Available online at www.sciencedirect.com

ScienceDirect

Underground Space 26 (2026) 282–304

www.keaipublishing.com/undsp

Research Paper

Development of a 3D-point cloud-based spatiotemporal series model for tunnel rock mass discontinuities prediction

Gang Yang [a,b,c], Tianbin Li [a,b,*]

[a] State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China
[b] College of Environment and Civil Engineering, Chengdu University of Technology, Chengdu 610059, China
[c] Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576, Singapore

Received 18 December 2024; received in revised form 16 June 2025; accepted 23 June 2025
Available online 13 November 2025

Abstract

Predicting the three-dimensional (3D) distributions of discontinuities within rock masses is crucial for evaluating tunnel stability. However, this task is challenging due to the inherent opacity of rock, which prevents the direct observation of discontinuities. Most current methods for predicting discontinuities are based on extracting the two-dimensional intersection lines of spatial discontinuities. In this paper, we propose a novel, purely visual approach to analyze and predict the 3D distributions of discontinuities in rock masses. In this method, a 3D model of the tunnel face is constructed based on motion prediction and multi-view stereo vision, and the development of discontinuities is then predicted. Each set of discontinuities is projected onto the virtual tunnel face using a convex hull algorithm, creating a virtual trace. A newly developed algorithm for predicting spatiotemporal sequences, which incorporates a self-attention mechanism and a zigzag recurrent transition mechanism, is then applied to predict the evolution of discontinuities. For testing and verification, we used smartphones to collect surface data on multiple sets of excavated rock from the Bimoyuan Tunnel in Sichuan, China. Extensive experiments involving these surface data demonstrated the effectiveness of our proposed method. The findings provide technical support for predicting tunnel collapse and ensuring tunnel safety.

Keywords: Convolutional long short-term memory; Self-attention; Zigzag recurrent transition mechanism; Spatiotemporal series prediction; Tunnel rock mass discontinuities; 3D scene reconstruction

1 Introduction

Discontinuities, including fractures, fissures, joints, faults, and bedding planes, represent separations within geological formations. These structures permeate rock masses, inducing uneven local stress distributions and anisotropic physical properties (Ju et al., 2022). Given their negligible or nonexistent tensile strength, discontinuities can destabilize tunnels and even cause catastrophic collapse. Understanding their spatial distribution is crucial for accurately assessing rock mass stability (Priest, 1985; Kulatilake et al., 2011; Zhan et al., 2017; Yan et al., 2023; Li et al., 2023).

Persistence, as defined by the International Society for Rock Mechanics (Ulusay, 2015), refers to the spatial extent of a discontinuity along a plane. Despite its fundamental importance, quantifying persistence remains challenging. Trace length serves as a rough proxy, and field geologists traditionally rely on visual sketches to document the two-dimensional (2D) distribution of discontinuities. Additional analytical approaches include one-dimensional (1D) borehole sampling and 2D window sampling. Borehole sampling provides localized, statistically analyzable measurements, while window sampling estimates persistence over a defined

* Corresponding author at: State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China.
E-mail address: ltb@cdut.edu.cn (T. Li).
Peer review under the responsibility of Tongji University

https://doi.org/10.1016/j.undsp.2025.06.009

area. However, these methods suffer from sampling errors, safety hazards, subjectivity, and high labor demands (Barton et al., 1974; Franklin et al., 1988).

Recent advancements have transformed discontinuity assessment, primarily through two approaches: the direct method and the observation-conjecture method.

Direct method: Geophysical techniques and sparse drilling assess rock mass discontinuities. Electromagnetic and seismic wave-based geologic prediction methods offer non-invasive detection (Longoni et al., 2012), but issues such as ambiguous results, limited resolution, high costs, and inefficiency hinder widespread adoption. Sparse drilling, despite providing localized information, is costly and inefficient, limiting its applicability to specific geological conditions.

Observation-conjecture method: This two-step approach involves observation and conjecture.

(1) Observation. Non-contact techniques, such as image analysis, have emerged as alternatives to traditional geological sketching. Some studies use 2D images to analyze discontinuity traces, where discontinuity intersections with an outcrop surface define the structural plane trace (Slob et al., 2007; Otoo et al., 2022). Image segmentation via deep learning has been applied to extract 2D discontinuities (Chen et al., 2021a). However, this approach struggles with sharp features such as fracture intersections, and the absence of depth information leads to misjudgments. To address this, researchers have used three-dimensional (3D) surface models (Giovanni and Nicola, 2011; Li et al., 2016; Roncella et al., 2005). Active 3D reconstruction methods employ depth sensors (e.g., LiDAR, structured light, time-of-flight) to construct models but face limitations in acquisition costs and computational efficiency (Zhu et al., 2021). Multi-view stereo reconstruction, which estimates depth from multiple images, offers a simpler data acquisition process but is computationally intensive (Yao et al., 2019). These techniques have been widely studied (Riquelme et al., 2014; Giovanni and Nicola, 2011; Chen et al., 2017; Dong et al., 2020; Kong et al., 2020; Matt et al., 2009).

(2) Conjecture. Discontinuity development in rock masses follows a random pattern, yet exhibits statistical regularities governing frequency and size. The discrete fracture network (DFN) model visualizes discontinuity persistence based on observed data, using probabilistic simulation to estimate discontinuity distribution (Zhang et al., 2022; Zhu et al., 2014). However, DFN models fail to capture hidden discontinuities beyond visible tunnel faces, limiting their predictive capability in engineering applications (Wang & Cai, 2020). To overcome this, researchers (Xue et al., 2023) have applied spatial-aware recurrent neural networks to predict discontinuity distribution, but challenges remain due to reliance on manually extracted traces and complex fracture interactions around tunnel faces.

Accurate discontinuity prediction is essential for excavation safety and optimized engineering design. Various methods—advanced geologic forecasting, 2D image analysis, and probability-based models—have been developed, yet each has inherent limitations.

Advanced geologic forecasting (e.g., tunnel seismic prediction) infers subsurface discontinuities via wave reflections. While non-destructive, these methods cannot always precisely locate rock mass discontinuities, reducing reliability in complex geological conditions. Figure 1 shows the results of tunnel seismic prediction.

2D imaging techniques offer valuable surface insights but lack depth information, limiting their ability to reconstruct full 3D discontinuity distributions. Additionally, rock edges can interfere with trace extraction (Fig. 2).

Probability-based approaches statistically predict discontinuity distribution at a macroscopic scale but cannot fully leverage the spatial arrangement of current discontinuities to forecast future formations.

In this paper, we introduce a nondestructive, purely vision-based method to predict the distribution of cracks (discontinuities) in rock masses. Our method involves the following steps: (1) Multi-view stereo vision (MVS) and structure from motion (SfM) are used to construct a 3D model of the tunnel face, which allows the spatial distribution of exposed discontinuities to be observed from a 3D perspective. (2) A clustering method is used to automatically group the discontinuities, and a convex hull algorithm is used to predict real discontinuities, solving the problem in which sharp features (e.g., edges) of the unstructured rock mass surface interfere with trace extraction. (3) A spatiotemporal sequence prediction algorithm combining a self-attention mechanism and zigzag recurrent transition mechanism (ZRTM) is used to predict discontinuities in the opaque rock mass.

2 Materials and methodology

As shown in Fig. 3, the Leshan–Xichang Expressway traverses Leshan City and Liangshan Prefecture in southern Sichuan Province, China. The Bimoyuan Tunnel, part of this expressway, has a right lane spanning 3595 m and a left lane of 3569 m. The rock mass in this region is highly fractured, with well-developed joints and fissures.

In this study, we utilized the tunnel face at mileage ZK91+988 to reconstruct a 3D scene and extract a trace map. To evaluate our method, we conducted predictions using tunnel faces at ZK91+927 and ZK91+988. The ZK91+988 tunnel face covers approximately 61.3 m^2, with a width of 8.8 m and a height of 8.5 m. Images were captured using a standard smartphone, with key parameters including a resolution of 5792 px × 4344 px, a principal point at (2896 px, 2172 px), and a focal length of 4229.071 px.

This section introduces our method for 3D tunnel face modeling and discontinuity prediction. The approach comprises two key steps: (1) constructing a spatiotemporal discontinuity prediction model and (2) preparing training and validation datasets.

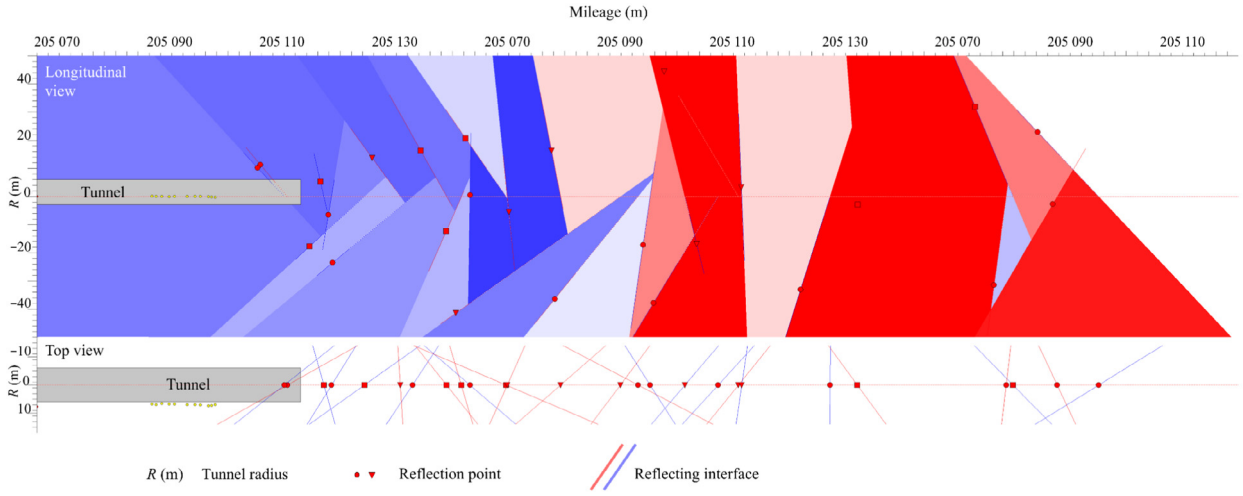


Fig. 1. Result of tunnel seismic prediction method.

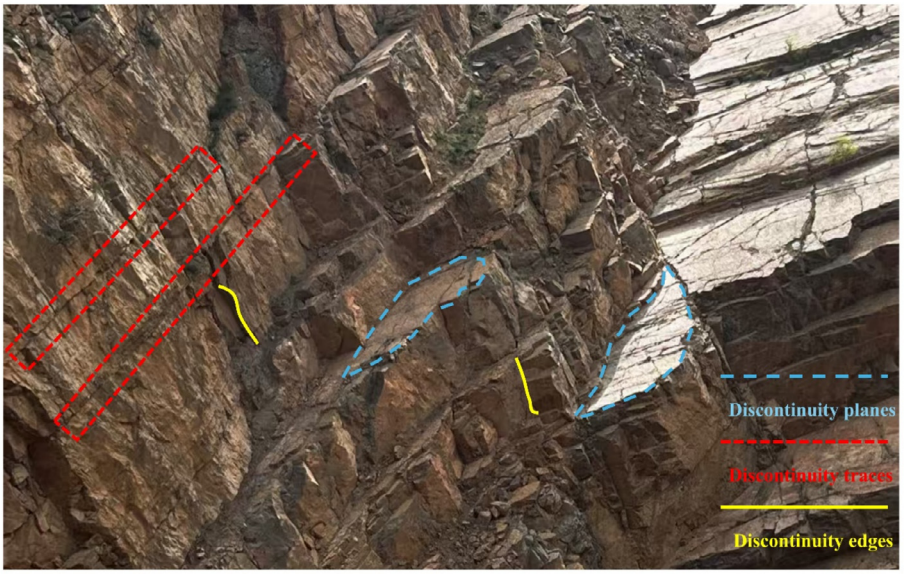


Fig. 2. Concepts of discontinuities and rock edges.

The 3D model of the tunnel face was used to identify discontinuities, which were automatically grouped based on their occurrence and projected onto a flat plane to generate 2D trace images. Due to the high cost of acquiring real 2D trace images, synthetic data were employed. The training dataset was generated by constructing a DFN model and slicing it accordingly. The workflow of the proposed method is illustrated in Fig. 4.

2.1 Three-dimensional reconstruction of the tunnel face

The first step of our method is 3D scene reconstruction. Given the tight schedules in tunnel engineering, traditional depth sensor-based correlation methods are impractical due to the direct trade-off between scanning time and point cloud accuracy. Instead, multi-view stereo (MVS) reconstruction, which generates a meshed model, is a more suitable choice under time constraints. The process involves two main steps: SfM and MVS.

SfM estimates camera parameters and reconstructs the scene structure by establishing correspondences between feature points. It determines the camera trajectory and parameters based on geometric constraints. Among the two main SfM approaches—global and incremental—incremental SfM is preferred for tunnel environments due to its higher accuracy and adaptability to complex camera distributions (Schönberger & Frahm, 2016). Therefore, we adopted an incremental SfM algorithm for scene reconstruction.

Key steps in incremental SfM include feature extraction, feature matching, initial perspective selection, triangulation, incremental reconstruction, perspective-n-point

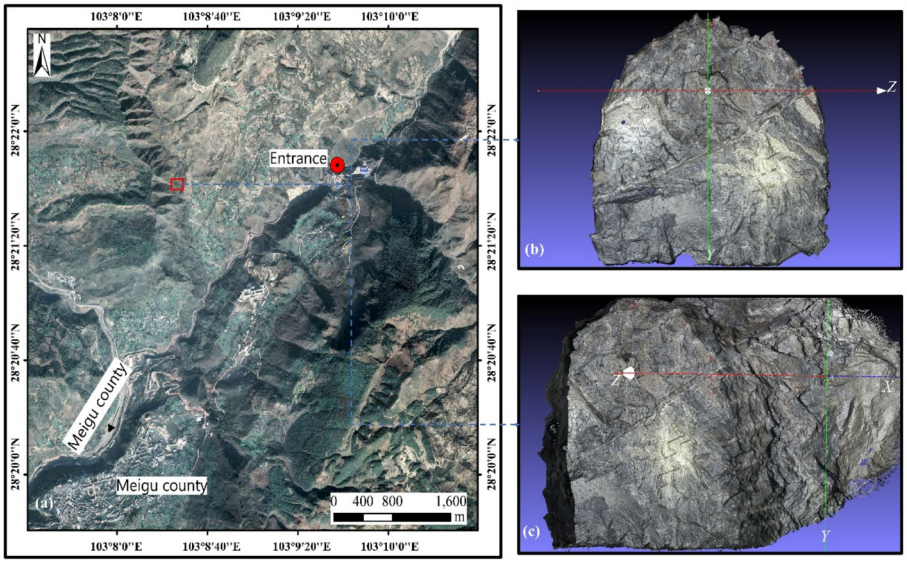


Fig. 3. (a) Site map of the Bimoyuan tunnel extracted from Google Earth, (b) the main view of the tunnel face, and (c) the side view of the tunnel face.

(PnP) estimation, and bundle adjustment optimization (Triggs et al., 2000). Advances in deep learning have further enhanced 3D reconstruction, particularly through depth space discretization, which transforms solid objects into classification problems (Okutomi & Kanade, 1993).

MVS generates dense point clouds from image sequences (Seitz et al., 2006) and serves as a post-processing step following SfM. MVS methods are categorized into four main types:

(1) Voxel-based methods (Paris et al., 2006; Pons et al., 2007) require a bounding box, limiting resolution.
(2) Deformable polygonal mesh methods (Zaharescu et al., 2007; Furukawa & Ponce, 2009) rely on good initialization, restricting their applicability.
(3) Depth map-based methods (Strecha et al., 2006; Bradley et al., 2008) have high computational complexity, making them unsuitable for large-scale scenes.
(4) Patch-based methods (Lhuillier & Quan, 2005; Furukawa & Ponce, 2010) use small patches to model surfaces, offering greater versatility.

Recent deep learning-based 3D reconstruction methods, combined with Collins' (1996) plane sweep algorithm, have shown potential for object-class reconstruction (Zhu et al., 2021). However, plane sweeping requires precise depth ranges, limiting its effectiveness in large-scale reconstructions like tunnel faces.

In the study, we adopted the PatchMatch method within the COLMAP framework, which is based on the PatchMatch algorithm proposed by Barnes et al. (2009), for depth estimation. The depth estimation pipeline includes matching cost construction, cost accumulation, depth estimation, and depth map optimization (Schönberger et al., 2016). Further details on the 3D tunnel face reconstruction method are provided in Section S1 in the supplementary material.

2.2 Validation dataset based on real trace maps obtained by 3D point cloud projection

In the second step of our method, 3D point clouds were used to generate trace maps for model validation. Since rock discontinuities result from geological factors (e.g., structural factors), similar planar features caused by tunnel construction rarely share the same orientation. By projecting the 3D point cloud onto a plane, we obtain a 2D trace map that minimizes interference from irrelevant factors such as rock mass edges, improving the accuracy of trace extraction.

Traditional image segmentation applied directly to 2D tunnel face images (Yang et al., 2024) primarily identifies linear or quasi-linear pixel regions, assuming they correspond to discontinuities. However, in a true 3D scene, discontinuities are better defined by clustering normal vectors of the point cloud, ensuring that points within the same discontinuity share similar orientations. Some linear features in 2D images may not correspond to actual 3D discontinuities. Using 3D point cloud projections ensures that extracted traces are based on geometric and structural consistency rather than purely image-based features, enhancing the reliability of discontinuity identification.

It is important to clarify: (1) The discontinuity depth information derived from 3D point data represents only plane-type discontinuities. The extraction of line-type discontinuities (traces) is detailed in our previous work (Yang et al., 2024). (2) In 3D computer vision, "depth information" refers to the distance along the axis perpen-

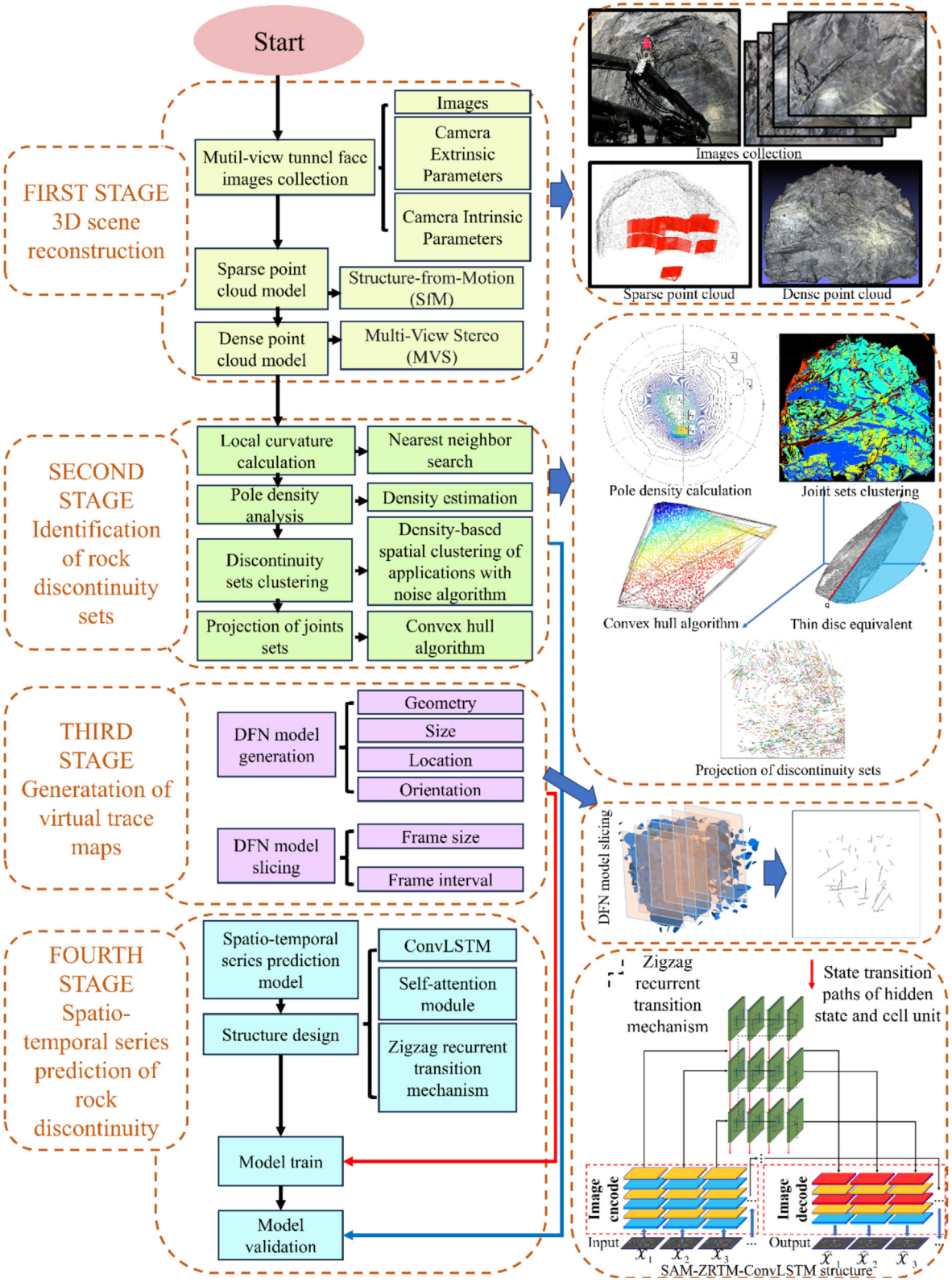


Fig. 4. Workflow of the proposed methods.

dicular to the tunnel face, with the camera as the reference point. This method captures only surface characteristics and does not indicate the extension of discontinuities into the hidden rock mass.

2.2.1 Identification of rock discontinuities from the 3D point cloud data

After reconstructing the 3D scene of the tunnel face rock mass, discontinuities were extracted from the 3D point

cloud data. Following Riquelme et al. (2014), the normal vectors of the point cloud were projected stereographically. A k-dimensional tree structure, combined with the k-nearest neighbors algorithm, was then used to identify principal normal vectors. The number of major discontinuities was determined using kernel density estimation, followed by density-based spatial clustering of applications with noise to identify spatial clusters within each discontinuity set. Finally, principal component analysis was applied to derive the mathematical expressions of the discontinuities, providing their corresponding plane equations.

2.2.2 Projection of 3D discontinuities

Accurately determining the true form of the discontinuity distribution within the rock mass is generally challenging (Zhang & Einstein, 2000). For mathematical convenience, we assumed a discontinuity to be a thin disk randomly distributed in space (Baecher et al., 1977; Kulatilake, 1993; Warburton, 1980). The discontinuity grouping results obtained based on the method in Section 2.2.1 are in the form of point clouds. The point clouds in the same group of the same plane form a polygon in space.

Referring to previous results (Chen et al., 2021b; Riquelme et al., 2018), the maximum Euclidean distance between polygon vertices was regarded as the diameter of the discontinuous disk (i.e., the equivalent trace length). The specific implementation method is as follows:

Step 1. Using a convex hull algorithm, the maximum Euclidean distance in a single point cloud cluster set can be obtained as follows:

$$\bar{l} = \max(l) = \sqrt{\left(X_{P_i} - X_{P_j}\right)^2 + \left(Y_{P_i} - Y_{P_j}\right)^2 + \left(Z_{P_i} - Z_{P_j}\right)^2}, \tag{1}$$

where $\bar{l}$ is the maximum length of all possible lengths set l between two points on the discontinuity. $P_i(X_{P_i}, Y_{P_i}, Z_{P_i})$ and $P_j(X_{P_j}, Y_{P_j}, Z_{P_j})$ are the two points on the discontinuity. It should be noted that the equivalent trace length is not the real trace length. The method made an approximation due to the inability to directly determine the persistence of the rock discontinuity. While this approach introduces some error in comparison to the actual trace length, it is similar across different tunnel faces and should provide a reliable approximation. Furthermore, this issue can be mitigated in subsequent analysis using the SAM-ZRTM-ConvLSTM model (in Sections 2.3–2.4), which captures the dynamic relationships between pixels and can help address the limitations of this approximation. That is, even if there is an error, the approximation process in the next frame will be re-approximated to replace the previous equivalent trace length, and finally reflect the persistence of the real discontinuity.

Figure 5 shows a schematic diagram of the extraction of boundary points using the convex hull algorithm.

Step 2. The midpoint position $M_{ij}(X_{P_m}, Y_{P_m}, Z_{P_m})$ (of $P_i(X_{P_i}, Y_{P_i}, Z_{P_i})$ and $P_j(X_{P_j}, Y_{P_j}, Z_{P_j})$) is recorded. The midpoint is the circle center M_{ij} of the discontinuity, which can be calculated as follows:

$$M_{ij}(X_{P_m}, Y_{P_m}, Z_{P_m}) = \left(\frac{X_{P_i} + X_{P_j}}{2}, \frac{Y_{P_i} + Y_{P_j}}{2}, \frac{Z_{P_i} + Z_{P_j}}{2}\right). \tag{2}$$

Step 3. Assuming a virtual plane perpendicular to the tunnel axis, the intersection line L is obtained using the normal vector of the virtual straight face and the normal vector of the point cloud set of the single discontinuity. The direction vector of the intersection line is $\overrightarrow{L}$.

Step 4. The point M_{ij} is projected onto the intersection line L along the gradient direction to obtain the projection point N_{ij}. Taking the projection point N_{ij} as the center, the points A and B with distances of $\frac{\bar{l}}{2}$ on the intersection line L are found. Points A, B, and N_{ij} are connected to obtain the equivalent trace AB of the single discontinuity point cloud set projected onto the virtual plane.

Steps 1–4 are repeated for different discontinuities of the same attitude of rock to obtain the equivalent trace of this group of discontinuities. Subsequently, according to different groups of discontinuities, the above process is repeated. Finally, the equivalent traces of different discontinuities of all different attitudes are obtained.

2.3 Training dataset based on virtual trace maps generated from DFN model slices

The third step of our method involves preparing the training dataset for the spatiotemporal series prediction model. Given the time constraints in tunnel construction, independently collecting a large number of tunnel face images, reconstructing 3D scenes using SfM and MVS, and extracting traces pose significant challenges. Additionally, acquiring tunnel engineering survey data is costly. In drilling and blasting operations, the footage per construction cycle (typically 3 m) is closely related to rock mass quality, with a maximum of three cycles per day allowing for multiple 3D reconstructions at different excavation stages. However, existing 3D point cloud models are insufficient for training neural networks due to limited data availability.

To address this limitation, we employed discrete fracture network (DFN) modeling, following the approach of Xue et al. (2023). DFN, a statistical modeling technique, effectively represents in situ fracture systems (Dershowitz & Einstein, 1988; Mayer et al., 2014). The DFN modeling and slicing process consists of four key steps: (1) defining statistical properties, (2) selecting a Monte Carlo simulation method, (3) generating discontinuity geometries, and (4) slicing the DFN model into trace maps (Fig. 6).

Constructing a DFN model typically requires statistical data on discontinuity orientation, size, and location, which

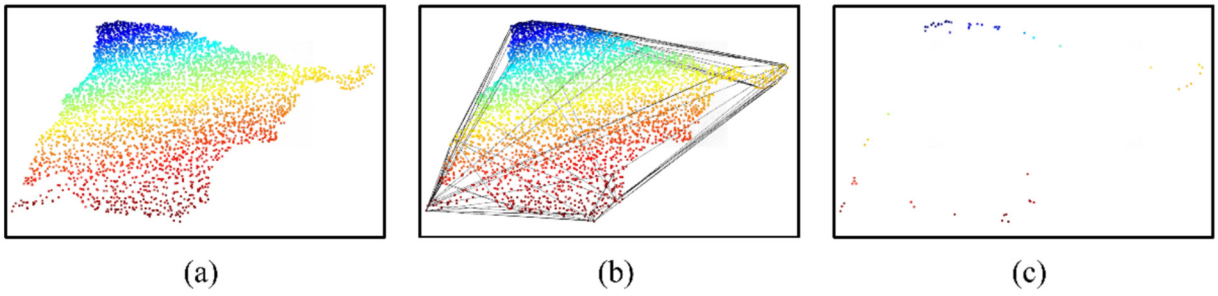


Fig. 5. Schematic diagram of using the convex hull algorithm to extract the boundary points of the point cloud plane. (a) Original point clouds, (b) triangle mesh from the convex hull algorithm, and (c) outline points.

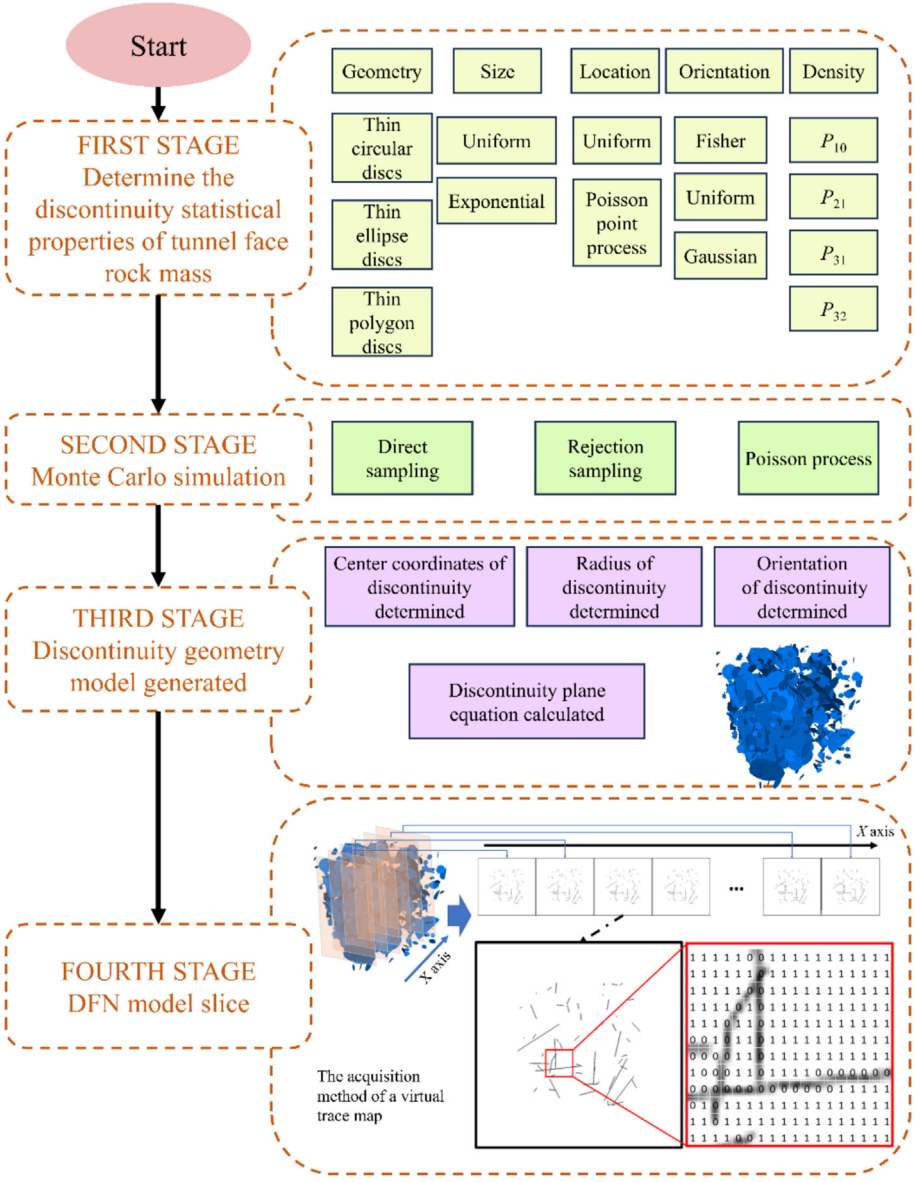


Fig. 6. A flowchart of Xue et al. (2023)'s DFN modeling and slicing.

are difficult and labor-intensive to obtain from real tunnel faces. In this study, we utilized randomly generated DFN models based on predefined probability density functions (details in Section S2 in the supplementary material). The SAM-ZRTM-ConvLSTM method (Section 2.4) primarily leverages transfer learning (Pan & Yang, 2010; Weiss

et al., 2016), training on synthetic trace maps, and predicting real tunnel face traces. By learning pixel movement patterns across sequential frames, the model forecasts future pixel disappearance without relying on probability density function constraints. Consequently, our approach does not impose strict statistical similarity between the training data (DFN-based trace maps) and the ground truth, enhancing its adaptability to real-world conditions.

2.4 Predicting the spatiotemporal evolution of rock mass discontinuities

The fourth step of our method involves constructing a model to predict the spatiotemporal evolution of discontinuities. Tunnel engineering, inherently linear in nature, involves creating underground passages with spatial continuity. As a result, the rock mass exposed during tunnel excavation also exhibits similar spatial continuity. Once the rock mass discontinuity traces of the tunnel face are obtained using the method described earlier, the evolution of discontinuities can be analyzed based on this continuity.

In this study, we employed the spatiotemporal image sequence method for analysis. This artificial intelligence-driven approach integrates image recognition with time series analysis, enabling the prediction of future discontinuities based on past unmarked spatiotemporal data. By leveraging patterns in sequential images, this method enhances the accuracy and reliability of discontinuity evolution forecasting.

The virtual trace map of discontinuities, obtained from multi-view rock mass images, was stored as multiple binary-sized matrices $\boldsymbol{X} \in \mathbb{R}^{H\times W}$, where W and H are the width and height of the image, respectively. A set of D virtual trace maps with a spatial sequence was observed as $\{\boldsymbol{X}_{t-D+1}, \cdots, \boldsymbol{X}_t\}$. The virtual trace maps to be inferred were assumed to have a corresponding serial number of $\left\{\hat{\boldsymbol{X}}_{T+1}, \hat{\boldsymbol{X}}_{T+2}, \cdots, \hat{\boldsymbol{X}}_{T+D}\right\}$, where D represents the number of sequences to be predicted. Trace map prediction required a deep learning model to learn the positional information of existing spatially continuous trace maps from data and infer the possible trace positions in subsequent excavation steps. This represents a spatiotemporal sequence prediction task (Shi et al., 2015), which can be formulated as

$$\hat{\boldsymbol{X}}_{t+1}, \cdots, \hat{\boldsymbol{X}}_{t+D} = \operatorname{argmax} p(\boldsymbol{X}_{t+1}, \cdots, \boldsymbol{X}_{t+D} | \boldsymbol{X}_{t-D+1}, \cdots, \boldsymbol{X}_t). \tag{3}$$

Overall, the basic framework adopted in this paper is the encoder-decoder architecture, which is capable of handling sequence-to-sequence prediction tasks effectively. The encoder-decoder architecture consists of three modules: the encoder, the spatio-temporal processing module, and the decoder. The encoder processes each frame of the input sequence through convolutional layers to extract spatio-temporal features. The spatio-temporal processing module processes the encoded feature sequence in a recursive manner using multiple layers of ConvLSTM. The decoder reconstructs the processed features into the output sequence through deconvolutional layers.

To achieve this spatio-temporal sequence task with a deep learning model, the first step is to solve the problem by using a sequence model. Long short-term memory (LSTM) is a type of recurrent neural network (RNN) designed to address the vanishing gradient problem, which often limits the ability of traditional RNNs to capture long-term dependencies in sequential data. LSTM overcomes this issue by introducing memory cells and gating mechanisms (input, forget, and output gates) that regulate information flow, allowing the model to retain relevant features over extended sequences (Hochreiter & Schmidhuber, 1997).

ConvLSTM is a structure that combines convolutional layers with LSTM architecture. While its fundamental structure remains similar to that of LSTM, ConvLSTM is particularly well-suited for handling spatiotemporal sequences as it retains temporal dependencies through memory cells while preserving spatial correlations. Using gate structures to regulate input, forget, and update operations, ConvLSTM effectively extracts time-series features. The integration of ConvLSTM aligns with the nature of tunnel excavation, where the spatial sequence of excavation can be modeled as a temporal sequence, making ConvLSTM a natural and effective choice for our application. By integrating ConvLSTM into our framework, we leverage the sequential excavation process of tunnel faces, which follows a spatial sequence analogous to a temporal sequence. This enables the model to effectively extract spatiotemporal features, predict the evolution of discontinuities, and improve the accuracy of trace line predictions. The mathematical formulation of ConvLSTM is as follows:

$$\begin{cases} i_t = \sigma(\boldsymbol{W}_{xi}\boldsymbol{X}_t + \boldsymbol{W}_{hi}\boldsymbol{H}_{t-1} + b_i) \\ f_t = \sigma(\boldsymbol{W}_{xf}\boldsymbol{X}_t + \boldsymbol{W}_{hf}\boldsymbol{H}_{t-1} + b_f) \\ g_t = \tanh\left(\boldsymbol{W}_{xg}\boldsymbol{X}_t + \boldsymbol{W}_{hg}\boldsymbol{H}_{t-1} + b_g\right) \\ \boldsymbol{C}_t = f_t \odot \boldsymbol{C}_{t-1} + i_t \odot g_t \\ o_t = \sigma(\boldsymbol{W}_{xo}\boldsymbol{X}_t + \boldsymbol{W}_{ho}\boldsymbol{H}_{t-1} + b_o) \\ \hat{\mathbf{H}}_t = o_t \odot \tanh(\boldsymbol{C}_t) \end{cases}, \tag{4}$$

where σ is the sigmoid activation function and tanh is the hyperbolic tangent activation function; $\odot$ indicates the Hadamard product; i_t, f_t, g_t, and o_t are the input gate, forget gate, update gate, and output gate, respectively; b_i, b_f, b_g, and b_o are the corresponding biases; $\boldsymbol{C}_t$ is the memory cell to save long-term memory; and $\boldsymbol{H}_{t-1}$ is the hidden state to save short-term memory; $\hat{\boldsymbol{H}}_t$ represents the next step hidden state; $\boldsymbol{W}_{xi}$ represents the weight corresponding to the input gate; $\boldsymbol{W}_{xf}$ represents the weight corresponding to the forget gate; $\boldsymbol{W}_{xo}$ represents the weight corresponding to the output gate; $\boldsymbol{W}_{xg}$ represents the weight corresponding to the update gate; $\boldsymbol{W}_{hi}$ represents the weight of the

hidden state in the input gate; W_{hf} represents the weight of the hidden state in the forget gate; W_{hg} represents the weight of the hidden state in the update gate; W_{ho} represents the weight of the hidden state of the output gate.

To preserve the features with long-range dependencies across spatial and temporal domains, enabling the model to focus on the key areas of the input while reducing the influence of irrelevant features, this paper introduces the self-attention module (SAM). The SAM consists of three parts: multi-head self-attention, position encoding, and gated fusion. Among them, multi-head self-attention allows the model to focus on different positions of the input in different representation subspaces; position encoding adds position information to the convolutional features to maintain the spatial structure semantics; gated fusion weights and fuses the self-attention output with the ConvLSTM output. The SAM aims to enhance the hidden state $\boldsymbol{H}$ through the self-attention mechanism and optimize the utilization of additional memory state $\boldsymbol{M}$. The SAM module in this paper designs two parallel self-attention calculations: hidden state self-attention which calculates spatial attention for the hidden state $\boldsymbol{H}_t$, enabling the model to focus on important spatial positions at the current moment; memory state self-attention which calculates spatial attention for the memory state $\boldsymbol{M}_{t-1}$, enabling the model to retrieve relevant information from historical memory. The process of updating $\boldsymbol{H}_t$ and $\boldsymbol{M}_{t-1}$ using SAM is illustrated in Fig. 7.

Hidden state self-attention mechanism:

$$\begin{cases} \boldsymbol{Q}_h = \mathrm{Conv}_{1\times1}(\boldsymbol{H}) \\ \boldsymbol{K}_h = \mathrm{Conv}_{1\times1}(\boldsymbol{H}) \\ \boldsymbol{V}_h = \mathrm{Conv}_{1\times1}(\boldsymbol{H}) \end{cases}. \tag{5}$$

Calculate the attention weights to measure the relationship between any two positions in $\boldsymbol{H}$.

$$\boldsymbol{A}_h = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}_h \boldsymbol{K}_h^{\mathrm{T}}}{\sqrt{d_k}}\right), \tag{6}$$

where $\sqrt{d_k}$ is a scaling factor.

Applying the weights $\boldsymbol{A}_h$, the enhanced hidden state output $\boldsymbol{Z}_h$ is obtained.

$$\boldsymbol{Z}_h = \boldsymbol{A}_h \boldsymbol{V}_h^{\mathrm{T}} \tag{7}$$

Cross-state attention (memory state) mechanism:

$$\begin{cases} K_m = \mathrm{Conv}_{1\times1}(\boldsymbol{M}) \\ V_m = \mathrm{Conv}_{1\times1}(\boldsymbol{M}) \end{cases}. \tag{8}$$

The cross-state attention weight $\boldsymbol{A}_m$ can be obtained as

$$\boldsymbol{A}_m = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}_h \boldsymbol{K}_m^{\mathrm{T}}}{\sqrt{d_k}}\right). \tag{9}$$

Applying the weights $\boldsymbol{A}_m$, the enhanced memory state output $\boldsymbol{Z}_m$ is obtained.

$$\boldsymbol{Z}_m = \boldsymbol{A}_m \boldsymbol{V}_m^{\mathrm{T}} \tag{10}$$

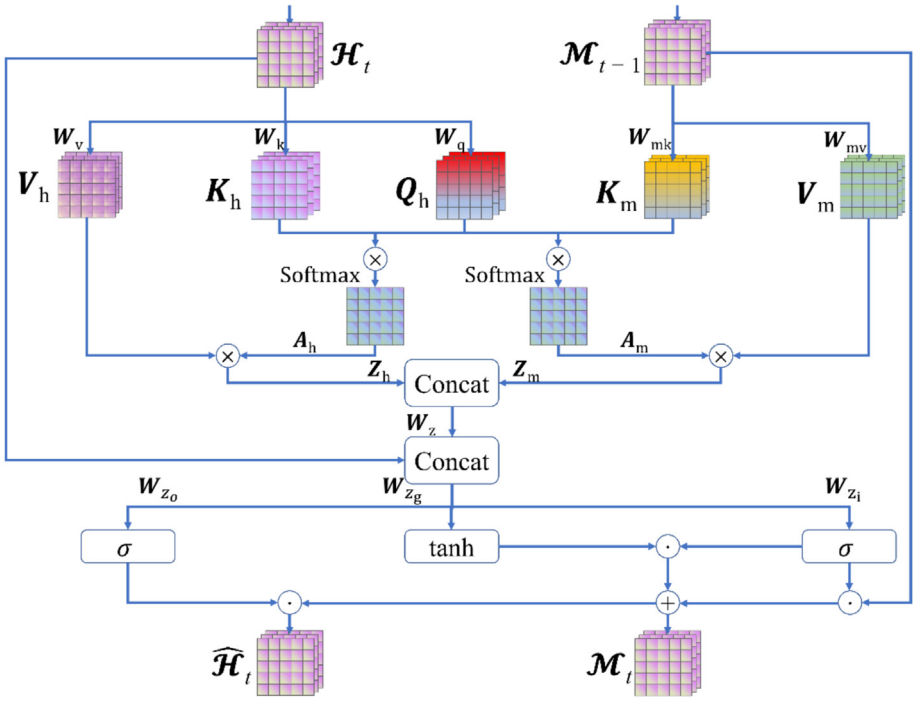


Fig. 7. Self-attention memory module to update $\boldsymbol{H}_t$ and $\boldsymbol{M}_{t-1}$.

Gated memory update: concatenates $\boldsymbol{Z}_{\mathrm{h}}$ and $\boldsymbol{Z}_{\mathrm{m}}$, and fuses them through a 1×1 convolution to obtain combined state $\boldsymbol{Z}$.

$$\boldsymbol{Z} = \mathrm{Conv}_{1\times1}(\mathrm{Concat}(\boldsymbol{Z}_{\mathrm{h}}, \boldsymbol{Z}_{\mathrm{m}}), \mathrm{dim} = 1) \tag{11}$$

In the ConvLSTM network, when iterating the frames, the states $\boldsymbol{H}_t^l$ and $\boldsymbol{C}_t^l$ carry numerous temporal features, but the delivery of the states is limited in the same level ConvLSTM module, which has the same subscript in Fig. 8(c). To take advantage of the deep and abstract features from a certain frame, a new delivery path of $\boldsymbol{M}_t$ is set (Wang et al., 2017).

The ZRTM enables the transfer of high-level features from the previous frame to lower-level processing stages in the subsequent frame, thereby expanding the inter-frame information exchange. This design allows multi-level features to propagate across frames in both

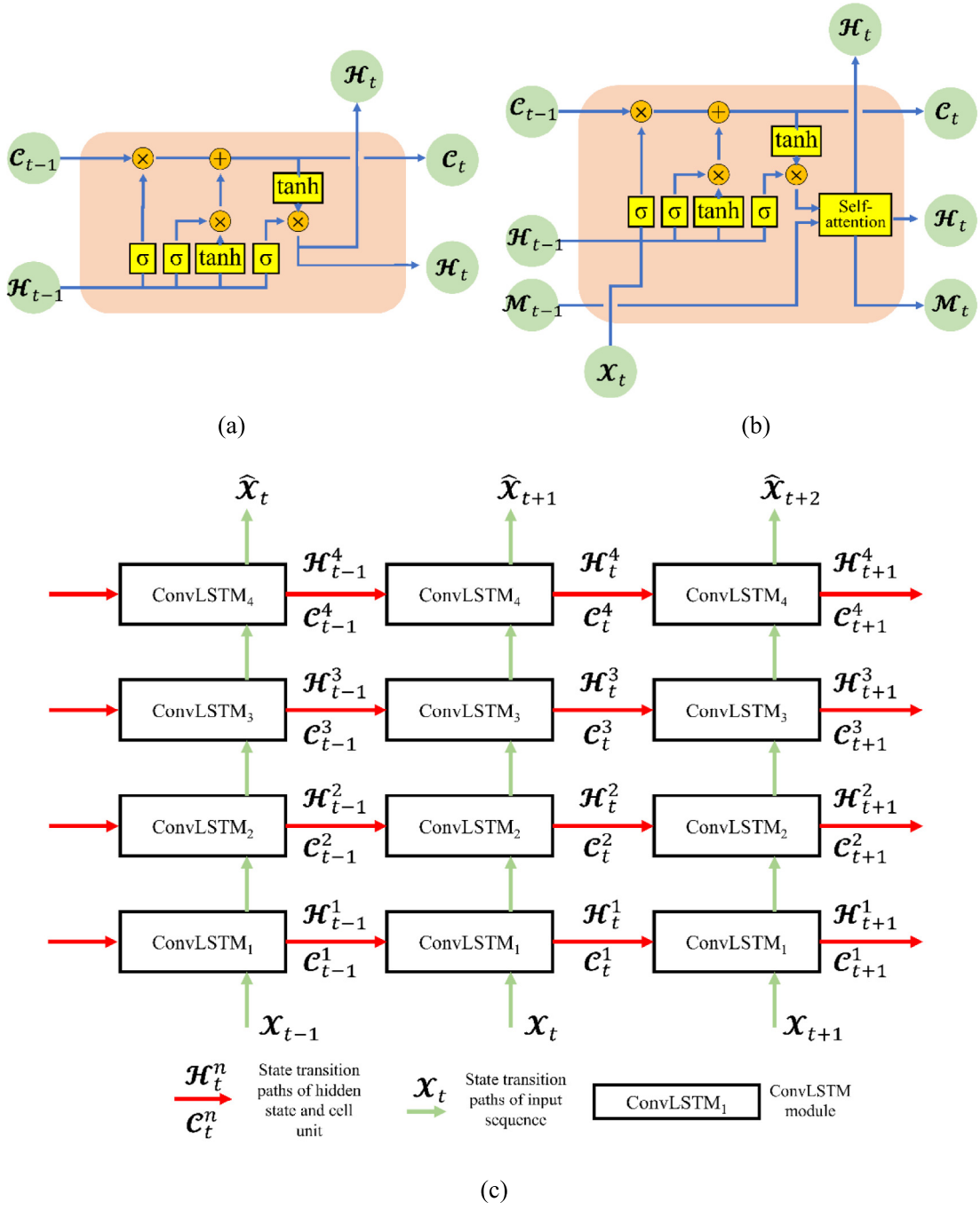
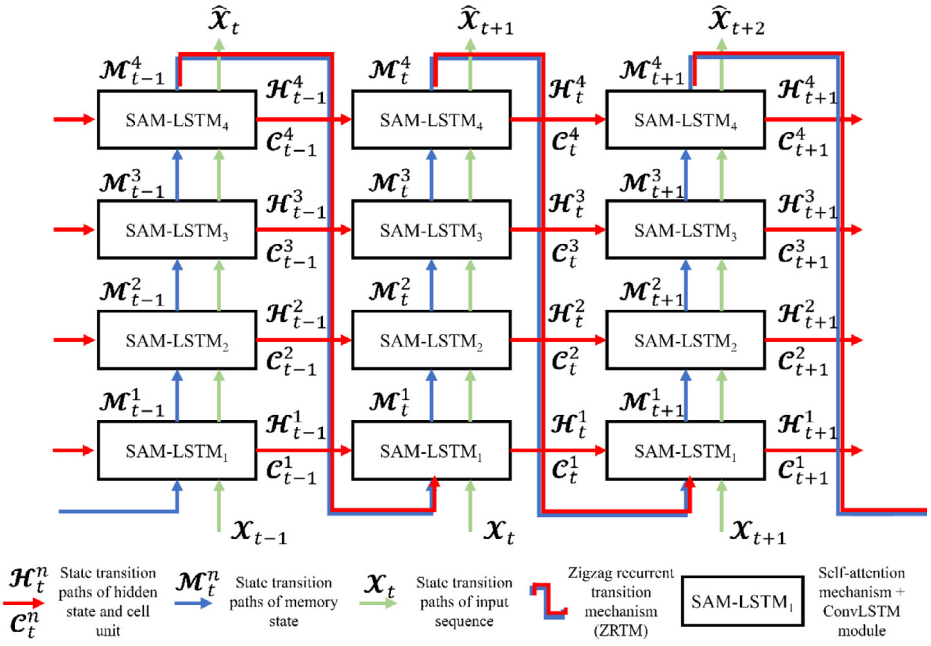


Fig. 8. (a) LSTM cell module, (b) SAM-LSTM cell module, (c) ConvLSTM architecture, and (d) SAM-ZRTM-ConvLSTM architecture.



(d)

Fig 8. *(continued)*

bottom-up and top-down directions, improving the spatiotemporal representation capability of the model. Specifically, the spatiotemporal memory state $\boldsymbol{M}_t$ is updated using a ZRTM. The first layer SAM-LSTM cell at time step t receives the spatiotemporal memory state $\boldsymbol{M}_t$ of the last layer SAM-LSTM cell of the previous time step $t-1$ and transmits it to the second layer until the last layer is reached. The SAM-LSTM cell module and the SAM-ConvLSTM architecture are shown in Fig. 8.

Finally, the memory update and output are formulated as follows:

$$\begin{cases} i'_t = \tanh\left(\boldsymbol{W}_{z_i}\boldsymbol{Z} + \boldsymbol{W}_{h_i}\boldsymbol{H}_t + b_i\right) \\ g'_t = \tanh\left(\boldsymbol{W}_{z_g}\boldsymbol{Z} + \boldsymbol{W}_{h_g}\boldsymbol{H}_t + b_g\right) \\ \boldsymbol{M}_t = \left(1 - i'_t\right) \odot \boldsymbol{M}_{t-1} + i'_t \odot g'_t \\ o'_t = \sigma(\boldsymbol{W}_{z_o} Z + \boldsymbol{W}_{h_o}\boldsymbol{H}_t + b_o) \\ \hat{\boldsymbol{H}}_t = o'_t \odot \boldsymbol{M}_t \end{cases}, \tag{12}$$

where i'_t represents the combined input gate, g'_t represents combined update gate, o'_t represents the combined output gate, $\boldsymbol{W}_{z_i}$ represents the combined weight corresponding to the input gate, $\boldsymbol{W}_{z_g}$ represents the combined weight corresponding to the update gate, $\boldsymbol{W}_{h_i}$ represents the hidden state weight corresponding to the input gate, $\boldsymbol{W}_{h_g}$ represents the hidden state weight corresponding to the update gate, $\boldsymbol{W}_{z_o}$ represents the combined weight corresponding to the output gate, $\boldsymbol{W}_{h_o}$ represents the hidden state weight corresponding to the output gate.

The final SAM-ZRTM-ConvLSTM architecture is shown in Fig. 9.

More details about the training process can be found in Section S3 in the supplementary material, and the evaluation indicators are presented in Section S4 in the supplementary material.

3 Results

3.1 Extraction of discontinuity sets

The initial point cloud contained a total of 1 980 212 data points. After classification, the processed point cloud retained 1 552 495 points, with 427 717 points remaining unassigned.

To compute the normal vectors for each point and their corresponding poles, a threshold value of 30 was set for determining the k-nearest neighbors. During the coplanarity test, a tolerance value of 0.2 was applied. Additionally, the number of bins for kernel density estimation was set as 64, while the minimum angle between the normal vectors of discontinuity sets was defined as 10°. The resulting pole density map is presented in Fig. 10. J_1–J_5 represent different types of joint sets.

To assign a unique discontinuity set to each point, a minimum angle of 30° was set between the normal vector of the discontinuity set and that of the point. Notably, all members within a given discontinuity set shared the same normal vector. Ultimately, all joint sets were obtained, as shown in Fig. 11.

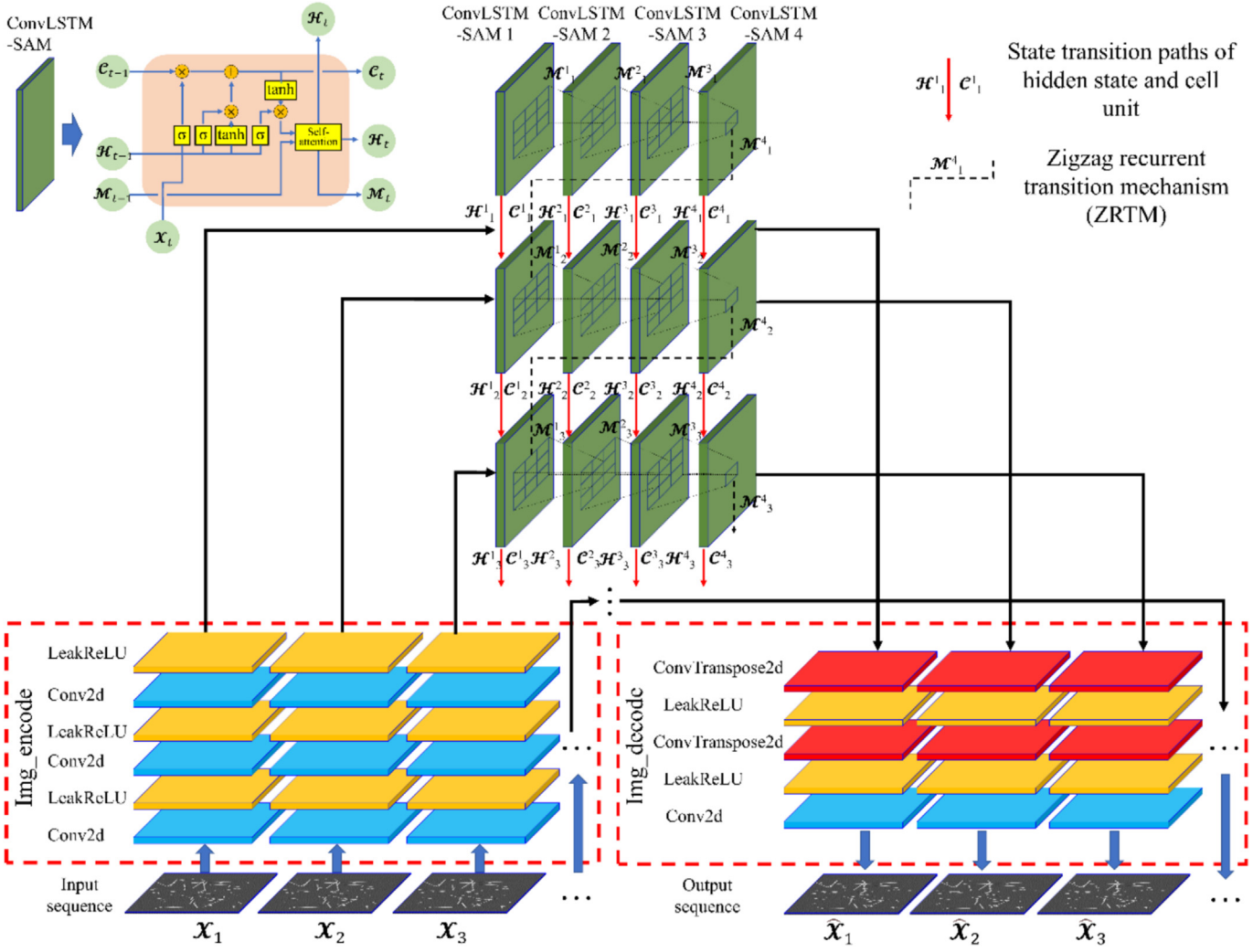


Fig. 9. Architecture of the SAM-ZRTM-ConvLSTM model.

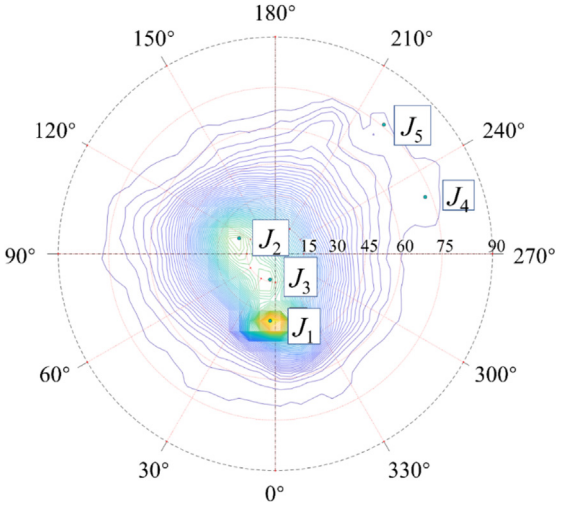


Fig. 10. Pole density map of five discontinuity sets.

The extracted discontinuity sets for the 3D scene of the studied tunnel face are listed in Table 1, in which dip orientation and dip are provided in degrees. The orientations and locations of the discontinuity sets are shown in Fig. 11(a)–(f).

3.2 Trace extraction from the 3D point cloud model projection

The trace maps of distinct point clusters were generated by projecting the different point cloud clusters onto the virtual tunnel face. The equivalent trace map was obtained by employing Steps 1–4 described in Section 2.2.2. While a one-to-one mapping relationship generally exists between the point cloud clusters and their corresponding traces, occasional slight angular deviations between the virtual and observed traces can occur. These visual discrepancies are attributed to distortions in the spatial positioning of

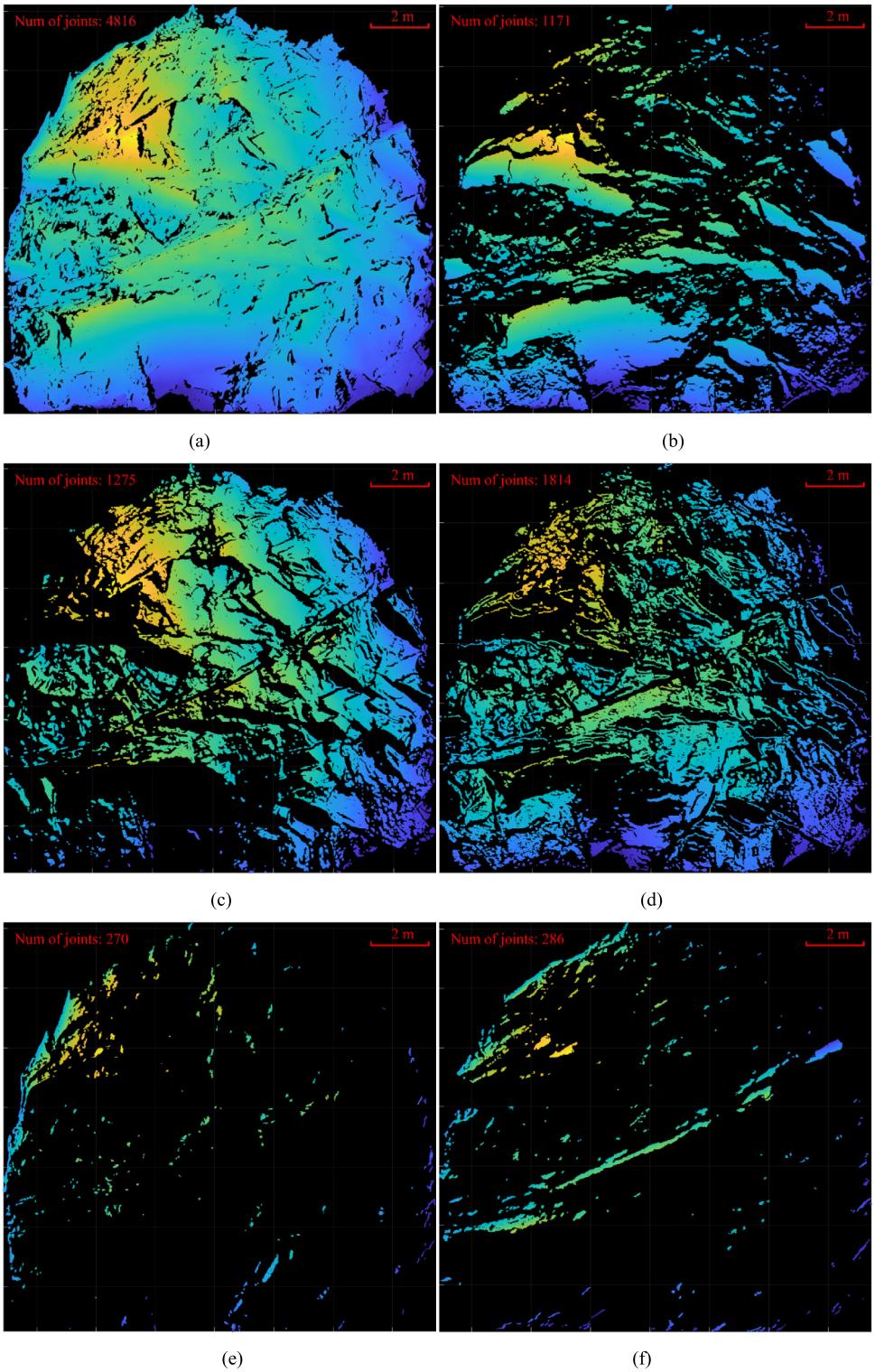


Fig. 11. (a) Point clouds of all joint sets, (b) joint set J_1 point cloud cluster, (c) joint set J_2 point cloud cluster, (d) joint set J_3 point cloud cluster, (e) joint set J_4 point cloud cluster, and (f) joint set J_5 point cloud cluster.

Table 1
Extracted discontinuity sets.

Joint set	Dip direction (°)	Dip (°)	Density	Percentage of point cloud quantity (%)
J_1	4.43	34.46	4.2771	29.17
J_2	113.19	20.56	2.8912	38.01
J_3	11.38	13.83	2.7514	26.81
J_4	249.23	72.86	0.0849	2.68
J_5	220.03	75.69	0.0613	3.33

Note: % is the number of assigned points to a discontinuity set over the total number of points.

the equivalent trajectory map, resulting from variations in normal vectors and projection directions across the point cloud clusters. All trace maps are shown in Fig. 12.

3.3 Trace map prediction

3.3.1 Model verification using DFN synthetic trace map data

To verify the performance of the method proposed in this work, we compared the results of our model with those obtained using the best spatiotemporal series prediction methods reported in recent years. Our entire training and validation process was conducted using a single NVIDIA GeForce RTX 4090 GPU with Intel i9-13900KF@3.00GHZ CPU. The models used for comparison are detailed as follows:

- ConvLSTM (Shi et al., 2015) is an encoding-forecasting structure based on ConvLSTM layers. ConvLSTM uses a fully connected LSTM (FC-LSTM) to model spatiotemporal relationships for nowcasting.
- MIM (Wang et al., 2019) is an improved memory transition function of recurrent neural networks. MIM stacks multiple MIM blocks to gradually stationarize the spatiotemporal process.
- PhyDNet (Le Guen & Thome, 2020) is a two-branch deep architecture that uses a recurrent physical cell termed PhyCell to perform PDE (partial differential equation)-constrained prediction in latent space.
- SimVP_gSTA (Gao et al., 2022) is a simple video prediction model based on CNN. This model avoids a complex set of auxiliary inputs and neural architecture designs.
- TAU (Tan et al., 2023) is an encoding–decoding structure that uses a spatial encoder and decoder to capture intra-frame features. The middle temporal module captures inter-frame correlations.
- MMVP (Zhong et al., 2023) is a two-stream video prediction framework. Motion matrices represent the temporal similarity of each pair of feature patches in the input frames and are the sole input of the MMVP motion prediction module. MMVP shows excellent performance for some public datasets.
- DMVFN (Hu et al., 2023) is a dynamic multi-scale voxel flow network based on a differentiable routing module. DMVFN enables good video prediction performance at a relatively low computational cost.
- SwinLSTM (Tang et al., 2023) is a spatiotemporal prediction network that integrates the Swin Transformer module with the simplified LSTM model. SwinLSTM shows strong performance in spatiotemporal series prediction tasks.

The model inference times are within a few seconds, with computational requirements remaining manageable. All training and validation experiments were conducted on a single NVIDIA RTX 4090 GPU, demonstrating the computational feasibility of our approach for real-world applications. We evaluated the performance of various methods on the DFN synthetic trace map dataset using MSE and SSIM metrics (see supplementary material for details on MSE and SSIM). Figure 13 illustrates the results of each method on the DFN synthetic trace maps dataset, which includes numerous complex movements. These complexities posed challenges, resulting in reduced prediction accuracy for the spatiotemporal evolution of rock discontinuities.

We trained the comparison methods without using pretrained models. MIM and MMVP performed poorly on this dataset, as these models were unable to produce accurate long-term predictions with the current data. SimVP_gSTA and TAU introduced potential Gaussian noise interference. ConvLSTM, DMVFN, and SAM-ZRTM-ConvLSTM demonstrated strong performance in modeling the spatiotemporal evolution of rock discontinuities. While ConvLSTM effectively captured most trace pixels, it also introduced some nonlinear artifacts. DMVFN achieved high prediction accuracy but failed to account for the dynamic characteristics of trace extension.

Although our proposed method performed well on the test dataset, it did not meet expectations under scenarios involving the emergence of new discontinuities. Additionally, as shown in Fig. 13, our method also introduced some non-linear traces. These limitations may be attributed to: (1) the insufficient representation of training images containing new discontinuity occurrences and corresponding ground-truth labels, and (2) the absence of linear output constraints. Future work will focus on addressing these issues to enhance the accuracy of spatiotemporal series predictions for rock discontinuities.

In addition to the visual comparison presented in Fig. 13, five supplementary experiments were conducted using different validation datasets. The corresponding

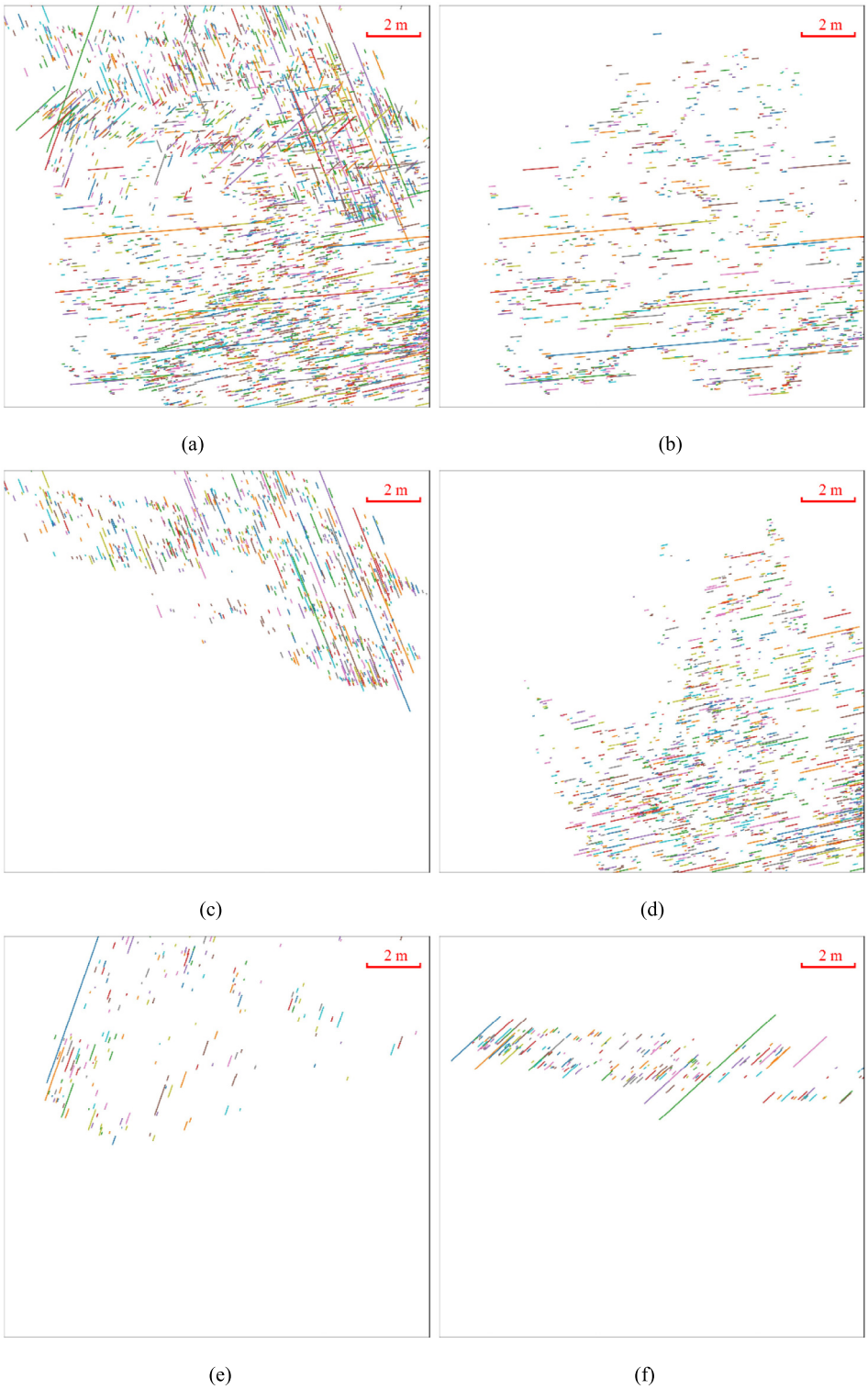


Fig. 12. Trace maps. (a) Trace map of all point cloud cluster, (b) trace map of the joint set J_1 point cloud cluster, (c) trace map of the joint set J_2 point cloud cluster, (d) trace map of the joint set J_3 point cloud cluster, (e) trace map of the joint set J_4 point cloud cluster, and (f) trace map of the joint set J_5 point cloud cluster.

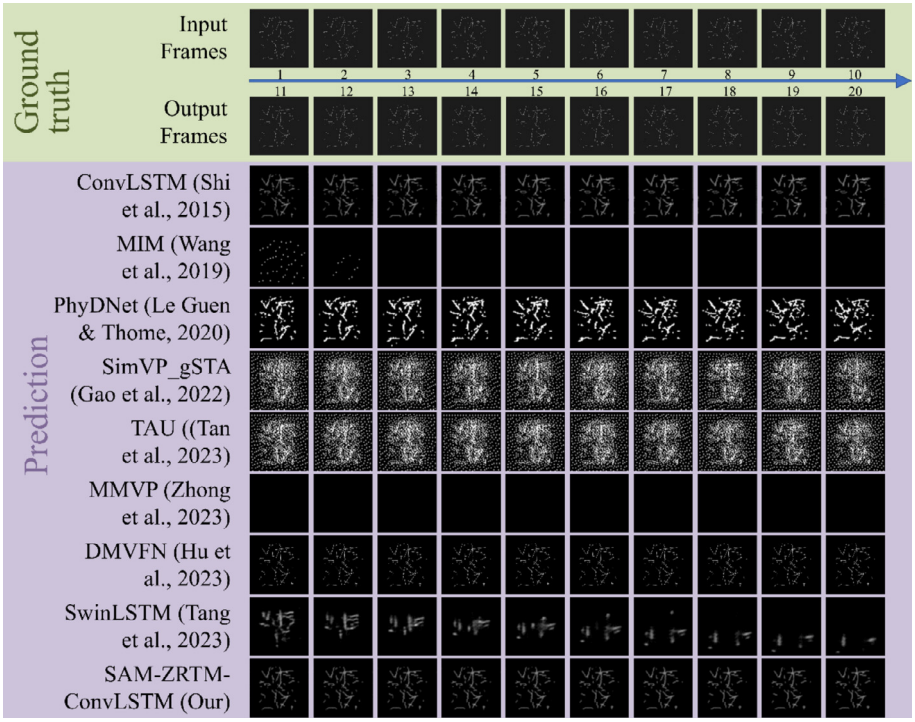


Fig. 13. Method comparison on the DFN synthetic trace maps dataset (Test 1). The other tests can be found in Section S7 in the supplementary material.

results are displayed in Figs. S12–S16 in the supplementary material.

3.3.2 Ablation study on the DFN synthetic trace map data

The advantages of the proposed method are evident in the above comparison. However, the impact of the model's internal structure on its performance remains unclear. To investigate this, we conducted ablation experiments using three different module integration strategies. For spatiotemporal prediction, we utilized a self-constructed virtual trace map dataset derived from the DFN model. To assess the effectiveness of the SAM-ConvLSTM module and the ZRTM component in the memory state $\boldsymbol{M}_t$, we performed an ablation study and visualized the results to highlight the superiority of our method.

Specifically, we designed three network structures: (1) a baseline ConvLSTM network with four stacked ConvLSTM modules; (2) a ConvLSTM network with SAM, replacing all ConvLSTM modules with SAM-ConvLSTM modules; and (3) a ConvLSTM network incorporating both SAM and ZRTM, maintaining the same module configuration. The results of the ablation study are presented in Tables 2 and 3. All parameter settings in this study are consistent with those used in the main experiment.

To illustrate the disparity between MSE and SSIM from a frame-wise perspective, we computed the MSE and SSIM values for each frame. Regardless of the model used, MSE gradually increased as the prediction range extended, whereas SSIM steadily declined (Tables 2 and 3).

The introduction of SAM reduced the MSE of the basic ConvLSTM model for predicted frames 11 to 20 by −63.13%, −20.44%, 3.77%, 12.99%, 21.38%, 29.45%, 35.78%, 41.55%, 43.92%, and 46.24%. Incorporating both SAM and ZRTM further reduced MSE for the same frames by 0.60%, 50.64%, 60.81%, 65.69%, 67.92%, 69.15%, 71.09%, 72.07%, 72.80%, and 72.32% compared to the basic ConvLSTM model. Notably, MSE increased more gradually in the SAM-ConvLSTM model than in the basic ConvLSTM model, while the SAM-ZRTM-ConvLSTM model exhibited an even slower MSE growth rate.

For SSIM, introducing SAM led to changes in the SSIM of the basic ConvLSTM model for predicted frames 11 to 20 by −0.49%, 0.82%, −0.64%, −0.51%, −0.91%, −1.23%, −1.05%, −0.75%, −0.04%, and 0.21%. Adding both SAM and ZRTM further increased SSIM for these frames by 0.13%, 17.17%, 1.25%, 3.07%, 4.83%, 6.92%, 9.13%, 11.39%, 13.52%, and 15.42% compared to the basic ConvLSTM model. While SSIM values for Frame 11 were similar across all models, they diverged significantly as the prediction range increased. The decline in SSIM was slower for the SAM-ConvLSTM model than for the basic ConvLSTM model, and even more gradual for the SAM-ZRTM-ConvLSTM model.

Overall, the MSE and SSIM results demonstrate that SAM effectively captures long-range spatial dependencies within the same frame. Furthermore, the inclusion of ZRTM enhances the extraction of spatial and temporal features from prior frames (Fig. 14).

Table 2
Ablation study of frame-wise comparison in mean square error.

Models	Frame-wise comparison in mean square error (MSE)									
	Frame 11(↓)	Frame 12(↓)	Frame 13(↓)	Frame 14(↓)	Frame 15(↓)	Frame 16(↓)	Frame 17(↓)	Frame 18(↓)	Frame 19(↓)	Frame 20(↓)
ConvLSTM	1.676	5.004	10.529	17.163	25.963	36.345	46.461	57.315	66.614	75.745
SAM-ConvLSTM	2.734	6.027	10.132	14.933	20.411	25.643	29.839	33.499	37.359	40.721
SAM-ZRTM-ConvLSTM	**1.666**	**2.470**	**4.126**	**5.888**	**8.330**	**11.211**	**13.434**	**16.007**	**18.116**	**20.963**

Table 3
Ablation study of frame-wise comparison in structural similarity index.

Models	Frame-wise comparison in structural similarity index (SSIM)									
	Frame 11(↑)	Frame 12(↑)	Frame 13(↑)	Frame 14(↑)	Frame 15(↑)	Frame 16(↑)	Frame 17(↑)	Frame 18(↑)	Frame 19(↑)	Frame 20(↑)
ConvLSTM	0.9937	0.7792	0.9788	0.9557	0.9329	0.9044	0.8769	0.8511	0.8254	0.8023
SAM-ConvLSTM	0.9888	0.7856	0.9725	0.9508	0.9244	0.8933	0.8677	0.8447	0.8251	0.8040
SAM-ZRTM-ConvLSTM	**0.995**	**0.913**	**0.991**	**0.985**	**0.978**	**0.967**	**0.957**	**0.948**	**0.937**	**0.926**

The prediction results of different models are visualized in Fig. 15. All three models demonstrated short-range spatiotemporal prediction capability. However, the SAM-ConvLSTM model produced less focused predictions, characterized by blurred trace edges, increased trace width, and lower pixel density. While the basic ConvLSTM model mitigated this issue, discrepancies remained between the predicted and ground-truth traces.

As the prediction range increased, the performance of all models declined, and the predicted traces gradually lost their linear characteristics. Notably, none of the models were able to predict newly emerging traces in the ground-truth output frame sequences. This limitation indicates that our proposed method experiences trace information loss during long-range prediction.

3.3.3 *Verification in a real tunnel engineering application*

To evaluate the effectiveness of the proposed method in predicting the spatiotemporal evolution of discontinuities, we applied it to a practical engineering project. Two working faces, located at mileage numbers ZK91+927 and ZK91+988, were selected for analysis.

In tunnel construction using the drilling and blasting method, maintaining equal spacing between exposed tunnel faces is challenging due to various influencing factors, such as geological conditions and explosion equivalent. To predict the trace map of a spatiotemporal series with the proposed method, we generated trace maps conforming to the spatiotemporal series by establishing different virtual planes near a specific tunnel face.

The acquisition of ground-truth input sequences for prediction involved the following steps: (1) reconstructing the 3D scene of the tunnel face at ZK91+927 using the SfM+MVS algorithm; (2) extracting shape outline points with the convex hull algorithm; (3) establishing parallel virtual planes; (4) obtaining the trace sequence for prediction.

Similarly, the ground-truth output sequences used for comparison and verification were acquired through: (1) reconstructing the 3D scene of the tunnel face at ZK91+988 using the SfM+MVS algorithm; (2) extracting con-

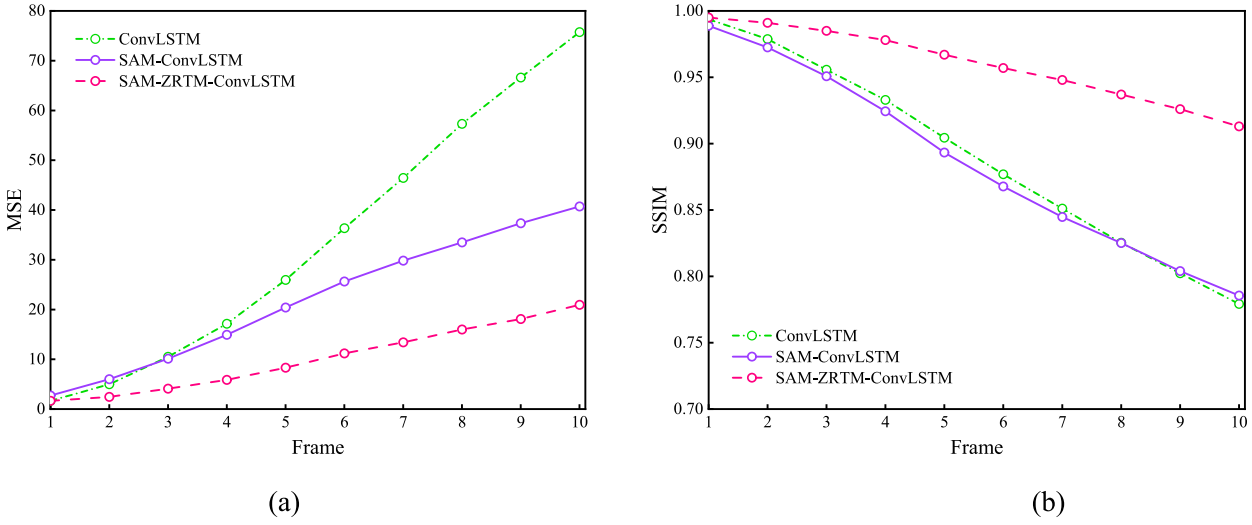


Fig. 14. Quantitative results of the ablation study. (a) MSE and (b) SSIM for ten-frame predictions.

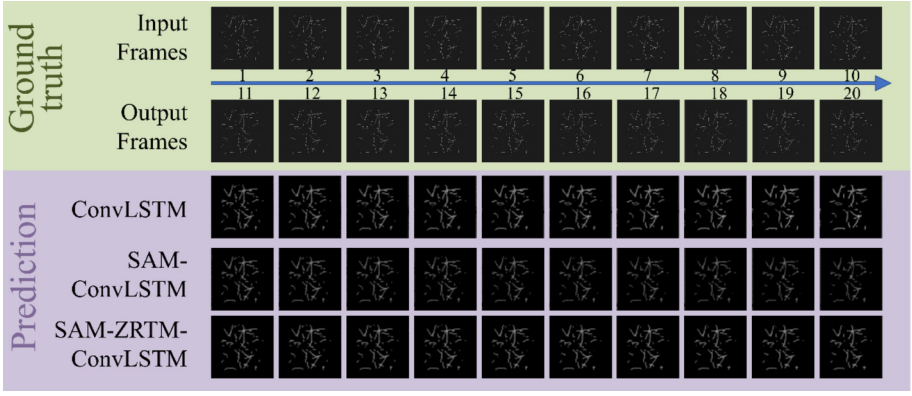


Fig. 15. Prediction examples on the DFN synthetic trace map dataset.

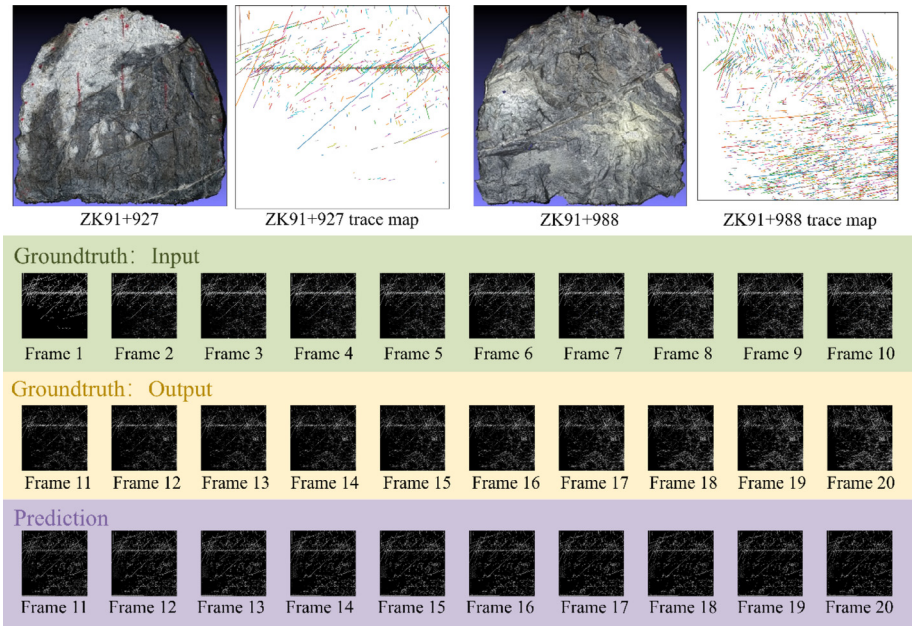


Fig. 16. Tunnel face trace map frames prediction from ZK91+988 to ZK92+127 of the Bimoyuan Tunnel.

tour points with the convex hull algorithm; (3) setting parallel virtual planes; (4) acquiring the trace map sequence for prediction.

The obtained sequence data were then fed into the proposed SAM-ZRTM-ConvLSTM model for prediction. The input sequence data and prediction results are presented in Fig. 16.

As shown in Fig. 16, a comparison between frame 20 of the "Ground Truth Output" and frame 20 of the "Prediction" demonstrates that our model accurately captured the movement and disappearance of most traces. However, its performance in predicting the appearance of new traces was limited, and predictions for frames with long prediction distances (i.e., the gap between the last input frame and the current output frame) deviated from the actual results.

The evaluation metrics, MSE and SSIM (Fig. 17), further highlight this issue. Consequently, while the proposed method is well-suited for short-range trace prediction, further improvements are necessary to enhance its performance in long-range prediction.

4 Discussion

In this study, we developed a nondestructive, purely visual method to predict crack (discontinuity) distribution in unexplored rock masses using a 3D model constructed during prior tunnel excavation. Our contributions are threefold: (1) integrating a 3D model with a 2D trace extraction technique to effectively suppress non-structural rock mass interference, (2) demonstrating the method's effectiveness, universality, robustness, and stability through

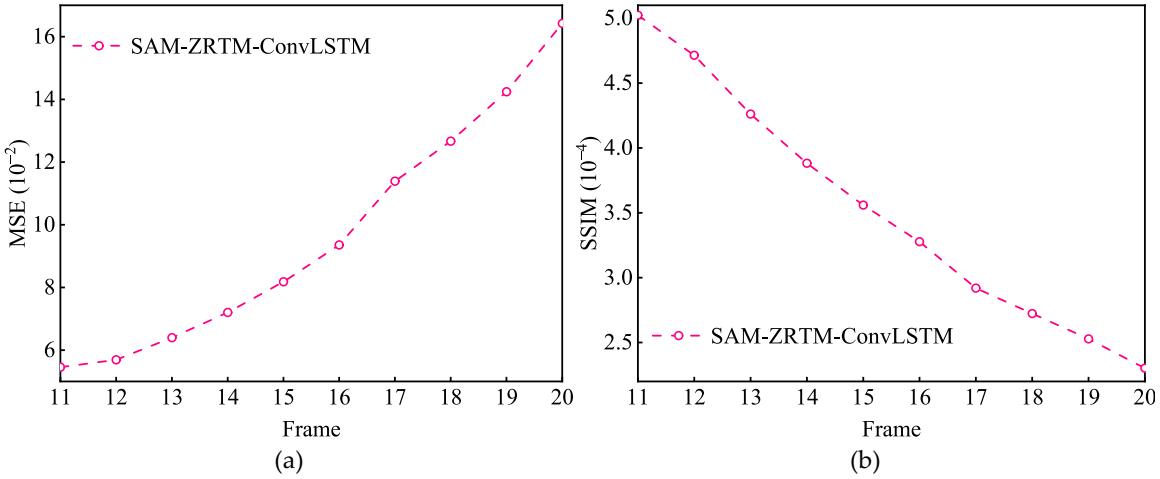


Fig. 17. Quantitative performance of the SAM-ZRTM-ConvLSTM model in a real-world tunnel engineering application. (a) MSE and (b) SSIM for ten predicted frames.

extensive evaluations on synthetic trace maps and real engineering scenes, and (3) investigating the impact of training image sizes and spatial intervals, offering valuable guidance for optimizing rock mass structure identification.

Our results highlight the significant influence of sequential data sizes and intervals on the predictive capabilities of the SAM-ZRTM-ConvLSTM model. Larger training images enhance trace prediction precision but reduce intricacy in predicted maps. Additionally, extending the prediction distance decreases both accuracy and precision, underscoring the need to balance predictive performance with computational constraints. Thus, selecting an appropriate training image interval and size based on available resources is crucial, as elaborated in Sections S5 and S6 in the supplementary material.

The practical implications of our method are substantial. By addressing key challenges in discontinuity prediction, our approach enhances model adaptability across diverse geological conditions, making it well-suited for real-world applications. Geotechnical engineers and tunnel construction professionals can leverage this method by capturing tunnel face images with a standard smartphone during excavation. These images are then automatically processed to generate continuous predictions of discontinuities, improving rock stability analysis and enabling early identification of geological hazards.

Moreover, accurate discontinuity prediction ahead of the tunnel face facilitates numerical simulations to assess tunnel stability. Engineers can pre-identify instability-prone zones and implement targeted support measures—such as pre-grouting or anchor bolt design—before instability occurs. This capability is particularly valuable for modern tunnel construction, where continuous monitoring and adaptive support design are critical for safety and performance. The system's intuitive design ensures usability even for personnel with minimal geological expertise, integrating seamlessly into routine site inspections and supporting proactive decision-making.

4.1 Limitations

Despite the promising results, several limitations remain. First, the proposed method is inherently a short-range prediction approach. It can infer the distribution of unexposed discontinuities based on existing patterns but cannot predict the emergence of entirely new structural planes due to the limited representation of geological variability in the training data. Second, the prediction accuracy of trace maps is highly dependent on the reliability of the DFN slice modeling. Any inaccuracies in the DFN input may propagate through the predictive pipeline. Third, the current DFN framework does not incorporate line-type discontinuities, which limits the completeness of the discontinuity representation. Fourth, the system still requires manual intervention during several stages, such as 3D point cloud reconstruction, trace extraction, and projection, which hinders its scalability in practical applications. Fifth, complex geological conditions—such as soft or weathered rock masses—reduce the visibility of discontinuities, while groundwater may introduce noise that degrades prediction quality. Finally, the impact of disturbed joints caused by excavation activities (e.g., blasting or mechanical clearing) has not yet been addressed in the current training framework, potentially affecting the model's generalization to real-world tunnel faces.

4.2 Future work

To overcome these limitations and enhance the model's applicability, future research should focus on the following directions:

(1) Extending prediction range with geological knowledge. Integrate principles from fracture mechanics and geological continuity to improve the model's ability to predict newly forming or previously unseen discontinuities. Refine the DFN generation process by incorporating site-specific geological constraints and validating outputs against real fracture network observations, thereby improving input data reliability. This will also help address the impact of variability in rock mass structural planes.
(2) Incorporating line-type discontinuities. Advance 2D line extraction techniques and integrate these features into the 3D DFN framework to achieve a more comprehensive structural representation.
(3) Addressing disturbed joints. Investigate the influence of secondary joints introduced by blasting or excavation tools. This includes identifying and filtering highly random or inconsistent joint traces—especially those differing significantly in orientation or appearing intermittently across sequential images—to improve model generalization.
(4) Developing a fully automated system. Build an end-to-end framework that automates 3D reconstruction, discontinuity detection, trace projection, and sequential prediction from multi-view images, reducing the need for manual processing.
(5) Enhancing robustness under complex conditions. Apply multi-scale fusion and advanced feature filtering mechanisms to better detect weakly visible features in soft or weathered rock and to mitigate the interference caused by groundwater.

In conclusion, our method offers a novel and practical approach for predicting rock mass discontinuities, yet further refinements are required to extend its predictive range and automate processing. Addressing these challenges will enhance predictive performance and operational efficiency, benefiting geotechnical engineering and tunnel construction projects.

5 Conclusions

In summary, authors introduced a novel, nondestructive, and purely visual method for predicting the distribution of discontinuities in unexplored rock masses. By integrating a 3D model with 2D trace extraction techniques, our approach effectively minimizes interference from non-structural features, enabling precise and reliable predictions during tunnel excavation. Extensive experimental evaluations, incorporating both synthetic trace maps and real-world engineering data, validate the method's robustness, universality, and stability. Furthermore, our analysis of training image size and spatial interval effects provides valuable insights for optimizing predictive performance, ensuring adaptability to diverse geological conditions.

Despite these promising outcomes, several limitations remain. The current model has a relatively short prediction distance and lacks the ability to anticipate the emergence of entirely new discontinuities. Additionally, the system still requires manual intervention due to its limited automation capabilities. Other challenges include accurately extracting discontinuities in soft or weathered rock masses and mitigating the effects of groundwater interference, both of which can compromise prediction accuracy. Addressing these challenges will require the development of a fully automated framework, the integration of fracture mechanics principles, the implementation of multi-scale fusion techniques, and the enhancement of feature filtering mechanisms to improve robustness and reliability.

Overall, our method offers a practical, accessible, and user-friendly solution for geotechnical engineers and tunnel construction professionals, with significant potential to improve rock mass stability analysis and early hazard detection. Future research aimed at overcoming the existing limitations is expected to further enhance the model's predictive capabilities and expand its applicability in real-world engineering projects, ultimately contributing to safer and more efficient tunnel excavation processes.

Data availability

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

CRediT authorship contribution statement

Gang Yang: Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tianbin Li:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

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

Acknowledgement

This research was financially supported by the National Natural Science Foundation of China (Grant Nos. 42130719 and U19A20111) and the Sichuan Science and Technology Project (Grant No. 2021YFS0317).

We sincerely appreciate the invaluable support provided by the journal editors and the three anonymous reviewers. Their insightful comments and constructive suggestions have significantly improved the quality of this manuscript. We are truly grateful for their time, effort, and dedication.

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.undsp.2025.06.009.

References

Baecher, G. B., Lanney, N. A., & Einstein, H. H. (1977). Statistical description of rock properties and sampling. *Proceedings of the 18th U. S. Symposium on Rock Mechanics (USRMS)*. .

Barnes, C., Shechtman, E., Finkelstein, A., & Goldman, B. D. (2009). PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics, 28*(3), 1–11.

Barton, N., Lien, R., & Lunde, J. (1974). Engineering classification of rock masses for the design of tunnel support. *Rock Mechanics, 6*(4), 189–236.

Bradley, D., Boubekeur, T., & Heidrich, W. (2008). Accurate multi-view reconstruction using robust binocular stereo and surface meshing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8). Anchorage, AK, USA.

Chen, J., Zhou, M., Huang, H., Zhang, D., & Peng, Z. (2021a). Automated extraction and evaluation of fracture trace maps from rock tunnel face images via deep learning. *International Journal of Rock Mechanics and Mining Sciences, 142*, 104745.

Chen, N., Cai, X., Li, S., Zhang, X., & Jiang, Q. (2021b). Automatic extraction of rock mass discontinuity based on 3D laser scanning. *Quarterly Journal of Engineering Geology and Hydrogeology, 54*(1), qjegh2020-054.

Chen, N., Kemeny, J., Jiang, Q., & Pan, Z. (2017). Automatic extraction of blocks from 3D point clouds of fractured rock. *Computers & Geosciences, 109*, 149–161.

Collins, R. T. (1996). A space-sweep approach to true multi-image matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 358–363). San Francisco, CA, USA.

Dershowitz, W. S., & Einstein, H. H. (1988). Characterizing rock joint geometry with joint system models. *Rock Mechanics and Rock Engineering, 21*(1), 21–51.

Dong, X., Xu, Q., Huang, R., Liu, Q., & Kieffer, D. S. (2020). Reconstruction of surficial rock blocks by means of rock structure modelling of 3D TLS point clouds: The 2013 Long-Chang rockfall. *Rock Mechanics and Rock Engineering, 53*(2), 671–689.

Franklin, J. A., Maerz, N. H., & Bennett, C. P. (1988). Rock mass characterization using photoanalysis. *International Journal of Mining and Geological Engineering, 6*(2), 97–112.

Furukawa, Y., & Ponce, J. (2010). Accurate, dense, and robust multi-view stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(8), 1362–1376.

Furukawa, Y., & Ponce, J. (2009). Carved visual hulls for image-based modeling. *International Journal of Computer Vision, 81*(1), 53–67.

Gao, Z., Tan, C., Wu, L., et al. (2022). SimVP: Simpler yet better video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 3160–3170). New Orleans, LA, USA.

Giovanni, G., & Nicola, C. (2011). Semi-automatic extraction of rock mass structural data from high-resolution LiDAR point clouds. *International Journal of Rock Mechanics and Mining Sciences, 48*(2), 187–198.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hu, X., Huang, Z., Huang, A., Xu, J., & Zhou, S. (2023). A dynamic multi-scale voxel flow network for video prediction. *arXiv preprint arXiv:2303.09875*.

Ju, Y., Wan, C., Fu, G., Ren, Z., Mao, L., & Chiang, F. P. (2022). Effects of near-fault stress evolution and surface asperities on rough fault slip: An evaluation based on photoelastic shear tests and additively printed models. *Engineering Geology, 305*, 106715.

Kong, D., Wu, F., & Saroglou, C. (2020). Automatic identification and characterization of discontinuities in rock masses from 3D point clouds. *Engineering Geology, 265*, 105442.

Kulatilake, P. (1993). Application of probability and statistics in joint network modeling in three dimensions. *Probabilistic Methods in Geotechnical Engineering*. .

Kulatilake, P., Wang, L., Tang, H., & Liang, Y. (2011). Evaluation of rock slope stability for Yujian River dam site by kinematic and block theory analyses. *Computers and Geotechnics, 38*(6), 846–860.

Le Guen, V., & Thome, N. (2020). Disentangling physical dynamics from unknown factors for unsupervised video prediction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 11471–11481). Seattle, WA, USA.

Lhuillier, M., & Quan, L. (2005). A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27*(3), 418–433.

Li, X., Chen, J., & Zhu, H. (2016). A new method for automated discontinuity trace mapping on rock mass 3D surface model. *Computers & Geosciences, 89*, 118–131.

Li, Y., Tang, H., Zhu, F., & Hu, Z. (2023). An anisotropic elastoplastic strong discontinuity model for shear failure in anisotropic rock masses. *Computers and Geotechnics, 176*, 106762.

Longoni, L., Arosio, D., Scaioni, M., Papini, M., Roncella, R., & Brambilla, D. (2012). Surface and subsurface non-invasive investigations to improve the characterization of a fractured rock mass. *Journal of Geophysics and Engineering, 9*(5), 461–472.

Matt, L., Mark, S., Diederichs, M., & Harrap, R. (2009). Optimization of LiDAR scanning and processing for automated structural evaluation of discontinuities in rock masses. *International Journal of Rock Mechanics and Mining Sciences, 46*(1), 194–199.

Mayer, J., Hamdi, P., & Stead, D. (2014). A modified discrete fracture network approach for geomechanical simulation. In *Proceedings of the International Conference on Discrete Fracture Network Engineering*. (pp. 12). Vancouver, Canada.

Okutomi, M., & Kanade, T. (1993). A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 15*(4), 353–363.

Otoo, J., Norbert, M., & Duan, Y. (2022). LiDAR and optical imaging for 3D fracture orientations. In *Proceedings of the 2011 NSF Engineering Research and Innovation Conference*. Atlanta, GA, USA.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering, 22*(10), 1345–1359.

Paris, S., Sillion, F. X., & Quan, L. (2006). A surface reconstruction method using global graph cut optimization. *International Journal of Computer Vision, 66*(2), 141–161.

Pons, J. P., Keriven, R., & Faugeras, O. (2007). Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. *International Journal of Computer Vision, 72*(2), 179–193.

Priest, S. D. (1985). Hemispherical projection methods in rock mechanics. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 22*(5), 124.

Riquelme, A., Abellán, A., Tomás, R., & Jaboyedoff, M. (2014). A new approach for semi-automatic rock mass joints recognition from 3D point clouds. *Computers & Geosciences, 68*, 38–52.

Riquelme, A., Tomás, R., Cano, M., Pastor, J., & Abellán, A. (2018). Automatic mapping of discontinuity persistence on rock masses using 3D point clouds. *Rock Mechanics and Rock Engineering, 51*(10), 3005–3028.

Roncella, R., Forlani, G., & Remondino, F. (2005). Photogrammetry for geological applications: Automatic retrieval of discontinuity orientation in rock slopes. In *In Proceedings of SPIE – The International Society for Optical Engineering* (pp. 5665).

Schönberger, J. L., Zheng, E., Frahm, J. M., & Pollefeys, M. (2016). Pixelwise view selection for unstructured multi-view stereo. In *European conference on computer vision* (pp. 501–518). Amsterdam, NL.

Schönberger, J. L., & Frahm, J. M. (2016). Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4104–4113). Las Vegas, NV, USA.

Seitz, S., Curless, B., Diebel, J., Scharstein, D., & Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 519–528). New York, NY, USA.

Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems, 28*, 802–810.

Slob, S., Hack, H., Feng, Q., Röshoff, K., & Turner, A.K. (2007). Fracture mapping using 3D laser scanning techniques. In *Proceedings of the 11th Congress of the International Society for Rock Mechanics*. (pp. 299–302). Lisbon, PT.

Strecha, C., Fransens, R., & Van Gool, L. (2006). Combined depth and outlier estimation in multi-view stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2394–2401). New York, NY, USA.

Tan, C., Gao, Z., Wu, L., Xu, Y., Xia, J., Li, S., & Li, S.Z. (2023). Temporal attention unit: Towards efficient spatiotemporal predictive learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 18770–18782). Vancouver, Canada.

Tang, S., Li, C., Zhang, P., & Tang, R. (2023). SwinLSTM: Improving spatiotemporal prediction accuracy using Swin Transformer and LSTM. In *IEEE/CVF International Conference on Computer Vision*. (pp. 13470–13479). Paris, France.

Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, W. A. (2000). Bundle adjustment: A modern synthesis. *In Lecture Notes in Computer Science, 1883*(1), 153–177.

Ulusay, R. (2015). *The ISRM suggested methods for rock characterization, testing and monitoring: 2007–2014*. Springer.

Wang, Y., Long, M., Wang, J., Gao, Z., & Yu, P. S. (2017). PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. In *Advances in neural information processing systems*. (pp. 879–888). Long Beach, CA, USA.

Wang, Y., Zhang, J., & Zhu, H. (2019). Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9146–9154). Long Beach, CA, USA.

Wang, X., & Cai, M. (2020). A DFN–DEM multi-scale modeling approach for simulating tunnel excavation response in jointed rock masses. *Rock Mechanics and Rock Engineering, 53*(3), 1053–1077.

Warburton, P. M. (1980). A stereological interpretation of joint trace data. *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 17*(4), 181–190.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data, 3*(1), 9.

Xue, Y., Cao, Y., Zhou, M., Zhang, F., Shen, K., & Jia, F. (2023). Rock mass fracture maps prediction based on spatiotemporal image sequence modeling. *Computer-Aided Civil and Infrastructure Engineering, 38*(4), 470–488.

Yan, J., Chen, J., Zhang, Y., Liu, Y., Zhao, X., Xue, J., Zhu, C., Mehmood, Q., & Wang, Q. (2023). Semi-automatic extraction of dangerous rock blocks from jointed rock exposures based on a discontinuity trace map. *Computers and Geotechnics, 156*, 105265.

Yang, G., Li, T. B., Tang, H., Xing, D., Hu, Y., & Li, S. (2024). Evaluation method for the intactness of the tunnel face surrounding rock based on tunnel face images. *Quarterly Journal of Engineering Geology and Hydrogeology, 57*(3), qjegh2024-018.

Yao, Y., Luo, Z., Li, S., et al. (2019). Recurrent MVSNet for high-resolution multi-view stereo depth inference. In *Proceedings of the*

IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 5520–5529). Long Beach, CA, USA.

Zaharescu, A., Boyer, E., & Horaud, R. (2007). TransforMesh: A topology-adaptive mesh-based approach to surface evolution. In *Asian Conference on Computer Vision* (pp. 166–175). China: Hong Kong.

Zhan, J., Chen, J., Xu, P., Han, X., Chen, Y., Ruan, Y., & Zhou, X. (2017). Computational framework for obtaining volumetric fracture intensity from 3D fracture network models using Delaunay triangulations. *Computers and Geotechnics, 89*, 179–194.

Zhang, L., & Einstein, H. H. (2000). Estimating the intensity of rock discontinuities. *International Journal of Rock Mechanics and Mining Sciences, 37*(5), 819–837.

Zhang, Q., Wang, X., Zhu, H., Zhang, K., & Li, X. (2022). Mixture distribution model for three-dimensional geometric attributes of multiple discontinuity sets based on trace data of rock mass. *Engineering Geology, 311*, 106915.

Zhong, Y., Liang, L., & Zharkov, I. (2023). MMVP: Motion-matrix-based video prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4250–4260). Vancouver, Canada.

Zhu, H., Zuo, Y., Li, X., Deng, J., & Zhuang, X. (2014). Estimation of the fracture diameter distributions using the maximum entropy principle. *International Journal of Rock Mechanics and Mining Sciences, 72*, 127–137.

Zhu, Q., Min, C., Wei, Z., Wei, Z., Chen, Y., & Wang, G. (2021). Deep learning for multi-view stereo via plane sweep: A survey. *arXiv preprint arXiv:2106.15328*.