



# 3D detection of pavement cracking utilizing a neural radiation field (NeRF) and semantic segmentation network

Changjie Yao<sup>a,b</sup>, Mingtang Chai<sup>a,b,\*</sup>, Guoyu Li<sup>c</sup>, Dun Chen<sup>c</sup>, Qingsong Du<sup>c</sup>,  
Shunshun Qi<sup>c</sup>, Kai Gao<sup>c</sup>

<sup>a</sup> School of Civil and Hydraulic Engineering, Ningxia University, 750021, China

<sup>b</sup> Key Laboratory of the Internet of Water and Digital Water Governance of the Yellow River, Ningxia University, 750021, China

<sup>c</sup> Key Laboratory of Cryosphere Science and Frozen Soil Engineering, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China

## ARTICLE INFO

### Keywords:

Pavement crack detection  
Computer vision  
Neural radiation field  
Deep learning  
Semantic segmentation  
PointNet++

## ABSTRACT

The automatic detection of pavement cracks is important for maintaining road surfaces and vehicle safety. However, the current two-dimensional image-based automatic pavement detection method has environmental limitations. A novel framework for automatically detecting pavement cracks was proposed that combines computer vision techniques with deep learning methodologies. Using the neural radiance field multi-view stereoscopic imaging algorithm (Point-NeRF), we established a road surface point-cloud model comprising color images, depth images, and color-depth overlapping images. This approach overcame the limitations of the traditional structure from motion-multi-view stereo (SfM-MVS) algorithm through its fast modeling capabilities. To improve the spatial feature extraction structure for semantic segmentation tasks while reducing the network training burden, our proposed deep learning framework was based on PointNet++ with integrated coordinate attention and aggregate attention pooling mechanisms. The results indicate that the average root mean square error of the point pair distance reconstructed by the Point-NeRF algorithm and the point cloud reconstructed by the SfM-MVS algorithm was < 1.4 cm, demonstrating a high level of accuracy. Furthermore, the models had strong similarity, indicating an excellent reconstruction quality and a tenfold improvement in modeling efficiency compared to the SfM-MVS algorithm. Our proposed method achieved an outstanding performance in semantic segmentation tasks for pavement crack detection in terms of overall accuracy (91.8 %), mean intersection over union (80.1 %), precision (89.2 %), recall (87.3 %), and F1 score (88.2 %). Errors in length, width, and depth calculations were within acceptable limits with deviations below 0.3, 0.02, and 0.7 cm, respectively, while error rates were below 20 %.

## 1. Introduction

Roads are essential for social and economic development. However, increasing traffic volumes have made managing road surfaces increasingly difficult [1,2]. Among the potential types of damage to pavements, cracks are particularly prevalent [3]. Extensive irregular cracking can occur through both heat accumulation on the pavement and freeze–thaw cycles in the roadbed, severely compromising the safe operation of roads in cold regions and significantly increasing maintenance expenses. Currently, the detection of pavement cracks primarily relies on skilled technicians and the use of expensive specialized equipment within the industry. However, for highways exposed to harsh conditions

in cold areas human inspection can be difficult and expensive, leading to an inability to meet real-time demands. An automated and efficient method for detecting pavement cracks is therefore crucial for minimizing labor, while also reducing time constraints and costs [4,5]. The utilization of two-dimensional (2D) images for pavement crack recognition has been extensively employed [6]. However, the conventional approach of crack identification through binary image extraction and analysis lacks generalizability and fails to effectively segment cracks using a universal algorithm [7]. The quantification of cracks relies on morphological operations and geometric calculations, which are contingent upon the accuracy of crack images and processing [8]. Furthermore, determining the precise depth of cracks is not possible

\* Corresponding author at: School of Civil and Hydraulic Engineering, Ningxia University, 750021, China.

E-mail address: [cmt620422@163.com](mailto:cmt620422@163.com) (M. Chai).

<https://doi.org/10.1016/j.measurement.2025.117336>

Received 15 December 2024; Received in revised form 18 February 2025; Accepted 17 March 2025

Available online 19 March 2025

0263-2241/© 2025 Elsevier Ltd. All rights reserved, including those for text and data mining, AI training, and similar technologies.

with 2D images alone, necessitating the acquisition of three-dimensional (3D) crack information for road maintenance and repair purposes [9]. 3D pavement detection technology has shown promise [10], as it enables crack depth information to be obtained while circumventing the information loss inherent in 2D images [11]. Therefore, conducting a quantitative analysis of highway pavement cracks at a 3D spatial level holds immense significance for crack detection as well as for meeting/maintenance objectives [12,13].

According to the theory of 3D reconstruction, acquiring a greater number of viewing angle images leads to a more precise reconstructed scene [14]. Ma et al [15] modeled the highway by high precision point cloud data acquired by vehicle-mounted LiDAR and applied unsupervised clustering method to monitor the target area on this basis. Ge et al [16] used a phase-shift terrestrial laser scanner (TLS) (FARO Focus3D S120) for detecting the concrete cracks. Zhang et al [17] proposed an automated pavement defect detection method based on 3D laser profile data with a final pavement injury detection error of 8.7 %. Nevertheless, in laser scanning-based systems, hardware requirements and costs are high, such as vehicles, high-resolution sensors and wide-angle, high-exposure motion lenses. In addition, the bumps in the body of the vehicle during driving and the difficulty in controlling the driving speed, all of which cause image distortion and ultimately lead to a decrease in the quality of reconstruction, the above circumstances make the practicality of vehicle-based detection poor. With the development of computational graphics, unmanned aerial vehicles (UAV) have emerged as the preferred choice for most researchers in this field due to their high feasibility and low cost [18]. Xiao et al [19] used UAV aerial photogrammetry to detect bridge cracks by projecting the reconstructed 3D model to 2D and quantifying the cracks using images with background information. Jeong et al [20] used UAV photogrammetry to complete three-dimensional reconstruction of a traditional wooden structure building, assessed the feasibility of UAV reconstruction measurement technology, and the study showed that the accuracy difference between the three-dimensional reconstruction model based on UAV photogrammetry and the traditional measurement method was 0.17 %. Feng et al [21] addressed the problem of needing a high-resolution panoramic image to analyze the damage of the pavement in the process of inspection and proposed a hybrid model of pavement damage segmentation, splicing and detection based on UAV remote sensing pavement images, using UAV stereo vision to complete three-dimensional reconstruction and combining with CE-Net model. The above technology utilize RGB images captured by UAV and employs the SFM algorithm along with the MVS algorithm to explicitly express the three-dimensional scene [22]. SFM aims at recovering the 3D structure and camera positions from multiple images with varying perspectives. It involves extracting feature points, estimating camera poses using point pairs, calculating three-dimensional coordinates through triangulation combined with geometric constraints between images, and ultimately obtaining a sparse model. MVS meticulously matches pixels and evaluates the corresponding points across different photos on a pixel-by-pixel basis, enabling the generation of dense point clouds. The SFM-MVS principle combines sparse and dense reconstructions, leveraging geometric constraints and pixel matching relationships between images to achieve high-precision 3D model reconstruction [23]. With the rapid advancements in deep learning, NeRF and its enhanced techniques are increasingly being employed. As a deep learning model for 3D implicit space representation, NeRF utilizes multi-layer perceptrons (MLP), making it an important branch of research in 3D reconstruction due to its superior visual quality and view synthesis capabilities. It is actively contributing to the field of computer vision [24,25].

Achieving high-precision crack segmentation from the massive point-cloud data obtained from the reconstructed model remains a significant challenge. The acquired point-cloud data has an uneven density distribution with missing sections, thereby increasing the difficulty of automatically extracting crack geometry information. The complex pavement conditions also introduce uncertainty and imperfections in

crack segmentation, consequently reducing model accuracy. Currently, extraction algorithms for pavement point-clouds can be broadly categorized into two groups. ① Traditional empirical threshold methods such as the plane grid method [26] and facet grid method [27]. However, these methods are constrained by fixed parameters that require manual configuration, making them unsuitable for complex and diverse pavement conditions [28]. ② Semantic segmentation algorithm: In the context of deep learning, this approach is used for classifying and segmenting pavement scenes. Some researchers have converted irregular point-cloud data into image information and then utilized convolutional neural networks to perform pixel segmentation without relying on an empirical threshold. Furthermore, they project 3D point clouds into multi-view images to construct convolutional neural networks for multi-view images [29,30]. However, this algorithm fails to fully exploit the spatial and structural information within point clouds. Additionally, errors are inevitably introduced during the process of re-projecting the segmentation results back into 3D space due to the variations in dimensions and division methods [31]. Therefore, a semantic segmentation method called PointNet, which was based on the original point cloud, was proposed and subsequently developed [32]. Building upon this foundation, improved algorithms such as PointNet++ [33], PointCNN [34], and PointSIFT [35] were introduced to further extract data features from the point cloud.

In summary, this study used UAV-captured images to reconstruct the road surface in the study area using multi-view stereoscopic imaging technology based on the structure from motion – multi-view stereo (SfM-MVS) and neural radiance field (NeRF) algorithms. The evaluation compared different reconstruction methods and selected Point-NeRF as the optimal method for collecting pavement point-cloud data. Additionally, the spatial feature extraction structure in PointNet++ point cloud semantic segmentation network was enhanced and validated by the open source point cloud set Semantic3D Dataset, which lead to the construction of a semantically rich digital pavement model and the extraction of the basic information of pavement cracks. Subsequently, visualization software and programming methods were used to segment and extract pavement crack point-cloud data, enabling the construction of a spatial model that calculated geometric properties such as the length, width, and depth of cracks. This established a comprehensive framework from road point-cloud analysis to integrated pavement crack detection, while significantly improving automation in point-cloud processing and enhancing the expressiveness and application potential. Moreover, it visually characterized road scenes with an accurate representation of cracks by providing an underlying platform supported by data for the digital representation of pavement crack information.

## 2. Methodology

### 2.1. The 3D reconstruction method

The 3D modeling approach employed in this study integrated a multi-view stereo vision algorithm for motion recovery structure with a neural radiation field algorithm. The reconstruction process is shown in Fig. 1(a). The specific steps are described below.

The reconstruction of the multi-view stereo vision algorithm began with the restoration of camera parameters based on polar geometry constraints, followed by the establishment of point correspondences between two images. Subsequently, through eigenmatrix decomposition, we derived the relative motion parameters of the camera and optimized the matching relationship between newly added camera orientations and existing cameras along with 3D point clouds. Finally, a point-cloud model was constructed via rendering and deep fusion.

The NeRF algorithm was used to acquire the implicit representation of a 3D scene. A multi-layer perceptron neural network was employed within the neural radiation field to implicitly express the scene, as shown in Fig. 1(b). The input encompassed both the angle and position along the ray, while the output comprised color information and spatial

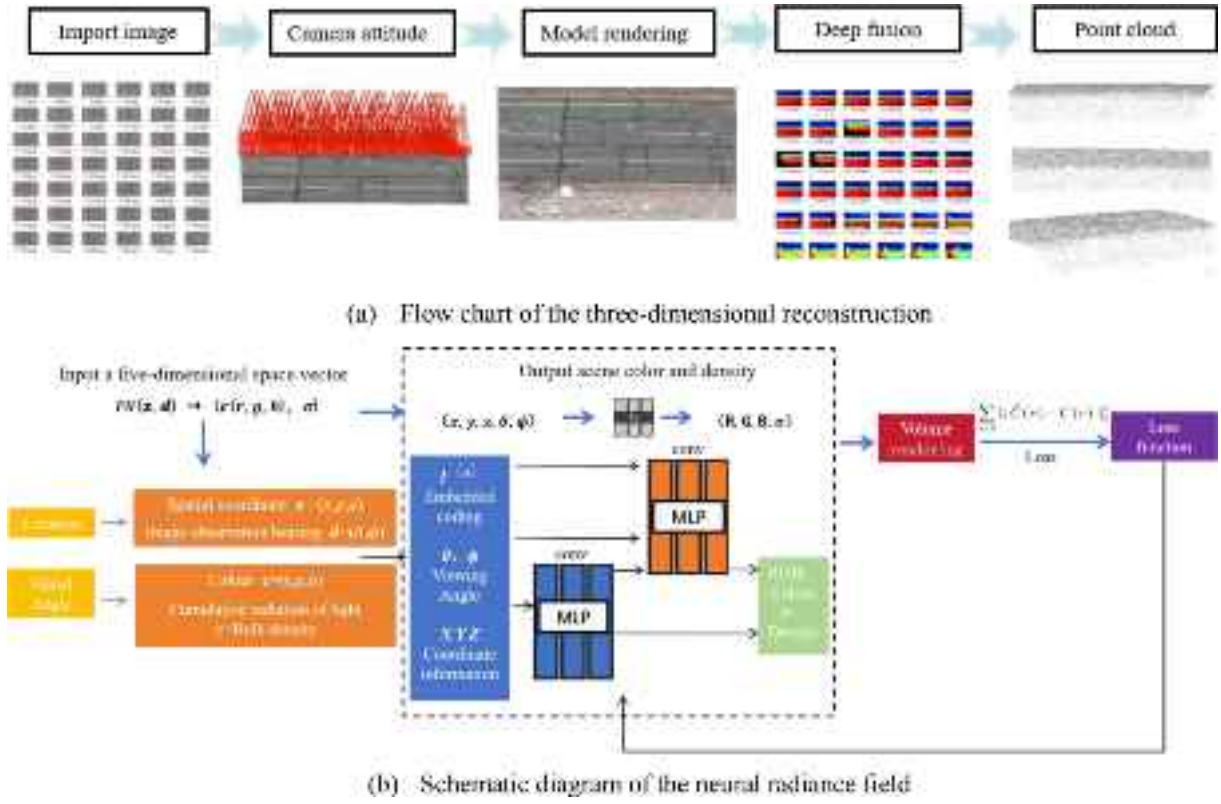


Fig. 1. 3D reconstruction flow chart and neural radiance field schematic diagram.

volume density corresponding to that particular position. Querying along emitted light facilitated a view synthesis, enabling the generation of novel perspective images during rendering. The static scene can be described as a continuous five-dimensional (5D) vector function using the following formula:

$$F(\theta) : (x, d) \rightarrow (C(r), \sigma) \quad (1)$$

where  $F(\theta)$  represents a neural network;  $x$  is the spatial coordinate of the point;  $d$  indicates the observation direction of the input image;  $C$  signifies the color information emitted from each point's position in the direction of  $d$ ; and  $\sigma$  corresponds to the volume density of each individual point.

For a given scene and observation angle,  $F(\theta)$  necessitates the use of a numerical integration method to approximate the real volume rendering process corresponding to the view. The discrepancy between the rendered and actual images was computed, followed by a comparison of loss functions, leading to the iterative optimization of rendering outcomes. Based on this foundation, an exceptionally realistic reality model was generated, as shown in Equation (2):

$$C(r) = \int_{t_n}^{t_f} T(t) \cdot \sigma(r(t)) \cdot c(r(t), d) dt \quad (2)$$

where  $t_n$  and  $t_f$  represent the proximal and distal boundaries of light, respectively;  $T(t)$  is the cumulative transmittance, signifies the probability of unobstructed passage for light through a particle;  $c(r(t))$  corresponds to the color at that specific point;  $\sigma(r(t))$  represents the density at that particular location.

## 2.2. Reconstruction quality assessment

The point-to-point distance of the reconstructed point-cloud model was calculated for each reconstruction method applied in this study, considering different image quantities. To quantify the similarity be-

tween the reconstructed models, we calculated the root-mean-square error ( $R_p$ ) of their point-to-point distances. Because the SFM-MVS and NeRF reconstruction models have different scales, it was necessary to align their coordinates before employing the iterative closest point (ICP) algorithm to register and calculate the distances between corresponding points in both models. For every point in the reconstructed model, we iteratively determined its nearest neighbor in the reference point cloud and calculated their Euclidean distance ( $D_p$ ). Using these pairwise distances, we evaluated  $R_p$  as a measure of similarity according to the following formulas:

$$D_p = \sqrt{(x_r - x_s)^2 + (y_r - y_s)^2 + (z_r - z_s)^2} \quad (3)$$

$$R_p = \sqrt{\frac{\sum_{k=1}^n D_k}{n}} \quad (4)$$

where  $(x_r, y_r, z_r)$  are the coordinates of a certain point in the reconstructed point cloud;  $(x_s, y_s, z_s)$  are the coordinates of the corresponding points in the reference point cloud;  $n$  is the number of point pairs.

## 2.3. The improved semantic segmentation network

### 2.3.1. Preprocessing of point-cloud data

Before constructing the semantic segmentation network, preprocessing of the point-cloud data is essential, which includes noise reduction and normalization. In this study, a radius filtering method was applied for point-cloud noise reduction to eliminate outliers. By adjusting the search radius and filtering threshold of nearest neighbor points, the identification of outliers was achieved, as shown in Fig. 2.

For the normalization of point clouds, this study used a combination of linear scaling and centralized processing to achieve data centralization and scale normalization. The formula was as follows:

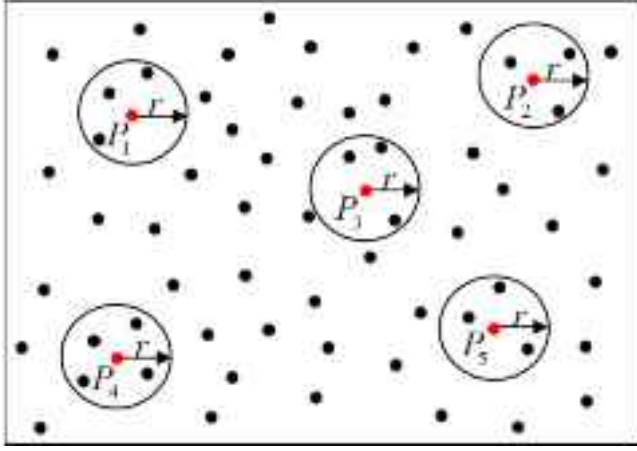


Fig. 2. Radius filtering diagram.

$$C = \frac{1}{N} \sum_{i=1}^N P_i \quad (5)$$

$$P' = P - C \quad (6)$$

where  $C$  is the centroid;  $N$  is the total number of data points;  $P$  is the original point cloud; and  $P'$  is the shifted point cloud.

The point cloud was then normalized to a standard scale through a linear transformation. This was achieved by determining the farthest point from the origin in the point cloud and using that distance as the scaling factor. The corresponding formula was as follows:

$$m = \max_{i=1}^N \sqrt{\sum_{j=1}^d (p'_{ij})^2} \quad (7)$$

$$P'' = \frac{P'}{m} \quad (8)$$

where  $m$  represents the maximum norm;  $d$  is the dimensionality of each point;  $p'_{ij}$  is to the  $J$ -dimensional coordinate of the  $i$ -th point post-translation; and  $P''$  signifies the scaled point cloud.

### 2.3.2. Enhanced spatial feature extraction structure

The module was used for extracting local and global features from point-cloud data, and primarily consisted of the local polar coordinate representation and attention pooling fusion. The attention pooling mechanism within the local feature extraction module was enhanced to improve the capability of extracting local features from the point cloud.

#### (1) Local polar representation of features

To mitigate the impact of model rotation on geometric feature extraction, a local feature invariant along the Z-axis was formulated

within the polar coordinate system. Each point  $p_i$  and its  $k$  neighboring points  $p_i^k$  were transformed into polar coordinates. The geometric distance between neighboring points was calculated, and the centroid  $p_i^m$  of the neighborhood was determined. Subsequently, the local direction was ascertained by subtracting the initial azimuth  $\varphi_i$  and  $\theta_i$  from the local azimuth  $\alpha_i$  and  $\beta_i$ , resulting in a relative azimuth  $\varphi_i^r$  and  $\theta_i^r$  as shown in Fig. 3. Finally, these relative azimuths and geometric distances were combined as shown in Fig. 4.

#### (2) Attention pooling

To enhance the accurate representation of spatial information in the point cloud and preserve its intricate geometric structure, we integrated the original coordinate attention mechanism with the aggregation attention pooling mechanism. This was achieved through parallel pooling operations along the same direction to aggregate input point-cloud features and strengthening spatial features based on both geometric distance and feature distance, as shown in Fig. 5. The specific process is described below.

First, the eigen distance  $d_{ij}^k$  was obtained by calculating the mean value of the eigenvector  $g(i)$  and the eigenvector  $g(k)$  of the  $k$ th nearest neighbor point surrounding it using the L1 norm. The formula can be expressed as follows:

$$d_{ij}^k = Ave(|g(i)| - |g(k)|) \quad (9)$$

where  $|\cdot|$  is the L1 norm;  $Ave(\cdot)$  is the mean value function; and  $g(i)$  and  $g(k)$  represent the eigenvectors of the  $i$  and  $k$  nearest neighbors, respectively.

The weight of attention needs to be determined based on the relationship between adjacent points, as reflected by the geometric distance  $d_{ig}^k$  and the feature distance  $d_{if}^k$ . To achieve this, we considered both negative geometric and feature distances, combining them through a weighted sum using the normalized exponential function softmax. This process yielded the distance feature  $d_i^k$ . Subsequently, we fused this distance feature with the local feature  $f_i^k$  to obtain the fused distance feature  $d_i^{k+}$ , which then underwent weight learning. The corresponding formula was as follows:

$$d_i^k = \exp(-d_{ig}^k) \oplus \exp(-d_{if}^k) \quad (10)$$

$$d_i^{k+} = d_i^k \oplus f_i^k \quad (11)$$

where  $d_{ig}^k$  and  $d_{if}^k$  are the geometric distance and characteristic distance, respectively;  $d_i^k$  and  $f_i^k$  are distance features and local features, respectively; and  $\oplus$  indicates the merge operation.

Finally, the aggregated local features were calculated, and the convolution and activation functions were employed to autonomously learn the attention weight for selecting salient features while discarding irrelevant ones. The acquired weight  $a_i^k$  and corresponding local feature  $f_i^k$  were linearly combined to calculate the local average feature  $f_{iLAve}$ , as

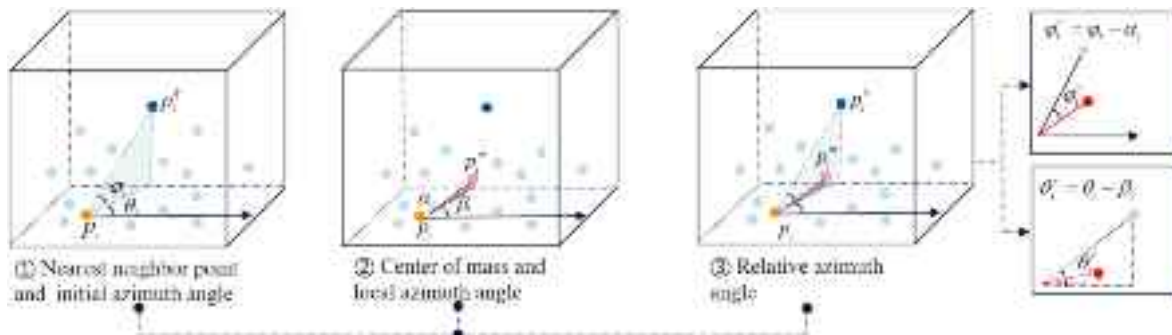


Fig. 3. The calculation of the relative azimuth angle.



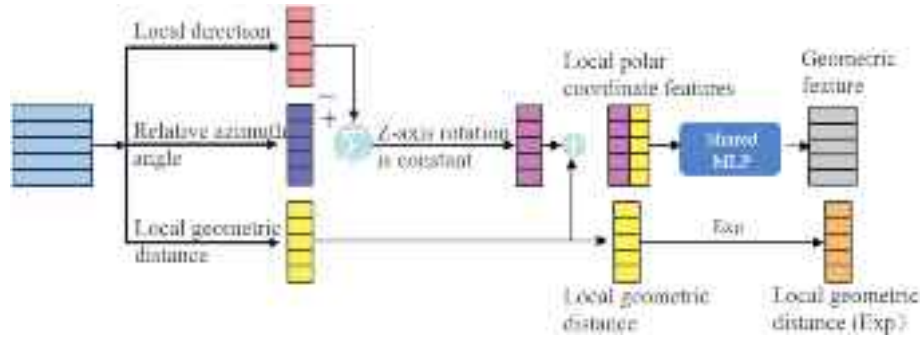


Fig. 4. Local polar representation of features.

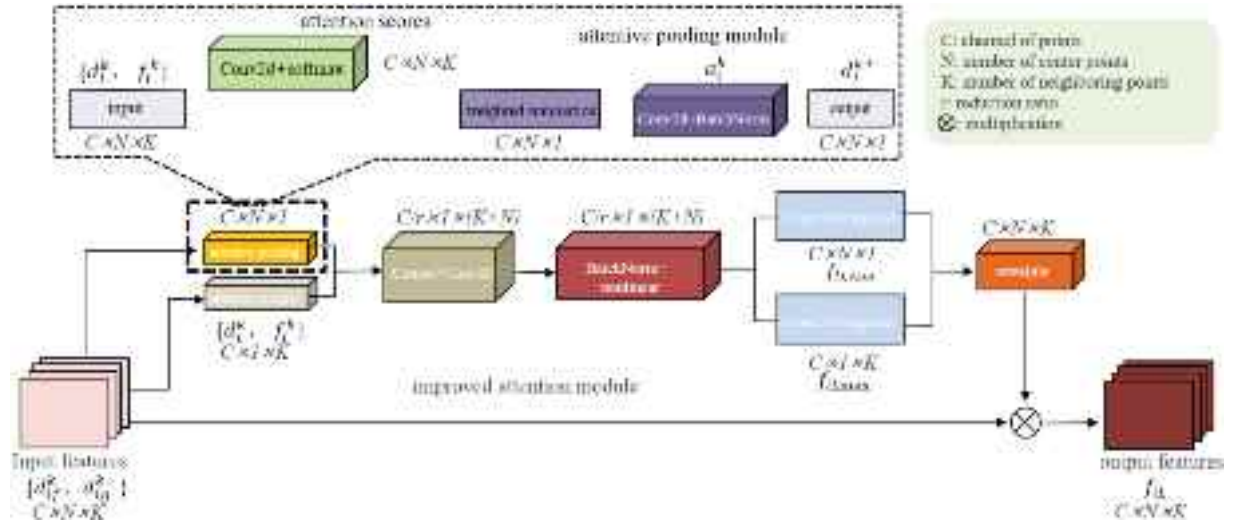


Fig. 5. The improved attention pooling fusion model.

shown in equation (12). Subsequently, the maximum value among  $k$  adjacent points was determined as the local maximum feature  $f_{iLmax}$ . Ultimately, these two features were fused to form the localized aggregation feature  $f_{iL}$ , as shown in equation (13):

$$f_{iLAve} = \sum_{k=1}^K \frac{(a_i^k f_i^k)}{K} \quad (12)$$

$$f_{iL} = f_{iLmax} \oplus f_{iLAve} \quad (13)$$

where  $f_{iL}$  is the feature of local polymerization;  $a_i^k$  is the weight of learning; and  $d_i^{k+}$  is the fused distance feature.

### 2.3.3. Semantic segmentation network structure

A hierarchical segmentation network structure was used in this study, in which the network model progressively expanded the local regions through classification and segmentation processes to extract higher-dimensional point-cloud features. The classification module

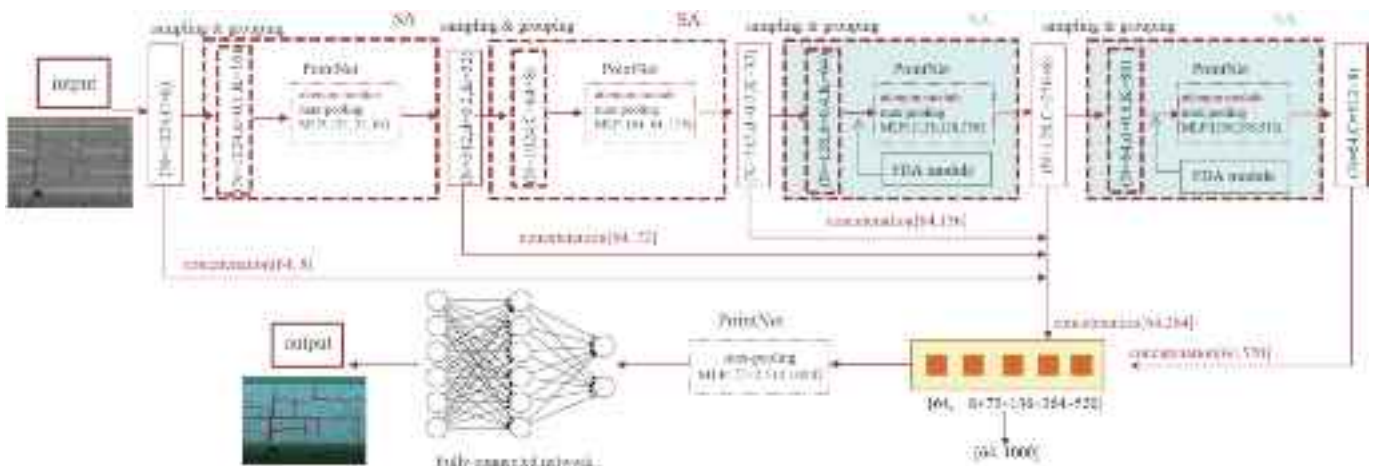


Fig. 6. Network classification structure.

consisted of four collection sampling layers (SA), as shown in Fig. 6. The input consisted of  $N \times C$  point clouds with an initial feature dimension  $C$  of 8. Farthest distance sampling was used for the sampling layer. The sampled points were then divided into local areas and the Fourier Domain Adaptation (FDA) module was incorporated to adjust the relative feature distribution among points. Subsequently, the PointNet layer was applied to obtain local features of point clouds, gradually expanding the feature dimension up to 1000 dimensions. Classification prediction scores were obtained via fully connected layers, enabling a determination of the input point-cloud's category based on the highest predicted score.

In the segmentation module, the encoder component remained consistent with that of the classification module; however, the decoder component applied interpolation and concatenation techniques, as shown in Fig. 7. Specifically, interpolation was conducted by calculating an inverse distance weighted average of the  $K$  nearest neighbor points to obtain up sampled point features. For each point in the up sampled layer, its corresponding features from the  $K$  nearest points in the preceding layer were identified and weighted based on their inverse distances. Additionally, skip connections were used to concatenate features related to points in both the encoder and decoder sections. Ultimately, this PointNet unit iteration continued until every point was assigned a predicted score for accurate categorization.

#### 2.3.4. Performance evaluation semantic segmentation network

In order to verify the effectiveness of the algorithms described in this study, the coordinate attention mechanism and the aggregated attention pooling mechanism were gradually introduced into the PointNet++ feature extraction module, and ablation experiments were conducted using Semantic3D and our dataset. The evaluation metrics used in this study were the overall accuracy (OA), mean intersection over union (MIoU), precision, recall, and F1 score. The corresponding formulas were as follows:

$$OA = \frac{TP + TN}{TP + TN + FN + FP} \quad (14)$$

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FN + FP} \quad (15)$$

$$precision = \frac{TP}{TP + FP} \quad (16)$$

$$recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1 = \frac{2 \times recall \times precision}{recall + precision} \quad (18)$$

where  $TP$  is the number of true positives, indicating the accurate classification of point clouds into this category;  $TN$  is the number of true negatives, signifying the correct classification of point clouds into other

categories;  $FN$  is the number of false negatives, representing the misclassification of point clouds into other categories; and  $FP$  is the number of false positives, denoting the misclassification of point clouds into this category.

#### 2.4. Crack extraction and calculation

Crack extraction refers to the secondary processing of crack point-cloud data in the road surface point-cloud model after training. This process includes pavement crack estimation, generation of a spatial crack map, and extraction of the crack surface contour and centerline [36,37]. The random sample consensus (RANSAC) algorithm was used to fit non-cracked pavement first, followed by fitting its contour and centerline based on the spatial characteristics of cracks. Finally, the geometric parameters of the cracks were quantified.

The crack length was determined by calculating the cumulative Euclidean distance between each pair of consecutive point clouds. The formula used for this calculation was as follows:

$$length = \sum_{i=2}^n \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \quad (19)$$

The crack width was determined using the width calculation method based on the crack centerline. A vertical line was drawn through the centerline at the observed section position of the crack, and the distance between this vertical line's intersection point and the crack edge then represented the specific observation point's crack width. The formula used for this calculation was as follows:

$$width_{(x_i, y_i)} = \sqrt{(x_{ir} - x_{il})^2 + (y_{ir} - y_{il})^2} \quad (20)$$

The RANSAC plane fitting algorithm was used to fit the surface plane, and the crack depth was determined by measuring the distance between the calculated points of the crack and the fitted pavement. The formula used for this calculation was as follows:

$$depth_{(x_i, y_i, z_i)} = \frac{|a x_i + b y_i + c z_i|}{\sqrt{a^2 + b^2 + c^2}} \quad (21)$$

Targeting the gaps in the synergistic application of Neural Radiance Fields (NeRF) and semantic segmentation technology within existing 3D detection research, this paper proposed a pavement crack detection method that integrated NeRF with 3D point cloud semantic segmentation. An innovative joint processing framework of “point cloud generation – feature enhancement – crack resolution” was constructed, which primarily consisted of three components, as illustrated in Fig. 8. Firstly, basing on NeRF's high-fidelity 3D reconstruction capabilities, the Point-NeRF algorithm was employed to achieve dense point cloud generation for the pavement scene, ensuring that the quality of the point cloud in the crack region meet the requirements for accurate crack morphology capture. Secondly, to improve upon the low training efficiency of traditional PointNet++ networks and insufficient

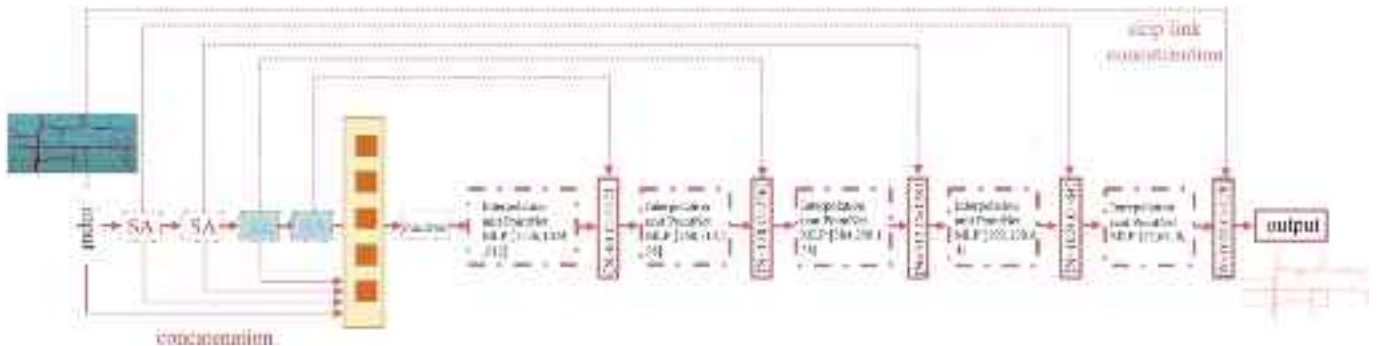


Fig. 7. Network segmentation structure.

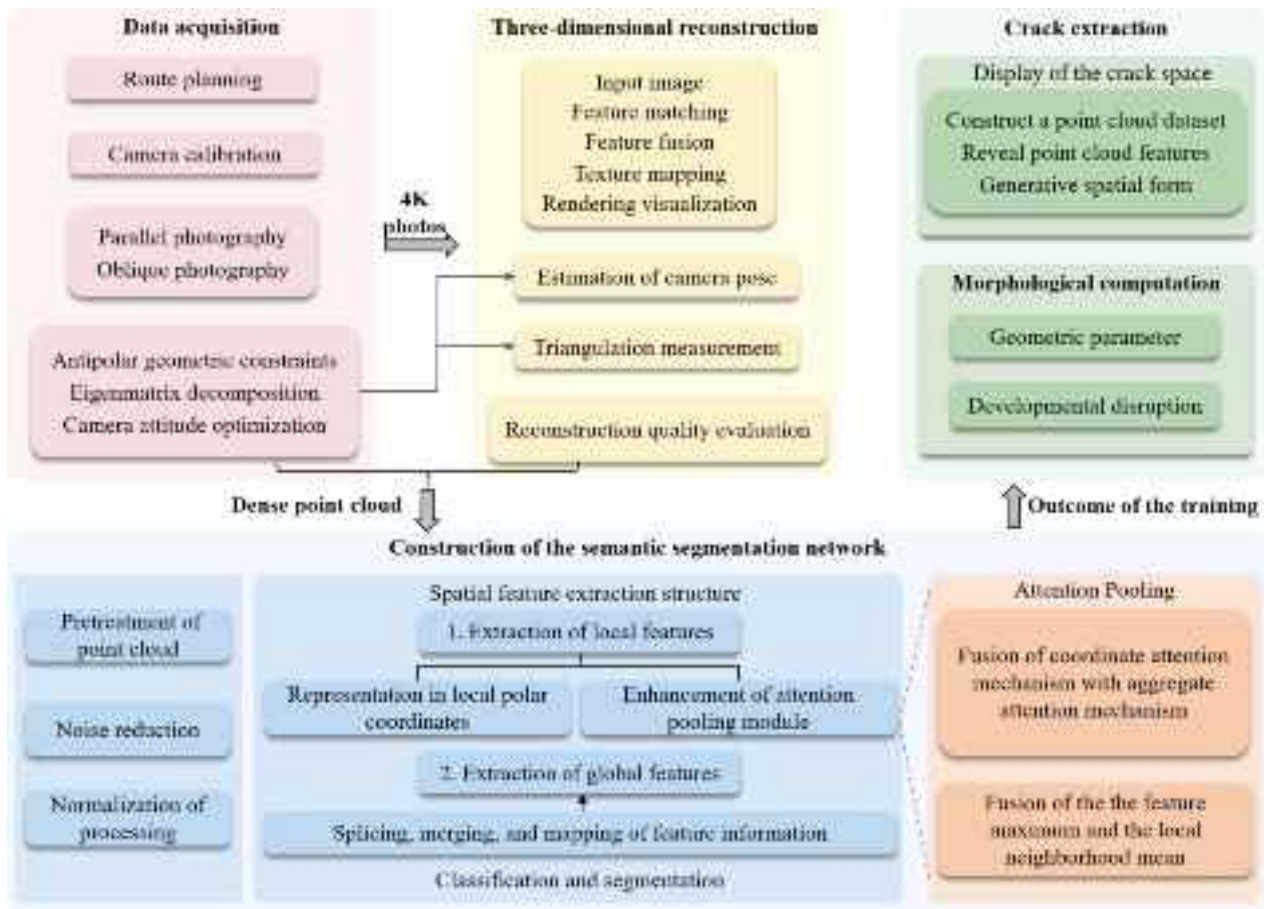


Fig. 8. Research method.

segmentation accuracy, a multi-scale attention fusion mechanism was introduced, this mechanism combined coordinate attention and aggregated attention pooling to enhance spatial feature extraction. Finally, through quantitative crack parameterization, the segmentation results were transformed into engineering parameters such as crack length and width, establishing a comprehensive “3D reconstruction-intelligent identification-quantitative analysis” technology chain.

### 3. Experimental results

#### 3.1. Data

By using a DJ M30 drone, different numbers of image data for the research area were obtained for subsequent research under various flight speeds at the same altitude. The details are presented in the Table 1 below.

#### 3.2. Analysis of reconstruction results

The reconstruction results were based on the reconstructed point

cloud obtained using the SFM-MVS algorithm as a reference. The quality of the reconstructed point cloud was visualized using Instant-Ng and Point-NeRF, which enabled the point-to-point distance to be assessed. As shown in Fig. 9, the blue color in the visualization indicates a good reconstruction quality, while green or yellow areas represent lower quality. Notably, compared to the use of the SFM-MVS algorithm for visualizing the reconstructed point cloud in our study area, the NeRF algorithm produced a larger blue area, suggesting a higher reconstruction quality. Furthermore, it was observed that the improved NeRF algorithm yielded superior visualization effects.

The efficacy of 3D reconstruction was further validated by comparing the efficiency and similarity of reconstructions. The time required to complete the process was used as the evaluation metric for reconstruction efficiency, while reconstruction similarity was assessed based on the root-mean-square error of point-to-point distances. Table 2 presents the results for both 3D reconstruction efficiency and similarity. In each reconstruction method, there was a positive correlation between the number of images used and the time taken for reconstruction; thus, obtaining a more accurate 3D model often requires a longer processing time. The experimental results demonstrated that when using 60 – 140 images for reconstruction, the SFM-MVS algorithm required 3120.83, 5106.42, 7623.52, 9841.08, and 11826.79 s, for successive reconstructions, whereas the improved Instant-Ng algorithm required 459.14, 729.47, 1178.36, 1476.83, and 1622.90 s. Similarly, the improved Point-NeRF algorithm required 281.45, 472.26, 693.73, 906.34, and 1093.17 s for successive reconstructions. In comparison, under identical conditions, the SFM-MVS algorithm required approximately 6.47- to 11.11-fold longer reconstruction time than the NeRF algorithm. It was evident that NeRF possessed distinct advantages over the traditional SFM-MVS algorithm in terms of reconstruction efficiency.

Table 1

Sources of research data.

UAV configuration	Speed (m/s)	Fly height (m)	Image number (pieces)	Image resolution (ppi)
DJM30	2	20	140	4000 × 3000
	4		120	
	6		100	
	8		80	
	10		60	



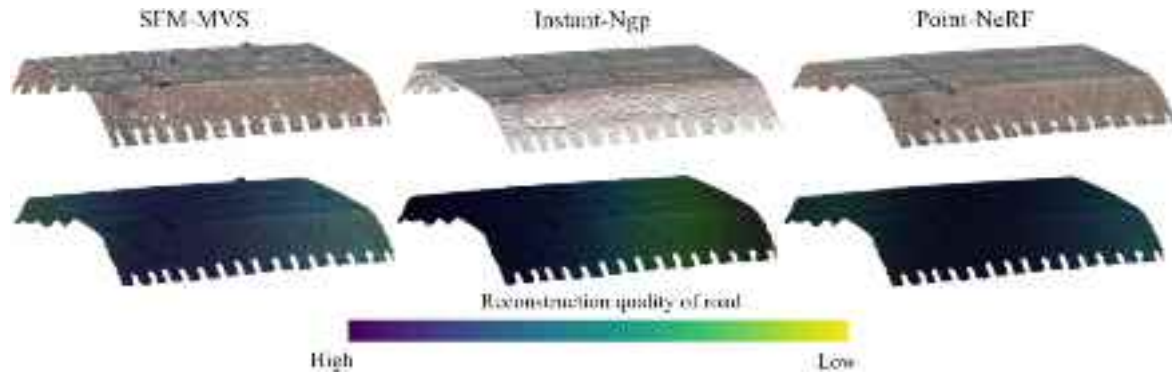


Fig. 9. Quality of model reconstruction.

Table 2

The efficiency of 3D modeling and similarity in reconstruction.

Image	SFM-MVS RT (s)	Instant-Ngp RT (s)	RMSE of point-to-point distances (cm)	Point-NeRF RT (s)	RMSE of point-to-point distances (cm)
60	3120.83	459.14	3.042	281.45	2.379
80	5106.42	729.47	3.256	472.26	2.352
100	7623.52	1178.46	2.864	693.73	2.168
120	9841.08	1476.83	2.701	906.34	2.149
140	11826.79	1622.90	2.892	1093.17	2.225

Note: RT-reconstruction time; RMSE- root mean square error.

The SFM-MVS method required a reconstruction time approximately 52–84 times that of the number of images when adopting the reconstruction strategy of 60–140 images. In contrast, the equivalent times for Instant-Ngp and Point-NeRF were only approximately 8–12 and 5–8 times higher, respectively. On average, Instant-Ngp reconstructed about 7 times faster than SFM-MVS, while Point-NeRF reconstructed about 10 times faster. Consequently, the NeRF algorithm significantly reduced the time cost in the 3D modeling stage, enabling rapid and efficient completion of complex road pavement reconstructions with detailed texture features.

The RMSE of point-to-point distances in the road surface point-cloud reconstructed by Instant-Ngp was only 3.256 cm, with a distribution within 2.8 cm, based on the point cloud of the 3D model reconstructed by SFM-MVS. Approximately 64 % of these distances were distributed within 2.5 cm, indicating a high level of similarity between reconstructions. For the pavement point cloud reconstructed by Point-NeRF, the RMSE of point-to-point distances was only 2.379 cm and had a distribution within 1.7 cm. Approximately 72 % of these distances fell within an even narrower range of 1.4 cm, demonstrating a superior similarity compared to the Instant-Ngp reconstruction results and reflecting the enhanced accuracy achieved through the Point-NeRF reconstruction methodology. Overall, both the SFM-MVS and NeRF algorithms yielded highly matched point clouds with NeRF exhibiting a significantly improved reconstruction speed ranging from 700 % to 1000 % when compared to the SFM-MVS method. Therefore, Point-NeRF was more practically valuable in terms of reconstruction efficiency and similarity.

### 3.3. Results and evaluation of semantic segmentation

In this study, the dataset comprised over 2 million points, each containing 3D coordinates, color information, and intensity data. The network architecture was implemented using Python 3.7 with an experimental setup consisting of an Ubuntu 20.04 operating system, NVIDIA RTX 4060TI GPU, CUDA 12.1 toolkit, and PyTorch 2.1.1 deep learning framework. During the experiment, the Ranger optimizer was used to train the network with specific parameters. This included

utilizing 2D convolutions with convolution kernels in a pixel grid size of  $3 \times 3$ ; setting an initial learning rate of 0.005; a batch size of 4; a decay\_rate of 0.0001; and the network underwent training for a total of 200 epochs using input batches comprising 4096 points at a time along with their respective neighboring points (8 in total).

The results of the two dataset ablation experiments are shown in Table 3, and the accuracy of PointNet++ detection was improved when each module is added. Among them, although each evaluation index of adding coordinate attention mechanism and aggregated attention pooling mechanism alone increased, the ability of aggregated attention pooling mechanism for point cloud recognition and extraction of complex features of point cloud was slightly higher than that of coordinate attention mechanism. When both attentional mechanisms were introduced into the PointNet++ feature extraction module, the increases in each metric for training the Semantic3D dataset were OA (+9.2 %), mIoU (+7.0 %), Precision (+6.0 %), Recall (+7.9 %), and F1 (+7.0 %); and the increases in each metric for training our dataset were OA (+10.4 %), mIoU (+9.4 %), Precision (+5.5 %), Recall (+9.6 %), and F1 (+6.6 %). Both datasets demonstrated significant improvements in detection accuracy. Although the proposed method introduces a slight increase in inference time and algorithmic complexity compared to single attention mechanisms, it effectively enhances the recognition, extraction, and fusion of complex features within point clouds. Therefore, this study integrates a combined approach of the coordinate attention mechanism and aggregated attention pooling mechanism into the PointNet++ network, offering a highly cost-effective solution for high-precision pavement detection tasks. Based on this, the results of semantic segmentation of our dataset point cloud are shown in Fig. 10.

The Semantic3D dataset and the self-built dataset were trained, and the training loss curves and validation loss curves in training were represented by train and eval, respectively. Analyzing the results from the two datasets (Fig. 11), the performance of the PointNet++ model before and after the enhancement was significantly different. The training and validation loss curves of the original model fluctuated greatly (Fig. 11(a)), the loss value decreased at a slow rate, the convergence was late and the convergence oscillation interval was located in the range of 0.020–0.025, which reflected the insufficient feature learning ability and lack of stability. In contrast, the fluctuation range of the loss curve of the enhanced model (Fig. 11(b)) was effectively reduced, the convergence oscillation interval was stable at 0.005–0.010, and the training/validation loss value accelerated the decline and reached convergence earlier. The experimental results demonstrated that the enhanced model effectively improved the feature extraction capability through structural optimization, and the improvement of the smoothness of the loss curve indicated the improvement of the model stability, and the reduction of the oscillation amplitude indicated the enhancement of the model robustness.

To comprehensively assess both the network processing efficiency and semantic segmentation accuracy, the OA, mIoU, precision, recall, and F1 score were used as metrics to evaluate the semantic segmentation

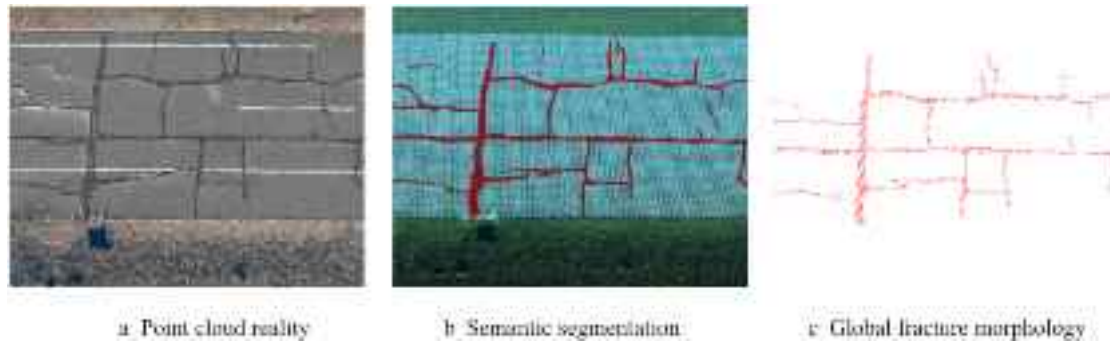


**Table 3**

Ablation experiment results of improved PointNet++.

Dataset	PointNet++		Epoch	Evaluation metric				
	CA	AAP		OA/%	mIoU/%	Precision/%	Recall/%	F1/%
Semantic3D	—	—	200	83.2	76.6	84.3	80.2	82.2
	✓	—	200	84.7	79.7	84.9	86.3	85.5
	—	✓	200	86.6	80.4	85.7	86.8	86.4
	✓	✓	200	92.4	83.6	90.3	88.1	89.2
This paper	—	—	200	80.4	74.8	83.5	78.7	82.0
	✓	—	200	84.2	77.2	83.9	82.6	83.2
	—	✓	200	85.9	79.2	86.4	86.1	86.2
	✓	✓	200	90.8	84.2	89.0	88.3	88.6

Note: CA- coordinate attention; AAP- aggregation attention pooling.

**Fig. 10.** Semantic segmentation visualization.

results. The corresponding outcomes are presented in Table 4 and Fig. 12.

According to the comparative analysis of Fig. 12(a)(c) and (b)(d), significant differences in segmentation performance were observed between the original and improved networks. The original model achieved overall segmentation accuracies exceeding 80 % on both datasets (with crack segmentation reaching 88.3 % and 84.6 %, respectively). However, the mean Intersection over Union (mIoU) for cracks was limited to 77.1 % and 74.8 %, while precision, recall, and F1-score ranged between 75 % and 80 %, indicating insufficient capability in capturing crack features. The improved network demonstrated enhanced performance: overall accuracy surpassed 87 % (cracks at 90.1 % and 91.8 %), mIoU for cracks increased significantly to 82.0 % and 80.1 %, and precision, recall, and F1-score all exceeded 80 %, with crack-related metrics surpassing 85 %. Training results from both datasets confirmed that the improved PointNet++ network outperformed the original version in crack segmentation. Specially, the overall accuracy of crack segmentation of the two datasets was increased by 1.8 % and 7.2 %, and the mean Intersection over Union were increased by 4.9 % and 5.3 %, respectively, the accuracy was increased by 5.8 % and 6.7 %, and the recall rate was increased by 4.8 % and 7.9 %, respectively. F1 value increased by 5.3 % and 7.3 % compared with before improvement. Additionally, crack segmentation metrics ranked highest among all five evaluation criteria in the enhanced network, fulfilling the requirements of this paper. These results demonstrated that the proposed semantic segmentation network, along with its optimized parameter configuration, successfully achieved high-precision extraction of crack features.

### 3.4. Crack extraction and calculation

The road plane obtained using the RANSAC algorithm is shown in Fig. 13 (a), while Fig. 13 (b) shows the morphology of the road crack. Leveraging the results of semantic segmentation, the crack point-cloud segmentation outcomes are incorporated into the pavement model for spatial visualization. Moreover, an analysis of both the transverse and longitudinal profiles of the pavement revealed the alterations in

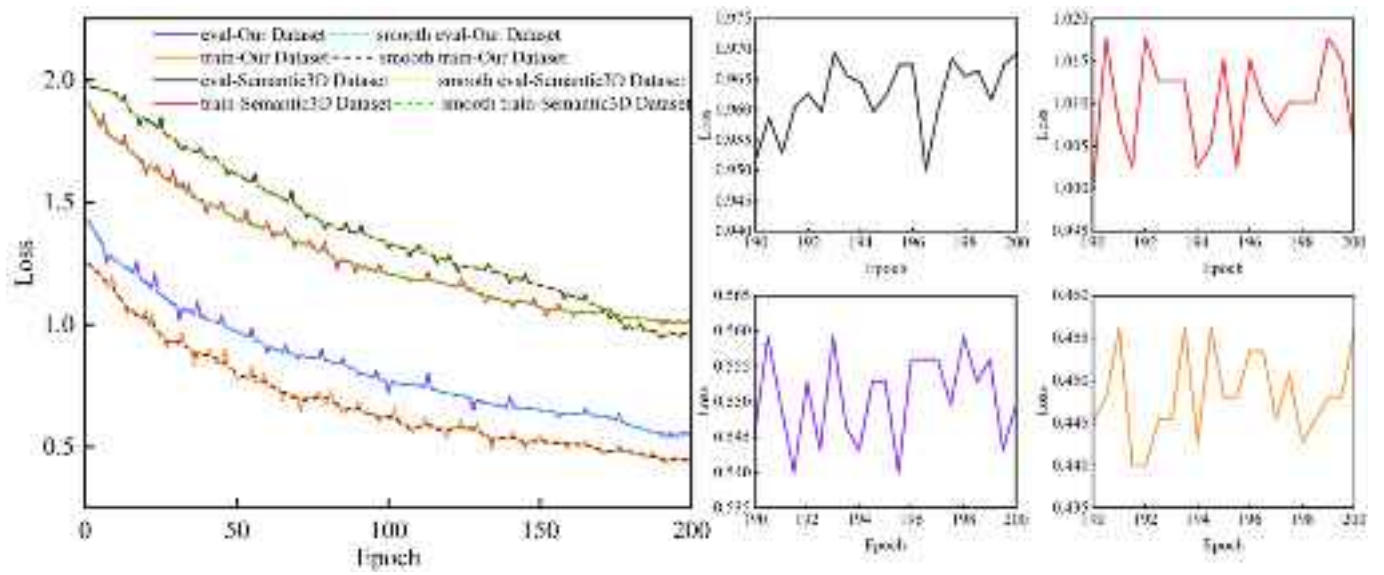
pavement characteristics, as shown in Fig. 14.

Based on Fig. 14, the elevation of the entire road surface varied by approximately 5 cm, encompassing both curvatures and depressions caused by cracks. To mitigate the impact of these irregularities, we used the RANSAC algorithm to establish a standardized road model line. An analysis of the road surface roughness map revealed that this model line exhibited minimal fluctuation in representing the relative depth along transverse profiles, with a maximum depth of 3.274 cm and an average depth of 2.131 cm. Conversely, there was significant variability in relative depth along longitudinal profiles for this model pavement line, with a maximum depth of 3.686 cm and an average depth of 2.436 cm. Because the crack distribution points were situated beneath the pavement level, transverse cracks exhibited smaller changes in average depth than longitudinal cracks.

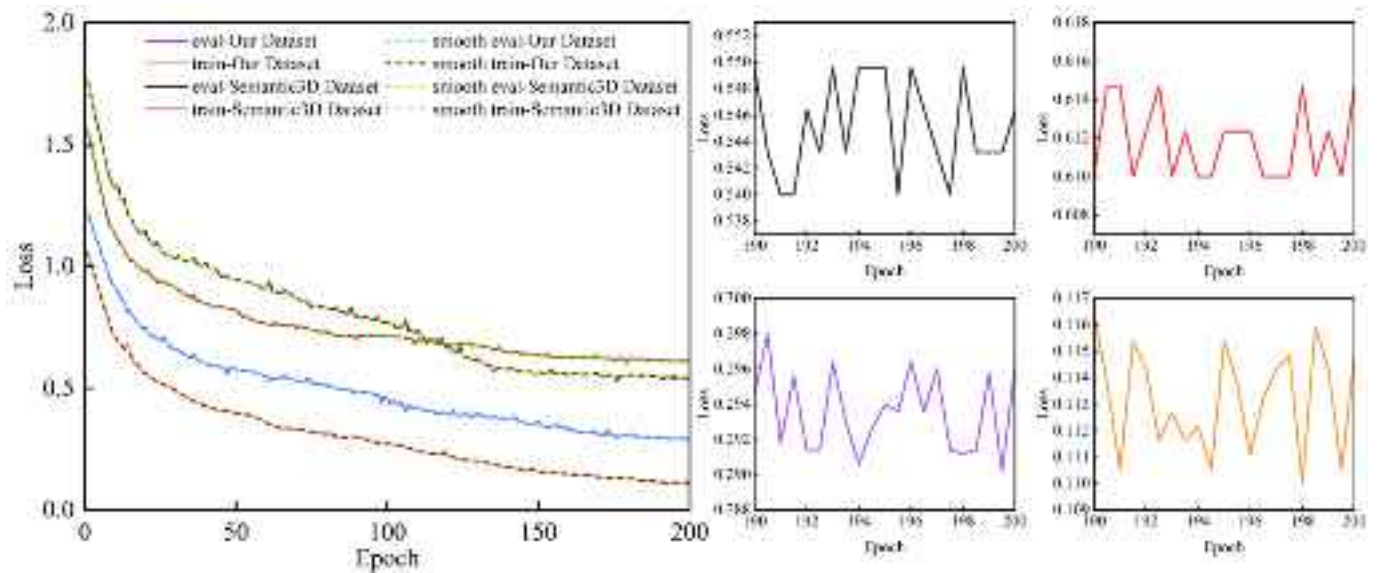
### 3.5. Calculation of crack parameters

The quantitative calculation of crack parameters requires the acquisition of both the surface contour and centerline. In this study, a curved surface was fitted to irregular crack contours, as shown in Fig. 15. The crack centerline was obtained through a least squares regression analysis, as shown in Fig. 16.

The results shown in Figs. 15 and 16 indicate a consistent development of the longitudinal crack profile. The fitted upper profile determination coefficient ( $R^2$ ) ranged from 0.9794 to 0.9906, while the root mean square error (RMSE) ranged from 0.008 to 0.016. The lower profile  $R^2$  value varied between 0.9813 and 0.9904, with the corresponding RMSE ranging from 0.009 to 0.014. The development of the transverse crack profile exhibited a higher volatility compared to that of the longitudinal crack. The fitted upper profile  $R^2$  value ranged from 0.9295 to 0.9526, while the RMSE ranged from 0.033 to 0.094. Additionally, the lower profile determination coefficient  $R^2$  ranged from 0.9186 to 0.9588 with the corresponding RMSE ranging from 0.027 to 0.108. The contour fitting and  $R^2$  values for both longitudinal and transverse cracks exceeded 0.9, indicating an excellent fit that accurately represented the actual crack contours.



(a) PointNet++ prior to enhancement



(b) PointNet++ later to enhancement

Fig. 11. Alteration of the loss value.

Table 4

Evaluation results of the road model for semantic segmentation.

Evaluation metric		OA/%		mIoU/%		Precision/%		Recall/%		F1/%	
PointNet++ Network enhancement											
Semantic3D Dataset	Condition	Prior	later	Prior	later	Prior	later	Prior	later	Prior	later
	Pavement	84.6	87.4	76.2	78.4	80.7	86.5	79.5	84.7	80.1	85.7
	Subgrade	85.2	89.6	75.3	77.2	78.2	85.9	74.6	80.0	76.4	82.8
	Crack	88.3	90.1	77.1	82.0	84.8	90.6	80.4	85.2	82.5	87.8
Our Dataset	Condition	Prior	later	Prior	later	Prior	later	Prior	later	Prior	later
	Pavement	82.4	88.6	78.2	78.7	82.1	85.8	80.1	82.4	81.1	84.1
	Subgrade	80.7	87.2	73.6	74.6	76.9	84.2	76.6	80.9	76.7	82.5
	Crack	84.6	91.8	74.8	80.1	82.5	89.2	79.4	87.3	80.9	88.2

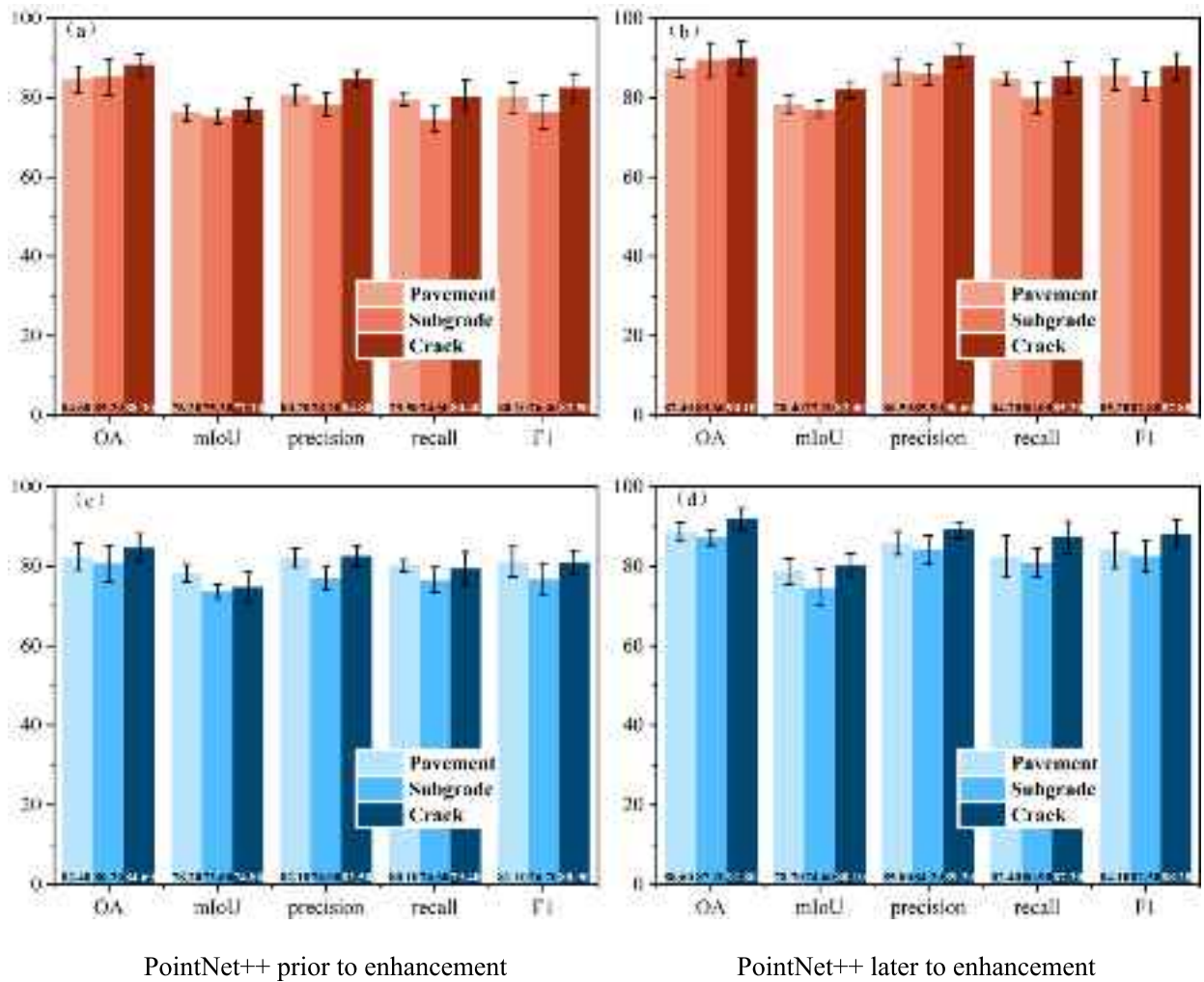


Fig. 12. Evaluation outcomes of the semantic segmentation (a, b-Semantic3D Dataset; c, d-Our Dataset).

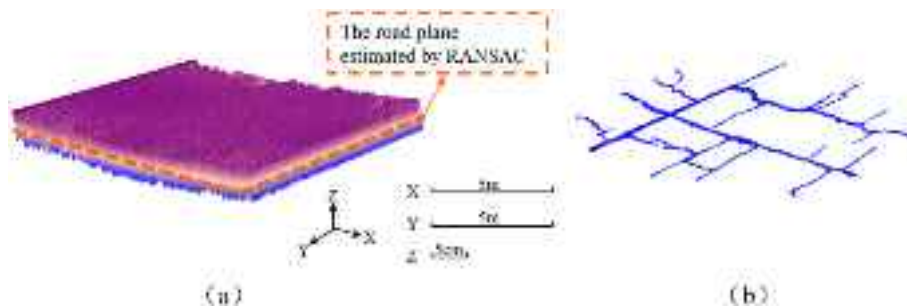


Fig. 13. Pavement fitting and cracking 3D image.

The fitting of the crack centerline also conformed to the established principles governing crack evolution, with the  $R^2$  value ranging from 0.9857 to 0.9964 for the longitudinal crack centerline and the RMSE ranging from 0.011 to 0.018. For the transverse crack centerline, the  $R^2$  value ranged from 0.9465 to 0.9719, while the RMSE ranged from 0.036 to 0.140. The longitudinal and transverse cracks exhibited a center line fitting with an  $R^2$  value exceeding 0.9, indicating an excellent fit that could be considered to represent the true crack center line.

In general, the fitting results of the crack contour and center line exhibited some errors; however, they were still strongly correlated with the actual situation. This observation reflects the modeling and segmentation accuracy in this study, thereby enabling its use for calculating

crack geometric parameters. The calculated crack geometric parameters are provided in Table 5.

It can be seen from the results in Table 5 that the error between the calculated crack length and the actual measurement was within 0.3 m, i.e., within 8 %. The error between the calculated crack width and the actual measurement was within 0.02 m, i.e., within 11 %. The error between the calculated crack depth and the actual measurement was within 0.7 cm, i.e., within 20 %. Overall, there was a strong correlation between the calculated results and the actual measurements and the errors were low.

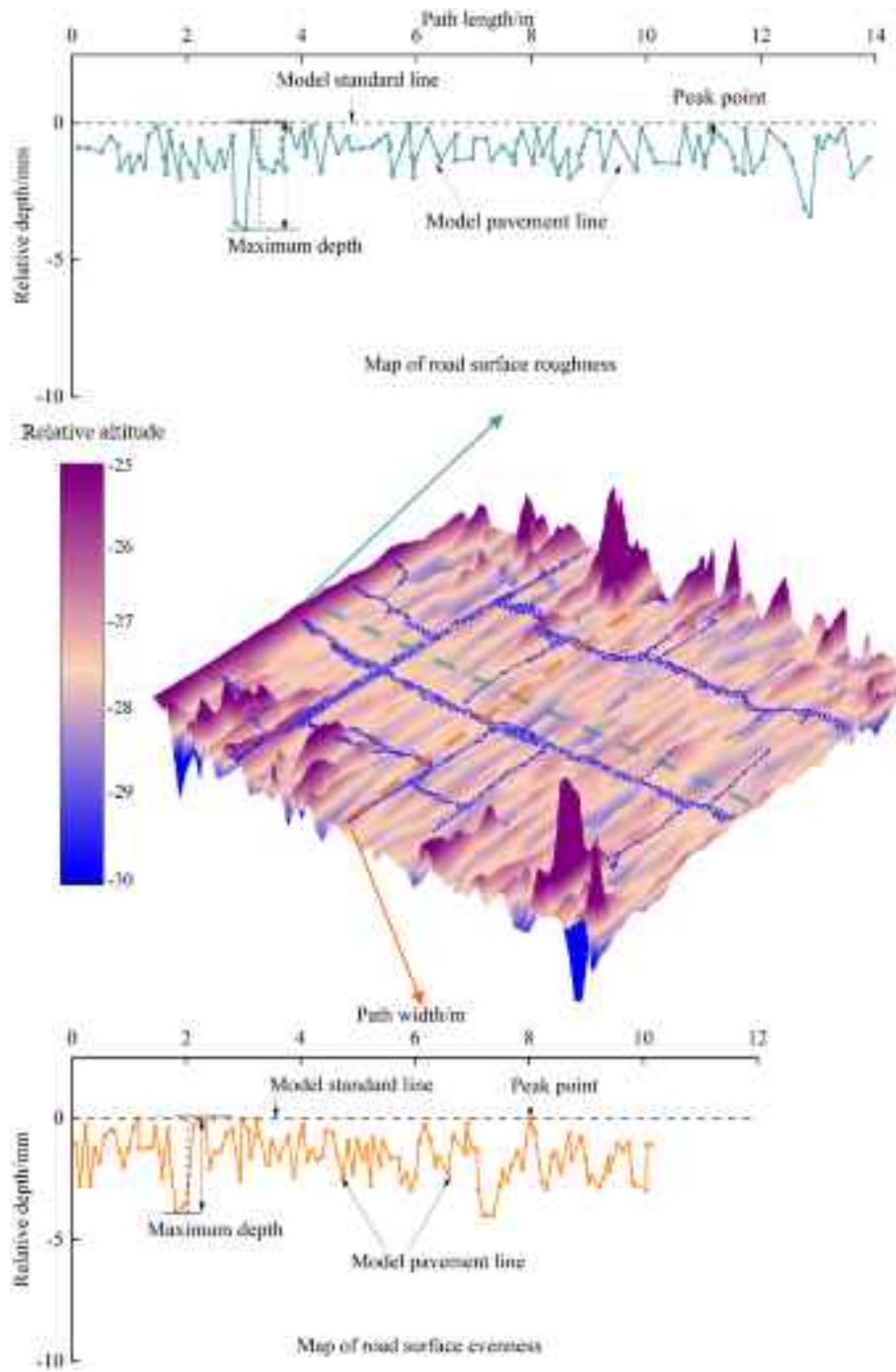


Fig. 14. Visualization of pavement spatial structure.

#### 4. Discussion

The SFM-MVS algorithm is currently widely used for road point-cloud model acquisition to extract accurate and cost-effective 3D models. However, its slow reconstruction speed limits its practical application. Moreover, changes in illumination conditions significantly influence the SFM-MVS reconstruction results. In contrast, NeRF-based models consist of self-luminous and self-absorbing particle clouds, minimizing the impact of ambient light [38,39]. Therefore, this study used the Point-NeRF algorithm to overcome the limitations of SFM-MVS by reducing environmental constraints to ensure precise 3D scene representation, which significantly decreased reconstruction time costs, and accelerated both the reconstruction process and point-cloud

acquisition.

An enhanced PointNet++ semantic segmentation network was proposed for crack detection and analysis, which improved operational efficiency and training accuracy by enhancing the local feature extraction module within the spatial feature extraction structure. This approach reduced the reliance on integrated channel features and classical image processing algorithms, and also mitigated the vulnerabilities of traditional semantic segmentation networks to factors such as image background, noise, shadow, and uneven illumination [40,41]. Meanwhile, the crack phenotype analysis technique based on 3D structure, combined with RANSAC plane fitting, surface fitting and least squares regression analysis, realized the calculation of the geometric parameters of the crack space, expanded the dimensions of the point



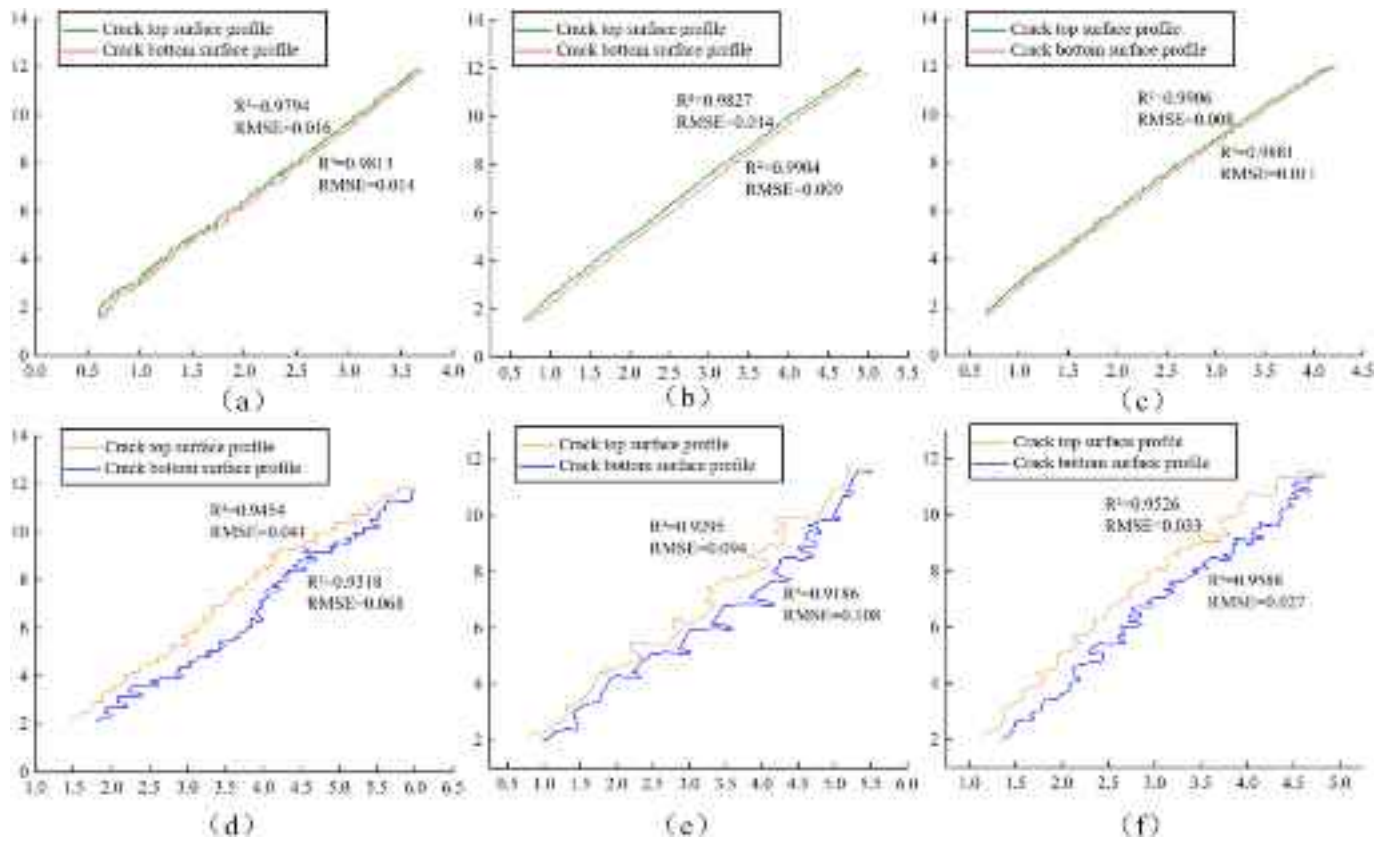


Fig. 15. Crack surface profiles (a to c indicates longitudinal cracks, d to f indicates transverse cracks).

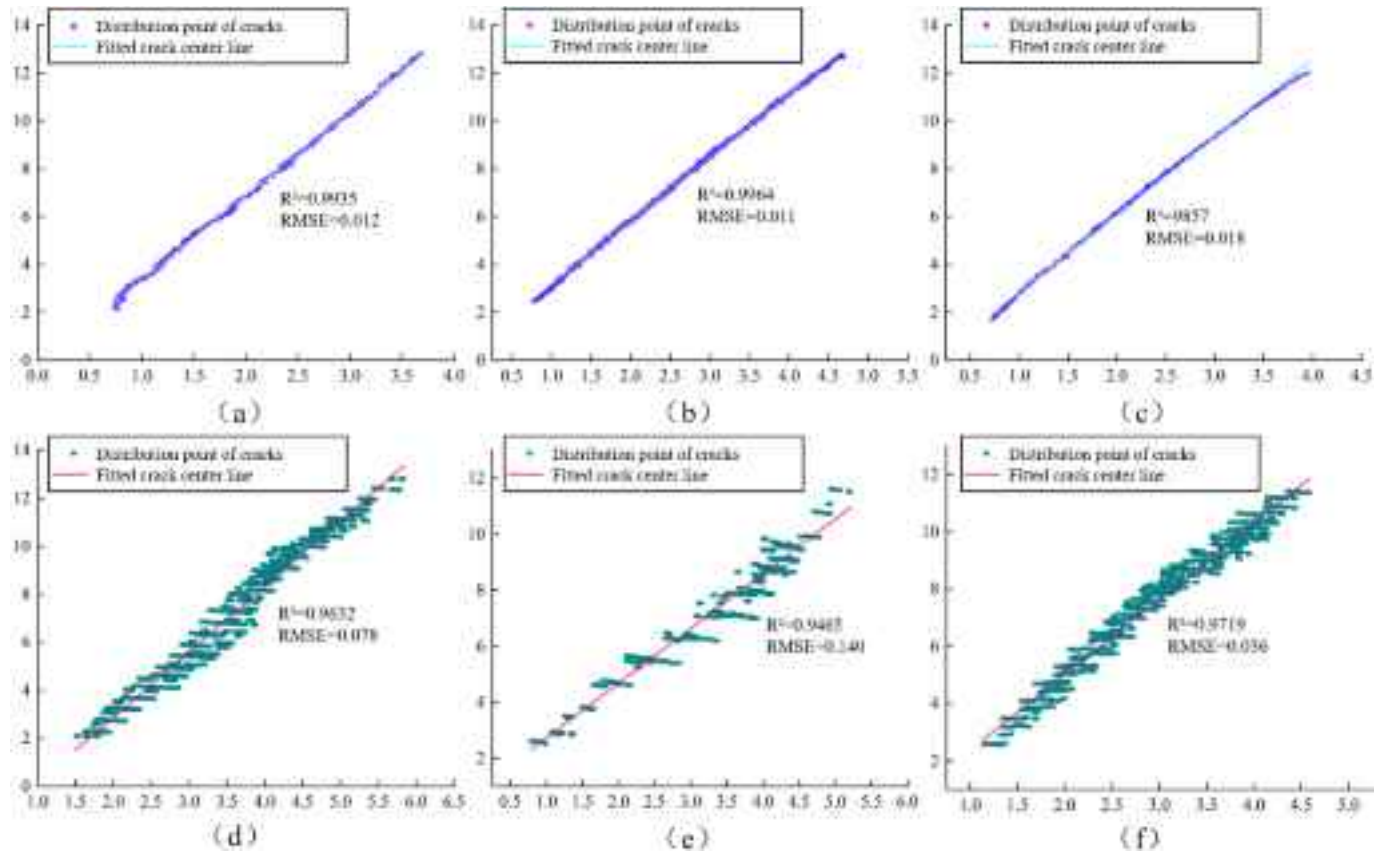


Fig. 16. Crack center lines (a to c indicates longitudinal cracks, d to f indicates transverse cracks).

**Table 5**

Results of the crack geometry parameter.

ID	CL /m	GL /m	RE /m	ER /%	CL /m	GL /m	RE /m	ER /%	CL /cm	GL /cm	RE /cm	ER /%
a	2.323	2.452	0.129	5.26	0.157	0.151	0.006	3.97	2.316	2.872	0.556	19.36
b	3.192	3.447	0.255	7.40	0.174	0.165	0.009	5.45	2.406	2.847	0.441	15.49
c	2.607	2.808	0.201	7.16	0.133	0.120	0.013	10.83	2.633	3.055	0.422	13.81
d	3.563	3.856	0.293	7.60	0.225	0.209	0.016	7.66	1.831	2.264	0.433	19.13
e	3.672	3.842	0.170	4.42	0.242	0.236	0.006	2.54	2.668	3.332	0.664	19.93
f	4.083	4.337	0.254	5.86	0.196	0.182	0.014	7.69	2.476	2.869	0.393	13.70

Note: CL- calculation length; GL- gage length; RE- relative error; ER- error rate.

cloud features, and made the point cloud spatial features richer and the calculation results more accurate.

In summary, this paper has realized the detection of pavement cracks based on Point-NeRF 3D reconstruction and improved PointNet++ semantic segmentation network. The method detects the pavement cracks from the three-dimensional level, avoiding the lack of information in the two-dimensional detection, making the crack detection results more complete; and the method has a high degree of automation, which can greatly reduce the manual intervention, lower the labor cost, and is suitable for large-scale detection of pavement cracks, which enhances its practicability in actual engineering; at the same time, the method can realize the monitoring of road conditions in remote or hard-to-reach areas, which greatly improves the efficiency and safety of road maintenance work. However, the method also has some limitations. In the 3D reconstruction, the interference of the complex background will lead to the reduction of the exposure and resolution performance of the camera carried by the UAV, which will cause motion blur, and reduce the quality of the reconstructed point cloud by the NeRF model; in the process of network segmentation, the crack segmentation results contain the real form of the cracks, but some of the cracks have been patched by asphalt, this will make the edges of the segmented cracks not uniform and smooth enough, which in turn affects the goodness of fit of the crack centerline and the crack contour line, and ultimately leads to a certain error in the results. Future improvements will focus on optimizing UAV-mounted camera systems for enhanced resolution and reduced false positive rates, refining network architectures through structural optimization, and integrating big data analytics to provide more reliable decision support for predictive maintenance planning and optimized road preservation strategies.

## 5. Conclusion

In this paper, we proposed a method for 3D pavement crack extraction and analysis based on neural radiation field and semantic segmentation network. A low-cost unmanned aerial vehicle (UAV) equipment was utilized to capture image sequences of the pavement in the study area. By comparing the modeling efficiency and reconstruction similarity of the Structural Multiview Stereo Vision with Motion Recovery (SFM-MVS) algorithm with two neural radiation field (NeRF) algorithms for 3D reconstruction, we selected Point-NeRF as a neural radiation field-based 3D modeling algorithm for pavements. Then combined with our improved semantic segmentation network PointNet++, the accurate segmentation of pavement cracks was realized. Finally, based on the segmentation results output from this network, the spatial geometric information of the cracks was computed using the correlation fitting method. The following findings are mainly drawn:

(1) The selected Point-NeRF algorithm can accurately and quickly complete the 3D modeling of road pavement, and the reconstructed pavement point cloud has a higher similarity between models than the traditional SFM-MVS algorithm reconstructed point cloud with an average point-to-point root-mean-square error of less than 1.4 cm. However, the Point-NeRF modeling process is lightweight on the basis of SFM-MVS, avoiding the secondary dense reconstruction, which reduces the modeling time to one-tenth of the original time and greatly reduces

the time cost.

(2) This paper reduced the errors and extra computational costs associated with image processing and geometric computation in traditional crack quantization methods. The quantitative results show that the OA, mIoU, Precision, Recall and F1 values of the improved PointNet++ network crack segmentation results have been improved by 7.2 %, 5.3 %, 6.7 %, 7.9 % and 7.3 %, respectively, compared with those before the improvement. Qualitatively, the improved Point-Net++ has better robustness for segmentation of crack point cloud data. Additionally, our method can learn effective features in a small amount of point cloud data, which reduces the dependence on large datasets and improves the detection efficiency.

(3) For crack quantification, we achieved this by fitting the crack contour and crack centerline, which were verified by the coefficient of determination ( $R^2$ ) and the root mean square error (RMSE). The results show that the  $R^2$  of the contour line of the cracks exceeds 0.9 and the RMSE is within 0.15, and the  $R^2$  of the center line of the cracks exceeds 0.9 and the RMSE is within 0.15, which makes the fitting results have high accuracy. The calculated crack length, width and depth are within 0.3 m, 0.02 m and 0.7 cm from the actual measurement results, and the relative error rates are within 8 %, 11 % and 20 %, which verifies the accuracy of crack quantification.

## CRedit authorship contribution statement

**Changjie Yao:** Writing – original draft, Validation, Software, Methodology, Investigation, Data curation. **Mingtang Chai:** Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Guoyu Li:** Methodology, Funding acquisition, Conceptualization. **Dun Chen:** Visualization, Data curation. **Qingsong Du:** Software, Resources. **Shunshun Qi:** Validation, Software. **Kai Gao:** Software.

## 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.

## Acknowledgments

This work was supported by the Second Tibetan Plateau Scientific Expedition and Research Program (No.2019QZKK0905), the Open Project of National Cryosphere Desert Data Center (E01Z790201), the Tibet Autonomous Region Science and Technology Key R&D (No. XZ202401ZY0040), the Foundation of the State Key Laboratory of Frozen Soil Engineering (Nos. SKLFSE-ZQ-202303 and SKLFSE-ZY-20), the University First-Class Discipline Construction Project of Ningxia (NXYLXK2021A03). We also thanked the editor and anonymous reviewers for very helpful comments that improved the quality of the manuscript significantly.

## Data availability

Data will be made available on request.

## References

- [1] M. Elkashef, R.C. Williams, E. Cochran, Investigation of fatigue and thermal cracking behavior of rejuvenated reclaimed asphalt pavement binders and mixtures, *Int. J. Fatigue* 108 (2018) 90–95, <https://doi.org/10.1016/j.ijfatigue.2017.11.013>.
- [2] Z.H. Pan, J.C. Guan, X. Yang, K. Fan, C.H.O. Jeremy, N.Q. Guo, X. Wang, One-stage 3D profile-based pavement crack detection and quantification, *Automat. Constr.* 153 (2023) 104946, <https://doi.org/10.1016/j.autcon.2023.104946>.
- [3] X. Yang, Z.P. You, J. Hiller, D. Watkins, Correlation analysis between temperature indices and flexible pavement distress predictions using mechanistic-empirical design, *J. Cold Reg. Eng.* 31 (2017), [https://doi.org/10.1061/\(ASCE\)CR.1943-5495.0000135](https://doi.org/10.1061/(ASCE)CR.1943-5495.0000135).
- [4] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, H.B. Ling, Feature pyramid and hierarchical boosting network for pavement crack detection, *IEEE Trans. Intell. Transp. Syst.* 21 (2020) 1525–1535, <https://doi.org/10.1109/TITS.2019.2910595>.
- [5] Yang, X., Guan, J.C., Ding, L., You, Z.P., Lee, V.C.S., Hasan, M.R.M., Chen, X.Y., Research and applications of artificial neural network in pavement engineering: a state-of-the-art review, *J. Traffic Transp. Eng. (Engl. Ed.)* 8 (6) (2021) 1000–1021, DOI: 10.1016/j.jtte.2021.03.005.
- [6] J.Q. Zhang, X. Yang, W. Wang, J.C. Guan, L. Ding, V.C.S. Lee, Automated guided vehicles and autonomous mobile robots for recognition and tracking in civil engineering, *Autom. Constr.* 146 (2023) 104699, <https://doi.org/10.1016/J.AUTCON.2022.104699>.
- [7] Q. Zhang, S.S. Chen, Y. Wu, Z.H. Ji, F. Yan, S.L. Huang, Y.Q. Liu, Improved U-net network asphalt pavement crack detection method, *PLoS One* 19 (5) (2024) e0300679–e, <https://doi.org/10.1371/JOURNAL.PONE.0300679>.
- [8] Y. Hou, Q.H. Li, C. Zhang, G.Y. Lu, Z.J. Ye, Y.H. Chen, L.B. Wang, D.D. Cao, The state-of-the-art review on applications of intrusive sensing, image processing techniques, and machine learning methods in pavement monitoring and analysis, *Engineering* 7 (6) (2020) 845–856, <https://doi.org/10.1016/J.ENG.2020.07.030>.
- [9] Y.M. Zhang, C. Chen, Q. Wu, Q. Lu, A Kinect-based approach for 3D pavement surface reconstruction and cracking recognition, *IEEE Trans. Intell. Transp. Syst.* 19 (12) (2018) 3935–3946, <https://doi.org/10.1109/TITS.2018.2791476>.
- [10] R. Gui, Q. Sun, W.Q. Wu, D.J. Zhang, Q.Q. Li, Transfer learning for cross-scene 3D pavement crack detection based on enhanced deep edge features, *Eng. Appl. Artif. Intell.* 123 (2023), <https://doi.org/10.1016/J.ENGAPPAI.2023.106452>.
- [11] 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 (2015) 2353–2362, <https://doi.org/10.1109/TITS.2015.2428655>.
- [12] A.L. Zhang, K.C.P. Wang, Y. Fei, Y. Liu, C. Chen, G.W. Yang, J.S.Q. Li, E.H. Yang, S. Qiu, Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network, *COMPUT-AIDED CIV. INF.* 34 (3) (2019) 213–229, <https://doi.org/10.1111/micc.12409>.
- [13] M.Q. Zhong, L.C. Sui, Z.H. Wang, D.M. Hu, Pavement crack detection from Mobile laser scanning point clouds using a time grid, *Sensors* 20 (15) (2020) 4198, <https://doi.org/10.3390/s20154198>.
- [14] Y.P. Liu, X.P. Yan, N. Wang, X. Zhang, Z. Li, A 3D reconstruction method of image sequence based on deep learning, *J. Phys. Conf. Ser.* 1550 (2020) 032051, <https://doi.org/10.1088/1742-6596/1550/3/032051>.
- [15] Y. Ma, Y.B. Zheng, S.Y. Wang, Y.D. Wong, S.M. Easa, Point cloud-based optimization of roadside LiDAR placement at constructed highways, *Autom. Constr.* 144 (2021) 104629, <https://doi.org/10.1016/j.autcon.2022.104629>.
- [16] Ge, Y.F., Liu, J., Zhang, X., Tang, H.M., Xia, X.L., Automated Detection and Characterization of Cracks on Concrete Using Laser Scanning, *Journal Of Computing In Civil Engineering*. 29 (2) (2023), DOI: 10.1061/JITSE4.ISENG-1936.
- [17] D.J. Zhang, Q. Zou, H. Lin, X. Xu, L. He, R. Gui, Q.Q. Li, Automatic pavement defect detection using 3D laser profiling technology, *Autom. Constr.* 96 (2018) 0926–5805, <https://doi.org/10.1016/j.autcon.2018.09.019>.
- [18] A.M. Saad, K.N. Tahar, Identification of rut and pothole by using multirotor unmanned aerial vehicle (UAV), *Measurement* 137 (2019) 647–654, <https://doi.org/10.1016/j.measurement.2019.01.093>.
- [19] J.L. Xiao, J.S. Fan, Y.F. Liu, B.L. Li, J.G. Nie, Region of interest (ROI) extraction and crack detection for UAV-based bridge inspection using point cloud segmentation and 3D-to-2D projection, *Autom. Constr.* 158 (2024) 105226, <https://doi.org/10.1016/j.autcon.2023.105226>.
- [20] Y.G. Jeong, N.T. Nguyen, K.D. Tran, B.H.H. Thi, Applying unmanned aerial vehicle photogrammetry for measuring dimension of structural elements in traditional timber building, *Measurement* 153 (2020) 107386, <https://doi.org/10.1016/j.measurement.2019.107386>.
- [21] S.D. Feng, M.X. Gao, X.W. Jin, T. Zhao, F. Yang, Fine-grained damage detection of cement concrete pavement based on UAV remote sensing image segmentation and stitching, *Measurement* 226 (2024) 113844, <https://doi.org/10.1016/j.measurement.2023.113844>.
- [22] W. He, Z.H. Ye, M.S. Li, Y.L. Yan, W. Lu, G.G. Xing, Extraction of soybean plant trait parameters based on SFM-MVS algorithm combined with GRNN, *Front. Plant Sci.* 14 (2023) 1181322, <https://doi.org/10.3389/FPLS.2023.1181322>.
- [23] Zhu, B.L., Liu, F.S., Zhu, J.Y., Guo, Y., Ma, Y.T., Three-dimensional quantifications of plant growth dynamics in field-grown plants based on machine vision method, *Transactions of the Chinese Society for Agricultural Machinery*. 49 (5) (2018) 256–262. (in Chinese), Doi: 10.6041/j.issn.1000-1298.2018.05.030.
- [24] B. Mildenhall, P.P. Srinivasan, M. Tancik, J.T. Barron, R. Ramamoorthi, R. Ng, Nerf: representing scenes as neural radiance fields for view synthesis, *CACM* 65 (1) (2020) 99–106, <https://doi.org/10.1145/3503250>.
- [25] Zhu, F., A review of 3D scene representation—NeRF: neural radiance fields, *Journal of Communication University of China (Science and Technology)*. 1673-4793 (2022) 05-0064-14. (in Chinese), DOI:10.16196/j.cnki.issn.1673-4793.2022.05.007.
- [26] P.S. Li, W.B. Li, Z.L. Feng, Q. Wan, H.R. Wang, Improved algorithm for point cloud data simplification based on GPU parallel, *Appl. Res. Comput.* 37 (9) (2020) 2730–2733, <https://doi.org/10.19734/j.issn.1001-3695.2019.04.0111>, in Chinese.
- [27] Y. Ma, Y.B. Zheng, S. Easa, J.C. Cheng, Semi-automated frame-work for generating cycling lane centerlines on roads with roadside barriers from noisy MLS data, *ISPRS J. Photogramm.* 167 (2020) 396–417, <https://doi.org/10.1016/j.isprsjprs.2020.07.009>.
- [28] Wang S.Z., Chen Y.R., Yang Y.T., Research on Digitalization Technology for Urban Road Infrastructure Based on Integration of Air and Ground, *Computer Measurement & Control*. 29(11) (2021) (in Chinese), Doi: 10.16526/j.cnki.11-4762/tp.2021.11.032.
- [29] Qi, C.R., Su, H., NieBner, M., Dai, A., Yan, M.Y., Guibas, L.J., 2016. Volumetric and Multi-view Convolutional Neural Networks (CNNs) for object classification on 3D data, *CVPR*, (2016) 5648-5656, Doi: 10.1109/CVPR.2016.609.
- [30] Su, H., Maji, S., Kalogerakis, E., Erik, G.L., Multi-view convolutional neural networks for 3D shape recognition, *ICCV*. New York: IEEE. (2015) 945-953, DOI: 10.48550/arXiv.1505.00880.
- [31] Wang, Y.C., Yu, B., Chen, X.Y., Zhang, Y.Q., Wang, S.Y., Extraction and Digital Modeling of Road Geometric Information Using LiDAR Data Point Clouds. *China J. Highw. Transp.* 36(03) (2023) 45-60. (in Chinese), Doi: 10.19721/j.cnki.1001-7372.2023.03.003.
- [32] Charles, R.Q., Su, H., Kaichun, M., Guibas, L.J., 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2017) 77-85, DOI: 10.48550/arXiv.1612.00593.
- [33] C.R. Qi, L. Yi, H. Su, L.J. Guibas, A.I. Claims, PointNet++: deep hierarchical feature learning on point sets in a metric space, *Adv. Neural Inf. Proces. Syst.* 30 (2017) 5099–5108, <https://doi.org/10.48550/arXiv.1706.02413>.
- [34] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, B. Chen, PointCNN: convolution on X-transformed points, *Adv. Neural Inf. Proces. Syst.* 31 (2018) 828–838, <https://doi.org/10.48550/arXiv.1801.07791>.
- [35] H.S. Zhao, D.D. Yang, J.K. Yu, 3D target detection using dual-domain attention and SIFT operator in indoor scenes, *Vis. Comput.* 38 (11) (2022) 3765–3774, <https://doi.org/10.1007/S00371-021-02217-Z>.
- [36] Y. Long, P.F. Xu, Calculation method for crack width of Earthen sites via directions of crack centerlines, *Transducer Microsyst. Technol.* 35 (09) (2016), [https://doi.org/10.13873/J.1000-9787\(2016\)09-0014-03](https://doi.org/10.13873/J.1000-9787(2016)09-0014-03) in Chinese.
- [37] N. Safaei, O. Smadi, A. Masoud, N. Safaei, An automatic image processing algorithm based on crack pixel density for pavement crack detection and classification, *Int. J. Pavement Res. Technol.* 15 (2022) 159–172, <https://doi.org/10.1007/s42947-021-00006-4>.
- [38] Martin-brualla, R., Radwan, N., Snavely, N., Barron, J.T., Sajjadi, N., M S M, et al. Nerf in the wild: neural radiance fields for unconstrained photo collections, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2021) 7210-7219, DOI: 10.1109/CVPR46437.2021.00714.
- [39] Srinivasan, P.P., Deng, B.Y., Zhang, X.M., Tancik, M., Mildenhall, B., Barron, J.T., Nerv: neural reflectance and visibility fields for relighting and view synthesis, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2020) 7495-7504, DOI: 10.48550/arXiv.2012.03927.
- [40] J.F. Wang, F. Liu, W.J. Yang, G.Y. Xu, Z. Tao, Pavement crack detection using attention u-net with multiple sources, *Pattern Recognition and Computer Vision: Third Chinese Conference, PRCV*. (2020) 664–672, [https://doi.org/10.1007/978-3-030-60639-8\\_55](https://doi.org/10.1007/978-3-030-60639-8_55).
- [41] Q. Zou, Y. Cao, Q. Li, Q.Z. Mao, S. Wang, CrackTree: Automatic crack detection from pavement images, *Pattern Recognit. Lett.* 33 (3) (2012) 227–238, <https://doi.org/10.1016/j.patrec.2011.11.004>.