International Journal of Mining Science and Technology 35 (2025) 1695–1711

Contents lists available at ScienceDirect

International Journal of Mining Science and Technology

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

An interactive framework integrating segment anything model and structure-from-motion for three-dimensional discontinuity identification in rock masses

Jiawei Wang [a,b], Jun Zheng [a,c,*], Jie Hu [a], Xiaojin Gong [d], Qing Lü [a], Ju Han [e], Jialiang Sun [f]

[a] Department of Civil Engineering, Zhejiang University, Hangzhou 310058, China
[b] Center for Balance Architecture, Zhejiang University, Hangzhou 310007, China
[c] The Architectural Design & Research Institute of Zhejiang University Co., Ltd., Hangzhou 310063, China
[d] College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
[e] China Construction First Group Construction & Development Co., Ltd., Beijing 100102, China
[f] Department of Land Management, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 10 June 2025
Received in revised form 30 August 2025
Accepted 3 September 2025
Available online 28 October 2025

Keywords:
Rock Mass
Discontinuity
Digital outcrop model (DOM)
Point clouds
Large-scale model (LSM)
Foundation model (FM)

ABSTRACT

The identification of rock mass discontinuities is critical for rock mass characterization. While high-resolution digital outcrop models (DOMs) are widely used, current digital methods struggle to generalize across diverse geological settings. Large-scale models (LSMs), with vast parameter spaces and extensive training datasets, excel in solving complex visual problems. This study explores the potential of using one such LSM, Segment anything model (SAM), to identify facet-type discontinuities across several outcrops via interactive prompting. The findings demonstrate that SAM effectively segments two-dimensional (2D) discontinuities, with its generalization capability validated on a dataset of 2426 identified discontinuities across 170 outcrops. The model achieves 0.78 mean IoU and 0.86 average precision using 11-point prompts. To extend to three dimensions (3D), a framework integrating SAM with Structure-from-Motion (SfM) was proposed. By utilizing the inherent but often overlooked relationship between image pixels and point clouds in SfM, the identification process was simplified and generalized across photogrammetric devices. Benchmark studies showed that the framework achieved 0.91 average precision, identifying 87 discontinuities in Dataset-3D. The results confirm its high precision and efficiency, making it a valuable tool for data annotation. The proposed method offers a practical solution for geological investigations.

1. Introduction

A rock mass is a complex fractured system and consists essentially of two constituents: discontinuities, i.e., fractures, fissures, joints, faults, bedding planes and micro-fissures, and intact rock [1]. Discontinuities play a significant role in the mechanical properties of rock masses, since they define the weak planes in a rock mass along which the rock blocks detach and fail [2]. Therefore, the characterization of discontinuities in rock exposures is an important step required to collect input information for further rock mechanics analysis and rock engineering design.

To characterize discontinuities visible at the surface of rock outcrops, a typical set of parameters often suggested for measurement includes orientation, spacing, persistence, and roughness [3–5]. This decisive information on rock discontinuities helps to predict the flow of groundwater, assesses rock mass quality, and provides a better stability condition for rock engineering [6,7]. Traditionally, these parameters are measured manually on-site at rock exposures, for example, measuring discontinuity orientation with a compass and an inclinometer [8]. However, such methods involve trade-offs between the number of samples and labor efficiency, and are often inaccessible and dangerous. To overcome these limitations, three-dimensional (3D) digital outcrop models (DOMs) [9] offer a high-resolution digital twin of rock outcrops.

DOMs are 3D representations of the outcrop surface in the form of point clouds, which can further generate polygon meshes [10]. Non-contact measurements, Light Detection and Ranging (LiDAR) and photogrammetry techniques, obtain the digital data needed to create a DOM [11]. LiDAR directly obtains the DOM, while photogrammetry first captures multiple images and then uses 3D

* Corresponding author.
E-mail address: zhengjun12@zju.edu.cn (J. Zheng).

https://doi.org/10.1016/j.ijmst.2025.09.005
2095-2686/

reconstruction techniques, such as Structure-from-Motion (SfM), to generate DOMs [12]. The primary challenge in discontinuity characterization from DOMs is identifying them. DOMs are nonintuitive, i.e., they do not contain any geological feature or any geometrical shape [13]. These characteristics often contribute to lengthy solutions, including the need for manual preprocessing, continuous parameter tuning or model training across different sites [14].

In addition to the complexity of DOMs, the complexity of geological and environmental conditions further hinders the generalizability of existing methods. Exogenic and endogenic geological processes, such as seismic activity, hydraulic forces, and long-term exhumation [15], influence the formation and evolution of rock masses, resulting in significant variations in discontinuity textures and geometry. Environmental factors, such as illumination, shadows, and vegetation noise, further complicate their identification. As shown in Table 1, the existing methods can be broadly categorized into two groups: (1) DOM-based identification, which directly utilizes DOM for discontinuity identification, and (2) vision-based identification, which relies on pattern information from raw images. These two approaches generally follow two strategies: rule-based and data-driven. The rule-based strategy uses specific rules to distinguish discontinuities from other features; in DOM-based segmentation, this includes methods such as plane fitting, region growing, and clustering, while in vision-based segmentation, image processing techniques like edge detection are commonly employed. The data-driven strategy utilizes labeled datasets and deep learning models to learn discontinuity features and identify them. However, both strategies have limitations: the rule-based approach is threshold-dependent, while the data-driven approach is data-dependent, with each model or each algorithm parameter set tailored to specific rock outcrops. Therefore, there is an urgent need for a framework capable of generalizing across multiple outcrops to overcome the limitations of the above approaches.

The emergence of large-scale artificial intelligence (AI) has revolutionized computer vision due to its remarkable generalization capabilities across a wide range of identification tasks, such as lesion monitoring in healthcare [16] and land-use recognition in agriculture [17]. While the limited amount of data has traditionally constrained the robustness of data-driven models, large-scale models (LSMs) excel in various recognition tasks with little or no additional training [18]. This is largely due to the vast parameter space of LSMs and their pretraining on extensive datasets, which allows a single model to address a broad spectrum of recognition challenges. Furthermore, as LSMs evolve, their generalization ability strengthens with an increase in dataset size. The above scenarios are similar to those encountered in discontinuity identification, where high precision is required to extract detailed features from data with high environmental complexity. This paper aims to use one LSM, Segment anything model (SAM) [19] as an example, to segment discontinuities across several outcrops.

In this study, we proposed a framework, called Discontinuity Identification for Geology (DIG), that integrates SAM with SfM for 3D discontinuity identification using data from different photogrammetric devices. The proposed framework builds upon earlier efforts to bridge two-dimensional (2D) and 3D in discontinuity analysis, such as Lee et al. [20], who focused on trace-type features by extracting joint traces in 2D and projecting them into 3D. This work extends this line of research, using facet-type discontinuities as an example, employs SAM to identify 2D discontinuities and leverages the relationship between image pixels and point clouds–an inherent connection in SfM that has often been overlooked in previous studies–to derive 3D discontinuities. This framework transforms a 3D point cloud segmentation problem into a 2D image segmentation task, allowing the use of

Table 1
Summary of the digital discontinuity identification methods.

Classification	Method	Feature
DOM-based identification	Manual measurement	• Manual extraction of 3D discontinuities in point cloud space • They are often used as ground truth in method comparison
	Plane fitting*	• Plane fitting algorithms, such as RANSAC, are used to extract planar discontinuities by iteratively identifying inliers that best fit the data • These algorithms are threshold-dependent, with optimization by adjusting parameters, such as the maximum distance and minimum group size • The iterative nature makes it suitable for small-sample data with distinct discontinuity features • These are point-wise computation algorithms that need manual noise pre-filtering
	Region growing*	• RG is used to segment discontinuities based on similarity criteria, which may include location, normal vector, color, curvature, geometry, or a combination of these • Like plane fitting, these are threshold-dependent, iterative, point-wise computation algorithms
	Clustering*	• Clustering is used to segment discontinuities, which involves seven major types, i.e., density-based methods (e.g., DBSCAN), partitioning methods (e.g., *k*-means, *k*-medoids), etc. • A hybrid type of clustering is commonly employed to improve effectiveness and robustness • Similar to plane fitting and RG, these are threshold-dependent, iterative, and point-wise computation algorithms.
	Deep learning**	• These methods directly extract 3D discontinuities using a data-driven approach by building point cloud datasets and training DL models • The performance of these models on out-of-sample data is limited. They are commonly used for facet-type discontinuities
Vision-based identification	Manual measurement	• Manual extraction of 3D discontinuities from images integrated with depth information layers
	Image processing*	• These methods process and analyze images to extract 2D discontinuities based on predefined rules or conditions. These rules are typically derived from pixel values, geometry, texture, or other attributes of the image • These are threshold-dependent algorithms, with limited application in discontinuity identification. They are commonly used for trace-type discontinuities
	Deep learning**	• These methods extract 2D discontinuities using a data-driven approach by building image datasets and training DL models. They project these discontinuities into 3D space, requiring additional calculations • The performance of these models on out-of-sample data is limited. They are commonly used for trace-type discontinuities
	Large-scale model**	• This method uses the general image segmentation capability of SAM without the need for model training • It intelligently extracts 3D discontinuities integrated with intrinsic 2D-3D information from SfM. • It is used for facet-type discontinuities in this paper

Note: Symbols in the table: * denotes a rule-based strategy; and ** a data-driven strategy.

advanced image segmentation tools for geological interpretation. This integration addresses the core challenge of interpreting unstructured point clouds and enables generalizable detection across diverse outcrop conditions.

2. Related works

2.1. LSMs vs. Other models or algorithms

Discontinuity identification in rock masses presents enduring challenges due to the geometric complexity, surface irregularity, and noise-prone nature of natural outcrops. Traditional algorithms, such as plane fitting, region growing, and clustering, typically rely on geometric cues extracted from sparse or dense point clouds. These algorithms operate within an unsupervised learning paradigm, applying fixed rules or thresholds without the need for labeled data (Fig. 1a). While such algorithms can be effective on clean datasets, their performance degrades significantly under real-world conditions involving weathering, vegetation, blast damage, or sensor noise. Moreover, they often require extensive manual pre-processing, including data denoising and parameter tuning, a form of "hidden labor" that is rarely accounted for in methodological evaluations. These methods also suffer from poor adaptability, requiring retuning or redesign when applied to new geological settings.

In contrast, supervised deep learning models promise higher accuracy by learning from labeled examples. However, in practice, they also face major limitations when applied to discontinuity detection. Unlike conventional objects in computer vision tasks, geological discontinuities are surface-like, irregular in shape, poorly bounded, and highly textured. As a result, traditional object detection architectures often fail to capture their patterns reliably [13,14]. Additionally, supervised models require large, high-quality labeled datasets for each new context, resulting in high annotation costs and limited transferability. Each new scene often demands its own training set and model, making the workflow inefficient and non-scalable in geological engineering applications.

LSMs (or more generally, foundation models), as illustrated in Fig. 1b, provide a complementary pathway. These models are typically trained on diverse datasets using self- or semi-supervised learning and are designed for flexible adaptation to various downstream tasks [21]. A distinctive feature of LSMs is the prompt-based interaction module, which allows users to guide segmentation through minimal input, a design concept similar to that of large language models (LLMs) like GPT-4. This design enables zero-shot or few-shot segmentation, reducing the dependence on extensive labeled data and improving adaptability to different field environments.

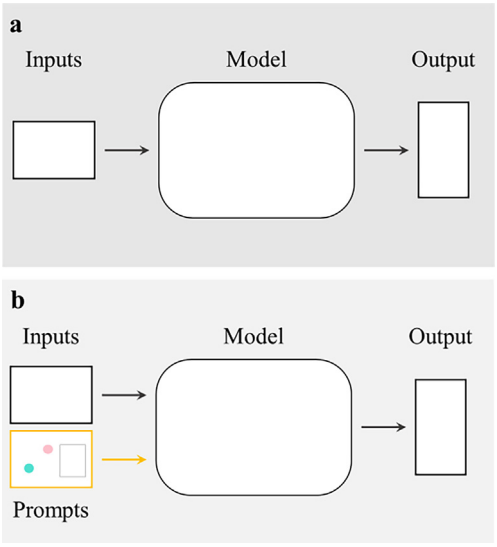


Fig. 1. Architectures comparison of tradition model and LSMs.

In the context of discontinuity identification, LSMs offer a pragmatic balance: expert prompting helps guide the model to interpret complex, noisy outcrop surfaces, while the underlying pretrained model accelerates the segmentation process without requiring full retraining. Although not a replacement for fully automated methods, the use of LSMs represents a complementary strategy: one that enables efficient, high-quality 3D discontinuity segmentation in scenarios where conventional or fully automated techniques may struggle.

2.2. SAM vs. Other LSMs

Within the broader category of LSMs, SAM is distinguished by its general segmentation capability. SAM has been used as an effective annotation tool across various domains, demonstrating its precise identification capability and ease of implementation. This capability is ensured by its training on a large-scale source of data. SAM was trained on SA-1B, a diverse dataset containing over 1 billion masks and 11 million images, using 256 A100 GPUs for 68 h, which is equivalent to driving about 11265 km in an average gasoline-powered passenger vehicle [19]. This scale and diversity of SAM's training data are exceptional compared to other LSMs (Table 2), which are largely attributed to its data engine.

The architecture of SAM provides a highly standardized solution for discontinuity identification tasks. By leveraging its prompt module, SAM enables high-precision segmentation of 2D discontinuities while offering an efficient annotation engine for automating this task. As large-scale annotated datasets for discontinuities are developed and models continue to evolve, the complex task of discontinuity identification will gradually move towards full automation.

Moreover, by fine-tuning the model with specific instructions, complete automation can be achieved in downstream tasks of discontinuity identification–specifically, discontinuity characterization and rock mass stability evaluation– with minimal human intervention. In terms of specific implementation paths, the future of automated discontinuity identification can draw inspiration from advances in medical image segmentation. SAM has been successfully applied to medical image segmentation through a variety of fine-tuning strategies [16]. These strategies include prompt-based adjustments, auxiliary prompt encoders, adapters, and decoder modifications. These methods not only enhance SAM's performance in medical image segmentation but also reduce the reliance on large-scale annotated datasets, offering valuable lessons for geological research.

3. Methodology

In digital geological investigations, traditional algorithms or models often fail in generalization across diverse outcrops. As a result, these models require customization for each outcrop, leading to excessive algorithm reuse and high development costs. To date, no unified model has effectively addressed discontinuity identification. However, the proposed DIG framework overcomes these limitations by leveraging the superior generalization capabilities of a pretrained SAM. This framework enables out-of-the-box 3D discontinuity identification without model retraining.

As illustrated in Fig. 2, the proposed DIG framework consists of two parallel processes: 2D discontinuity detection and 3D reconstruction, represented in blue and orange lines, respectively. The

Table 2
A summary of key information about LSMs for image segmentation (listed chronologically by publication date).

LSM	Backbone	Prompt	Dataset	Dataset size
OneFormer [22]	Transformer	text	Integrated 3 datasets	149 thousand images
Uni-Perceiver v2 [23]	CNN or Transformer	text	Integrated 6 datasets	29.9 million images
SAM [19]	ViT	Points, boxes, text	SA-1B	11 million images, 1.1 billion masks
SegGPT [24]	ViT	Learnable image prompt	Integrated 12 datasets	0.3 million images
SEEM [25]	Transformer	Points, boxes, scribbles, images, text	COCO	118 thousand images

Note: CNN denotes "convolutional neural network", and ViT the "vision transformer".

methodology is divided into four sections, described sequentially below.

Image Selection or Composition. Photogrammetric data acquisition systems are categorized into mobile camera platforms (e.g., UAVs) and fixed-position setups (e.g., tripod-mounted or gimbal-stabilized systems). While extensive overlap in images benefits photogrammetric workflows, it introduces substantial image redundancies for interactive discontinuity detection. To mitigate this, the proposed automated algorithm selects or composites images based on their acquisition modes, thereby reducing the volume of images requiring detection. These strategies are detailed in Section 3.1.1.

Interactive discontinuity detection. SAM's prompt module enables interactive discontinuity segmentation. This approach achieves real-time and precise 2D discontinuity identification, with a response of just 0.055 s. Further details are provided in Section 3.1.2.

Pixel-point relationship generation. SfM matches pixel features between images and calculates the poses (positions and orientations) of each image. As a result, it produces a sparse 3D point cloud and preserves pixel-point relationships without additional computational cost. This is discussed in Section 3.2.1.

Three-dimensional discontinuity indexing. The pixel-point relationships act as a bridge between image segmentation and point cloud segmentation. Therefore, automated indexing is applied to obtain 3D discontinuities. For mobile camera platforms, a single indexing links images and their point cloud. For fixed-position setups, dual indexing connects images, their composite image, and the corresponding point cloud. This is discussed in Section 3.2.2.

3.1. Two-dimensional discontinuity detection

3.1.1. Image selection or composition

In photogrammetry, regardless of the equipment used (e.g., drones, cameras), it is essential to ensure an overlap rate of more than 70% between photos to facilitate successful 3D reconstruction [26]. However, if interactive discontinuity detection is directly performed on every raw image, it would significantly increase identification costs. To mitigate this detection redundancy, this section introduces an automated procedure to either select or composite images based on different image acquisition modes.

Image acquisition modes in rock mass outcrop photogrammetry can be categorized into two types: mobile camera platforms and fixed-position setups. Mobile camera platforms typically include drones, which allow flexible positioning and dynamic viewpoints. Fixed-position setups commonly use tripod-mounted or gimbal-stabilized cameras, providing stable and controlled imaging conditions. However, the large volume of raw images (e.g., hundreds to thousands of images per outcrop) imposes a heavy workload for interactive discontinuity detection. For images captured by fixed-position equipment, we detect discontinuities from panoramic composites. Panoramas enable holistic discontinuity identification across the entire slope, drastically reducing image redundancy for detection.

Fixed-position equipment adjusts the pitch angle at each setup location, using a gimbal to automatically capture tens to hundreds of images. These overlapped images with varying pitch angles from multiple fixed positions are then used for 3D reconstruction. By performing interactive discontinuity detection on only one panoramic composite image per position, the number of processed images is drastically reduced from hundreds or thousands to a few. We utilize the Image Composite Editor software [27] to generate high-resolution panoramas from overlapping photos of the same camera position, while explicitly preserving the relationship between original and composited images. Section 3.2.2 will detail how this relationship is further utilized to obtain 3D discontinuities.

Unfortunately, images captured by mobile camera platforms cannot be composited into panoramas due to significant variation in shooting positions [28]. In such cases, we reduce the number of images for detection through an image selection process based on spatial coverage and redundancy reduction. We develop a step-by-step automated image selection algorithm to streamline the tedious selection process. The pseudo-code is shown in Fig. 3a. The algorithm first computes the number of effective pixels (i.e.,

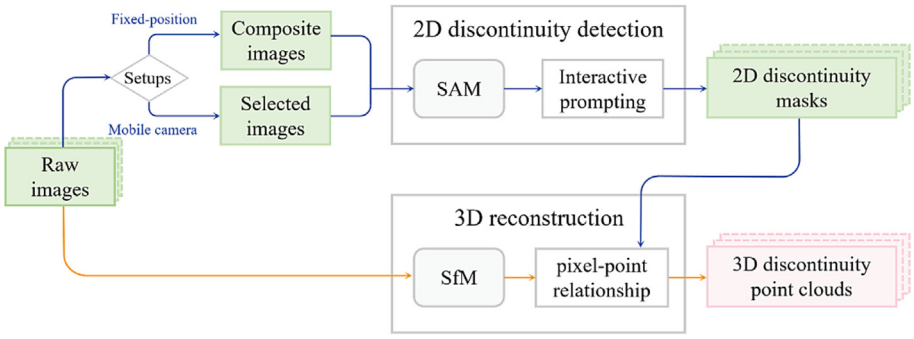


Fig. 2. The flowchart of the proposed DIG framework for discontinuity identification.

pixels with valid 3D point correspondences) for each image, ranks them in descending order, and selects the image with the maximum effective pixels (Fig. 3b, selected image I). Additional images are then iteratively selected based on viewpoint diversity and coverage (Fig. 3b, selected image II), following the principle of maximizing effective pixels while minimizing redundancy. The pipeline outputs a reduced image set, streamlining the selection process and supporting efficient interactive discontinuity detection.

3.1.2. Interactive discontinuity detection

For 2D discontinuity detection, traditional data-driven methods, such as deep learning, require dataset construction and model training, which necessitate large amounts of labeled data. However, pretrained SAM reduces the need for retraining and lowers the barrier to geological investigations. Therefore, this approach offers a unified solution for discontinuity detection in various geological contexts by leveraging SAM's generalization capabilities.

The overall design of SAM focuses on efficiency, enabling seamless interactive use. As shown in Fig. 4, the model architecture consists of three core components: an image encoder, a prompt encoder, and a mask decoder. The image encoder computes each input image once, generating a reusable image embedding in approximately 0.15 s. The prompt encoder processes input prompts, and the lightweight mask decoder combines these two information sources to predict segments within just 55 ms per prompt, enabling real-time responsiveness without redundant computation. Additionally, SAM operates effectively on input images of arbitrary resolutions without requiring preprocessing, further enhancing its practical utility in field applications.

We use two types of prompts: points (foreground points for adding areas, marked in green, and background points for removing areas, marked in red) and boxes. As illustrated in Fig. 5, we interactively click foreground and background points to guide the model in detecting complex discontinuities. The example highlights the detection of discontinuity No. 6, with complex features due to its blurred boundary, partial occlusion, and location at the intersection of light and shadow. The discontinuity No. 6 is successfully segmented in approximately 1–3 s using 5 foreground and 5 background points.

Unless otherwise specified, this paper uses the largest model type of SAM, the pre-trained ViT-H image encoder with its publicly released checkpoint, and does not modify or adapt any parameters for evaluation. The model was implemented on a Linux Ubuntu 20.04 platform with an NVIDIA TITAN RTX GPU and an Intel® Core™ i9-10920X CPU.

3.2. Three-dimensional discontinuity identification

3.2.1. Pixel-point relationship generation

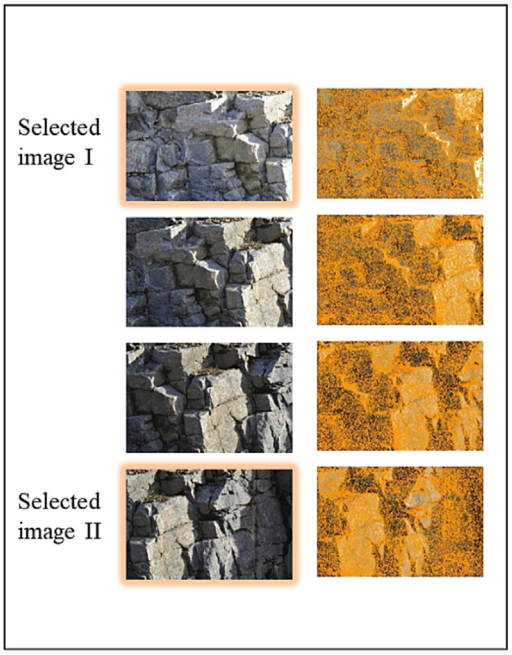
SfM is widely used in geological surveys for its flexibility in 3D reconstruction [29]. Unlike stereo photogrammetry, another photogrammetric technique that requires pre-calibrated cameras and a fixed baseline, SfM processes without these constraints and performs feature matching on overlapping images to reconstruct 3D structures. For real-world scaling, SfM requires real-time kinematic global positioning system (RTK-GPS), ground control points (GCPs), or known object dimensions. This framework uses UAVs equipped with high-accuracy RTK-GPS for mobile platforms, and keeps baseline consistency in fixed-position setups to reduce the likelihood of spatial distortion. For GCP-based real-world scaling, it is recommended to deploy multiple GCPs in large-scale studies to mitigate localized deformation.

SfM is highly automated, and its pipeline is illustrated in Fig. 6. The input consists of unordered overlapping images. The outputs include three components: a colored sparse point cloud model, camera poses, and pixel-point relationships. In this study, we use Meshroom [30], an open-source SfM tool with GPU acceleration and high-quality reconstruction; other software such as Agisoft PhotoScan [31] is also applicable. The process includes five main steps: (1) Feature extraction identifies and describes distinctive points in images for matching. We adopt a hybrid of

Input:
A set of images $I = \{I_1, I_2, \ldots, I_n\}$
A 3D point cloud $P = \{P_1, P_2, \ldots, P_m\}$ reconstructed from the images
Pixel-point relationships $R = \{(I_i, P_i)\}$, indicating which 3D points are covered by which images
Output:
A set of selected images or image regions $S = \{S_1, S_2, \ldots, S_k\}$

Procedure:
1. **Initialize** an empty list for selected images: $S = []$
2. **Initialize** a set to track covered 3D points: $CoveredPoints = \{\}$
3. For each image I_i in I:
   a. **Calculate** the effective pixel coverage for I_i:
      EffectivePixels (I_i) = Count the number of unique 3D points in P covered by I_i
   b. **Store** the image and its effective pixel coverage in a list:
      ImageCoverage = $[(I_i, \text{EffectivePixels}(I_i))]$
4. **Sort** ImageCoverage in **descending order** based on EffectivePixels:
      ImageCoverage = Sort(ImageCoverage, by=EffectivePixels, descending=True)
5. For each $(I_i, \text{EffectivePixels}(I_i))$ in ImageCoverage:
   a. **Initialize** a flag to indicate if the image is selected:
      IsSelected = False
   b. For each 3D point P_i covered by I_i:
      - **If** P_i is not in CoveredPoints:
        Add P_i to CoveredPoints
        Set IsSelected = True
   c. **If** IsSelected:
      - **Add** I_i to the list of selected images: S.append(I_i)
6. **Return** the list of selected images or image regions: S

(a) Pseudo-code of the selection process



(b) Visual illustration

Fig. 3. The automated image selection algorithm.

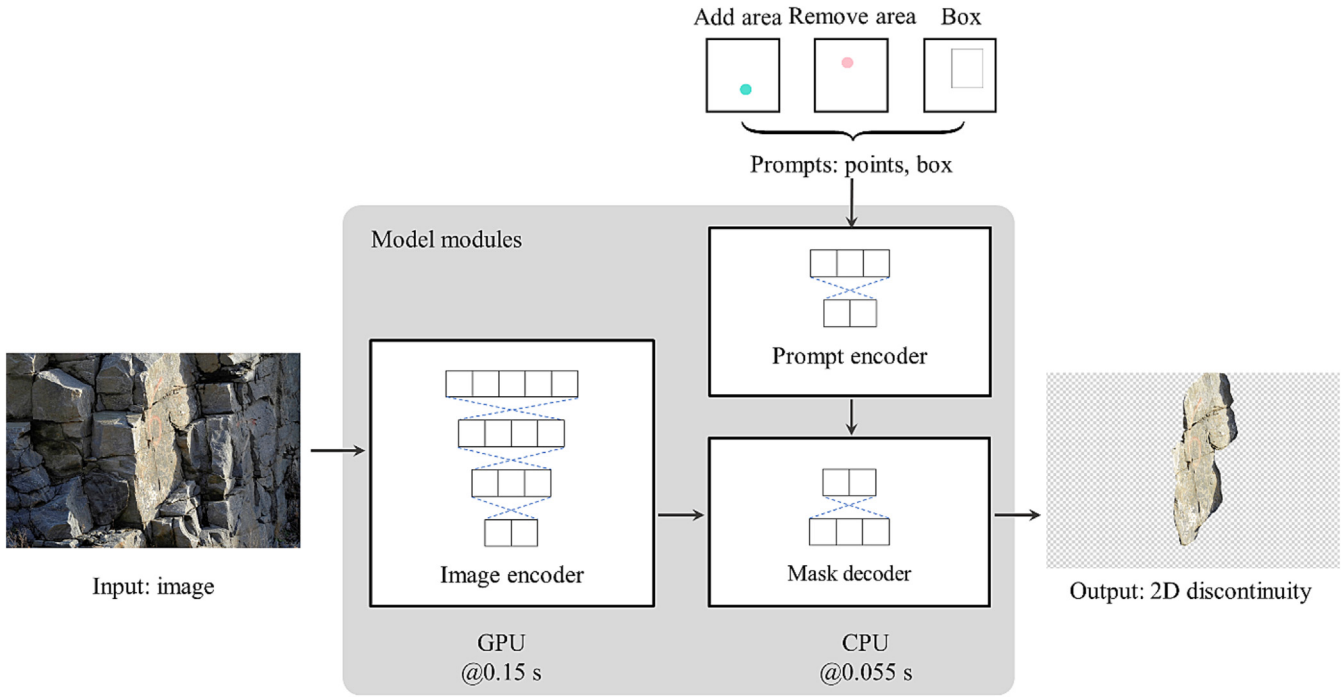


Fig. 4. The architecture of SAM for interactive discontinuity detection, modified after Ref. [19].

Fig. 5. The step-by-step demonstration of interactive discontinuity detection for a complex discontinuity.

scale-invariant feature transform (SIFT) [32] and accelerated-KAZE (A-KAZE) [33], which provides robustness to illumination/scale changes and effective delineation of natural rock mass boundaries. (2) Image matching establishes geometric relationships by comparing features, accelerated using a vocabulary tree for efficient candidate pair retrieval. (3) Camera pose estimation determines camera positions and orientations using essential matrices, with RANSAC applied to remove outliers. (4) Triangulation computes 3D point coordinates from multi-view correspondences, establishing the initial pixel–point relationships. (5) Bundle adjustment jointly optimizes 3D points and camera parameters by minimizing reprojection error, with GPS or GCPs used as soft constraints for geo-referencing. Key parameter settings and recommended values for improving pixel-point corre-

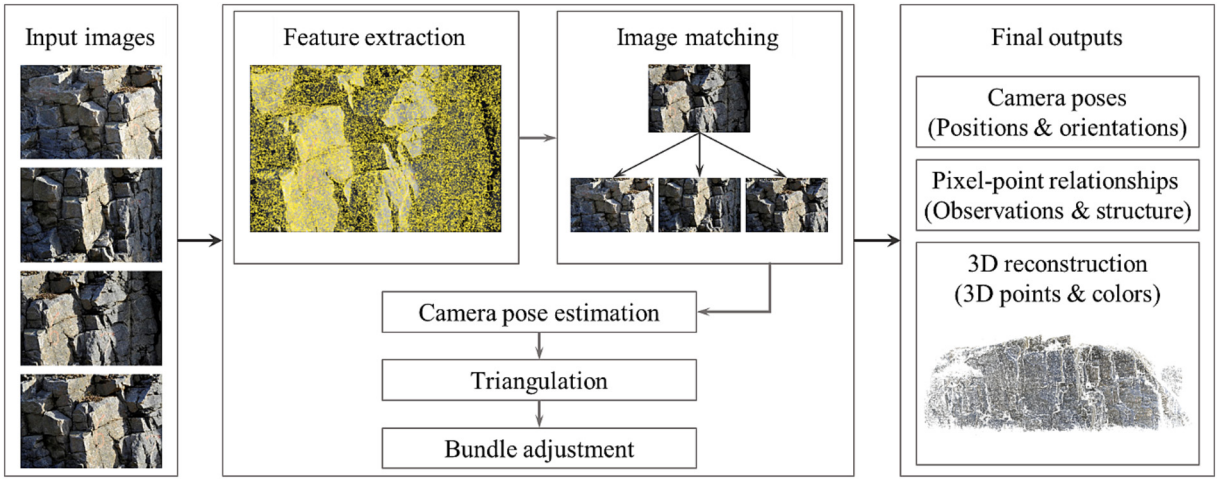


Fig. 6. The input, key steps, and outputs of SfM for 3D rock mass reconstruction.

spondences and overall reconstruction quality are summarized in Table 3.

The pixel-point relationship is a fundamental data structure in SfM, dynamically updated throughout the pipeline. The schematic representation and data format of pixel-point relationship are illustrated in Fig. 7. It records feature matching pixels (referred to as "observations") and their associated 3D spatial coordinates (referred to as "structure"). The finalized data of pixel-point relationships can be exported using functions such as the ConvertSfM function in Meshroom, or alternatively, via the exportMatches function in Agisoft PhotoScan.

3.2.2. Three-dimensional discontinuity indexing

This section utilizes the 2D-3D pixel-point geometric relationships to extract corresponding 3D point clouds from 2D image masks. After carefully selecting parameters to achieve satisfactory SfM modeling results, the pixel-point relationships are utilized to convert 2D discontinuity segmentation results into 3D point cloud data. This conversion is achieved through automated indexing, enabling the spatial localization of discontinuities based on the established 2D-3D relationships. Since each observation corresponds to a uniquely determined 3D structure point, the conversion process is straightforward and precise.

For fixed-position devices, dual indexing is employed to connect images, their composite image, and the corresponding point cloud. The segmentation results are identified on the composite image, allowing for a unique determination of which pixel in which original image corresponds to a given composite image pixel. Additionally, image observations are indexed to their 3D points, which are uniquely determined through the pixel-point relationships. Fig. 8 demonstrates the process of obtaining a 3D discontinuity using automated indexing for fixed-position devices.

For mobile camera platforms, only the second automated indexing step is processed. Unlike fixed-position setups, where composite panoramic images are used as input, mobile systems directly process discontinuity detection results from individual images. This approach, described in Section 3.1.1, avoids the issue of overlapping image areas. The effective observations extracted from the segmentation results in the images are then indexed to their respective 3D points.

4. Method validation

We designed a two-step validation pipeline to demonstrate that DIG can reliably detect discontinuities across diverse geological contexts, while maintaining high accuracy and low cost. First, we developed a 2D discontinuity dataset gathered from the internet and self-captured images to evaluate SAM's performance in identifying 2D discontinuity. This validates the generalization capability of SAM, the core component of 2D detection, in real-world geological scenarios. This is discussed in Section 4.1. Second, we assembled a 3D discontinuity benchmark by generating one specific DOM and manually labeling discontinuities from the DOM to eval-

Table 3
Key parameter recommended settings for 3D reconstruction of rock masses.

Step	Parameter	Recommended setting	Notes
Feature extraction	Describer type	SIFT and A-KAZE	Recommended for robustness and efficiency
	Describer density	High or normal	
	Describer quality	High or normal	
Image matching	Matching method	Sequential and vocabulary tree	Combines two methods for efficiency in sequential inputs
	Photometric matching	ANN L2	Uses L2 norm for approximate nearest neighbor matching
Pose estimation	Geometric estimator	Acransac	Improved RANSAC to enhance performance
	Geometric filter type	Fundamental matrix	Describes geometric relationships between views
Triangulation	Observation constraint	Scale	Uses reprojection error relative to feature scale
Bundle adjustment	Local bundle adjustment	Enabled	Reduces reconstruction time, especially for large datasets
	Compute structure color	Enabled	Computes color for each 3D point
	Transformation method	From_GPS	Scales the SfM model to real-world size using GPS data

Note: The settings in this table are based on Meshroom [30]. For other software, refer to similar parameters and adjust accordingly to achieve optimal SfM reconstruction results.

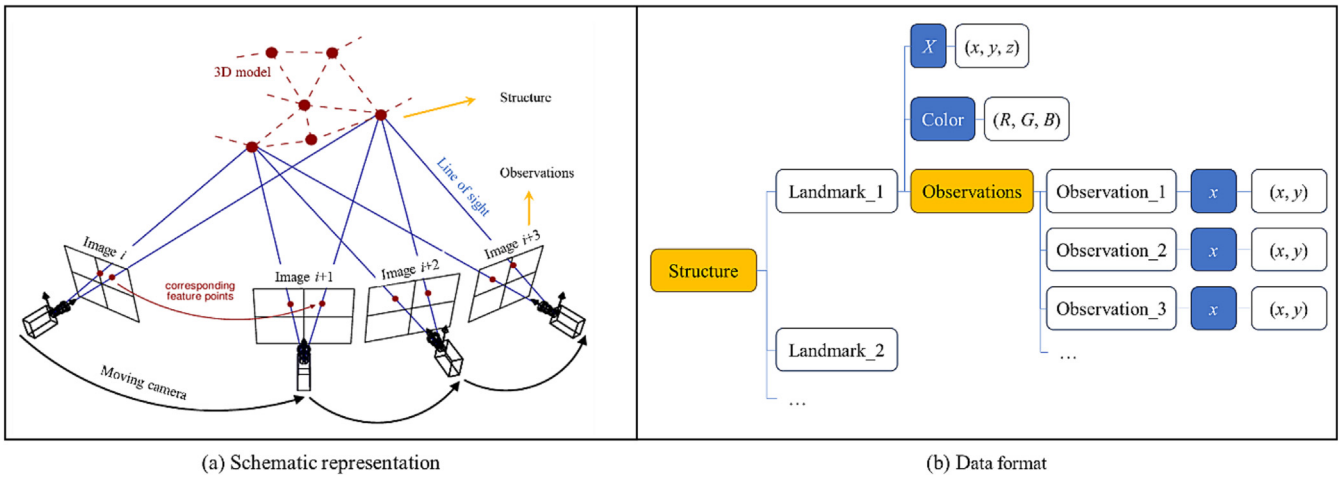


Fig. 7. The pixel-point relationships in SfM, modified from Ref. [34].

uate the geometric accuracy of 3D discontinuity identification results in the DIG framework. Details are provided in Section 4.2.

4.1. Validation of generalization in 2D discontinuity detection

4.1.1. The discontinuity image dataset

To validate generalization in 2D discontinuity detection, we need data from different rock outcrops. Unfortunately, there is no dataset available for this specific task; to address this, we introduce Dataset-2D, a newly constructed dataset with diverse outcrop types and discontinuity patterns for evaluation. Dataset-2D comprises 2426 identified 2D discontinuities across 170 images, including 90 sourced from the internet and 80 self-captured images. We included a wide range of natural and man-made outcrops. These include slopes, river-cut banks, sea cliffs and roadcuts, tunnel walls (with both adequate and poor lighting conditions), open-pit mine faces, and foundation excavation walls. This collection captures varying viewpoints (Figs. 9a and b), lighting conditions, degrees of fracturing, and geometric complexity, particularly encompassing dark underground tunnels (Fig. 9c) and heavily fractured rock outcrops (Fig. 9d).

The discontinuities in Dataset-2D were manually annotated using LabelMe [35] as ground truth. Further details about the dataset and annotation process are provided in Supplementary data §A. Dataset-2D captures multiple forms of discontinuity textures and geometries, including partially occluded features and small-scale discontinuities (Fig. 9e), visually highlighted by semi-transparent masks in different colors overlaid on the source outcrop images.

4.1.2. The generalization evaluation pipeline

The evaluation pipeline is a crucial tool for validation. We evaluate discontinuity segmentation by measuring IoU and precision metrics between predicted masks and ground truth masks. The IoU quantifies the overlap between the predicted mask after N point prompts and the ground truth mask, where N ranges from 1 to 20. Precision measures the proportion of correctly predicted pixels within the predicted mask. The mean IoU (mIoU) and average precision (AP) are calculated by averaging the per-instance IoU and precision across all discontinuities in the dataset. This dual-metric approach allows us to assess both spatial overlap and the reliability of prediction.

The evaluation pipeline follows established practices in interactive segmentation [36], using simulated user prompts to assess how well segmentation improves as more points are added. During evaluation, point prompts are deterministically sampled using ground truth masks to ensure consistency and reproducibility across runs. The first foreground point is selected as the location farthest from the boundary of the annotated region. Subsequent points are selected from error regions between predictions and ground truth masks, alternating between foreground and background based on dominant error type (under- or over-segmentation). A distance transform prioritizes the most representative error regions, while previously sampled points are tracked to avoid redundancy. In cases where SAM produces multiple masks, we retain the highest-confidence prediction.

This simulation protocol is used only for controlled evaluation and does not assume the availability of ground truth during actual deployment. In real-world applications, users can interactively provide point prompts by visually identifying discontinuities or suspected regions of interest. This deployment strategy allows the DIG framework to operate without requiring any ground truth.

4.1.3. Results of generalization validation

We use all 170 image samples and 2426 identified 2D discontinuities in Dataset-2D for evaluation. SAM achieves 0.78 mIoU and 0.86 AP with 11-point prompts, after which performance stabilizes

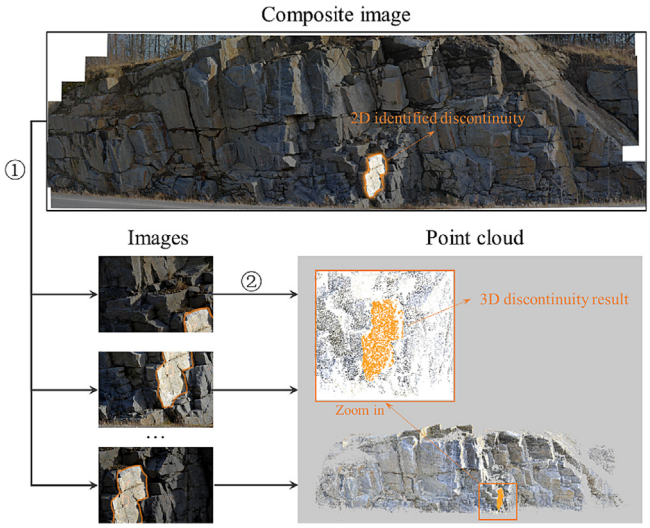


Fig. 8. The process of extracting a 3D discontinuity using automated indexing for fixed-position devices.

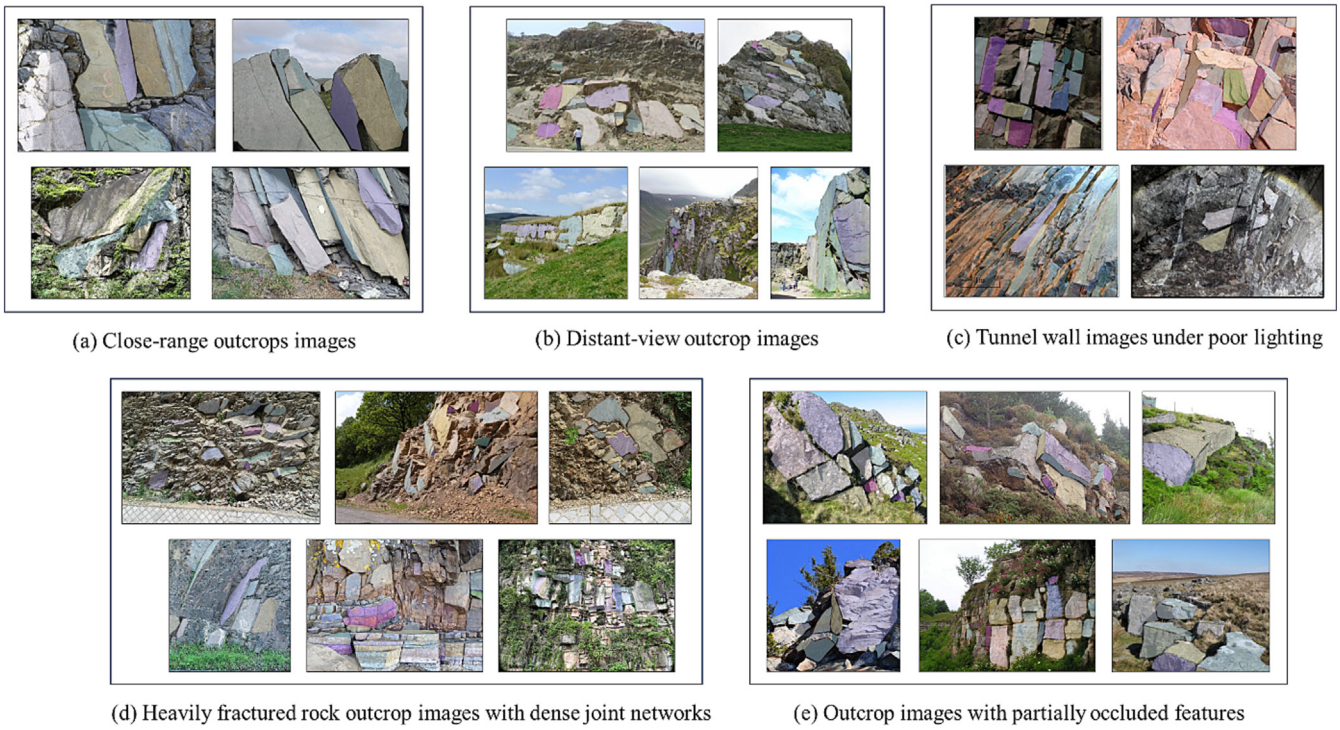


Fig. 9. Example images with overlaid masks from Dataset-2D. Masks are shown in semi-transparent colors.

despite increasing prompt numbers, as shown in the boxplot of IoU distribution (Fig. 10a). SA-1B is a large-scale dataset released together with SAM and contains 1.1 billion high-quality masks for open-world images [19]. Comparative results in Fig. 10b demonstrate SAM's zero-shot capability on discontinuity identification, with mIoU decreasing from 0.85 (SA-1 billion dataset) to 0.76 (Dataset-2D) at 9-point prompting. This confirms SAM's strong generalization across unseen discontinuity images despite not being specifically trained on discontinuities.

SAM enables efficient segmentation under experts' interaction. Its performance should be viewed as complementary to human expertise. SAM demonstrates advantages in two areas: (1) noise suppression, such as distinguishing discontinuities from vegetation (Fig. 11a) and man-made elements (Fig. 11b); and (2) precise boundary delineation for certain complex geometries (Fig. 11c–f). However, humans remain more capable of interpreting ambiguous patterns and contextual cues, particularly when supported by field observations. The DIG framework bridges both strengths by combining SAM's segmentation power with expert oversight for optimal accuracy. However, failure cases also reveal two primary limitations in SAM's discontinuity identification: (1) segmentation errors occur under challenging conditions—strong lighting artifacts (e.g., tunnel light reflections, Fig. 11g), dense vegetation occlusion and weathered surfaces lead to fragmented boundaries or false holes (Fig. 11h); and (2) discrepancies arise between SAM's predictions and manual annotations due to ambiguous weathering patterns (Fig. 11i).

4.2. Validation of accuracy in 3D discontinuity identification

4.2.1. The 3D discontinuity dataset

To validate the geometric accuracy of 3D discontinuity identification in the proposed DIG framework, we constructed a 3D benchmark (Dataset-3D, see Fig. 12) by generating a specific DOM and manually annotating discontinuities using CloudCompare [37] as ground truth. Detailed annotation guidelines are provided in Supplementary data §B. In this study, we demonstrate 3D validation using a fixed-position setup, which involves more elaborate procedures and richer indexing steps compared to mobile camera platforms, offering a clearer illustration of the DIG framework's capabilities.

The images were obtained using a Nikon D300s camera (4256 pixel × 2832 pixel) at the Sunbury RockBench [38], focusing on a granite rock mass exposed at a road cut in Canada. A total of 471 images were taken from three fixed positions, with 146, 168, and 157 images captured at each position, respectively. Using the method described in Section 3.2.1, pixel-point relationships were established consisting of 392229 points. Point cloud resolution, typically defined as the average distance between neighboring points, was computed using Eq. (1), where M represents the number of point pairs and d_i denotes the distance between the i–th closest pair of points. Manual annotation identified 87 discontinuities of varying sizes and orientations, establishing a benchmark for the DIG framework.

$$\text{Resolution} = \frac{1}{M}\sum_{i=1}^{M} d_i \tag{1}$$

4.2.2. The accuracy validation method

We evaluate discontinuity segmentation by measuring the IoU and precision between predicted and ground truth point clouds. IoU, a commonly used metric in segmentation tasks, quantifies the overlap between the predicted and ground truth point clouds. Its values range from 0 to 1, where 1 indicates perfect alignment and 0 indicates no overlap. Precision, on the other hand, assesses the reliability of the model's predictions by measuring the proportion of correctly identified points among all predicted points. While IoU reflects the overall accuracy of segmentation, precision provides insight into the framework's ability to minimize false positives.

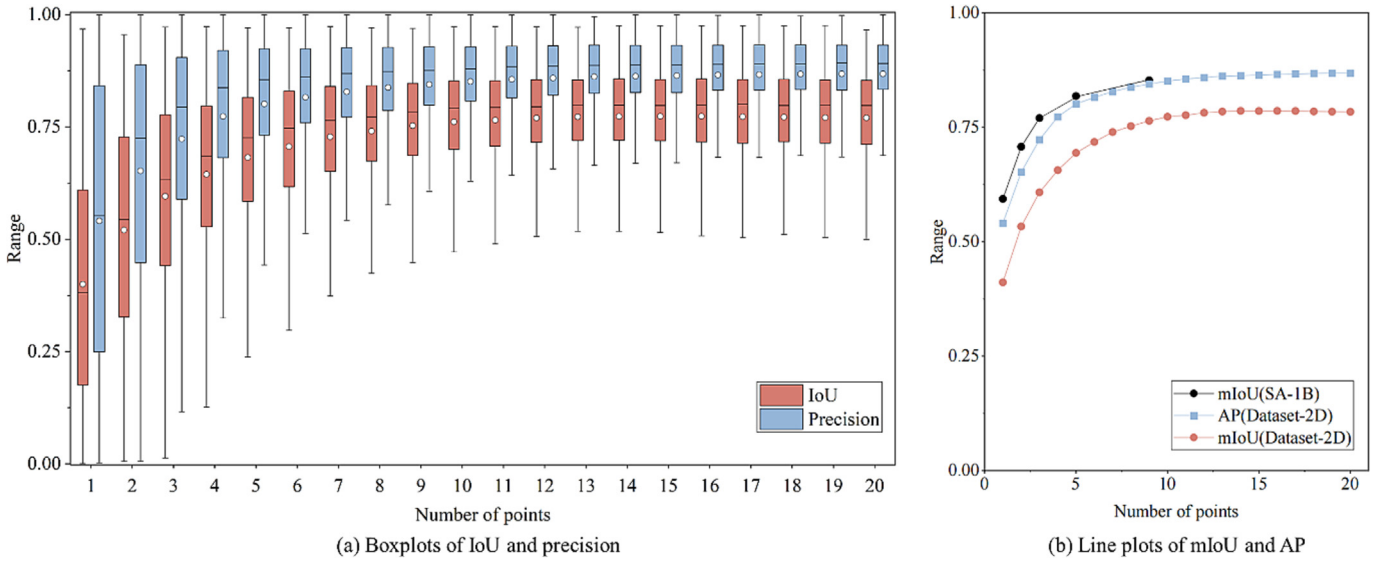


Fig. 10. IoU and precision versus the number of points on dataset-2D.

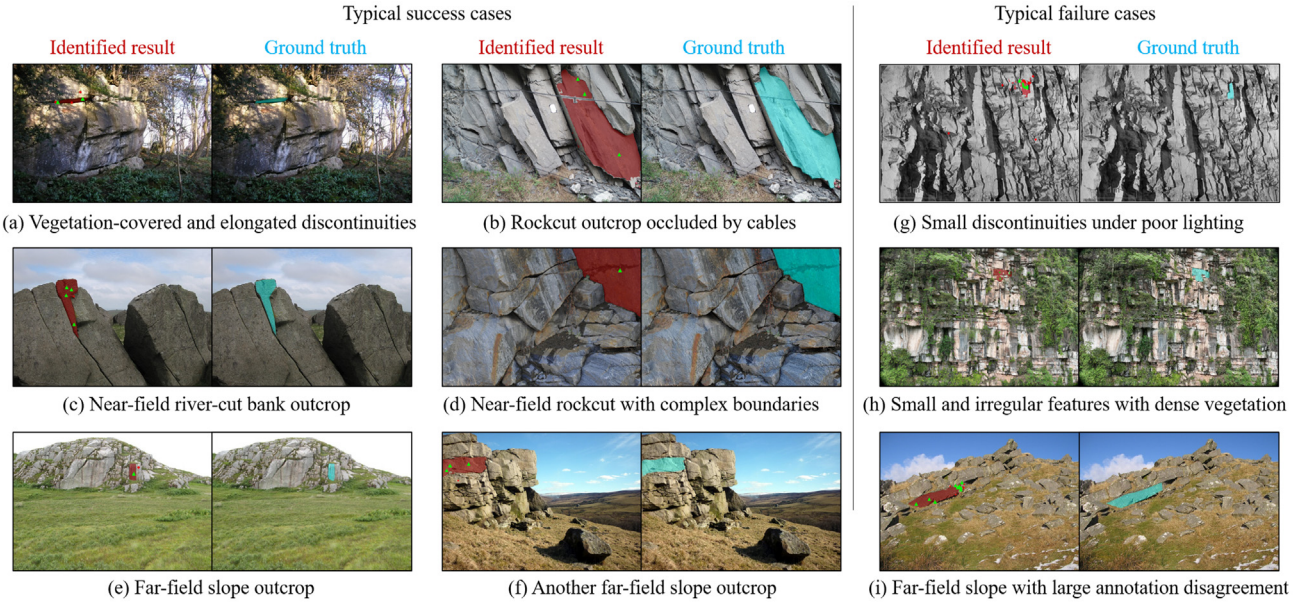


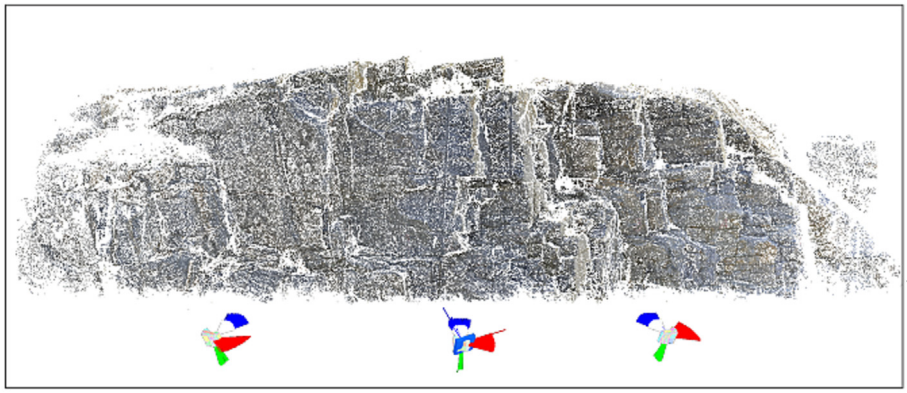
Fig. 11. Segmentation results under standard interactive segmentation prot ocols. Triangular markers: foreground prompts. circular markers: background prompts.

To further assess segmentation performance, we calculate mIoU and AP. mIoU is obtained by averaging the IoU scores across all discontinuities in the dataset, offering a global evaluation of the segmentation quality. AP is derived from the precision-recall curve and measures the framework's performance across different confidence thresholds, providing a comprehensive assessment of its predictive capability.

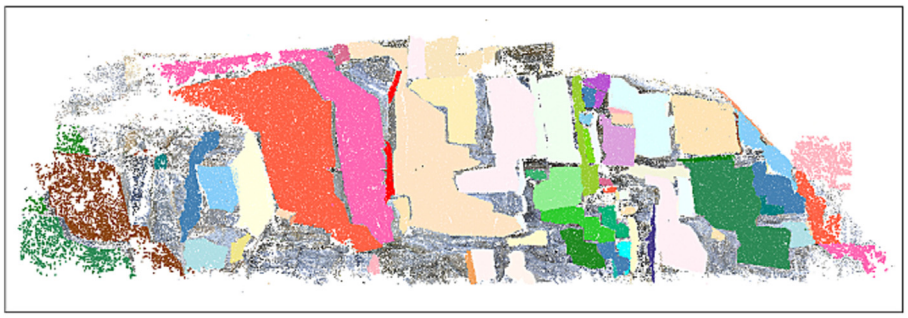
4.2.3. Results of accuracy validation

The DIG framework successfully identified 87 discontinuities (Fig. 13a), achieving an AP of 0.91 and a mIoU of 0.70 for 3D discontinuity identification. Figs. 13c and d present the IoU and precision distributions using half-box-and-half-violin plots, highlighting extrema, the interquartile range, median, and mean values. Approximately 70% of the identified 3D discontinuities exhibit a precision above 0.90, and about 75% achieve an IoU exceeding 0.60. These results demonstrate that the DIG framework maintains high accuracy in 3D discontinuity identification, offering a practical and computationally efficient tool for large-scale geological annotation.

Fig. 13b illustrates the false positive points (marked in black) alongside the complete annotated point cloud. False positives (black) are predominantly clustered near annotated discontinuities, indicating strong alignment between segmentation results and ground truth. The primary sources of these errors are inaccuracies in 2D interactive discontinuity detection and subjectivity in manual 3D annotation. Direct segmentation of discontinuities within point clouds often necessitates frequent viewpoint adjustments, making manual annotation both time-consuming and prone to boundary ambiguity. Human-induced discrepancies, particularly in weathered discontinuities or transition zones between discontinuities, contribute to these differences. Fundamentally, the DIG framework operates as an image segmentation-based approach, whereas ground truth annotations are generated



(a) 3D reconstruction result



(b) Annotated 3D discontinuities

Fig. 12. Overview of Dataset-3D.

through direct point cloud segmentation. Both methods leverage 3D reconstruction results, but the DIG framework integrates 3D reconstruction with image segmentation, whereas ground truth segmentation is performed sequentially after 3D reconstruction. This distinction inherently explains the similarity between the two segmentation results.

Further analysis of 3D discontinuity identification results is conducted by examining No. 83, the discontinuity with the largest pixel area among those with lower precision. Fig. 14a highlights the 2D interactive discontinuity detection results, with No. 83 specifically marked in blue. The images used for reconstruction was captured from only three fixed-position camera locations, and No. 83 is nearly parallel to the camera orientations (see schematic diagram in Fig. 14b). This results in significant reconstruction gaps in the point cloud. Additionally, when the 2D results identified by SAM slightly exceed the actual discontinuity, this leads to the erroneous inclusion of adjacent discontinuities (marked by a circle in Fig. 14c). These inaccuracies originate from data acquisition limitations, suggesting that additional data collection in these regions is necessary. For improved 2D detection, image selection should prioritize those captured at a large angle relative to the outcrop surface orientation.

The relationship between pixel area and precision is further examined in Fig. 14d, indicating that the occurrence of low precision values decreases as discontinuity size increases. This reflects a geometric property whereby larger discontinuities are less affected by minor boundary deviations or local misalignments. As a result, the segmentation precision for larger structures tends to concentrate within a higher value range, reducing variance and improving overall reliability in evaluation. Conversely, the identification of smaller discontinuities requires greater caution and refinement in 2D discontinuity detection, due to their sensitivity to segmentation errors.

Finally, we used CloudCompare to automatically calculate the orientation of discontinuities. Each discontinuity orientation is converted into an upper unit normal vector (with $\boldsymbol{u}_z \geq 0$). Subsequently, the angle difference between identified discontinuities and manually annotated discontinuities, i.e., the error of orientation γ, is calculated using Eq. (2), where $\boldsymbol{n}_{\text{proposed}_{\text{f}}\text{ramework}}$ is the upper unit normal vector of identified discontinuities and $\boldsymbol{n}_{\text{ground}_{\text{t}}\text{ruth}}$ is the upper unit normal vector of manually annotated discontinuities [39].

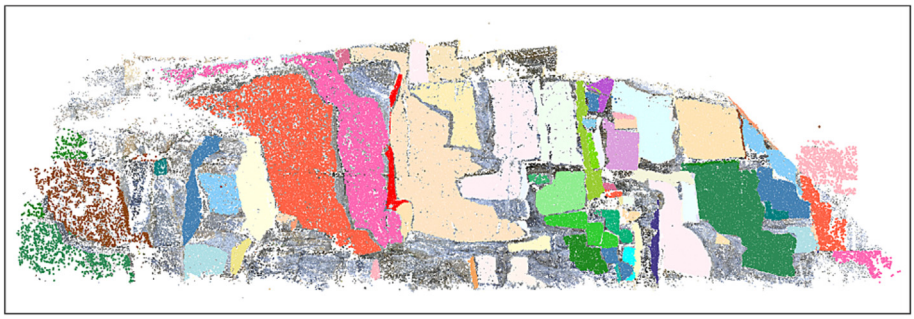
$$\gamma = \sin^{-1}\left(\frac{\left|\boldsymbol{n}_{\text{proposed}_{\text{f}}\text{ramework}} \times \boldsymbol{n}_{\text{ground}_{\text{t}}\text{ruth}}\right|}{\left|\boldsymbol{n}_{\text{proposed}_{\text{f}}\text{ramework}}\right| \cdot \left|\boldsymbol{n}_{\text{ground}_{\text{t}}\text{ruth}}\right|}\right) \tag{2}$$

Supplementary data §B Illustrates how to measure discontinuity orientation from 3D discontinuities in point cloud form Supplementary data §C presents a detailed comparison of the identified results and annotated results. Fig. 15 shows the frequency distribution and box plots of the error of orientation (the unit normal vector differences). From Fig. 15, the error range of 0° to 3.14° indicates a relatively small deviation between the identified and manually annotated discontinuities, which is comparable to the results discussed earlier. This suggests that the method is capable of accurately capturing the orientation of discontinuities. The underlying reason for the discrepancies is that the point clouds of the discontinuities are not entirely consistent. This result further validates the accuracy of the proposed framework

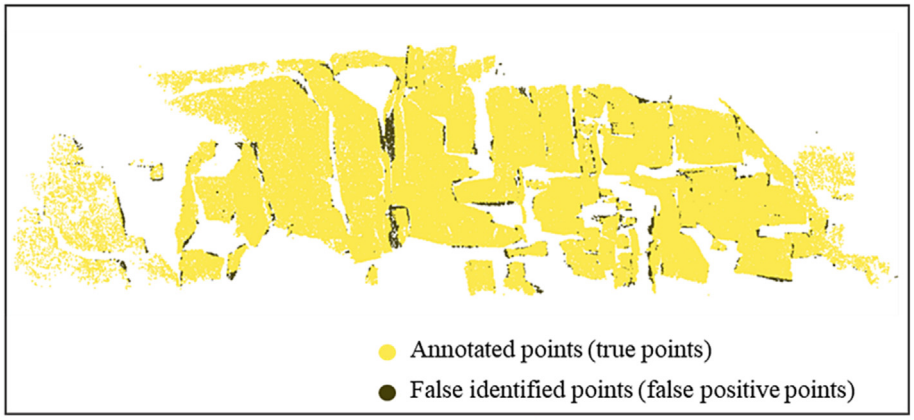
5. Discussion

5.1. Key considerations

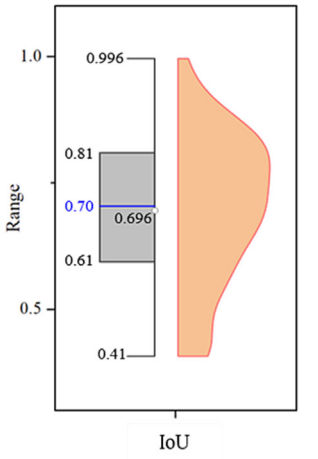
Three-dimensional discontinuity identification in photogrammetry follows three distinct pathways, as shown in Fig. 16. Unlike



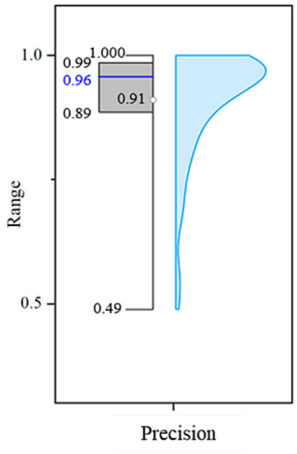
(a) 3D identified discontinuity results in point cloud forms



(b) False positive points are displayed on top of the annotated results



(c) Half-box-and-half-violin plot of IoU



(d) Half-box-and-half-violin plot of precision

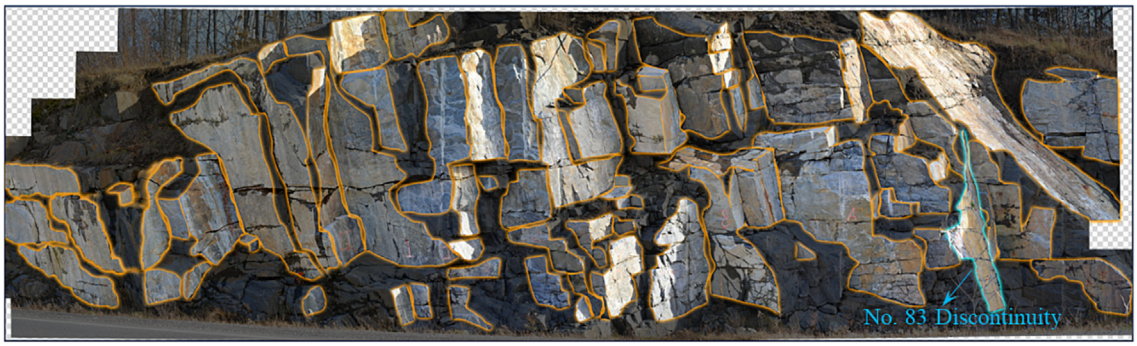
Fig. 13. 3D Discontinuity identification results.

LiDAR, which directly provides 3D point clouds, photogrammetry relies on 2D images as the raw data. For LiDAR, discontinuity identification is typically achieved through point segmentation. However, existing point segmentation methods are limited due to the lack of 3D annotated data, which restricts both their performance and generalization capabilities. Therefore, we opt for image segmentation using LSMs instead of point segmentation, as further discussed in Section 5.1.1.

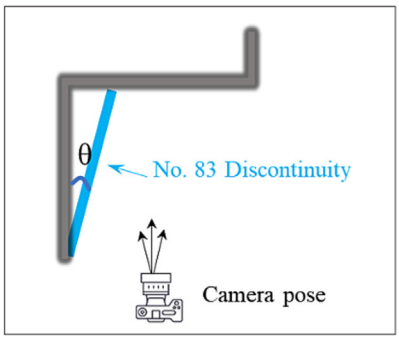
The requirements for 3D discontinuity identification include precision, feasibility, generalizability, and preferably low cost, along with timely response. Unlike autonomous vehicles, which often use RGB-D cameras for depth information, photogrammetry techniques require 3D reconstruction to obtain spatial data. Current 3D reconstruction software typically employs SfM and Multi-View Stereo (MVS) to generate depth maps. However, we find that MVS may not be suitable for rock mass reconstruction due to its inherent limitations. Additionally, the depth fusion process in MVS complicates the establishment of 2D-3D relationships. Therefore, we prioritize SfM as the 3D reconstruction module in the DIG framework, as discussed in Section 5.1.2. Given these considerations, the proposed DIG framework adopts the third pathway, combining the zero-shot generalizability of SAM with the precision of the pixel-point relationships generated by SfM.

5.1.1. The rationale for choosing image segmentation over point segmentation

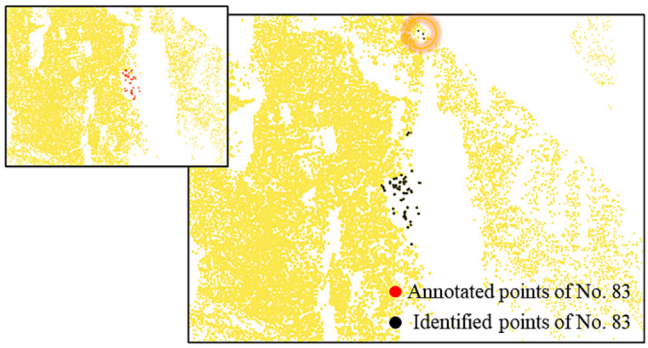
PointNet is utilized respectively for 3D discontinuity identification, employing these deep learning models to extract high-



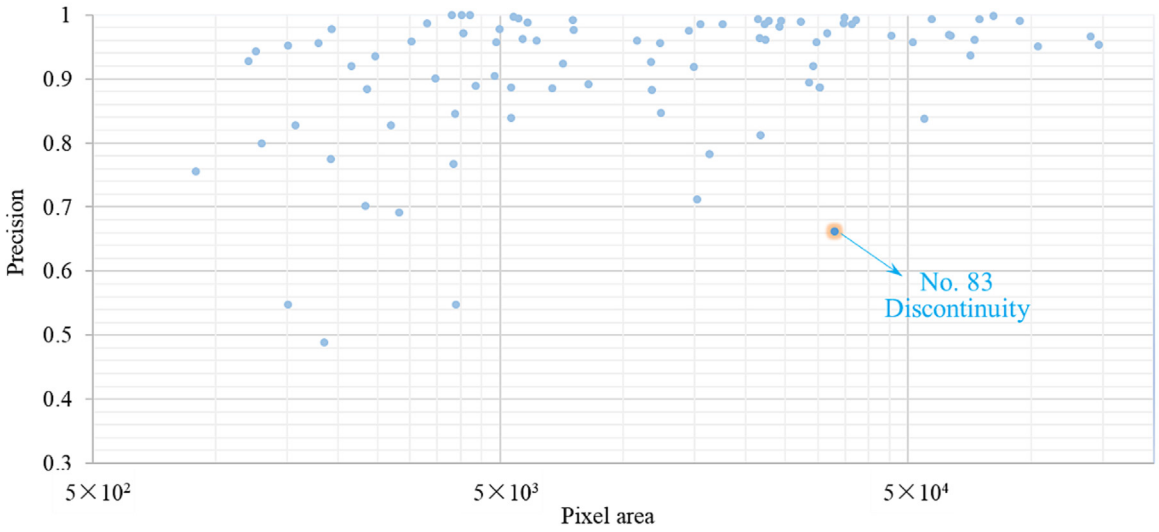
(a) 2D identified discontinuity results



(b) Schematic diagram of No.83 discontinuity and camera pose (top view)



(c) Identified points not belonging to No.83 discontinuity



(d) Scatter plot of precision vs. pixel area

Fig. 14. Analysis of 3D discontinuity identification results.

dimensional features from DOM and perform direct point cloud segmentation [40]. While effective on specific datasets, these methods struggle to accurately segment boundaries. Furthermore, their generalization capability is constrained by the limited availability of annotated 3D discontinuity data.

The inherent complexity of point clouds, compared to images, has hindered point segmentation from achieving the success of image segmentation. Challenges such as inconsistent data formats, poor model scalability, and insufficient annotated masks further limit the performance of 3D models. These issues prevent them from attaining results comparable to SAM. Consequently, we avoid direct point segmentation and instead adopt image segmentation. As previously mentioned, the DIG framework not only performs recognition tasks but also acts as a data engine for collecting large-scale annotated data from discontinuity point clouds.

5.1.2. The rationale for not using MVS in reconstruction

MVS, which uses calibrated multi-view images to reconstruct dense 3D geometry through image overlap, is referred to as dense reconstruction. It is typically implemented after SfM in the 3D reconstruction workflow. Depth map-based MVS is the most commonly implemented in reconstruction software [14]. As shown in Fig. 17, it uses patch matching to estimate the depth maps for indi-

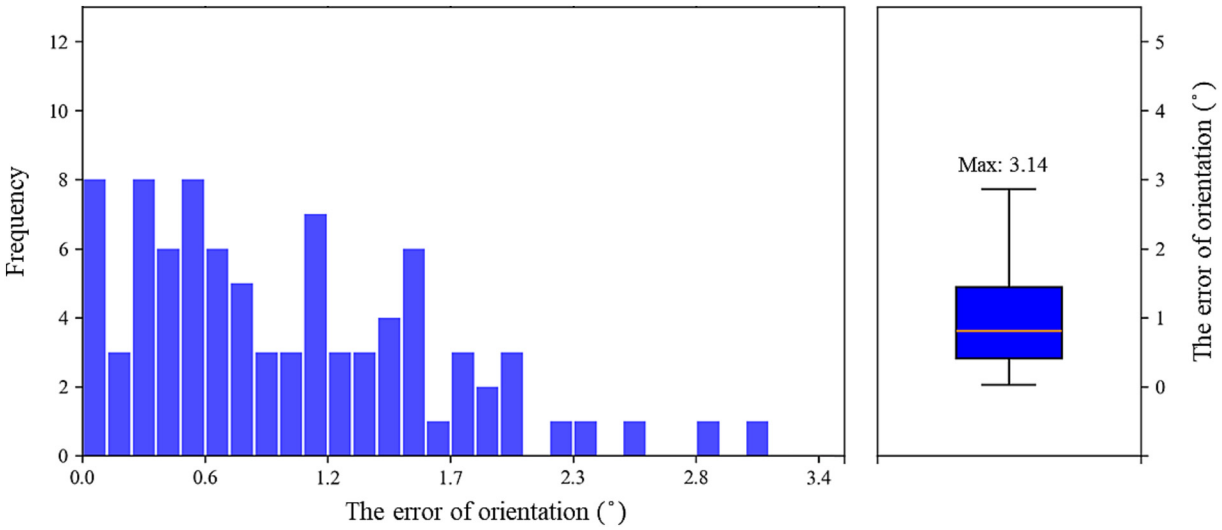


Fig. 15. Frequency distribution and boxplot of the error of orientation (the angle differences of the upper unit normal vectors) between identified discontinuities and manually annotated discontinuities.

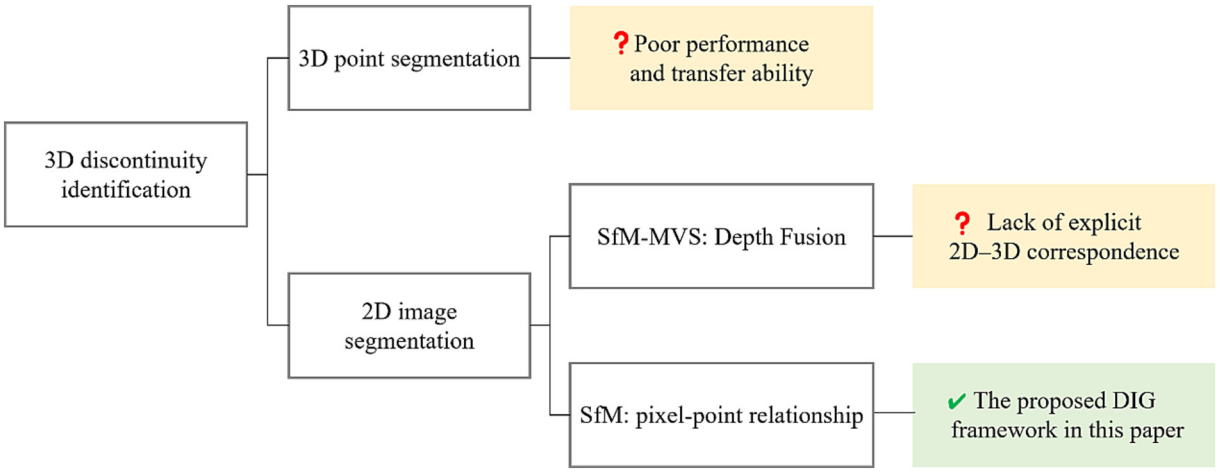


Fig. 16. Overview of the three pathways for 3D discontinuity identification.

vidual images and then fuses these depth maps into a dense point cloud [41].

Despite its widespread use in rock mass characterization (e.g., through dense point clouds to characterize surface roughness [42]), depth map-based MVS still exhibits inherent instability in dense reconstruction. This is primarily because it relies on "photometric consistency", which suffers under conditions such as illumination changes, non-Lambertian reflection surfaces (i.e., non-uniform reflection surfaces), and low-to-non-textured regions [43]. Specifically, outdoor strong sunlight, shadow variations, and near-mirror reflective surfaces (e.g., quartz surfaces) cause modeling errors in dense reconstructions [44]. In contrast to MVS, which relies on photometric consistency, SfM leverages geometric consistency and focuses on the positional changes of local visual features (e.g., SIFT, A-KAZE) across different views. This makes SfM inherently more robust under the aforementioned conditions, and is one of the key reasons we adopt SfM in the DIG framework.

Additionally, while depth map-based MVS generates depth maps for each image, the depth information does not directly align with the final dense point cloud after fusion. Fusion algorithms, which use weighted averaging, maximum distance criteria, or other post-processing techniques, aim to minimize errors in the dense point cloud [45]. As a result, MVS lacks explicit and preserved pixel-point relationships. In contrast, SfM maintains the generation and tracking of pixel-to-point relationships throughout the reconstruction process, and this mapping is explicitly preserved in the final output. Moreover, the overall workflow of SfM is lighter and less computationally intensive than that of SfM-MVS. While SfM alone yields a sparse point cloud and does not reconstruct non-feature-based regions, this limitation can be mitigated by capturing higher-spatial-resolution images and using stricter parameters for feature descriptor density and quality, thereby increasing the number and coverage of matched features and improving spatial completeness. Therefore, the proposed DIG framework prioritizes SfM as the 3D reconstruction module to ensure reliable geometric mapping and reduce model complexity without compromising discontinuity projection accuracy.

5.2. Advantages and limitations

5.2.1. Advantages

The DIG framework offers a generalizable, "one-for-all" solution that can be applied across a variety of real-world outcrops without the need for parameter tuning or model retraining. This is particu-

larly advantageous for geological surveys, as it shifts the focus from discontinuity identification to more specialized tasks, such as discontinuity characterization or stability assessment. Additionally, this approach supports efficient dataset construction, enabling the development of specialized LSMs, which can improve identification accuracy in the long run.

Another advantage of the DIG framework lies in its parallel architecture. 2D discontinuity detection via SAM and 3D reconstruction via SfM proceed in parallel, which helps eliminate the cumulative errors typically associated with traditional serial methods and also enhances operational efficiency. By interactively refining 2D masks and adjusting 3D reconstruction parameters (Table 3), robust results can be achieved efficiently. In terms of runtime, the total computational load is dominated by SfM modeling (hours-level), while discontinuity identification via indexing operates in seconds. This makes the DIG framework highly suitable for field applications.

Moreover, the DIG framework prioritizes ease of implementation, requiring minimal expertise. In the interactive discontinuity detection phase, users need only basic familiarity with discontinuities and simple mouse interactions to generate responses in real-time, significantly lowering the operational barriers.

Additionally, the framework provides an intuitive method for assessing result accuracy and model performance. Compared to direct comparisons of 3D discontinuities in point clouds, the use of image-based IoU metrics offers a more feasible evaluation approach. This is especially beneficial in complex geological scenarios, where manual annotation in point clouds may not be practical.

Furthermore, the framework generates high-quality discontinuity results that are free from noise and exhibit clearly defined boundaries. The mIoU and AP reach 0.70 and 0.91, respectively, between the detected 3D discontinuities and the manually annotated ground truth in the proposed Dataset-3D. This facilitates subsequent tasks and ensures data reusability, providing a solid foundation for further analyses.

5.2.2. Limitations

There are certain limitations to the DIG framework. First, its performance can be constrained in specific scenarios, such as when occlusions or heavy vegetation interfere with visual data, making 2D discontinuity detection less effective.

Moreover, the interactive point-prompt process (based on zero-shot prompting with SAM) can be labor-intensive in geologically complex scenes. This is primarily because discontinuity features differ significantly from the object categories used during SAM's original training (e.g., the SA-1B dataset), resulting in a modest performance drop in these domain-specific applications. As a result, more user prompts are often needed to achieve accurate segmentation of discontinuities, which highlights the limitations of applying general-purpose models in specialized domains. The proposed DIG framework is a necessary transition step toward robust automatic solutions. Future development will focus on reducing manual interaction through model fine-tuning and semi-automated interfaces, as detailed in Section 5.3.

Another limitation of the framework is its reliance on sufficient image-viewing angles. When discontinuity surfaces are nearly perpendicular to the camera orientation, as in a certain case illustrated in Fig. 14c, segmentation and reconstruction accuracy can degrade due to limited surface visibility. This issue could be partially addressed by capturing images from more diverse angles. Recent work by Mehrishal et al. [46] introduces an alternative approach by incorporating virtual physical cameras and image rendering from textured 3D models and generating multiple views with minimal input images, which offers a promising solution to improve coverage and accuracy in future iterations of the DIG framework.

5.3. Future work and potential applications

The DIG framework's evolution toward full automation presents two distinct implementation pathways. The first focuses on advancing an automated version based on 2D image datasets, aiming to reduce the need for manual prompts. As part of this, we plan to develop an API integration with SAM to improve labeling efficiency. The second, more ambitious DIG2.0 variant aims to achieve direct segmentation on 3D point clouds without relying on image segmentation. Realizing this capability requires the development of specialized, labeled 3D discontinuity datasets and the design of point-cloud-based model architectures.

While the concept of full automation is appealing, it remains currently infeasible due to the model's limited capacity to interpret complex discontinuities without human guidance. This human oversight enables the model to adapt to the specific visual and geometric challenges of each discontinuity. To achieve robust automation, future development should rely on establishing a diverse, well-annotated discontinuity dataset capable of supporting fine-tuning large-scale models.

The scalability of the DIG framework can be extended to additional geological targets, including discontinuities (e.g., complex trace-type features) and in-situ blocks (rock-mass blocks). This could be achieved by further exploring the capabilities of LSMs in fine-grained pattern recognition. Recent work on in-situ block intelligent identification (SA4B-3D) shows that LSMs can also be adapted for large-scale block extraction and volumetric analysis

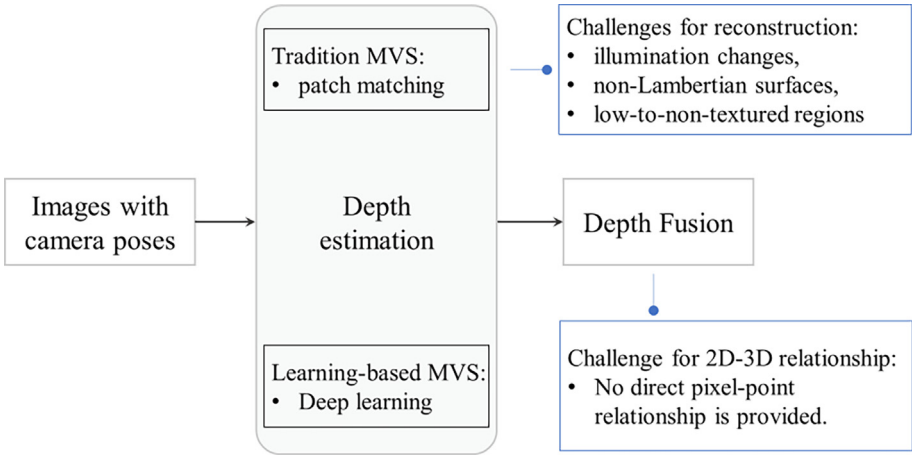


Fig. 17. Workflow diagram of depth map-based MVS method.

[47]. Moreover, the DIG framework is expected to evolve beyond identification, toward detailed discontinuity characterization and stability assessment. The automated extraction of discontinuity parameters from 3D point clouds would support stability evaluation and decision-making in geological engineering.

6. Conclusions

We recognize that the essence of the 3D discontinuity identification problem is segmentation. Numerous identification methods have emerged, yet they are constrained by the complexity of rock masses and their digital twins (i.e., DOMs). The existing methods fall short in addressing this complexity. Moreover, deep learning models with limited data cannot resolve the complex out-of-domain identification problem. This results in a situation where, despite the development of a vast array of algorithms and models, the challenges of discontinuity identification remain largely unresolved. In this paper, we propose a unified framework, DIG, that integrates SAM with SfM for 3D discontinuity identification in diverse real-world outcrops. The concluding remarks are summarized as follows:

(1) We systematically compare DIG with conventional DOM-, vision-, and ML-based methods (Table 1) and highlight the paradigm shift brought by large-scale models (Table 2). SAM was selected as a representative LSM due to its zero-shot capability, scale of training data, and robust segmentation accuracy, making it the top priority for geological discontinuity identification.

(2) To validate the proposed framework, we construct novel benchmark datasets for both 2D and 3D discontinuities. We introduce a two-step standard validation protocol to assess the effectiveness of the proposed framework, focusing on the generalization of 2D discontinuity detection and the accuracy of 3D discontinuity identification, yielding quantitative results of mIoU=0.78/AP=0.86 for 2D discontinuities and mIoU=0.70/AP=0.91 for 3D discontinuities. Runtime analysis further shows that interactive detection can be achieved in seconds, demonstrating its practicality.

(3) DIG demonstrates efficiency and scalability by transforming 3D point cloud segmentation into 2D image segmentation, preserving pixel-point correspondences for robust projection into 3D. This framework is expected to evolve toward recognition of more subtle forms (e.g., traces, blocks) and automated discontinuity characterization. The continuous integration of general-purpose LSMs with big data engines will further accelerate the move toward standardized, automated geological methods incorporating spatial intelligence.

Acknowledgments

The authors would like to thank Dr. Matthew J. Lato for providing raw data of the Roadcut in Canada, as well as Dr. Hamid Daghigh for his support in dataset preparation. This study was funded by National Natural Science Foundation of China (Nos. 42422704 and 52379109), Opening the fund of State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (Chengdu University of Technology) (No. SKLGP2024K028), and Science and Technology Research and Design Projects of China State Construction Engineering Corporation Ltd. (No. CSCEC-2024-Q-68).

Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmst.2025.09.005.

References

[1] Hudson JA, Priest SD. Discontinuities and rock mass geometry. Int J Rock Mech Min Sci Geomech Abstr 1979;16(6):339–62.
[2] Zhu HH, Azarafza M, Akgün H. Deep learning-based key-block classification framework for discontinuous rock slopes. J Rock Mech Geotech Eng 2022;14(4):1131–9.
[3] Zheng J, Guo JC, Wang JC, Sun HL, Deng JH, Lv Q. A universal elliptical disc (UED) model to represent natural rock fractures. Int J Min Sci Technol 2022;32(2):261–70.
[4] Azarafza M, Ghazifard A, Akgün H, Asghari-Kaljahi E. Development of a 2D and 3D computational algorithm for discontinuity structural geometry identification by artificial intelligence based on image processing techniques. Bull Eng Geol Environ 2019;78(5):3371–83.
[5] Azarafza M, Koçkar MK, Faramarzi L. Spacing and block volume estimation in discontinuous rock masses using image processing technique: A case study. Environ Earth Sci 2021;80(14):471.
[6] Zare Naghadehi M, Jimenez R, KhaloKakaie R, Jalali SE. A new open-pit mine slope instability index defined using the improved rock engineering systems approach. Int J Rock Mech Min Sci 2013;61:1–14.
[7] Zare M, Jimenez R. On the development of a slope instability index for open-pit mines using an improved systems approach. In: Proceedings of the ISRM Regional Symposium-EUROCK 2015. Salzburg: ISRM; 2015:ISRM-EUROCK-2015-170.
[8] Palmstrom A. Measurement and Characterization of Rock Mass Jointing. In: Proceedings of the In-Situ Characterization of Rocks. London: A.A. Balkema Publishers; 2001.p.97.
[9] Xu XM, Aiken CLV, Bhattacharya JP, Corbeanu RM, Nielsen KC, McMechan GA, Abdelsalam MG. Creating virtual 3-D outcrop. Lead Edge 2000;19(2):197–202.
[10] Bellian JA, Kerans C, Jennette DC. Digital outcrop models: Applications of terrestrial scanning lidar technology in stratigraphic modeling. J Sediment Res 2005;75(2):166–76.
[11] Riquelme AJ, Abellán A, Tomás R, Jaboyedoff M. A new approach for semi-automatic rock mass joints recognition from 3D point clouds. Comput Geosci 2014;68:38–52.
[12] Haneberg WC. Using close range terrestrial digital photogrammetry for 3-D rock slope modeling and discontinuity mapping in the United States. Bull Eng Geol Environ 2008;67(4):457–69.
[13] Daghigh H, Tannant DD, Daghigh V, Lichti DD, Lindenbergh R. A critical review of discontinuity plane extraction from 3D point cloud data of rock mass surfaces. Comput Geosci 2022;169:105241.
[14] Battulwar R, Zare-Naghadehi M, Emami E, Sattarvand J. A state-of-the-art review of automated extraction of rock mass discontinuity characteristics using three-dimensional surface models. J Rock Mech Geotech Eng 2021;13(4):920–36.
[15] Wang MY, Wang EZ, Liu XL, Wang CC. Scale-space effect and scale hybridization in image intelligent recognition of geological discontinuities on rock slopes. J Rock Mech Geotech Eng 2024;16(4):1315–36.
[16] Ma J, He YT, Li FF, Han L, You CY, Wang B. Segment anything in medical images. Nat Commun 2024;15(1):654.
[17] Osco LP, Wu QS, de Lemos EL, Gonçalves WN, Ramos APM, Li J, Junior JM. The Segment Anything Model (SAM) for remote sensing applications: from zero to one shot. Int J Appl Earth Obs Geoinf 2023;124:103540.
[18] Awais M, Naseer M, Khan S, Anwer RM, Cholakkal H, Shah M. Foundation models defining a new era in vision: a survey and outlook. IEEE Trans Pattern Anal Mach Intell 2025;47(4):2245–64.
[19] Kirillov A, Mintun E, Ravi N, Mao HZ, Rolland C, Gustafson L, Xiao TT, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. Segment anything. arXiv:230402643 [Preprint]. 2023 [cited 2025 Jun 1]: [30 p.]. Available from: https://arxiv.org/abs/2304.02643.
[20] Lee YK, Kim J, Choi CS, Song JJ. Semi-automatic calculation of joint trace length from digital images based on deep learning and data structuring techniques. Int J Rock Mech Min Sci 2022;149:104981.
[21] Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, Brynjolfsson E, Buch S, Card D, Castellon R, Chatterji N. On the opportunities and risks of foundation models. arXiv:210807258 [Preprint]. 2022 [cited 2025 Jun 1]: [214 p.]. Available from: https://arxiv.org/abs/2108.07258.
[22] Jain J, Li JC, Chiu MT, Hassani A, Orlov N, Shi H. OneFormer: One transformer to rule universal image segmentation. arXiv:221106220 [Preprint]. 2022 [cited 2025 Jun 1]: [19 p.]. Available from: https://arxiv.org/abs/2211.06220.
[23] Li H, Zhu JG, Jiang XH, Zhu XZ, Li HS, Yuan C, Wang XH, Qiao Y, Wang XG, Wang WH, Dai JF. Uni-Perceiver v2: A generalist model for large-scale vision and vision-language tasks. arXiv:221109808 [Preprint]. 2022 [cited 2025 Jun 1]: [14 p.]. Available from: https://arxiv.org/abs/2211.09808.
[24] Wang XL, Zhang XS, Cao Y, Wang W, Shen CH, Huang TJ. SegGPT: Segmenting everything in context. arXiv:230403284 [Preprint]. 2023 [cited 2025 Jun 1]: [12 p.]. Available from: https://arxiv.org/abs/2304.03284.
[25] Zou XY, Yang JW, Zhang H, Li F, Li LJ, Wang JF, Wang LJ, Gao JF, Lee JY. Segment everything everywhere all at once. arXiv:230406718 [Preprint]. 2023 [cited 2025 Jun 1]: [13 p.]. Available from: https://arxiv.org/abs/2304.06718.
[26] Moulon P, Monasse P, Marlet R. Adaptive structure from motion with a contrario model estimation. Computer Vision – ACCV 2012. Berlin, Heidelberg: Springer, 2013:257–70.

[27] Microsoft. Image Composite Editor [software]. [software].2015 Dec 02 [cited 2025 Jun 1]. Available from: http://research.microsoft.com/en-us/um/redmond/projects/ice/.
[28] Gledhill D, Tian GY, Taylor D, Clarke D. Panoramic imaging: A review. Comput Graph 2003;27(3):435–45.
[29] Bemis SP, Micklethwaite S, Turner D, James MR, Akciz S, Thiele ST, Bangash HA. Ground-based and UAV-Based photogrammetry: A multi-scale, high-resolution mapping tool for structural geology and paleoseismology. J Struct Geol 2014;69:163–78.
[30] Griwodz C, Gasparini S, Calvet L, Gurdjos P, Castan F, Maujean B, De Lillo G, Lanthony Y. In: AliceVision Meshroom: An open-source 3D reconstruction pipeline. Istanbul Turkey: ACM; 2021:. p. 241–7.
[31] Agisoft LLC. Agisoft Metashape [software]. 1.2.6 [software][cited 2025 Jun 1]. Available from: https://www.agisoft.com/.
[32] Rey OI. Anatomy of the SIFT method. Doctoral dissertation Paris: Ecole Normale Supérieure de Cachan 2015.
[33] Alcantarilla PF, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces 2013;13:.1–13.11..
[34] Sweeney C, Hollerer T, Theia TM. In: A fast and scalable structure-from-motion library. Brisbane Australia: ACM; 2015. p. 693–6.
[35] Russell BC, Torralba A, Murphy KP, Freeman WT. LabelMe: A database and web-based tool for image annotation. Int J Comput Vis 2008;77(1):157–73.
[36] Xu N, Price B, Cohen S, Yang JM, Huang T. In: Deep interactive object selection. Las Vegas: IEEE; 2016. p. 373–81.
[37] CloudCompare. CloudCompare [software]. Version 2.13.2 [software].2024 Jul 11 [cited 2025 Jun 1]. Available from: https://www.cloudcompare.org/.
[38] Lato M, Kemeny J, Harrap RM, Bevan G. Rock bench: establishing a common repository and standards for assessing rockmass characteristics using LiDAR and photogrammetry. Comput Geosci 2013;50:106–14.
[39] Zhang BH, Li WK, Zheng J, Kulatilake PHSW, Qing L, Duan YS. A numerical simulation procedure for evaluating the accuracy of 3-dimensional photogrammetric models and its application to geometric parameters of discontinuities in rock masses. Phys Chem Earth Parts A/B/C 2023;129:103364.
[40] Battulwar R, Emami E, Naghadehi MZ, Sattarvand J. Automatic extraction of joint orientations in rock mass using PointNet and DBSCAN. advances in Visual Computing. Cham: Springer International Publishing; 2020. p. 718–27.
[41] Xu QS, Tao WB. In: Multi-scale geometric consistency guided multi-view stereo. Long Beach: IEEE; 2019. p. 5478–87.
[42] Paixão A, Muralha J, Resende R, Fortunato E. Close-range photogrammetry for 3D rock joint roughness evaluation. Rock Mech Rock Eng 2022;55(6):3213–33.
[43] Xiang X, Wang ZY, Lao SS, Zhang BC. Pruning multi-view stereo net for efficient 3D reconstruction. ISPRS J Photogramm Remote Sens 2020;168:17–27.
[44] Yang WQ, Chen GY, Chen CF, Chen ZF, Wong KK. PS-NeRF: Neural inverse rendering for multi-view photometric stereo. In: Computer Vision – ECCV 2022. Cham: Springer Nature Switzerland; 2022:266–84.
[45] Wang FJH, Zhu QT, Chang D, Gao QK, Han JL, Zhang T, Hartley R, Pollefeys M. Learning-based multi-view stereo: a survey. arXiv preprint 2024: arXiv:2408.15235.
[46] Mehrishal S, Kim J, Song JJ, Sainoki A. A semi-automatic approach for joint orientation recognition using 3D trace network analysis. Eng Geol 2024;332:107462.
[47] Wang JW, Zheng J, Hu J, Jimenez R, Zhu AL, Lü Q, Wang JC. In-situ block intelligent identification: Large-scale artificial intelligence (AI) models enable efficient geological surveys via image segmentation. Eng Geol 2025;356:108286.