decoder model, which can effectively improve the effectiveness of 3D point cloud data segmentation of potholes.

4.4.2. Comparative experiments of segmentation networks

The proposed network are compared with the state-of-the-art 3D point cloud segmentation networks to verify the effectiveness of the Trans-3D-Seg model performance. PointNet++ [43], PointRCNN [44], PointCNN [45] and the proposed Trans-3D-Seg are trained separately on the same dataset and experimental environment. In PointNet++, the set abstraction structure was used to extract the features layer by layer and the sampled points fused with the feature information of the neighborhood points [43]. In addition, only half of the point cloud is retained in each sampling, and the receptive field of the subsequent network layer gradually expands. Finally, the feature representation of the global point cloud in the target region is obtained. PointRCNN is the network of two-stage structure [44]. In the first stage, feature extraction is carried out through PointNet++. Based on the extracted features, the segmentation

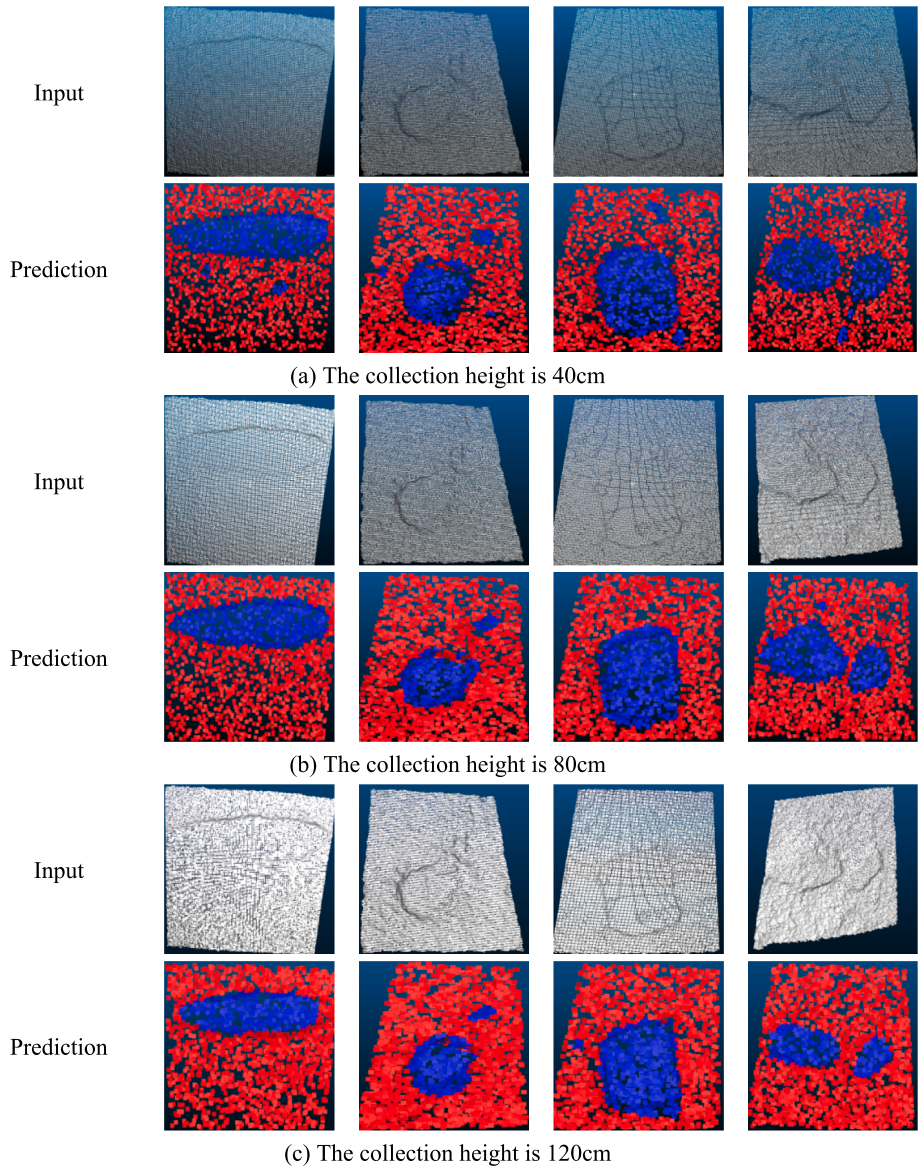


Fig. 10. Example of segmentation results for different acquisition heights.

Table 3
Segmentation effect under different segmentation height.

Heights	Number	Accuracy	F1-score
All	165	93.17%	92.31%
40 cm	55	91.25%	90.75%
80 cm	55	94.38%	93.52%
120 cm	55	91.36%	89.87%

of foreground and background is performed. In the second stage, refine is further performed based on the proposal extracted in the previous step. The spatial local correlation in the image is utilized through PointCNN, which solves the problem of loss of local features [45]. The change of accuracy on verification set is shown in Fig. 13, and the proposed Trans-3D-Seg has better accuracy.

Graphical analysis is an intuitive method of evaluation. Precision and recall are expected to be as high as possible, which means that the closer the PR curve is to the upper right, the better the performance. The P-R curves of the above four models are shown in Fig. 14. Compared with the other three models, the P-R curve of the Trans-3D-Seg model is closer to the upper right corner. As shown in Table 6, the performance metrics are calculated on the test set. The accuracy of PointNet++, PointRCNN, PointCNN and Trans-3D-Seg models are 89.31%, 90.48%, 90.27% and 93.44%, respectively. In addition, the F1-score are 90.17%, 90.15%, 89.65% and 92.58%, respectively. It proves that the proposed model in this study is more effective for segmenting the 3D point cloud of pavement potholes. The proposed system has a better effect for the following reasons:

First, the proposed system is improved through transformer. Transformer can sense pixels at a distance to learn a more comprehensive feature representation. Therefore, the precision of segmentation can be improved.

Secondly, Feature maps obtained from upconvolution are fused with symmetric feature maps in encoder. In other words, low-order features are fused with high-order features to enhance the correlation between features and improve the accuracy of feature extraction.

5. Conclusions

The integrity of the pavement was important for the transportation of goods and the movement of people. By regularly detecting pavement potholes, pavement structure breakdowns can be effectively avoided.

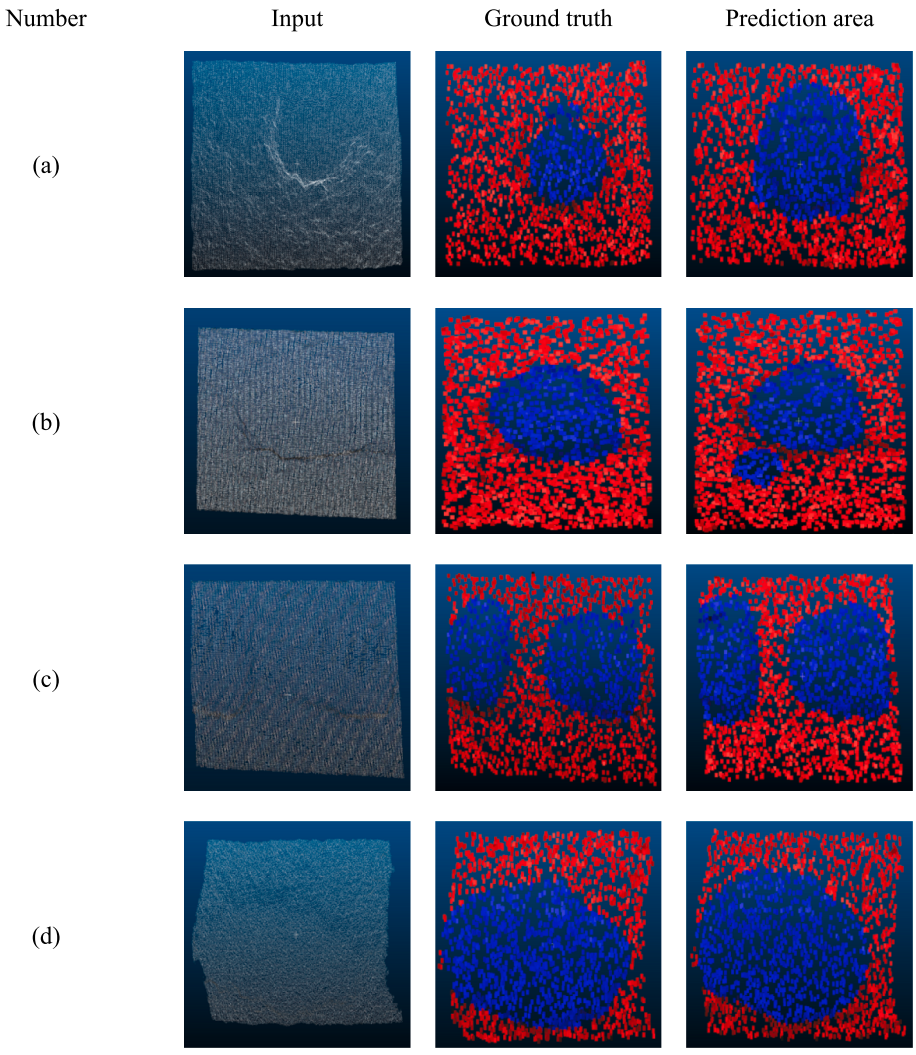


Fig. 11. Example of segmentation results under dark light conditions.

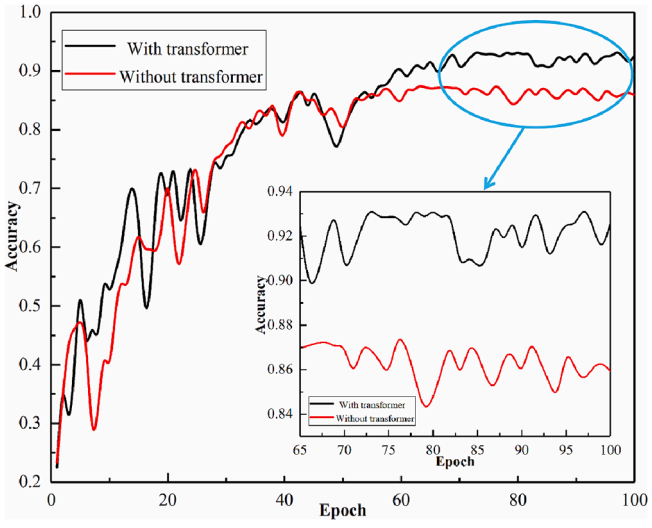


Fig. 12. Comparison of change curves of accuracy.

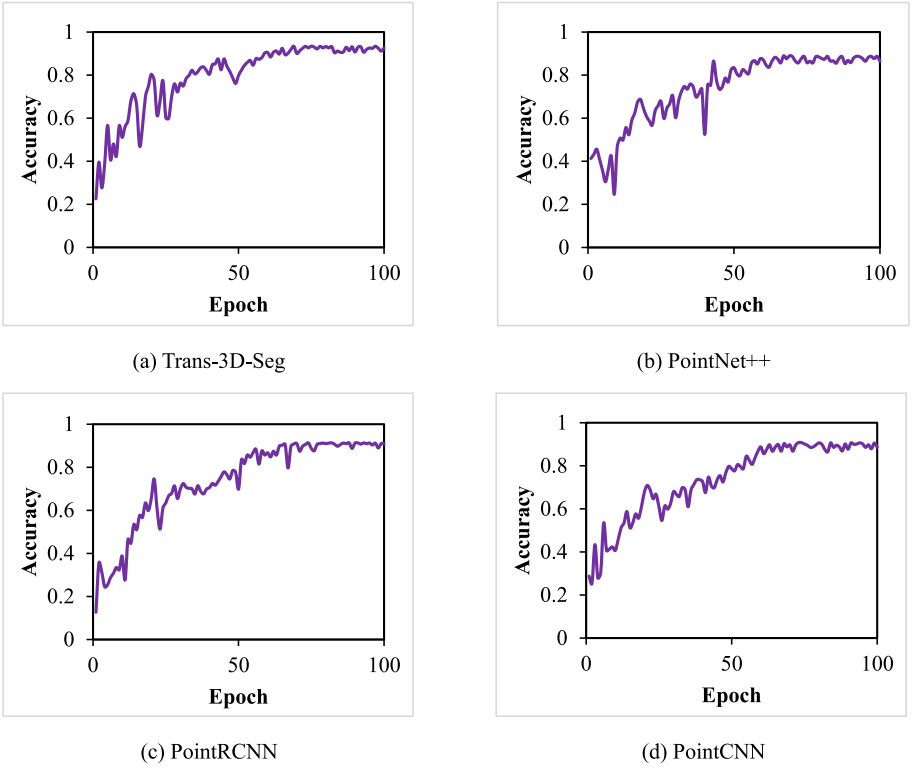
Image-based pavement damage detection methods have achieved excellent results. However, two-dimensional images cannot visualize three-dimensional information such as depth and volume. In addition,

Table 4
Accuracy comparison of models with and without transformer.

Network	Data type					
	All	Light condition		Height of acquisition		
		Normal	Dark light	40 cm	80 cm	120 cm
Trans-3D-Seg	93.44%	95.77%	91.86%	91.25%	94.38%	91.36%
Without transformer	87.98%	90.58%	86.15%	87.43%	91.66%	88.04%

Table 5
F1-score comparison of models with and without transformer.

Network	Data type					
	All	Light condition		Height of acquisition		
		Normal	Dark light	40 cm	80 cm	120 cm
Trans-3D-Seg	92.58%	93.67%	92.13%	90.75%	93.52%	89.87%
Without transformer	88.53%	90.25%	87.54%	88.13%	91.24%	87.58%



(a) Trans-3D-Seg (b) PointNet++

(c) PointRCNN (d) PointCNN

Fig. 13. Accuracy variation curves of different models in the validation set.

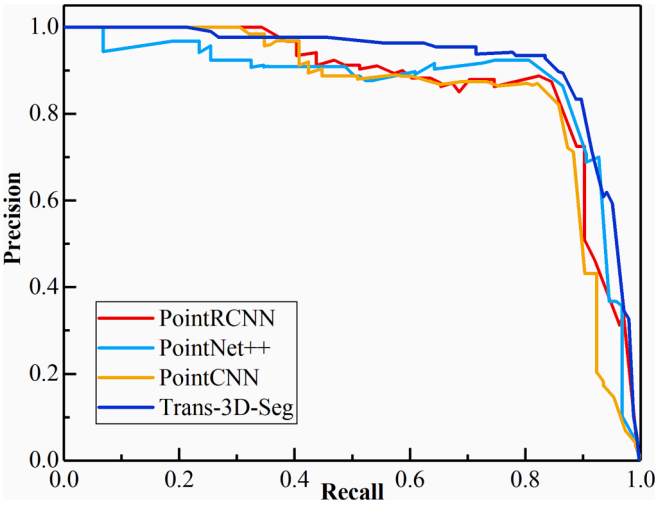


Fig. 14. P-R curve comparison.

Table 6
Model comparison results.

Model	Accuracy	Precision	Recall	F1-score
PointNet++ [43]	89.31%	90.47%	89.87%	90.17%
PointRCNN [44]	90.48%	92.73%	87.71%	90.15%
PointCNN [45]	90.27%	88.44%	90.86%	89.65%
Trans-3D-Seg	93.44%	91.68%	90.51%	92.58%

the existing 3D reconstruction methods based on laser point cloud cameras were costly. Therefore, a 3D reconstruction and segmentation system of pavement potholes based on PP-SFM and Trans-3D-Seg was proposed. Firstly, a point cloud reconstruction method based on PP-SFM was constructed to realize three-dimensional intensive reconstruction of multi-view images. This method solves the difficult problem that 3D laser point clouds were expensive and not easy to operate. Then, a Trans-3D-Seg-based 3D reconstruction method was proposed for efficient and accurate segmentation of 3D point cloud data of pavement potholes.

First, pavement pothole 3D point cloud data were generated by the proposed PP-SFM method. The Trans-3D-Seg was trained on the generated 3D point cloud dataset in a server terminal. Next, eight different working conditions were studied and the best hyperparameters were filtered. In addition, the segmentation accuracy was 94.55% on the validation set. Then, the minimum accuracy and F1-score of point cloud segmentation were 90.36% and 89.87% under three shooting heights of 40 cm, 80 cm and 120 cm, respectively. Under the dark light, the segmentation accuracy was 91.86% and F1-score was 92.13%. The robustness and generalization ability of Trans-3D-Seg model were proved. In addition, the effectiveness of the transformer technique for model improvement was discussed. Compared to the segmentation model without transformer, the accuracy and F1-score were improved by 5.46% and 4.05%, respectively. Compared with other deep learning methods (PointNet++, PointRCNN and PointCNN), the Trans-3D-Seg method has the highest accuracy and F1-score of 93.44%, 92.58%, respectively. In addition, the P-R curve of the proposed model was closest to the upper right. This proves that the performance of Trans-3D-Seg was the best in the field of 3D point cloud segmentation of pavement potholes.

The PP-SFM proposed in this study achieves the 3D reconstruction of pavement potholes. However, the method was semi-automatic. In the future, we will propose a 3D reconstruction method based on deep learning to improve the intelligence and efficiency of 3D reconstruction. In addition, pavement damage such as cracks and map crack also affects the safety of road operation. In the future work, the proposed system will be improved to accurately segment more types and complex damages.

Funding

This research was supported by the National Key Research and

Development Program of China (No. 2022YFC3801000), the National Natural Science Foundation of China (No. 51978630, 52108289), the Program for Innovative Research Team (in Science and Technology) in University of Henan Province (No. 23IRTSTHN004), the Program for Science & Technology Innovation Talents in Universities of Henan Province (No. 23HASTIT006), the Postdoctoral Science Foundation of China (No. 2022TQ0306), the Key Scientific Research Projects of Higher Education in Henan Province (No. 21A560013), the Open Fund of Changjiang Institute of Survey, Lanning, Design and Research (No. CX2020K10). The authors would like to thank for these financial supports.

CRediT authorship contribution statement

Niannian Wang: Conceptualization, Methodology. **Jiuxiu Dong:** Writing – original draft, Supervision. **Hongyuan Fang:** Writing – review & editing, Funding acquisition, Project administration. **Bin Li:** Writing – review & editing, Validation. **Kejie Zhai:** . **Duo Ma:** Writing – review & editing, Investigation. **Yibo Shen:** . **Haobang Hu:** .

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.

Data availability

The authors do not have permission to share data.

References

[1] M.R. Jahanshahi, S.F. Masri, A new methodology for non-contact accurate crack width measurement through photogrammetry for automated structural safety evaluation, Smart Mater. Struct. 22 (3) (2013), 035019, https://doi.org/10.1088/0964-1726/22/3/035019.
[2] A.K. Pandey, R. Iqbal, T. Maniak, C. Karyotis, S. Akuma, V. Palade, Convolution neural networks for pothole detection of critical road infrastructure, Comput. Electr. Eng. 99 (2022) 107725, https://doi.org/10.1016/j.compeleceng.2022.107725.
[3] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, I. Pitas, Pothole detection based on disparity transformation and road surface modeling, IEEE Trans. Image Process. 29 (2020) 897–908.
[4] S. Mathavan, K. Kamal, M. Rahman, A review of three-dimensional imaging technologies for pavement distress detection and measurements, IEEE Trans. Intell. Transp. Syst. 16 (5) (2015) 2353–2362, https://doi.org/10.1109/TITS.2015.2428655.
[5] H.-C. Dan, G.-W. Bai, Z.-H. Zhu, X. Liu, W. Cao, An improved computation method for asphalt pavement texture depth based on multiocular vision 3D reconstruction technology, Constr. Build. Mater. 321 (2022) 126427, https://doi.org/10.1016/j.conbuildmat.2022.126427.
[6] R. Fan, H. Wang, Y. Wang, et al., Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms, IEEE Trans. Image Process. 30 (2021) 8144–8154. https://doi.org/10.48550/arXiv.2109.02711.
[7] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, P. Fieguth, A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure, Adv. Eng. Inf. 29 (2) (2015) 196–210.
[8] Y.C. Tsai, A. Chatterjee, Pothole detection and classification using 3D technology and watershed method, J. Comput. Civ. Eng. 32 (2) (2018) 04017078, https://doi.org/10.1061/(ASCE)CP.1943-5487.0000726.
[9] J. Eriksson, L. Girod, B. Hull, et al., The pothole patrol: using a mobile sensor network for road surface monitoring, in: ProceedIngs of the 6th International Conference on Mobile Systems, Applications, and Services, 2008, pp. 29–39, https://doi.org/10.1145/1378600.1378605.
[10] K. De Zoysa, C. Keppitiyagama, G.P. Seneviratne, et al., A public transport system based sensor network for road surface condition monitoring, in: ProceedIngs of the 2007 Workshop on Networked Systems for DevelopIng regions, 2007, pp. 1–6, https://doi.org/10.1145/1326571.1326585.
[11] A. Dhiman, R. Klette, Pothole detection using computer vision and learning, IEEE Trans. Intell. Transp. Syst. 21 (8) (2019) 3536–3550, https://doi.org/10.1109/TITS.2019.2931297.
[12] R. Agrawal, Y. Chhadva,S. Addagarla et al., Road surface classification and subsequent pothole detection using deep learning, *In 2021 2nd International Conference for Emerging Technology (INCET). IEEE* (2021) 1-6, https://doi.org/10.1109/INCET51464.2021.9456126.
[13] H. Chen, M. Yao, Q. Gu, Pothole detection using location-aware convolutional neural networks, Int. J. Mach. Learn. Cyb. 11 (4) (2020) 899–911, https://doi.org/10.1007/s13042-020-01078-7.
[14] M.R. Jahanshahi, F. Jazizadeh, S.F. Masri, B. Becerik-Gerber, Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor, J. Comput. Civ. Eng. 27 (6) (2013) 743–754.
[15] C. Zhang, A. Elaksher, An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses, Comput. Aided Civ. Inf. 27 (2) (2012) 118–129, https://doi.org/10.1111/j.1467-8667.2011.00727.x.
[16] R. Fan, X. Ai, N. Dahnoun, Road surface 3D reconstruction based on dense subpixel disparity map estimation, IEEE Trans. Image Process. 27 (6) (2018) 3025–3035, https://doi.org/10.1109/TIP.2018.2808770.
[17] L. Inzerillo, G. Di Mino, R. Roberts, Image-based 3D reconstruction using traditional and UAV datasets for analysis of road pavement distress, Autom. Constr. 96 (2018) 457–469, https://doi.org/10.1016/j.autcon.2018.10.010.
[18] J. Chen, X. Huang, B. Zheng, R. Zhao, X. Liu, Q. Cao, S. Zhu, Real-time identification system of asphalt pavement texture based on the close-range photogrammetry, Constr. Build. Mater. 226 (2019) 910–919.
[19] B. Yang, W. Xu, Z. Dong, Automated extraction of building outlines from airborne laser scanning point clouds, IEEE Geosci. Remote S. 10 (6) (2013) 1399–1403, https://doi.org/10.1109/LGRS.2013.2258887.
[20] J.K. Anochie-Boateng, J.J. Komba, G.M. Mvelase, Three-dimensional laser scanning technique to quantify aggregate and ballast shape properties, Constr. Build. Mater. 43 (2013) 389–398, https://doi.org/10.1016/j.conbuildmat.2013.02.062.
[21] J. Liang, J. Zhang, K. Deng, et al., A new power-line extraction method based on airborne LiDAR point cloud data, *2011 International Symposium on Image and Data Fusion. IEEE* (2011) 1-4, https://doi.org/10.1109/ISIDF.2011.6024293.
[22] S. Dong, S. Han, Q. Zhang, X. Han, Z. Zhang, T. Yao, Three-dimensional evaluation method for asphalt pavement texture characteristics, Constr. Build. Mater. 287 (2021) 122966, https://doi.org/10.1016/j.conbuildmat.2021.122966.
[23] A. Yu, W. Mei, M. Han, Deep learning based method of longitudinal dislocation detection for metro shield tunnel segment, Tunn. Undergr. Sp. Tech. 113 (2021), 103949, https://doi.org/10.1016/j.tust.2021.103949.
[24] S. Dong, S. Han, Y. Yin, Z. Zhang, T. Yao, The method for accurate acquisition of pavement macro-texture and corresponding finite element model based on three-dimensional point cloud data, Constr. Build. Mater. 312 (2021) 125390, https://doi.org/10.1016/j.conbuildmat.2021.125390.
[25] S. Dong, S. Han, C. Wu, O. Xu, H. Kong, Asphalt pavement macrotexture reconstruction from monocular image based on deep convolutional neural network, Comput. Aided Civ. Inf. 37 (13) (2022) 1754–1768.
[26] J.H. Lee, C.S. Kim, Monocular depth estimation using relative depth maps, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 9729–9738, https://doi.org/10.1109/CVPR.2019.00996.
[27] B. Li, C. Shen, Y. Dai, et al., Depth and surface normal estimation from monocular images using regression on deep features and hierarchical crfs, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1119–1127, https://doi.org/10.1109/CVPR.2015.7298715.
[28] V. Guizilini, R. Ambrus, S. Pillai, et al., 3d packing for self-supervised monocular depth estimation, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 2485–2494, https://doi.org/10.1109/CVPR42600.2020.00256.
[29] C.J. Hawley, P.J. Gräbe, Water leakage mapping in concrete railway tunnels using LiDAR generated point clouds, Constr. Build. Mater. 361 (2022), 129644, https://doi.org/10.1016/j.conbuildmat.2022.129644.
[30] F. Bosche, C.T. Haas, Automated retrieval of 3D CAD model objects in construction range images, Autom. Constr. 17 (4) (2008) 499–512, https://doi.org/10.1016/j.autcon.2007.09.001.
[31] J. Lam, M. Greenspan, On the repeatability of 3d point cloud segmentation based on interest points, in: 2012 Ninth Conference on Computer and Robot Vision. IEEE, 2012, pp. 9–16, https://doi.org/10.1109/CRV.2012.9.
[32] F. Li, M. Lehtomäki, S. Oude Elberink, G. Vosselman, A. Kukko, E. Puttonen, Y. Chen, J. Hyyppä, Instance-aware semantic segmentation of road furniture in mobile laser scanning data, IEEE Trans. Intell. Transp. Syst. 154 (2019) 98–113.
[33] J. Valença, I. Puente, E. Júlio, H. González-Jorge, P. Arias-Sánchez, Assessment of cracks on concrete bridges using image processing supported by laser scanning survey, Constr. Build. Mater. 146 (2017) 668–678.
[34] N. Saovana, N. Yabuki, T. Fukuda, Automated point cloud classification using an image-based instance segmentation for structure from motion, Autom. Constr. 129 (2021), 103804, https://doi.org/10.1016/j.autcon.2021.103804.
[35] Y. Ben-Shabat, M. Lindenbaum, A. Fischer, 3dmfv: Three-dimensional point cloud classification in real-time using convolutional neural networks, IEEE Robot. Autom. Lett. 3 (4) (2018) 3145–3152, https://doi.org/10.1109/LRA.2018.2850061.
[36] Y. Kim, C.H.P. Nguyen, Y. Choi, Automatic pipe and elbow recognition from three-dimensional point cloud model of industrial plant piping system using convolutional neural network-based primitive classification, Autom. Constr. 116 (2020), 103236, https://doi.org/10.1016/j.autcon.2020.103236.
[37] D. Bobkov, S. Chen, R. Jian, M.Z. Iqbal, E. Steinbach, Noise-resistant deep learning for object classification in three-dimensional point clouds using a point pair descriptor, IEEE Robot. Autom. Lett. 3 (2) (2018) 865–872.
[38] H. Kim, J. Yoon, S.H. Sim, Automated bridge component recognition from point clouds using deep learning, Struct. Control Health Monit. 27 (9) (2020) e2591.
[39] C. Harris, M. Stephens, A combined corner and edge detector, Alvey Vision Conference 15 (50) (1988) 10–5244, https://doi.org/10.5244/C.2.23.

[40] D. Lowe, Distinctive image features from scale-invariant key points, Int. J. Comput. Vis. 20 (2003) 91–110, https://doi.org/10.1023/B:VISI.0000029664.99615.94.
[41] K. Liao, G. Liu, Y. Hui, An improvement to the SIFT descriptor for image representation and matching, Pattern Recogn. Lett. 34 (11) (2013) 1211–1220, https://doi.org/10.1016/j.patrec.2013.03.021.
[42] A. Ahmed, M. Ashfaque, M.U. Ulhaq, S. Mathavan, K. Kamal, M. Rahman, Pothole 3d reconstruction with a novel imaging system and structure from motion techniques, IEEE Trans. Intell. Transp. Syst. 23 (5) (2022) 4685–4694.
[43] C.R. Qi, L. Yi, H. Su, et al., Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst. 30 (2017). https://doi.org/10.48550/arXiv.1706.02413.
[44] S. Shi, X. Wang, H. Li, PointRCNN: 3D Object Proposal Generation and Detection From Point Cloud. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2019.
[45] Y. Li, R. Bu, M. Sun, et al., Pointcnn: Convolution on x-transformed points, Adv. Neural Inf. Process. Syst. 31 (2018). https://doi.org/10.48550/arXiv.1801.07791.