Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Automatic fracture recognition and steady-state quantification in forward-looking borehole videos via deep learning

Wenhui Zhang
China University of Mining and Technology
Dongsheng Zhang
dshzhang123@126.com
China University of Mining and Technology
Gangwei Fan
China University of Mining and Technology
Yanzi Miao
China University of Mining and Technology
Zhanglei Fan
China University of Mining and Technology

Article

Keywords: fracture recognition, forward-looking borehole video, deep learning, steady-state quantification

Posted Date: March 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8964613/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Automatic fracture recognition and steady-state quantification in forward-looking borehole videos via deep learning

Wenhui Zhang [a], Dongsheng Zhang [a,b,d,*], Gangwei Fan [a,b], Yanzi Miao [c], Zhanglei Fan [a,b]

[a] School of Mines, China University of Mining and Technology, Xuzhou, 221116, China

[b] State Key Laboratory for Fine Exploration and Intelligent Development of Coal Resources, Xuzhou, 221116, China

[c] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221116, China

[d] Mining Engineering and Geology Collegy, Xinjiang Institute of Engineering, Urumqi, 830023, China

* Correspondence: dshzhang123@126.com; Tel.: 13182307782

Abstract: Automatic recognition and quantitative characterization of rock mass fractures are critical for surrounding rock classification, stability analysis, and geological hazard risk assessment. To address inconsistencies in frame-by-frame measurements caused by variations in visible scale and local morphology in forward-looking borehole videos, as well as geometric distortion from panoramic unwrapping that limits the accuracy of full-length quantification, this study introduces a raw video-domain approach for automatic fracture recognition and steady-state parameter quantification. A joint detection–segmentation strategy is implemented to obtain pixel-level fracture masks. In the video domain, continuous cross-frame trajectories of individual fractures are established by integrating feature matching and optical-flow tracking with a consistency verification criterion. Multi-frame fusion is then performed to convert instantaneous single-frame geometric measurements into steady-state video-domain outputs. Validation experiments on a self-constructed dataset of 55 borehole videos demonstrate that, on the validation set, the detection mP@0.5 and mR@0.5 reach 90.76% and 88.20%, respectively, while the segmentation Dice score is 82.51%. The cross-frame association success rate remains stable at 94.5%, with a drift error below 2.6 pixels, enabling continuous and reliable fracture trajectories. The fused geometric parameters exhibit strong agreement with manual measurements ($R^2 > 0.995$), and the quantification error decreases and converges as the number of fused frames increases. Compared with three panoramic image-domain quantification methods, the proposed approach reduces mean quantification errors in length, width, and orientation by 52.9%, 58.1%, and 45.2%, respectively, effectively mitigating distortion-induced errors from panoramic unwrapping. These results indicate that the proposed method enables stable video-domain fracture recognition, consistent cross-frame modeling, and steady-state parameter output, thereby providing reliable inputs for engineering fracture quantification and safety assessment decision-making.

Keywords: fracture recognition; forward-looking borehole video; deep learning; steady-state quantification

1. Introduction

Accurate recognition and quantitative characterization of fracture geometric features are essential for preliminary assessments in geotechnical and mining engineering [1-4]. The degree of fracture development and its geometric morphology directly affect the integrity and load-bearing capacity of the surrounding rock [5,6] and serve as critical evidence for rock mass classification [7], stability analysis [8], and geological hazard risk assessment [9]. Therefore, once reliable fracture recognition is achieved, robust and stable quantification of geometric parameters, which enables the transition from qualitative recognition to quantitative description, provides significant engineering value.

Advancements in in-borehole imaging technologies, including borehole television and endoscopy, have established borehole imaging as a primary method for acquiring information on rock fractures [10]. Forward-looking borehole video [11,12] is widely used in fracture investigations due to its efficient data acquisition, continuous recording, and cost-effectiveness. However, these videos are affected by limited viewing angles, variations in illumination, and changes in camera pose [13,14]. As a result, the visible scale and local morphology of the same fracture can vary considerably across frames, leading to temporal inconsistencies in recognition and quantification outcomes and complicating the stable quantification of fractures.

Research on fracture recognition and quantification has evolved from traditional image processing methods to deep-learning-based approaches [15-17]. Conventional methods typically rely on edge, texture, or geometric cues [18-20] extracted from panoramic borehole wall images to detect fractures and estimate their parameters, but often lack robustness in complex backgrounds, low-texture regions, and noisy environments. In contrast, deep learning methods enhance fracture representation through object detection [21-23] or semantic segmentation [24,25] and have been increasingly

applied to forward-looking borehole imaging, resulting in notable improvements in fracture extraction accuracy. Despite these advancements, most progress has focused on fracture recognition, while reliable quantification for underground engineering applications remains insufficiently addressed. Cross-frame fluctuations and inter-frame inconsistencies in quantified parameters continue to affect real-world data.

To enable full-interval visualization and manual logging, many studies convert borehole videos into unfolded panoramic wall images before performing recognition and quantification [26-28]. However, the unwrapping and stitching process often introduces geometric distortion, texture stretching, and detail loss [29,30], which systematically alter fracture morphology and propagate errors into the estimation of length, width, and orientation. These issues limit the accuracy of quantification, particularly under scale variations, lens eccentricity, and nonuniform illumination. Even when recognition and measurement are performed directly on raw video frames [31,32], the same fracture may exhibit abrupt parameter variations due to changes in viewing distance and fluctuations in segmentation boundaries, reducing the reliability of the results for engineering analysis. Overall, most existing studies remain at the frame-by-frame recognition and measurement stage, lacking (i) a cross-frame association mechanism to link observations of the same fracture consistently and (ii) a reproducible pipeline and criteria for transforming multi-frame observations into steady-state parameter outputs. This gap represents a key bottleneck in the engineering deployment of fracture recognition and quantification.

To address these challenges, this study proposes a deep learning-based approach for automatic fracture recognition and steady-state parameter quantification in raw forward-looking borehole videos. The method initially applies a joint detection and segmentation strategy to generate pixel-level fracture masks, providing a stable morphological basis for geometric measurement. Subsequently, it constructs continuous trajectories of the same fracture across video frames through cross-frame association, reducing mismatches and drift via consistency-based filtering. Finally, multi-frame fusion converts instantaneous single-frame measurements of fracture length, width, and orientation into steady-state video-domain outputs, enabling reliable full-interval quantification without the need for panoramic unwrapping.

2. Methods

2.1. Video-domain automatic fracture recognition

Reliable quantification of fracture geometric parameters in forward-looking borehole videos requires pixel-level representations of fractures with well-defined boundaries and stable morphology. Because width and angle measurements are sensitive to segmentation boundaries, increasing the per-frame detection rate alone does not ensure stable quantification [33]. Therefore, the detection model is initially employed to robustly localize the fracture area within complex backgrounds, followed by the segmentation model, which generates a high-fidelity mask. This approach provides a stable morphological template for consistent association of skeletons and widths across frames.

2.1.1. YOLOv8-CAPS detection

During the detection stage, the objective is to generate candidate bounding boxes and to identify fracture candidates with high recall and minimal missed detections for subsequent segmentation and temporal association. This approach prevents fracture track breakage over time, which may result in underestimated lengths and failure of steady-state fusion. Accordingly, a Capsule [34] (CAPS) module is integrated into the multi-scale feature fusion path of YOLOv8 [35] to enhance the separability of minor and weak-texture fractures and to enable spatially continuous responses.

As illustrated in Figure 1, at the feature fusion node following upsampling from P4 to P3, the upsampled features (40×40×512) are first reduced in channel dimension by a 1×1 convolution and then concatenated with the P3 features (80×80×256) to form a fused feature map (80×80×512). The C2f module processes this fused feature map to extract local features, which are subsequently input into the CAPS module. The CAPS module performs vectorized encoding of fused features [36], enabling dynamic binding of spatial and semantic information through a two-level structure comprising Primary Caps and Routing Caps. This configuration allows the model to generate a continuous, stable, high-response feature band in fracture regions.

The computation of the CAPS module [37] is depicted in Figure 2. Each CAPS module consists of two levels: Primary Caps and Routing Caps. Primary Caps generates 32 primary capsules of eight dimensions using a 1×1 convolution, while Routing Caps produces 32 higher-level capsules of eight dimensions through a dynamic routing mechanism with r=3 iterations. The squash nonlinearity compresses the output capsules, which are then restored to 256 channels via a 1×1 convolution, thereby preserving the structure of the original YOLOv8 detection head. With only an 8% increase in computational cost, the proposed module improves the model's fracture recognition performance and spatial continuity.

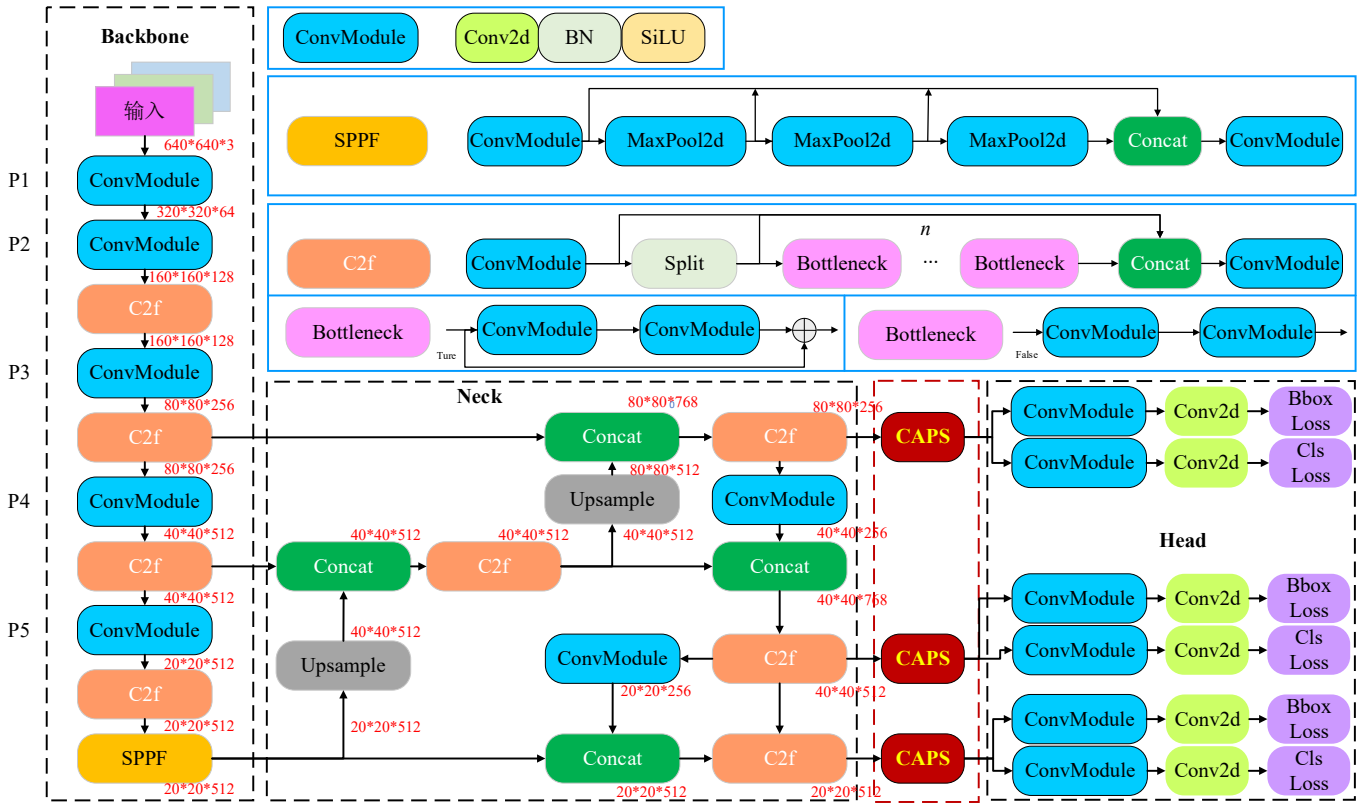


Figure 1. The structure of YOLOv8-CAPS.

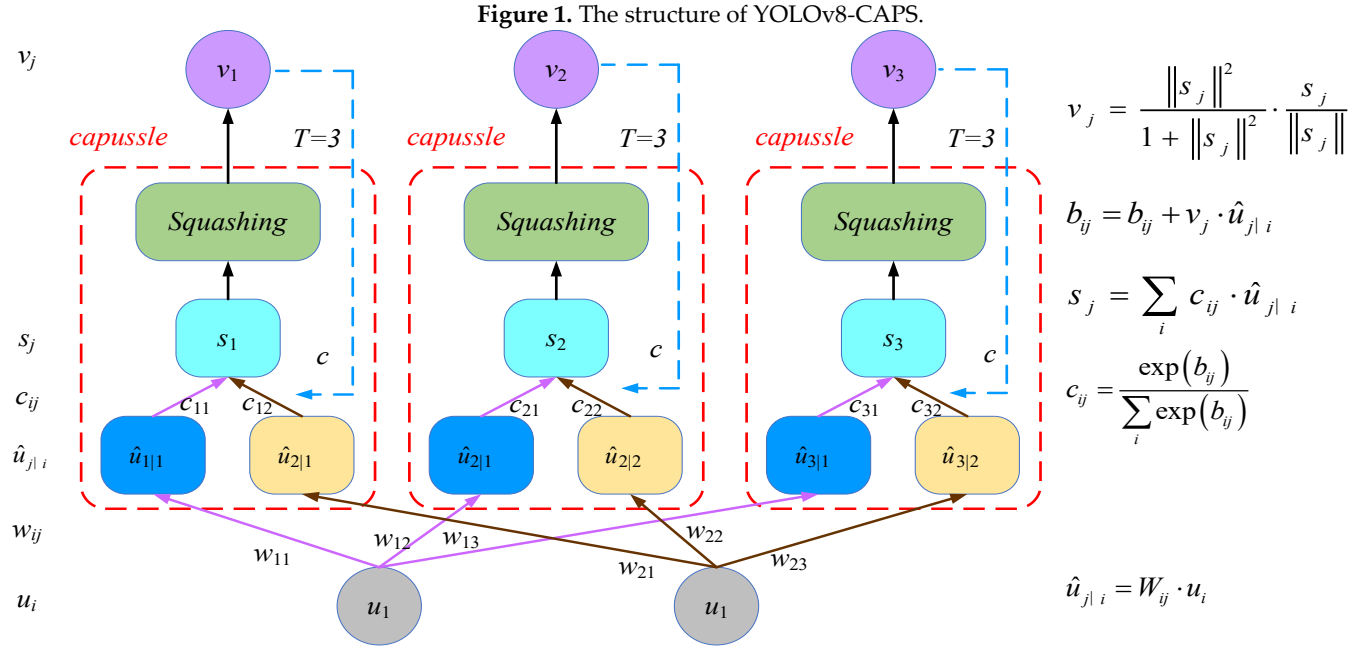


Figure 2. The structure of the capsule network.

2.1.2 Mask R-CNN segmentation with boundary-weighted supervision

The segmentation stage significantly affects the stability of the mask boundary and is critical for minimizing width jitter and skeleton drift. To address boundary breakage resulting from the low proportion of boundary pixels and their grayscale similarity to the background, a Weighted Binary Cross-Entropy (WBCE) boundary-weighted loss is incorporated into the mask branch of Mask R-CNN. By strengthening supervision in boundary regions, this loss function enhances mask continuity and geometric accuracy, thereby reducing systematic errors in subsequent steady-state quantification.

Mask R-CNN [38] extracts features through region pooling, utilizes a classification head and a regression head for object detection and localization, and employs a mask head to generate precise semantic segmentation masks, thereby enabling pixel-level instance segmentation. The overall network architecture is illustrated in Figure 3.

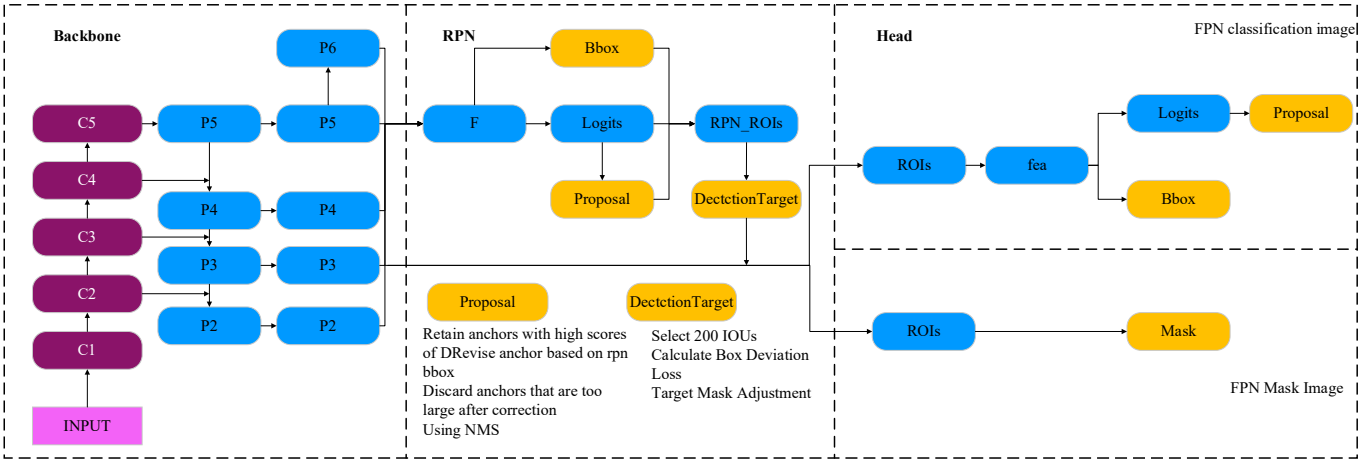


Figure 3. The structure of Mask R-CNN.

Fracture boundary pixels are typically sparse and susceptible to misclassification as background, which often results in blurred and discontinuous boundaries when only standard cross-entropy loss is applied. To address this issue, a weighted boundary cross-entropy (WBCE) loss [39] is introduced. This approach utilizes a weight matrix to assign greater importance to fracture-boundary regions, thereby enhancing the model's focus on these areas and improving segmentation accuracy. The WBCE loss is defined as follows:

$$L_{WBCE} = -\frac{1}{N}\sum_{i=1}^{N} w_i \left[y_i \log p_i + (1-y_i)\log(1-p_i) \right] \tag{1}$$

where LWBCE denotes the boundary-weighted cross-entropy loss; N is the total number of pixels used in the loss computation; y_i is the ground-truth label of pixel i (0 for background and 1 for fracture); p_i is the predicted probability that pixel i belongs to the fracture class; and w_i is the weight assigned to pixel i in the weight matrix, $w_i=W(u_i, v_i)$, with (u_i, v_i) representing the coordinates of pixel i.

The weight matrix enhances boundary supervision by assigning greater loss weights to pixels situated on or near the fracture boundary. The binary ground-truth (GT) mask is denoted as $Y(u_i, v_i)\in\{0,1\}$. A boundary map $B(u_i, v_i)\in\{0,1\}$ derived from Y using a neighborhood-consistency criterion. For each pixel (u_i, v_i), the number of foreground pixels within its 3×3 neighborhood is calculated as follows:

$$S(u_i, v_i) = \sum_{m=-1}^{1}\sum_{n=-1}^{1} Y(u_i+m, v_i+n) \tag{2}$$

Here, $S(u_i, v_i)\in[0,9]$ measures the local foreground occupancy around (u_i, v_i); a larger value indicates that the local region is more consistently inside the fracture mask.

A boundary map $B(u_i, v_i)\in\{0,1\}$ is extracted from the GT mask by identifying foreground pixels whose neighborhood is not entirely occupied by fracture pixels:

$$B(u_i, v_i) = \begin{cases} 1, & Y(u_i, v_i) = 1 \,\&\, S(u_i, v_i) < 9 \\ 0 \end{cases} \tag{3}$$

In this context, $B(u_i, v_i)=1$ indicates fracture-boundary pixels located on the inner boundary of the GT mask, whereas $B(u_i, v_i)=0$ represents non-boundary pixels.

Using $B(u_i, v_i)$, a distance transform is performed to calculate the Euclidean distance from each pixel to its nearest boundary pixel:

$$\mathrm{dist}(u_i, v_i) = \min_{(x,y):B(x,y)=1} \sqrt{(u_i-x)^2 + (v_i-y)^2} \tag{4}$$

In this context, $\mathrm{dist}(u_i, v_i)\geq 0$ quantifies the distance between pixel (u_i, v_i) and the nearest ground truth (GT) boundary, where $\mathrm{dist}(u_i, v_i)=0$ for pixels located directly on the boundary.

To ensure a stable boundary emphasis range, the distance is limited by a predefined radius R (measured in pixels) and normalized to the interval [0,1] as follows:

$$D(u_i, v_i) = \frac{\min(\mathrm{dist}(u_i, v_i), R)}{R} \tag{5}$$

Where r controls the spatial extent of boundary emphasis: pixels farther than R pixels from the boundary are assigned $D(u_i, v_i)=1$. Therefore, $D(u_i, v_i)\in[0,1]$, where smaller values indicate pixels closer to the boundary.

The pixel-wise weight is then defined as

$$w_i(u_i, v_i) = 1 + \alpha\left(1 - D(u_i, v_i)\right), w_i(u_i, v_i) \in [1, 1+\alpha] \tag{6}$$

where α>0 is a tuning parameter controlling the boundary emphasis strength. According to Eq. (6), boundary pixels satisfy $D(u_i, v_i)$=0 and thus receive the maximum weight $w_i(u_i, v_i)$=1+α, whereas pixels far from the boundary ($D(u_i, v_i)$=1) revert to the baseline weight $w_i(u_i, v_i)$=1.

2.2 Steady-state quantification via multi-frame fusion

In raw forward-looking borehole videos, observations of the same fracture across frames are influenced by variations in viewing distance and attitude, resulting in abrupt frame-to-frame jumps in single-frame measurements. To derive steady-state parameters suitable for engineering applications, per-frame fracture masks are processed to extract the skeleton and centroid, estimate a reliable fracture width, and infer the three-dimensional orientation. Cross-frame consistent association and fracture grouping are then performed using a two-stage gating strategy that integrates positional continuity with geometric consistency. For each group, estimates from multiple frames for the length, width, and direction are merged to produce stable outputs, thereby reducing the influence of abnormal frames on the final results.

2.2.1 Single-frame geometric measurement from fracture masks

In each frame, the fracture segmentation mask functions as a pixel-level, boundary-consistent morphological carrier for geometric measurements. From this mask, the following single-frame attributes are extracted: (i) the distance field and skeleton; (ii) robust width statistics and a prior for the fracture center position; (iii) the two-dimensional principal-axis orientation; and (iv) the three-dimensional fracture attitude according to the camera calibration model.

(1) Distance field and skeleton extraction

A distance-field function $H(u_i, v_i)$ is defined within the fracture mask. For each fracture pixel (u_i, v_i), the Euclidean distance to the nearest background pixel (u_b, v_b) is calculated as follows:

$$H(u_i, v_i) = \min_{(u_b, v_b) \in B} \sqrt{(u_i - u_b)^2 + (v_i - v_b)^2} \tag{7}$$

where $H(u_i, v_i)$ denotes the distance from pixel (u_i, v_i) to its nearest background pixel; (u_b, v_b) is the coordinate of a background pixel; and B denotes the background set.

(2) Width estimation and center cues

Local maxima of the distance field $H(u_i, v_i)$ indicate fracture skeleton points. At a skeleton point (u_s, v_s), the distance-field value $H(u_s, v_s)$ denotes the "half-width", which is the distance from the skeleton to the fracture boundary. The local width at this point is defined as twice this distance:

$$w(u_s, v_s) = 2 \cdot H(u_s, v_s) \tag{8}$$

where $w(u_s, v_s)$ denotes the local width at the skeleton point (u_s, v_s); and $H(u_s, v_s)$ denotes the corresponding "half-width", i.e., the distance from the skeleton to the fracture boundary.

Direct averaging across the entire skeleton is susceptible to errors caused by branching endpoints, mask fragmentation, and specular-induced dilation. To improve robustness, the median of skeleton half-widths within a neighborhood centered on the centroid is used as the representative single-frame width, calculated as follows:

$$w_t = 2 \cdot median\left\{H_{t,i}(p^*) \middle| p^* \in \Omega_{t,i}\right\} \tag{9}$$

where w_t denotes the representative width at the fracture centroid; p^* denotes the skeleton point closest to the mask centroid; $\Omega_{t,i}$ denotes the centroid-neighborhood skeleton set, constructed by taking 12 skeleton pixels on each side along the skeleton; and $H(p)$ represents the corresponding "half-width", i.e., the distance from the skeleton point to the fracture boundary.

After camera calibration and establishment of pixel-to-space mapping, the scale factor at the centroid can be determined. This enables conversion of pixel-wise width into physical width as follows:

$$W_c = W_t \cdot S_t(c_s) \tag{10}$$

where w_c denotes the fracture width in physical units, and $s_t(c_s)$ denotes the lens-based scale factor at the centroid, obtained from calibration experiments.

To enhance the spatial stability of cross-frame association, a center-location prior is developed for subsequent gating and trajectory updates by calculating both the mask centroid cm and the skeleton centroid c_s. Given their differing sensitivities to boundary defects and local occlusions, a refined centroid is obtained through weighted fusion, as calculated below:

$$c_t = \lambda c_m + (1-\lambda) c_s \tag{11}$$

where c_t denotes the refined centroid location, and λ is the weighting coefficient set to 0.6.

(3) 2D orientation estimation

The in-plane extension direction of a fracture is characterized by calculating the second-order central moments of the fracture pixel set, from which the principal-axis orientation angle in the two-dimensional image plane is derived as follows:

$$\theta = \frac{1}{2}\arctan 2\left(2\mu_{11}, \mu_{20} - \mu_{02}\right) \tag{12}$$

where θ denotes the principal-axis orientation angle in the 2D plane, and (μ_{20}, μ_{02}, μ_{11}) are the second-order central moments computed from the set of fracture pixels.

(4) Optional 3D dip estimation

In a calibrated camera model, each skeleton pixel (u_s, v_s) is mapped to a 3D point (x, y, z), resulting in a set of 3D points representing all skeleton pixels. A plane is then fitted to this set using RANSAC [40] in combination with singular value decomposition (SVD), which produces the normal vector of the fracture plane, n = (n_x, n_y, n_z).

The borehole axis k = (0,0,1) is used as the reference to calculate the actual dip angle of the fracture plane relative to the borehole axis as follows:

$$\delta = \arccos\left(|n \cdot k|\right) = \arctan\left(\frac{\sqrt{n_x^2 + n_y^2}}{|n_z|}\right) \tag{13}$$

where δ denotes the actual dip angle of the fracture plane relative to the borehole axis; k is the borehole axis vector, k = (0,0,1); and n is the fracture-plane normal vector, n = (n_x, n_y, n_z).

The fracture strike is defined by the direction of the line formed at the intersection of the fracture plane and the horizontal plane. This value is determined by projecting the normal vector onto the horizontal plane, as outlined below:

$$\varphi = \arctan 2\left(n_x, -n_y\right) \tag{14}$$

where φ denotes the fracture strike.

2.2.2 Cross-frame consistent association in raw videos

Single-frame measurements often exhibit discontinuities caused by specular highlights, occlusions, and camera jitter. Consequently, establishing stable correspondences for the same fracture across adjacent frames is essential for accurate steady-state estimation and multi-frame fusion. Reliable cross-frame correspondences are obtained by integrating mask-constrained feature matching, optical-flow propagation, and RANSAC-based motion filtering. A two-stage gating scheme determines whether the fracture trajectory requires updating, and a failure-recovery mechanism mitigates drift arising from accumulated errors.

(1) Feature extraction and match initialization

SURF key points are extracted within the fracture-mask region, and feature descriptors are constructed using Haar-wavelet responses. For two adjacent frames, t and t+1, descriptor similarity is quantified by Euclidean distance. Candidate matches are filtered using the KNN-ratio [41] criterion. For each key point, the nearest and second-nearest neighbor distances, d_1 and d_2, are retained. If the ratio d_1/d_2 is less than τ_{ratio} (with k=2 and τ_{ratio}=0.75), the match is considered a reliable initial correspondence for subsequent optical-flow tracking.

(2) Optical-flow tracking and feature filtering

Building on the initial matches, the Lucas–Kanade optical flow method is employed to estimate feature-point motion from frame t to t+1, thereby refining and extending the trajectories. Points with weak texture are susceptible to drift due to illumination changes or specular reflections. To address this, a Hessian-based stability screening strategy is implemented, retaining only key points with significant Hessian determinants as stable. This approach reduces the risk of optical flow drift and enhances the quality of subsequent motion model fitting. The determinant is calculated as follows:

$$\mathrm{Det}\left(H\right) = L_{xx}L_{yy} - \left(L_{xy}\right)^2 \tag{15}$$

where Det(H) denotes the determinant of the Hessian matrix; L_{xx} and L_{yy} are the second-order partial derivatives of the image at point (x, y) and scale σ along the x- and y-directions, respectively; and L_{xy} is the mixed second-order derivative at (x, y) and scale σ.

(3) Motion-model fitting and anchor-point update

Based on the optical-flow–propagated matches, an inter-frame motion model is fitted using RANSAC to reject outliers and obtain an inlier set. The inlier correspondences identified by RANSAC are defined as cross-frame consistent anchor points. Linking these anchors over time forms a continuous image-domain trajectory of the fracture. This trajectory constrains updates to the fracture center location and establishes stable correspondences among multi-frame observations of the same fracture.

For subsequent evaluation, two motion-consistency metrics are defined as follows:

$$r_{in} = \frac{\left|L_{t\to t+1}\right|}{\left|P_{t\to t+1}\right|} \tag{16}$$

where r_{in} denotes the inlier ratio; $L_{t\to t+1}$ refers to the initial set of matched correspondences between frames t and t+1; and $P_{t\to t+1}$ designates the corresponding RANSAC inlier set, which serves as the anchor correspondences following motion-model fitting.

For each inlier, the pixel-wise error between the model-predicted position and the observed position is computed as follows:

$$e_{re} = \text{median}\left\{e_j\right\} \tag{17}$$

where ere denotes the median reprojection error, and e_j denotes the pixel error between the predicted and observed positions for the j-th correspondence.

(4) Cross-frame consistency decision

To mitigate the risk of mismatches and accumulated drift, the fracture-tracking trajectory is updated using a two-stage gating scheme consisting of coarse screening and fine screening.

Stage I: spatial overlap filtering.

The intersection-over-union (IoU) between fracture masks in two consecutive frames is calculated as follows:

$$IoU\left(M_t, M_{t+i}\right) = \frac{\left|M_t \cap M_{t+1}\right|}{M_t \cup M_{t+1}} \tag{18}$$

where IoU(M_t, M_{t+1}) denotes the intersection-over-union between the fracture masks in frames t and t+1, with Mt and M_{t+1} are the respective fracture masks for these frames.

If the correspondence fails the cross-frame consistency check ($IoU\left(M_t, M_{t+i}\right) < \tau_{\text{iou}}$), the trajectory is not updated in the current frame and the instance is recorded as a miss. This initial screening efficiently eliminates target switches, incorrect associations between adjacent fractures, and significant mask misalignments. As a result, both the risk of false positives and the computational cost of subsequent fine screening are reduced.

Stage II: motion–morphology joint validation.

For candidate matches that pass Stage I, a joint decision criterion is applied, integrating both motion consistency and morphological consistency as described below:

i) Motion consistency: Based on the RANSAC-fitted motion model, let r_{in} denote the inlier ratio and $e$re represent the median reprojection error. The following condition must be satisfied:

$$r_{in} \geq \tau_{in}, e_{re} \leq \tau_{re} \tag{19}$$

ii) Morphological consistency: Based on the single-frame measurements described in Section 2.2.1, let the 2D principal-axis orientations in two consecutive frames be (θ_t, θ_{t+1}). Because orientations are equivalent modulo 180°, the morphological change is defined as follows:

$$\Delta\theta_t = \left|\theta_{t+1} - \theta_t\right| \leq \tau_\theta \tag{20}$$

The distributions of these metrics for correct and incorrect associations on the validation set are analyzed to select default thresholds that balance trajectory continuity with the suppression of false associations. Table 1 summarizes the cross-frame consistency measures used for trajectory updating, including their decision rules, default values, and respective roles.

Table 1 Cross-frame consistency criteria for trajectory update.

Constraint type	Criterion	Decision rule	Default threshold
Spatial consistency	IoU(M_t, M_{t+1})	$\geq \tau_{iou}$	0.75
Motion consistency	r_{in}	$\geq \tau_{in}$	0.60

	e_{re}	$\le \tau_{re}$	2.0 pixels
Morphological consistency	$\Delta\theta$	$\le \tau_{\theta}$	5°

(5) Failure recovery

To mitigate drift accumulation resulting from incorrect updates during cross-frame association failures, and to address short-term occlusions and strong specular reflections frequently observed in forward-looking videos, a memory-based failure-recovery and track-termination mechanism is implemented. This approach transforms error propagation from accumulated drift into a controlled process comprising failure freezing, limited reconnection, and thresholded termination. As a result, brief mismatches do not disrupt trajectory continuity, while long-term mismatches are prevented from causing false associations and parameter distortion. If either stage of the two-level gating is not satisfied, the current frame is classified as an association failure, and the following procedures are executed:

Freeze update: The fracture track states, including the anchor set, refined centroid, width, and orientation, are not updated. Instead, the most recent reliable state that satisfied the consistency checks is retained as the latest valid estimate.

Limited-window reconnection: Reassociation with the current track is attempted for the subsequent five frames. Reconnection is permitted only when all decision criteria are met. Upon successful reconnection, the failure counter is reset to zero, and standard track updating resumes.

Track termination: If five consecutive association failures occur, the fracture is considered no longer observable in the current field of view. The track is terminated, and steady-state quantification is initiated. Multi-frame fusion is performed on the accumulated valid frames to generate steady-state fracture parameters, and the track is closed to prevent subsequent erroneous reconnections.

New track initialization: For fracture observations that appear in subsequent frames and cannot be associated with any existing track based on the consistency checks, a new track is initialized as a new target. This is followed by independent cross-frame association and steady-state quantification.

2.2.3 Steady-state segment determination and multi-frame fusion quantification

Instantaneous single-frame measurements are converted into stable, engineering-ready parameters by treating the cross-frame trajectory of each fracture as a set of observations. The trajectory is maintained using the cross-frame consistent association and two-stage gating described in Section 2.2.2, which ensures that all observations in the set correspond to the same fracture. For each frame within a trajectory, width, orientation, and three-dimensional geometric quantities are derived from the single-frame measurements detailed in Section 2.2.1. Robust multi-frame fusion is then performed at the trajectory level for width, orientation, and length, thereby reducing the influence of outlier observations arising from specular reflections, occlusions, and segmentation-boundary fluctuations on the final outputs.

(1) Three-dimensional integration for fracture length

As described in Section 2.2.2, we extract the set of fracture anchor points along the temporal sequence and back-project the anchor pixels into the 3D coordinate system, denoted as (X_j, Y_j, Z_j), j=1, 2, … , M. To estimate the 3D projected length of the fracture, these points are sorted along the fracture extension direction, and the length is calculated by summing the Euclidean distances between adjacent points as follows:

$$L = \sum_{j=1}^{M-1} \sqrt{(X_{j+1} - X_j)^2 + (Y_{j+1} - Y_j)^2 + (Z_{j+1} - Z_j)^2} \tag{21}$$

where L denotes the maximum fracture length obtained after tracking and fitting refinement; M is the number of sampled points along the trajectory; and (X_j, Y_j, Z_j), j=1, 2, … , M are the M 3D points ordered along the fracture trace in space.

(2) Robust multi-frame fusion for fracture width

Robust fusion is performed on the physical-width sequence within each trajectory by applying a trimmed mean, in which 10% of the samples are excluded from each end as follows:

$$W = \mathrm{TrimMean}\left(\{w_c\}, 0.1\right) \tag{22}$$

where W denotes the steady-state fracture width obtained by averaging across multiple frames.

(3) Cross-frame fusion for fracture orientation

The per-frame orientation θ_t is calculated using Eq. (12); however, this measure is susceptible to local boundary defects and segmentation variability. To obtain a stable global orientation, each frame-wise angle is converted into a unit direction vector $v_t = (\cos\theta_t, \sin\theta_t)$. Principal component analysis (PCA) is applied to the set of direction vectors,

and the first principal-component unit vector (v_y, v_x) is identified as the dominant projected direction of the fracture. The overall orientation is then computed as follows:

$$\theta_q = \arctan 2(v_y, v_x) \tag{23}$$

where θ_q denotes the overall fracture orientation; (v_y, v_x) is the unit vector of the first principal component obtained from principal component analysis; and arctan2 is the two-argument inverse tangent function, which returns the vector angle with the correct quadrant.

3 Experimental preparation

3.1 Experimental environment and hyperparameters

Table 2 presents the hardware and software configurations utilized for model training and testing.

Table 2 Experimental environment parameters.

Component	Model
CPU	AMD Ryzen 7 3700X 8-Core Processor
GPU	NVIDIA GTX 1660 Super
Memory	32 GB
System	Windows 11
GPU acceleration	CUDA 11.2
Python	Python 3.9

The key hyperparameters for the YOLOv8-CAPS detection module and the Mask R-CNN-WBCE segmentation module are systematically specified to promote reproducibility and training stability.

The primary hyperparameters for YOLOv8-CAPS include a momentum of 0.937, an initial learning rate of 0.001, a weight decay of 0.0001, a batch size of 16, and 150 training epochs. Within the capsule module, both the Primary Caps and Routing Caps contain N_{caps}=32 capsules, the capsule dimension is set to do=8, and the number of dynamic routing iterations is r=3.

The Mask R-CNN model was trained using the Adam optimizer with a learning rate of 0.0005 and a batch size of 8. The loss function for the mask branch utilizes a hybrid approach, combining weighted boundary cross-entropy (WBCE) and Dice loss:

$$L = \lambda L_{WBCE} + (1-\lambda) L_{Dice} \tag{24}$$

Here, λ was set to 0.7. In the default configuration, the boundary-weight strength α was set to 0.55. This value specifies the maximum weight assigned to pixels on the ground-truth boundary and decreases progressively with increasing distance from the boundary.

3.2 Establishment and pre-processing of fracture dataset

Borehole videos were obtained from multiple mines located in the coal-producing regions of Shandong, Henan, Shanxi, Guizhou, and Xinjiang, as illustrated in Figure 4a–b, using a mine electronic imaging trajectory logging device. Acquisition parameters are detailed in Table 3. The dataset comprises videos from 55 boreholes. One key frame was extracted every 50 frames, resulting in 2,172 valid video frames containing borehole fractures.

To mitigate data leakage resulting from highly correlated frames within the same borehole, the dataset was partitioned at the borehole level using a stratified random sampling strategy based on mining areas. Boreholes, rather than individual frames, were allocated to the training, validation, and test sets to ensure that frames from a single borehole were not distributed across multiple subsets.

To more accurately simulate real field conditions, data augmentation techniques such as horizontal and vertical flipping, random geometric transformations, color adjustment, and noise or blur perturbations were applied. These methods expanded the dataset to 12,100 images, with 9,075 allocated for training, 1,210 for testing, and 1,815 for validation.

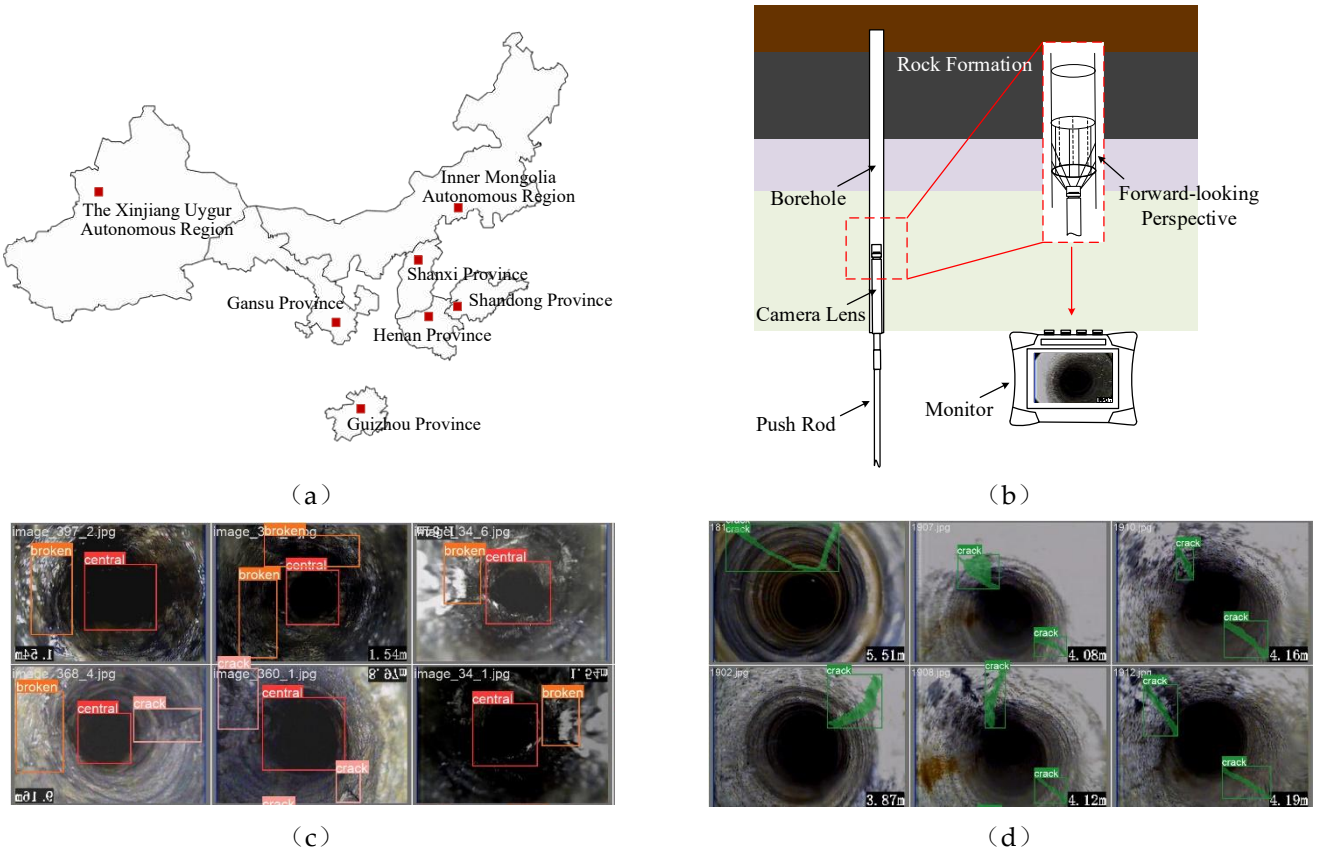
Table 3 Technical parameters of electronic imaging trajectory detection device.

Probe diameter	Output resolution	Detection accuracy	Image pixels	Aperture
40 mm	800×600	0.1 mm	2 million	50 mm

The LabelMe tool was employed to manually annotate fracture-related targets in forward-looking borehole images for the YOLO dataset. A K-means clustering algorithm was subsequently used to pre-classify image regions into three categories: invalid central region, fractures, and broken zones. Accordingly, the YOLO label set was defined as central (invalid central region), fracture (fracture), and broken (broken zone), with representative annotations provided in Figure 4c.

Labels for the Mask R-CNN dataset were also generated using LabelMe. The resulting JSON files were converted to the COCO dataset format, and annotation results are presented in Figure 4d. To ensure consistency, all images were annotated under the supervision of three experts in mining engineering and image processing, and a cross-review procedure was implemented to verify label accuracy.

Throughout this study, the term "crack" is used for labelling consistency in the dataset and training process, whereas "fracture" is reserved for the theoretical and descriptive sections to maintain alignment with fracture mechanics terminology.



（a） （b）

（c） （d）

Figure 4. Image acquisition system and dataset annotation examples: (a) collection sites; (b) schematic of borehole video acquisition; (c) example annotation in the YOLO dataset; (d) example annotation in the Mask R-CNN dataset.

3.3 Evaluation metrics

To comprehensively assess the overall performance of the proposed models in fracture detection and segmentation, seven evaluation metrics are employed: Precision, Recall, F1-score, Intersection over Union (IoU), Dice coefficient (Dice), number of parameters (Params), and floating-point operations (FLOPs).

The YOLOv8 model is designed for object detection, so its performance is primarily evaluated using object-level Precision, Recall, and mean Average Precision (mAP), which indicate the ability to localize and identify fracture targets. In contrast, Mask R-CNN is intended for pixel-level instance segmentation; therefore, evaluation focuses on pixel-level IoU and Dice coefficients to quantify the overlap between predicted fracture masks and ground-truth (GT) masks. Additionally, Params and FLOPs are reported to assess model complexity and computational cost quantitatively.

For detection, a predicted bounding box is considered a true positive if it matches a ground-truth (GT) box with an IoU above a specified threshold (for example, 0.5 for mP@0.5 and mR@0.5). Unmatched predictions are classified as false positives, while unmatched GT boxes are classified as false negatives. Precision represents the proportion of correctly predicted fracture objects, Recall represents the proportion of GT fracture objects that are successfully detected, and the F1-score is the harmonic mean of Precision and Recall. For segmentation, IoU quantifies the overlap ratio between predicted and GT pixels, whereas Dice provides a similar measure but is more sensitive to small fracture regions. Params refers to the total number of trainable parameters in the model, and FLOPs denotes the number of floating-point operations required for a single forward inference. The corresponding formulations are as follows:

$$\text{Precision}=\frac{\text{TP}_{\text{det}}}{\text{TP}_{\text{det}}+\text{FP}_{\text{det}}} \tag{25}$$

$$\text{Recall}=\frac{\text{TP}_{\text{det}}}{\text{TP}_{\text{det}}+\text{FN}_{\text{det}}} \tag{26}$$

$$\text{F}_1=\frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{27}$$

$$\text{IoU}=\frac{\text{TP}_{\text{pix}}}{\text{TP}_{\text{pix}}+\text{FP}_{\text{pix}}+\text{FN}_{\text{pix}}} \tag{28}$$

$$\text{Dice}=\frac{2\cdot\text{TP}_{\text{pix}}}{2\cdot\text{TP}_{\text{pix}}+\text{FP}_{\text{pix}}+\text{FN}_{\text{pix}}} \tag{29}$$

where TP_{det}, FP_{det}, and FN_{det} represent the numbers of true-positive, false-positive, and false-negative detected objects, specifically fracture bounding boxes, under a specified intersection over union (IoU) threshold; and TP_{pix}, FP_{pix}, and FN_{pix} refer to the numbers of true-positive, false-positive, and false-negative pixels within the predicted segmentation masks.

4. Results and discussion

4.1 Visualization of Training Process

The effectiveness of the fracture recognition models is evaluated using loss curves and performance metrics on the validation set. Figure 5 presents the training and validation performance of YOLOv8 and Mask R-CNN. Both models demonstrate strong convergence and stability. The training loss decreases rapidly during the initial epochs and subsequently stabilizes, while the validation metrics show consistent improvement before reaching a plateau.

For YOLOv8, both box_loss and dfl_loss converge to low values after 150 epochs. The mean precision (mP@0.5) and mean recall (mR@0.5) stabilize at 90.76% and 88.20%, respectively, reflecting high detection accuracy and comprehensive identification of fracture targets. In contrast, Mask R-CNN demonstrates significant reductions in seg_loss and mask_loss, with the Dice coefficient and mIoU reaching 82.51% and 76.33%. These results indicate improved geometric consistency in boundary delineation and enhanced pixel-level segmentation accuracy within fracture regions.

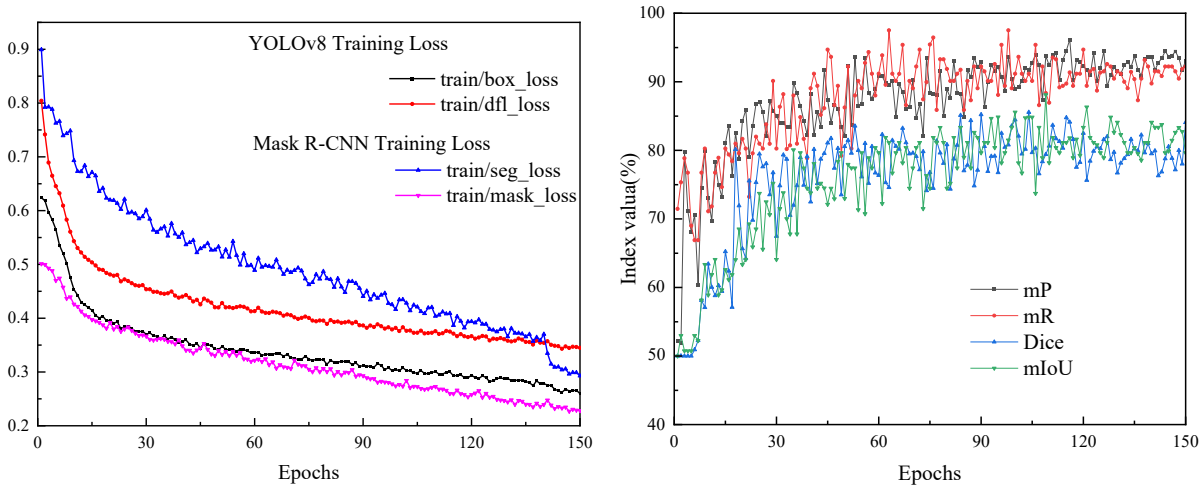
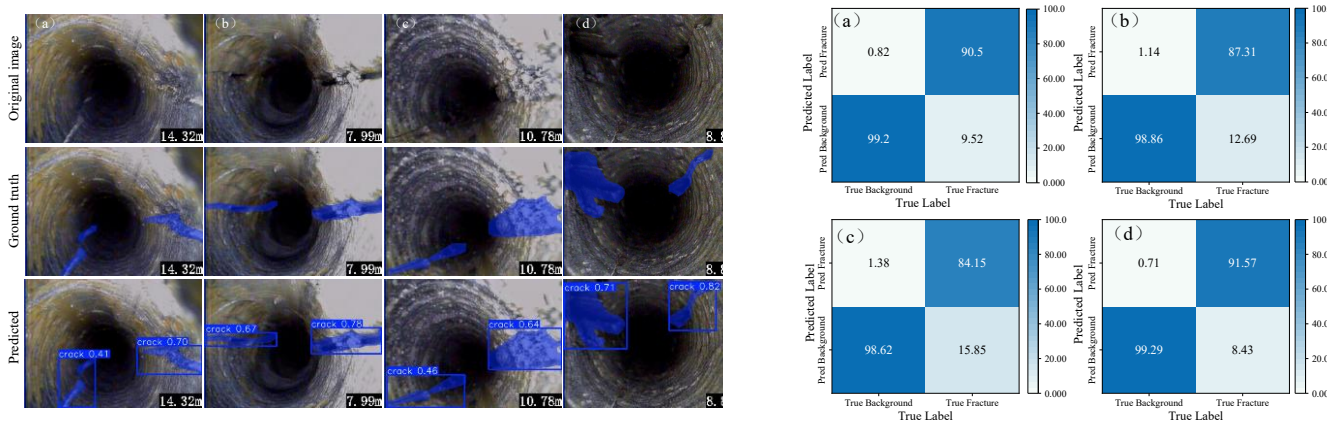


Figure 5. Training and validation performance of YOLOv8 and Mask R-CNN: (a) training loss; (b) validation performance.

4.2 Recognition quality for geometry-oriented quantification

Four representative single-frame images from the forward-looking perspective were selected to illustrate fracture recognition performance in detail. Figure 6 presents detection and segmentation results for fractures of varying scales and morphologies. The predictions correspond closely with the ground-truth labels, and the extracted features, including fracture skeleton, width, and orientation, are consistent with the observed morphology. These results demonstrate the effectiveness of the proposed method. Additionally, the confusion matrix indicates that the model maintains high detection stability in complex scenarios.



Recognition results **Confusion matrix**

Figure 6. Fracture recognition results and the corresponding confusion matrix.

Variations in illumination have a significant impact on recognition performance. In Figure 6b, the fracture exhibits grayscale values similar to the background, and partial edge breakage is observed in some regions, reducing the recognition rate to 87.3%. Nevertheless, the overall fracture area remains correctly identified. In Figure 6c, intense illumination and reflection interference result in blurred and discontinuous mask boundaries, leading to a lower recognition rate (84.2%) and a slight increase in false detections in the confusion matrix (1.38%). These anomalies do not directly influence the final output, as the subsequent cross-frame consistency screening identifies them as abnormal frames and either removes them or halts updates. This process ensures that the final steady-state parameters remain repeatable and robust for engineering applications.

Figure 7 displays the feature-extraction results for four representative fractures. The model simultaneously generates the fracture skeleton, centroid, pixel-level width, and orientation parameters. For average fracture widths between 7 and 31 pixels and orientation angles from 40° to 80°, the geometric features are accurately estimated, supporting stable feature extraction in the presence of complex fracture morphologies.

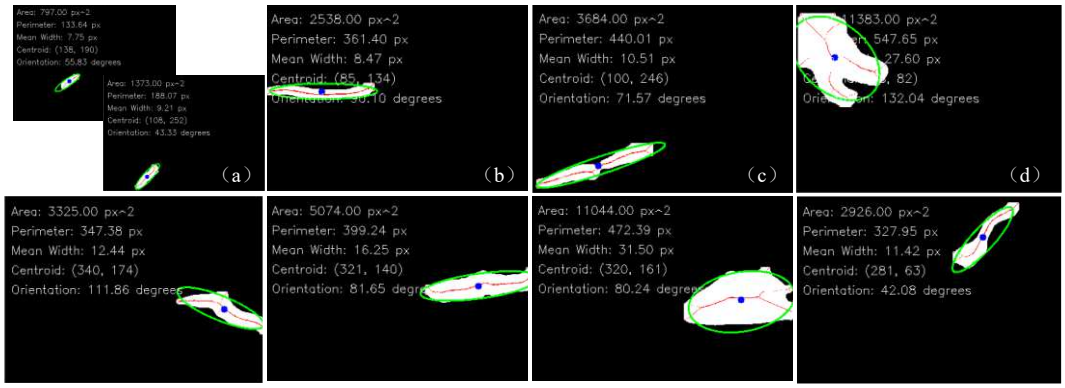


Figure 7. Complex fracture recognition and feature extraction: (a) small-scale fracture; (b) typical fracture; (c) large-scale fracture; (d) irregular fracture.

4.3 Module effectiveness

4.3.1 CAPS ablation experiment

YOLOv8-s served as the baseline for ablation studies, with four functionally similar alternative modules—CBAM [42], CoordAtt [43], DCNv3 [44], and C3-Transformer [45]—introduced for comparison. Table 4 presents the detailed results of the ablation experiments on the CAPS module within YOLOv8. Under identical training conditions and comparable model capacity, the CAPS module demonstrated superior performance. Compared with the baseline, mean precision and mean recall increased by 12.65% and 34.21%, respectively, while the F1-score reached 89.45%, surpassing the other evaluated models.

Table 4 Results of the YOLOv8 model ablation experiment.

Models	mP@0.5	mR@0.5	F1-score
YOLOv8(Baseline)	80.57%	65.72%	72.34%
YOLOv8+CBAM	83.88%	72.84%	77.92%
YOLOv8+CoordAtt	85.03%	75.94%	80.24%

YOLOv8+DCNv3	84.02%	77.02%	80.38%
YOLOv8+C3-Transformer	86.51%	73.51%	79.45%
YOLOv8+CAPS	90.76%	88.20%	89.45%

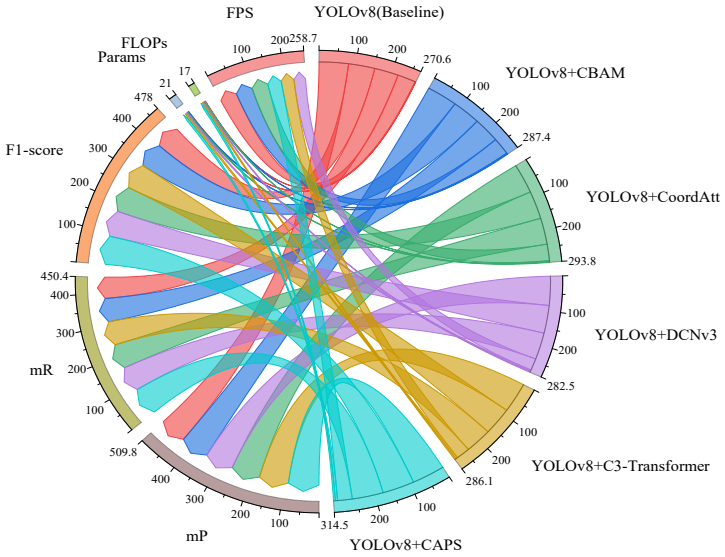


Figure 8. The comparison of the Params, FLOPs, FPS, and model size for the six models.

An additional comparison was conducted to evaluate parameter count, computational cost, and inference speed across the models. Figure 8 presents the results for the six previously discussed models. While the number of parameters and floating-point operations (FLOPs) increase slightly and frames per second (FPS) decreases, Table 4 demonstrates that the improvements in mean precision at 0.5 (mP@0.5) and mean recall at 0.5 (mR@0.5) are substantially greater. This suggests an overall advantage in balancing enhanced performance with computational requirements.

Figure 9 provides additional visualization of heatmap distributions before and after the integration of the CAPS module. The results indicate that, following the incorporation of CAPS, the network generates a more spatially continuous high-response band in small-scale fracture regions. This enhancement improves the network's ability to perceive and localize tiny fractures, thereby demonstrating the substantial effectiveness of the CAPS module in feature extraction and fracture recognition.

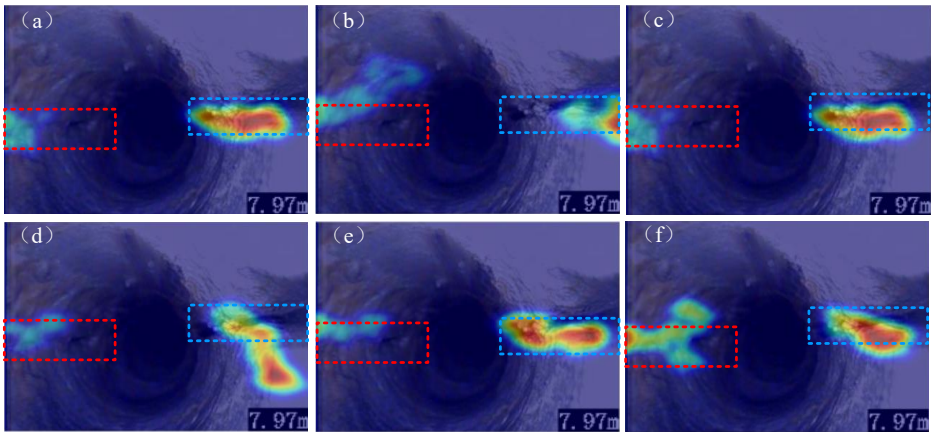


Figure 9. Activation heatmaps of different models: (a) YOLOv8; (b) CBAM; (c) CoordAtt; (d) YOLOv8+DCNv3; (e) C3-Transformer; (f) CAPS.

4.3.2 WBCE comparative experiment

To further assess the impact of loss-function optimization on segmentation performance, a comparative experiment was conducted using the weighted boundary cross-entropy (WBCE) loss within the Mask R-CNN framework. Under strictly controlled conditions, including identical network architecture, data split, training epochs, and hyperparameters, two models were trained: one using the standard cross-entropy (CE) as the mask-branch loss, and the other employing WBCE, which assigns greater weights to fracture boundary pixels to improve boundary supervision.

Table 5 presents the Mask R-CNN segmentation performance under different loss functions. The results indicate that the WBCE-based model outperforms the baseline across all metrics. Specifically, the Dice coefficient increases from 79.62% to 82.51%, and mIoU improves from 72.84% to 76.33%. These findings indicate that assigning higher weights to fracture boundary pixels directs the model's attention toward boundary regions during training, thereby enhancing its capacity to learn edge-related features.

Table 5 Comparison of segmentation performance of Mask R-CNN with different loss functions.

Models	Dice	mIoU	F1-score
Mask R-CNN+CE	79.62%	72.84%	79.62%
Mask R-CNN+WBCE	82.51%	76.33%	83.40%

Figure 10 presents the visualization results for complex fracture segmentation. Compared to alternative approaches, the model employing WBCE produces more accurate segmentation in areas with irregular fracture morphology and increased boundary curvature, which improves the continuity and completeness of the predicted masks.

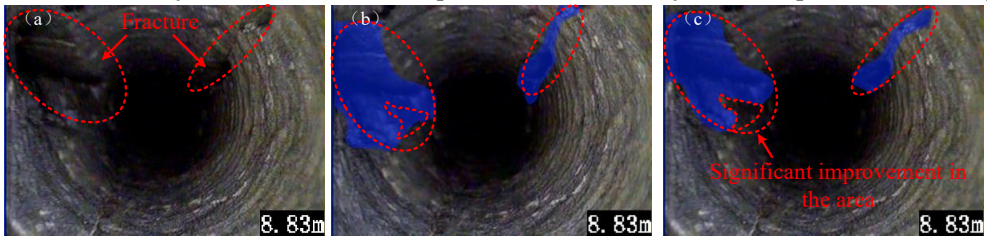


Figure 10. Visualization of fracture segmentation: (a) original image; (b) Mask R-CNN+CE; (c) Mask R-CNN+WBCE.

4.4 Cross-frame consistent association and trajectory reliability

The stability of the proposed method in the video domain is evaluated through the process illustrated in Figure 11, which outlines fracture feature extraction and dynamic tracking. Figures 11a–d show the frame-by-frame variations of fracture masks in consecutive frames, while Figures 11e–h illustrate the distributions of feature points extracted from the masks in each frame. Figure 11i presents the inter-frame matching results of these feature points. The feature points are primarily distributed along the fracture skeleton and in regions of high boundary curvature, effectively capturing the geometric characteristics of the fractures. By applying optical-flow constraints and local descriptor matching, the method establishes stable spatial correspondences between adjacent frames.

As the camera progresses, fracture morphology changes continuously. Consequently, recognition depends primarily on the continuity of the fracture structure rather than single-frame accuracy. Building on the preceding spatial recognition results, this study further examines the model's temporal steady-state behavior in the video domain and identifies the primary sources of error. Figure 12 presents the trends in feature-matching success rate, drift error, and orientation consistency across consecutive frames. Matching success is evaluated using the RANSAC inlier ratio rin, as defined in Eq. (16). A consecutive frame pair (t, t+1) is considered successful if $r_{in}(t) \geq \tau_{in}$, and the success rate is calculated as the percentage of successful pairs among all adjacent pairs. The reported value is based on a representative 120-frame sequence.

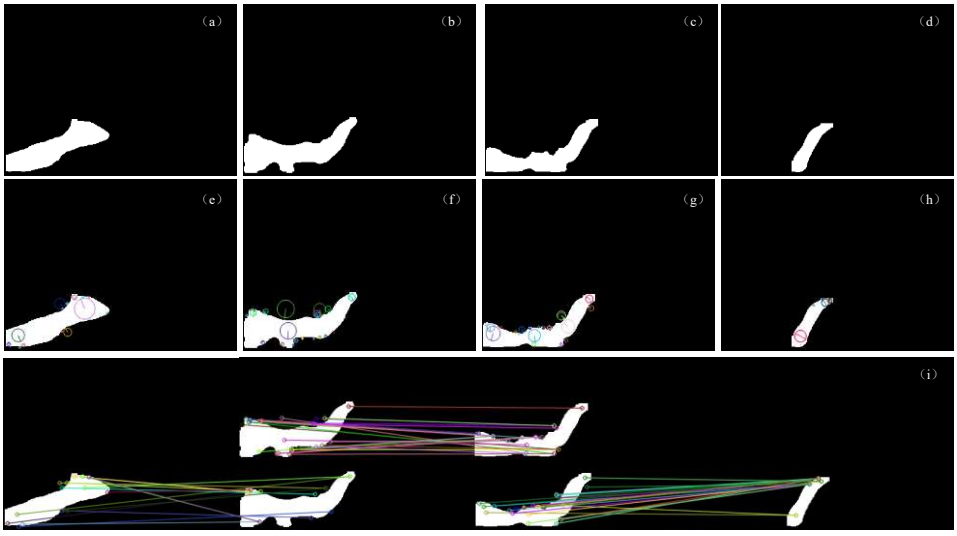


Figure 11. Fracture feature extraction and dynamic tracking for parameter quantification: (a–d) fracture masks; (e–h) extracted key points; (i) inter-frame key points matching result.

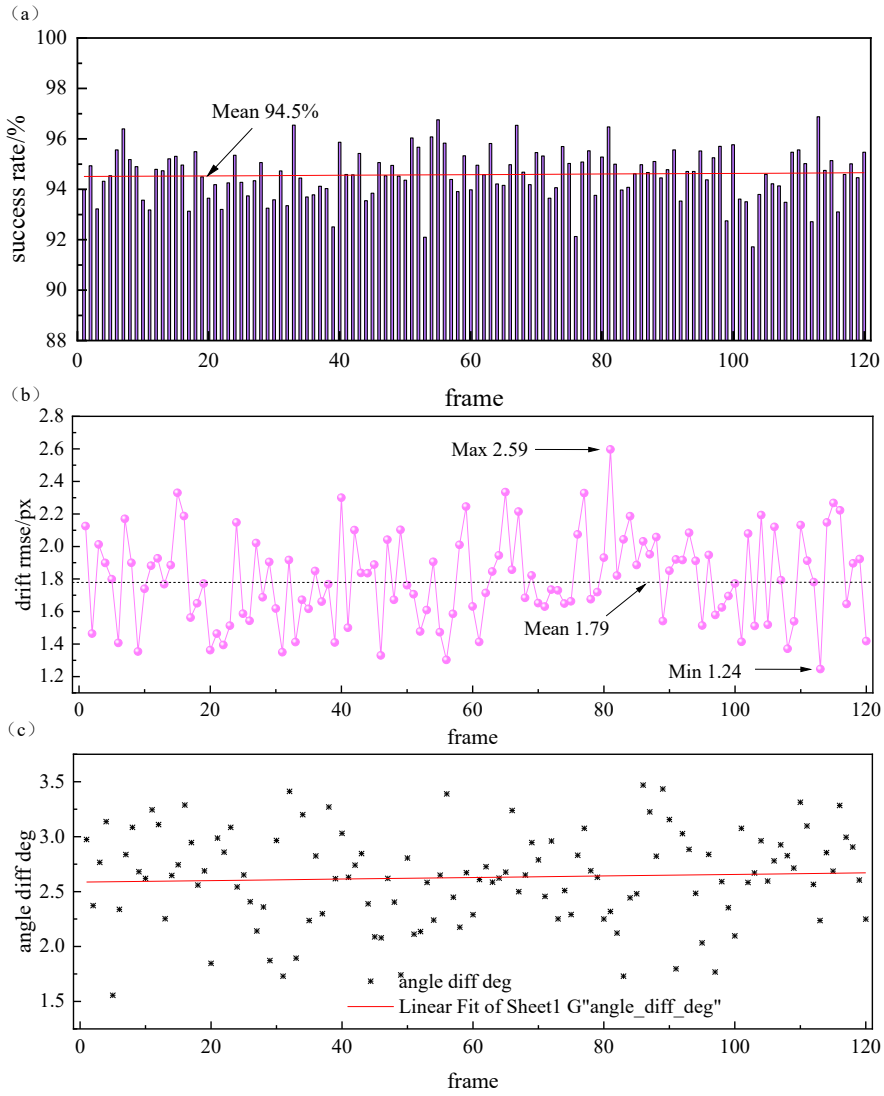


Figure 12. Stability evaluation of dynamic fracture-feature tracking: (a) matching success rate; (b) drift error distribution; (c) scatter plot and linear fit of the angle between the matching direction and the fracture principal axis.

Figure 12a demonstrates that, despite adverse factors such as illumination reflections and scale perturbations, the matching success rate remains consistently high at 94.5%. This result indicates that the extracted features are highly discriminative across frames and that the fracture's central axis forms a temporally continuous and stable geometric trajectory. Additionally, the drift error is consistently constrained within 1.24–2.59 pixels without cumulative deviation (Figure 12b). This finding suggests that the optical flow constraint and local geometric consistency together establish a temporal correction mechanism that continuously absorbs and offsets instantaneous positional disturbances. Furthermore, the matching direction maintains an average included angle of only 2.6° relative to the fracture's central axis, and the fitted slope is close to zero (Figure 12c). These results imply that cross-frame tracking does not induce trending-orientation drift and that the geometric structure remains temporally consistent.

In summary, cross-frame feature matching provides reliable temporal association and reduces drift propagation through continuous dynamic correction. This approach ensures steady-state consistency of fracture structures in the video domain and establishes a foundation for stable convergence in subsequent process quantification.

4.5 Steady-state quantification via multi-frame fusion

Figure 13 presents the quantitative characterization results of fracture parameters for a representative video segment from the dataset. Four fractures are identified in the test video. Specifically, Figure 13a illustrates the physical lengths of the inter-frame tracking trajectories for typical fractures. For instance, the cumulative trajectory length of Fracture 1 is 672.03 mm, which demonstrates greater stability and completeness compared to estimations based on single-frame panoramic images.

Figure 13b uses violin plots to depict the distributions of inter-frame widths for each fracture, capturing both frame-to-frame fluctuations and overall statistical trends. For instance, Fracture 1 exhibits an average width of 13.77 mm with relatively high variability, while Fracture 3 shows a narrow distribution centered at 13.17 mm, indicating a more stable structural feature. Figure 13c displays the principal-axis orientation distributions of Fractures 1–4 in polar coordinates, revealing their dominant orientation ranges and development characteristics. These results underscore the significance of cross-frame fusion and optimization in fracture-parameter estimation.

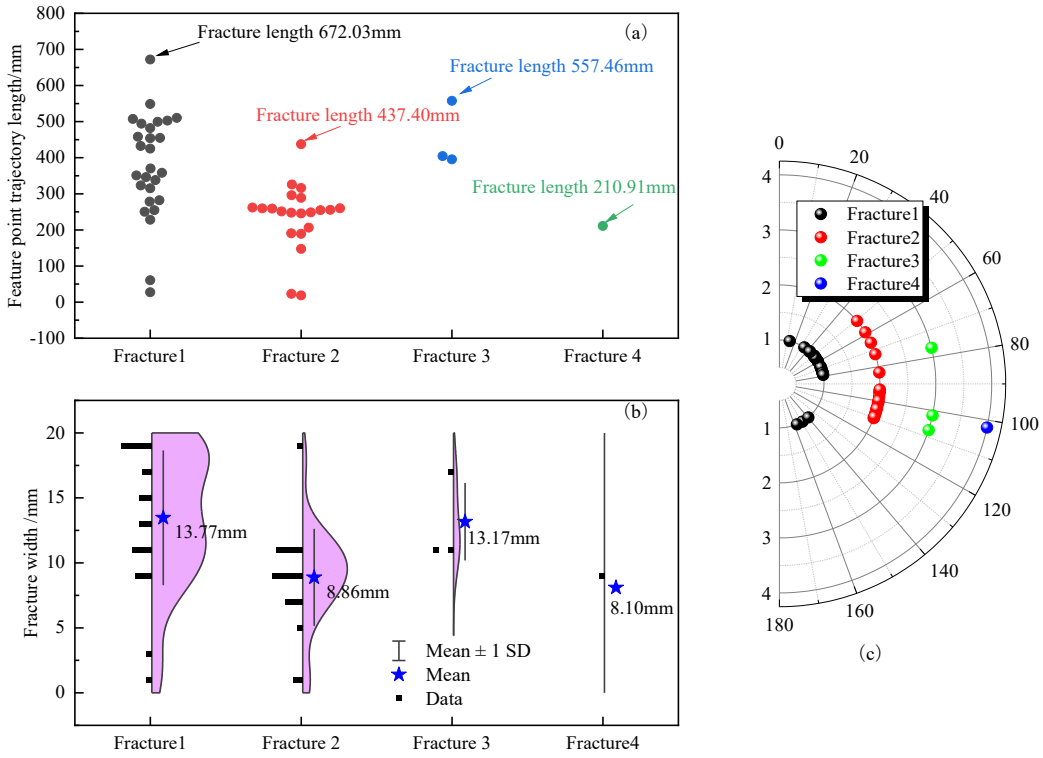


Figure 13. Characterization of fracture parameters in borehole video frames: (a) length; (b) width; (c) orientation.

To further assess the reliability of the proposed recognition and quantification method, the predicted results for three geometric parameters—fracture length, width, and orientation angle—are compared with ground-truth measurements from the validation set, as illustrated in Figure 14. The horizontal axis represents the manually measured ground-truth values, while the vertical axis corresponds to the model predictions. Each scatter point denotes an individual sample. The diagonal line indicates the ideal agreement (y = x), and the solid line represents the linear regression fit.

The results indicate that all three parameters exhibit strong agreement, with coefficients of determination (R^2) exceeding 0.995. Fracture length demonstrates the highest linearity ($R^2 = 0.9992$), suggesting that the model can accurately estimate fracture size across different scales. The predicted width and orientation angle are also highly consistent with the ground-truth values, with deviations of less than 0.1 mm and 3°, respectively. These findings confirm that the proposed method provides high accuracy and stability in quantifying fracture geometric parameters, thereby enabling reliable characterization of fracture morphology.

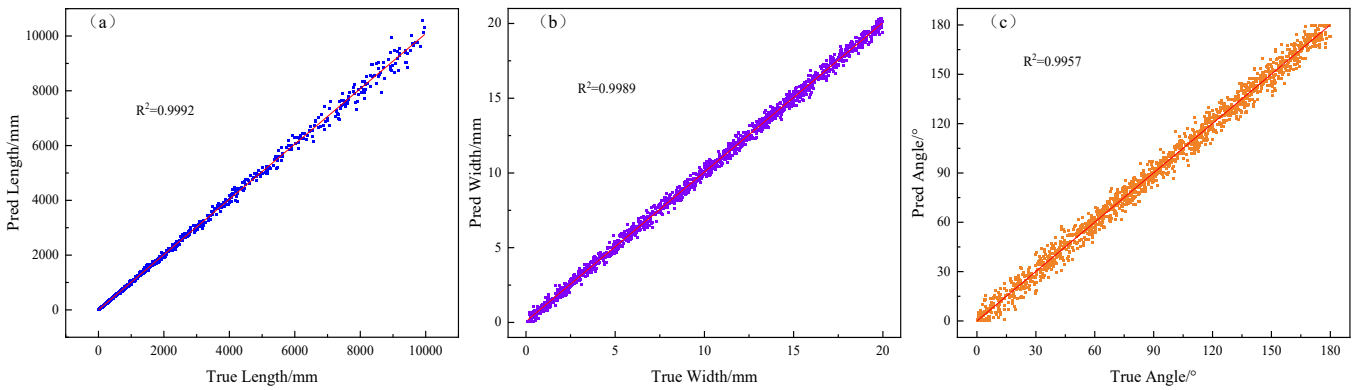


Figure 14. Quantitative results of fracture parameters: (a) length; (b) width; (c) orientation.

4.6 Comparison with panorama-domain quantification pipelines

To objectively evaluate the advantages of the proposed method for fracture recognition and parameter quantification, a comparative analysis was performed between conventional panorama-based recognition techniques and the proposed raw-video-based approach. Quantification error was used as the primary metric. Tests were conducted on the same borehole video sample and its corresponding panoramic image to compare the accuracy of fracture-parameter extraction across different methods. The methods evaluated were: (1) a conventional semi-automatic process combining threshold segmentation and Canny edge detection [46]; (2) a U-Net semantic segmentation approach based on panoramic images, representative of early deep-learning segmentation methods [47]; (3) a DeepLabV3+ panoramic-image segmentation approach employing multi-scale atrous convolutions, representing a widely used high-accuracy panorama-based model [48]; and (4) the proposed integrated detection–segmentation–quantification framework based on raw videos.

Figure 15 compares recognition results across methods on a representative borehole-fracture sample. The traditional threshold-based approach is significantly affected by uneven illumination and unwrapping distortion, often failing to extract fracture boundaries. While U-Net and DeepLabV3+ improve region extraction, both methods are constrained by scale distortion introduced during the unwrapping or mapping process. In contrast, the proposed method detects fractures directly in the raw video domain, which avoids spatial deformation and edge information loss associated with panoramic unwrapping. This approach preserves the integrity and continuity of fracture geometry and yields quantification results that more accurately represent the actual physical feature dimensions.

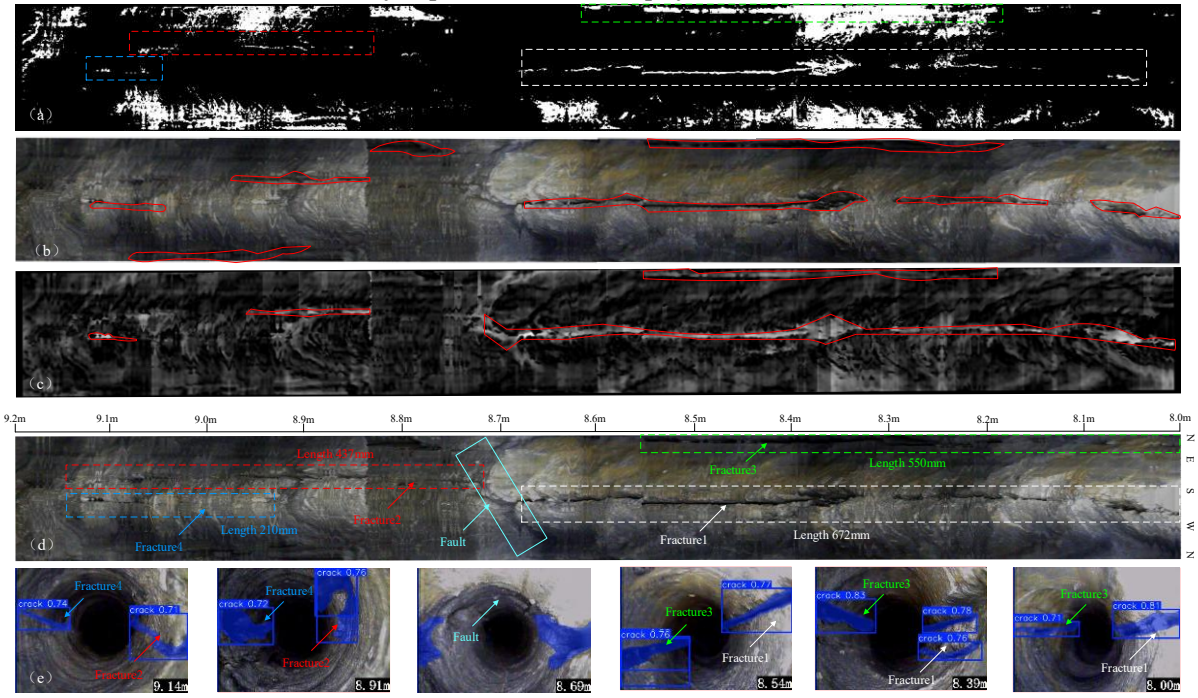


Figure 15. Comparison of recognition results: (a) thresholding + edge detection; (b) U-Net; (c) DeepLabV3+; (d) expert interpretation; (e) proposed method.

Table 6 provides a comparative analysis of errors in fracture-parameter quantification on the validation set for various methods. The proposed method demonstrates the lowest quantification errors across all parameters. In comparison to conventional panorama-based approaches, the mean errors for length, width, and orientation angle are reduced by 52.9%, 58.1%, and 45.2%, respectively, while the overall relative error decreases by 48.0%.

Table 6 Comparison of fracture-parameter quantification errors.

Method	Data domain	Length error/ mm	Width error/ mm	Angle error/ °	Relative error/ %
Thresholding + Edge detection	Panoramas	4.32	3.82	3.32	4.64
U-Net		3.68	3.62	3.43	3.68
DeepLabV3+		3.04	3.42	3.53	3.12
Proposed method	Video	2.03	1.60	1.82	2.41

Figure 16 illustrates, from the perspective of the temporal fusion process, the dynamic mechanism by which the proposed method's error decreases with increasing frame count, thereby clarifying the significant error reduction

reported in Table 6. In Figure 16a, the length error exhibits an exponential decay, decreasing from 5.7 mm in a single frame to 2.3 mm. This trend indicates that the initial scale bias present in early frames is progressively corrected through feature rematching in subsequent frames. This dynamic convergence demonstrates that the cross-frame continuity of the fracture central axis provides temporal stability, ensuring that errors diminish rather than accumulate over time.

A comparable mechanism is evident for the width and orientation parameters. As shown in Figures 16b and 16c, the width error decreases by 42.8 percent, and the orientation error is reduced to 1.5 degrees. The convergent trend in width error across frames reflects the suppression of local noise through sliding-window filtering and robust estimation using RANSAC. In contrast, the removal of outliers from the orientation-error boxplot indicates that a stable dominant axis direction has been established over time. Consequently, cross-frame fusion not only reduces numerical errors but also enforces structural inter-frame consistency constraints on fracture geometric features within the video domain.

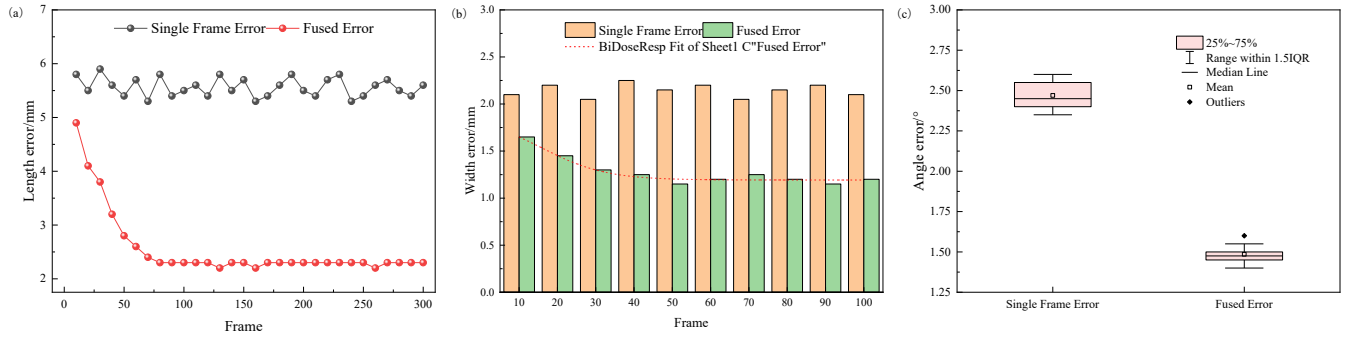


Figure 16. Statistics of fracture-parameter quantification errors: (a) length; (b) width; (c) angle.

From a methodological perspective, the concept of "statistical dissipation" of error underscores the primary advantage of raw-video-based fracture recognition and quantification compared to static panorama-based approaches. Although individual frames may exhibit boundary discontinuities or reflection artifacts, these errors can be mitigated during temporal fusion through more reliable observations from subsequent frames. Consequently, the proposed method achieves stable quantification under challenging conditions, including complex illumination, rough lithology, or weak-texture fractures, and provides a more robust explanatory framework for forward-looking borehole video–based fracture recognition than approaches relying on unwrapped panoramic images.

4.7 Practical considerations and limitations

The proposed fracture recognition and quantification method is designed for forward-looking borehole imaging and applies to linear or band-shaped fractures that develop along the direction of video progression. Such fractures commonly occur in bedding cuts and water-conducting fracture zones, where their spatial extension is approximately parallel to the borehole axis. This orientation results in stable, linear, high-contrast textures in forward-looking videos. For these fractures, the model demonstrates high detection completeness and temporal stability, supporting continuous quantification of geometric parameters such as fracture length, width, and orientation.

The proposed method is limited in its capacity to address nonlinear structures, including joint planes, circumferential fractures, and broken zones. In forward-looking imaging, these features typically manifest as planar, arc-shaped, or scattered patterns, with texture orientations that differ significantly from those of linear fractures. As a result, they are classified as "broken" regions during the YOLO detection stage. In the absence of a well-defined linear central axis, the geometric parameters of these fractures are difficult to compute directly, which limits the applicability of the current method.

5. Conclusions

Accurate fracture recognition in forward-looking borehole videos and quantification of their geometric parameters are essential for evaluating surrounding rock stability and mitigating geohazards. This study proposes a video-domain pipeline for automatic fracture recognition and steady-state parameter quantification, validated on a self-constructed borehole video dataset. The primary conclusions are as follows:

(1) Proposal of a video-domain fracture recognition and steady-state quantification method: A deep learning-based detection–segmentation scheme is introduced to localize fractures and extract pixel-level masks automatically. Feature matching and optical flow tracking, combined with consistency screening, establish cross-frame consistent associations for the same fracture in raw videos. Multi-frame fusion is then applied to convert single-frame measurements of fracture length, width, and orientation into steady-state estimates.

(2) Dataset construction and verification of recognition performance: A dataset comprising videos from 55 boreholes is constructed for ablation and comparative experiments. On the validation set, the proposed method achieves mP@0.5 and mR@0.5 of 90.76% and 88.20%, respectively, and a segmentation Dice score of 82.51%. These results demonstrate accurate recognition and robust mask extraction for multi-scale fractures with complex morphologies.

(3) Evaluation of tracking stability and quantification accuracy with panorama-domain comparisons: Cross-frame tracking achieves a matching success rate of 94.5% with a drift error below 2.6 pixels, enabling reliable fracture trajectories for fusion. The fused length, width, and orientation parameters show strong agreement with manual measurements ($R^2 > 0.995$), and quantification error decreases as the number of fusion frames increases. Compared with three mainstream panorama-domain quantification methods, the mean quantification errors of length, width, and angle are reduced by 52.9%, 58.1%, and 45.2%, respectively. These findings confirm that the proposed approach effectively suppresses distortion-induced errors from panoramic unwrapping and provides reliable inputs for engineering evaluation.

Appendix A

Pseudocode for video-domain fracture recognition.

Innovations: A raw video-domain "detection–segmentation" pipeline is adopted. Detection provides high-recall fracture candidates; instance segmentation refines pixel-level masks with boundary-weighted supervision (WBCE) to improve boundary continuity and geometric fidelity.

Stage 1: Detection (YOLOv8-CAPS)

1. **Input frame**: I_t // Forward-looking borehole video frame at time t.
2. **YOLOv8-CAPS inference**: $B_t \leftarrow D(I_t; \tau_{conf}, \tau_{nms})$ // Generate candidate boxes with high recall.
3. **Class filtering**: retain $b \in B_t$ with class = fracture // Discard irrelevant classes (e.g., central/broken) if not needed.

Stage 2: Instance segmentation (Mask R-CNN with WBCE supervision)

4. **RoI cropping**: $I_t, b \leftarrow C_{rop}(I_t, b)$ // Extract region-of-interest for each fracture box.
5. **Mask prediction**: $M_t, b \leftarrow S(I_t, b)$ // Mask head outputs pixel-level fracture mask within RoI.
6. **Paste-back mapping**: $M_t(k) \leftarrow$ Paste Back(M_t, b) // Map RoI mask to full-frame coordinates.
7. **Mask set aggregation**: $M_t \leftarrow M_t \cup \{M_t(k)\}$ // Collect all fracture instance masks in frame t.

Training-time boundary-weighted supervision (WBCE, applied to mask branch only)

8. **Neighborhood occupancy**: $S(u, v)=\sum\sum Y(u+\Delta u, v+\Delta v)$ // 3×3 foreground count.
9. **Boundary extraction**: $B(u, v)=1$ if $Y(u, v)=1$ and $S(u, v)<9$; else 0 // Inner boundary of GT mask.
10. **Distance to boundary**: $\text{dist}(u, v)=\min(u_b, v_b) \in \Omega B \,//\, (u, v)-(u_b, v_b)2 \quad // \Omega B=\{(u, v) \mid B(u, v)=1\}$.
11. **Clipped normalisation**: $D(u, v)=\min(\text{dist}(u, v), R)$ // Radius R in pixels controls emphasis range.
12. **Weight map**: $w(u, v)=1+\alpha(1-D(u, v))$ // Boundary pixels get maximum weight $1+\alpha$.
13. **WBCE loss**: $L_{WBCE}=-1/N \cdot \sum w(u, v)[Y\log P+(1-Y)\log(1-P)]$.
14. **Mask-branch loss**: $L_{mask}=\lambda L_{WBCE}+(1-\lambda)L_{Dice}$ // $\lambda=0.7$ (Sec. 3.1).

Note: The WBCE weight map is computed from the GT mask (boundary + distance transform), ensuring consistency with Sec. 2.1.2 and avoiding gradient-based edge heuristics.

Loss: Detection loss (YOLOv8 family) + L_{mask} for instance segmentation.

Function: Produce boundary-consistent fracture masks M_t for subsequent cross-frame association and steady-state quantification.

Appendix B.

Pseudocode for steady-state quantification via cross-frame association and multi-frame fusion.

Innovations: Cross-frame trajectories are established by integrating mask-constrained feature matching and optical-flow tracking with a two-stage gating scheme (spatial IoU + motion–morphology consistency). Multi-frame fusion converts per-frame measurements into steady-state parameters.

1. **Inputs:** frames $\{I_t\}$; instance masks $\{M_t\}$; optional scale factor $s_t(\cdot)$; thresholds τ_{iou}, τ_{in}, τ_{re}, τ_{θ} (Table 1); $\tau_{iou}=0.75$; $W_{reconnect}=5$; $m_{max}=5$; trim rate $\eta=10\%$.

2. **Outputs:** per track q: {L_q, W_q, θ_q} (optional dip/strike if enabled).

3. **Single-frame measurement from masks (Sec. 2.2.1)**

4. **Distance field**: $H_t(k) \leftarrow$ Distance Transform $M_t(k)$ // inside-mask distance to background.

5. **Skeleton extraction**: $S_t(k) \leftarrow$ Skeletonize $M_t(k)$.

6. **Centroids**: $c_m \leftarrow$ Mask Centroid $M_t(k)$, $c_s \leftarrow$ Skeleton Centroid $S_t(k)$.

7. **Refined centroid**: $c \leftarrow \lambda c_m + (1-\lambda) c_s$, λ=0.6 // Stabilize center cue for gating.

8. **Representative width**: $p^* \leftarrow$ Nearest Skeleton Pixel($S_t(k)$, c); $\Omega \leftarrow$ Skeleton Neighborhood($S_t(k)$, p^*, radius=12);

$w_t(k) \leftarrow$ median{$2H_t(k)(p) : p \in \Omega$} // Robust to endpoints/branching.

9. **Physical width**: $w_c(k) \leftarrow w_t(k) \cdot s_t(c_s)$ if $s_t(\cdot)$ available.

10. **2D orientation**: $\theta_t(k) \leftarrow$ Principal Axis Angle $M_t(k) \in [0^\circ, 180^\circ)$.

11. **Cross-frame association and trajectory update (Sec. 2.2.2)**

12. **Stage I gate (spatial overlap)**:

Compute IoU($M_{t-1}(q)$, $M_t(k)$).

Candidate set: C = {k | IoU $\geq \tau_{iou}$} // Suppress adjacent-fracture switches.

13. **Mask-constrained matching initialisation**: SURF key points/descriptors within mask; KNN ratio test $d_1/d_2 < \tau_{ratio}$.

14. **Optical-flow refinement**: Lucas–Kanade tracking to refine correspondences.

15. **Hessian stability filter**: retain key points with significant Hessian determinant // Mitigate illumination/specular drift.

16. **RANSAC motion fitting**: fit inter-frame motion model; obtain inliers (anchors).

17. **Stage II gate (motion + morphology)**:

r_{in} = | inliers | /max(| matches | ,1); e_{re}=median reprojection error (pixels); $\Delta\theta$=min(| $\theta_{t-1} - \theta_t$ | ,180° − | $\theta_{t-1} - \theta_t$ |)

Update only if $r_{in} \geq \tau_{in}$ and $e_{re} \leq \tau_{re}$ and $\Delta\theta \leq \tau_\theta$; otherwise freeze.

18. **Failure recovery (Sec. 2.2.2)**

19. **Freeze update**: keep last valid state when gating fails; increment miss counter.

20. **Limited-window reconnection**: attempt re-association for next $W_{reconnect}$=5 frames under the same gates; reset miss on success.

21. **Track termination**: if consecutive failures reach m_{max}=5, terminate the track and trigger fusion.

Steady-state fusion (Sec. 2.2.3)

22. **Steady segment selection**: keep frames satisfying gating (optional T^*).

23. **Length fusion**: back-project anchor pixels to 3D (if available), order along dominant extension direction, accumulate polyline distances to obtain L_q.

24. **Width fusion**: $W_q \leftarrow$ Trimmed Mean({$w_c(t)$}$t \in T^*$, η=10%) // Remove tail outliers.

25. **Orientation fusion (PCA)**:

map θ_t to v_t=(cosθ_t, sinθ_t), perform PCA, take first principal component vPCA,

set θq=atan2(vPCA_y, vPCA_x) and map to [0°,180°).

Note: Candidate selection should be deterministic (e.g., max IoU then min centroid distance) to ensure reproducibility.

Loss: Not applicable (inference/estimation stage); gating thresholds follow Table 1.

Function: Establish reliable fracture trajectories and transform instantaneous per-frame measurements into steady-state outputs (L_q, W_q, θ_q).

References

[1] Liu, J., AlDajani, O., Li, B. Q., Einstein, H. & Li, Z. Automatic and accurate characterization of rock fractures based on Deep Learning. *Engineering Applications of Artificial Intelligence* **163**, 112950, doi:https://doi.org/10.1016/j.engappai.2025.112950 (2026).

[2] Qiu, Y., Zhou, J. & Li, C. Autonomous Fracture Trace Characterization in Underground Hard-Rock Environments via a Multi-scale Hierarchical Transformer Network. *Rock Mechanics and Rock Engineering*, doi:https://doi.org/10.1007/s00603-025-04905-x (2025).

[3] Bae, D. S., Kim, K. S., Koh, Y. K. & Kim, J. Y. Characterization of joint roughness in granite by applying the scan circle technique to images from a borehole televiewer. *Rock Mechanics and Rock Engineering* **44**, 497-504, doi:https://doi.org/10.1007/s00603-011-0134-9 (2011).

[4] Chen, Z. *et al.* Quantitative characterization of fracture surface undulations and gas-guiding patterns in fractured rocks under steady loading. *International Journal of Mining Science and Technology* **35**, 1753-1773, doi:https://doi.org/10.1016/j.ijmst.2025.08.017 (2025).

[5] Cui, L. *et al.* Bearing capacity of composite tunnel lining in soft rocks: Large-scale model test. *Journal of Rock Mechanics and Geotechnical Engineering*, doi:https://doi.org/10.1016/j.jrmge.2025.06.046 (2025).

[6] Ma, K. *et al.* Determination of the rock mass bearing mechanism following excavation of circular tunnels in layered surrounding rock. *Tunnelling and Underground Space Technology* **164**, 106863, doi:https://doi.org/10.1016/j.tust.2025.106863 (2025).

[7] Wang, Y., Wang, J., Li, B. & Liu, T. Dynamic classification method for TBM cutting face rock mass boreability based on tunneling characteristics and multi-criteria decision analysis. *Results in Engineering* **29**, 108614, doi:https://doi.org/10.1016/j.rineng.2025.108614 (2026).

[8] Wang, J. *et al.* A novel multi-scale structural equivalent method for jointed rock masses and its application to slope stability analysis. *International Journal of Rock Mechanics and Mining Sciences* **197**, 106329, doi:https://doi.org/10.1016/j.ijrmms.2025.106329 (2026).

[9] Benà, E. *et al.* A new perspective in radon risk assessment: Mapping the geological hazard as a first step to define the collective radon risk exposure. *Science of The Total Environment* **912**, 169569, doi:https://doi.org/10.1016/j.scitotenv.2023.169569 (2024).

[10] Wang, C., Han, Z., Wang, Y. & Wang, J. Study on in situ measurement and analysis method of rock pores based on borehole camera technology. *Canadian Geotechnical Journal* **62**, 1-13, doi:https://doi.org/10.1139/cgj-2023-0550 (2025).

[11] Jiang, T., Meng, F. & Wan, L. A study on the automation of borehole image auto-recognition based on instance segmentation. *Engineering Applications of Artificial Intelligence* **141**, 109789, doi:https://doi.org/10.1016/j.engappai.2024.109789 (2025).

[12] Xu, E. *et al.* Study on the development pattern of continuous aquitard water-conducting fracture zones and rock layer detection borehole camera technology. *Scientific Reports* **15**, 40418, doi:https://doi.org/10.1038/s41598-025-13349-w (2025).

[13] Wei, L. *et al.* Web-based visualisation for look-ahead ground imaging in tunnel boring machines. *Automation in Construction* **105**, 102830, doi:https://doi.org/10.1016/j.autcon.2019.04.025 (2019).

[14] Xie, H. M., Qiu, H. & Yang, X. B. in *FUZZY SYSTEMS AND DATA MINING V (FSDM 2019)* Vol. 320 746-759 (2019).

[15] Li, M., Chen, M., Lu, W., Yan, P. & Tan, Z. Automatic extraction and quantitative analysis of characteristics from complex fractures on rock surfaces via deep learning. *International Journal of Rock Mechanics and Mining Sciences* **187**, 106038, doi:https://doi.org/10.1016/j.ijrmms.2025.106038 (2025).

[16] Ji, Y. *et al.* Automatic identification of rock fractures based on deep learning. *Engineering Geology* **345**, 107874, doi:https://doi.org/10.1016/j.enggeo.2024.107874 (2025).

[17] Naik, D. L. & Khan, R. Identification and characterization of fracture in metals using machine learning based texture recognition algorithms. *ENGINEERING FRACTURE MECHANICS* **219**, doi:https://doi.org/10.1016/j.engfracmech.2019.106618 (2019).

[18] Młynarczuk, M., Habrat, M. & Skoczylas, N. The application of the automatic search for visually similar geological layers in a borehole in introscopic camera recordings. *Measurement* **85**, 142-151, doi:https://doi.org/10.1016/j.measurement.2016.02.043 (2016).

[19] Yu, M., Chen, W. & Hou, J. A lightweight network for concrete crack detection in complex environments based on cloud-edge collaboration. *Expert Systems with Applications* **299**, 130204, doi:https://doi.org/10.1016/j.eswa.2025.130204 (2026).

[20] Li, E. *et al.* Lithology identification using borehole images by contrast-limited adaptive histogram equalization and machine learning models. *Journal of Rock Mechanics and Geotechnical Engineering*, doi:https://doi.org/10.1016/j.jrmge.2025.08.002 (2025).

[21] Zhang, L., Zheng, A., Sun, X. & Sun, Z. Enhanced YOLOv11-Based River Aerial Image Detection Research. *IEEE Geoscience and Remote Sensing Letters* **22**, 1-5, doi:https://doi.org/10.1109/LGRS.2025.3576640 (2025).

[22] Chang, H., Bassir, D., Barrios, A. & Chen, G. Fourier-Mixture of Experts YOLO for concrete crack segmentation with visual

interpretability. *Automation in Construction* **179**, 106452, doi:https://doi.org/10.1016/j.autcon.2025.106452 (2025).

[23] Huang, W. *et al.* Multi-condition dam crack segmentation based on a novel lightweight CEE-YOLO model. *Advanced Engineering Informatics* **68**, 103697, doi:https://doi.org/10.1016/j.aei.2025.103697 (2025).

[24] Lu, Y. *et al.* Dynamic snake vision mask r-cnn: a method for microfracture detection in electrical logging imaging. *Earth Science Informatics* **18**, 544, doi:https://doi.org/10.1007/s12145-025-02014-5 (2025).

[25] Alzubaidi, F. *et al.* Automatic fracture detection and characterization from unwrapped drill-core images using mask R–CNN. *Journal of Petroleum Science and Engineering* **208**, 109471, doi:https://doi.org/10.1016/j.petrol.2021.109471 (2022).

[26] Deng, Z., Cao, M., Geng, Y. & Rai, L. Generating a cylindrical panorama from a forward-looking borehole video for borehole condition analysis. *Applied Sciences (Switzerland)* **9**, doi:https://doi.org/10.3390/app9163437 (2019).

[27] Liu, C. *et al.* Intelligent recognition and identification of fracture types and parameters for borehole images based on developed convolutional neural networks and post-processing. *Engineering Fracture Mechanics* **292**, 109624, doi:https://doi.org/10.1016/j.engfracmech.2023.109624 (2023).

[28] Kudasik, M. *et al.* Investigation of strata fractures during longwall mining: Original introscopic probe and image analysis methods. *Journal of Rock Mechanics and Geotechnical Engineering* **18**, 1049-1059, doi:https://doi.org/10.1016/j.jrmge.2025.04.034 (2026).

[29] Ge, Y., Liu, G., Wang, H., Tang, H. & Zhao, B. Rock Joint Detection from Borehole Imaging Logs Using a Convolutional Neural Networks Model. *Journal of Earth Science* **36**, 1700-1716, doi:https://doi.org/10.1007/s12583-024-1989-5 (2025).

[30] Zong, N., Al-Nuaimy, W. & Worthington, P. Borehole Inner Surface Visualization System with Vibration Cancellation and Trajectory Smoothing Based on Optical Monocular Video Camera. *SN Computer Science* **6**, 453, doi:https://doi.org/10.1007/s42979-025-03967-6 (2025).

[31] Liu, Q., Huo, J., Tang, X. & Xue, M. Image stitching method for CMOS grayscale cameras in industrial applications. *Optics & Laser Technology* **181**, 111874, doi:https://doi.org/10.1016/j.optlastec.2024.111874 (2025).

[32] Wang, J., Fang, X., Song, Y. & He, D. Fracture detection in tunnel borehole images using MSFDNet: An interpretable multi-scale network with partial convolution and bi-level routing attention. *Tunnelling and Underground Space Technology* **169**, 107292, doi:https://doi.org/10.1016/j.tust.2025.107292 (2026).

[33] Liu, G., Ding, W., Shu, J., Strauss, A. & Duan, Y. Two-Stream Boundary-Aware Neural Network for Concrete Crack Segmentation and Quantification. *Structural Control and Health Monitoring* **2023**, doi:https://doi.org/10.1155/2023/3301106 (2023).

[34] Song, S., Lu, M., Su, Y. & Liu, T. CCMPIP: Cross-attention and capsule network-based multi-feature fusion for proinflammatory peptide prediction. *Computational Biology and Chemistry* **121**, 108846, doi:https://doi.org/10.1016/j.compbiolchem.2025.108846 (2026).

[35] Wen, C. *et al.* YOLOv8-DSRI:an improved YOLOV8 instance segmentation algorithm for identifying silkworms in dense environments. *Pattern Recognition*, 113017, doi:https://doi.org/10.1016/j.patcog.2025.113017 (2025).

[36] Wei, Y. *et al.* MCN: A mixture capsule network for authentic blind image quality assessment. *Knowledge-Based Systems* **331**, 114840, doi:https://doi.org/10.1016/j.knosys.2025.114840 (2026).

[37] Yan, J. *et al.* A novel adaptive capsule network with dual-branch feature extraction for multi-source partial discharge diagnosis in gas-insulated switchgear. *International Journal of Electrical Power & Energy Systems* **173**, 111369, doi:https://doi.org/10.1016/j.ijepes.2025.111369 (2025).

[38] Zheng, R. *et al.* Mask R-CNN algorithm on detecting plantable urban roofs from aerial photography images in Central London, UK. *Remote Sensing Applications: Society and Environment* **40**, 101791, doi:https://doi.org/10.1016/j.rsase.2025.101791 (2025).

[39] Yang, S. *et al.* Optimizing foreign fiber segmentation performance with DeepLab V3+ and GAN in industrial IoE environments. *Digital Communications and Networks*, doi:https://doi.org/10.1016/j.dcan.2025.03.005 (2025).

[40] Liu, C., Liu, C., Zhang, Y., Wu, Z. & Guo, J. Efficient RANSAC in 4D Plane Space for Point Cloud Registration. *Graphical Models* **141**, 101289, doi:https://doi.org/10.1016/j.gmod.2025.101289 (2025).

[41] Liu, R. *et al.* A multi-method for identifying and spatial modeling of groundwater nitrate natural background levels in an

urbanized delta in China: Insight from Cl/Br ratios combining with KNN algorithm and EBK. *Water Research* **289**, 124953, doi:https://doi.org/10.1016/j.watres.2025.124953 (2026).

[42] Song, S. *et al.* Research on intelligent segmentation method of coal body CT image fracture based on CBAM-UNet. *Results in Engineering* **27**, 106528, doi:https://doi.org/10.1016/j.rineng.2025.106528 (2025).

[43] Deepti Raj, G. & Prabadevi, B. Enhancing surface detection: A comprehensive analysis of various YOLO models. *Heliyon* **11**, e42433, doi:https://doi.org/10.1016/j.heliyon.2025.e42433 (2025).

[44] Li, J., Yuan, C., Wang, X., Chen, G. & Ma, G. Semi-supervised crack detection using segment anything model and deep transfer learning. *Automation in Construction* **170**, 105899, doi:https://doi.org/10.1016/j.autcon.2024.105899 (2025).

[45] Chen, Z. & Shen, Z. Precise tunnel lining crack detection using multimodal fusion transformer and metaheuristic-optimized ResNet-50. *Alexandria Engineering Journal* **134**, 527-541, doi:https://doi.org/10.1016/j.aej.2025.12.022 (2026).

[46] Jinchao, W. *et al.* Monitoring method for borehole images of changes in overlying rock structure during coal mining. *Scientific Reports* **15**, 15495, doi:https://doi.org/10.1038/s41598-025-00513-5 (2025).

[47] Letshwiti, T. M., Shahsavar, M., Moniri-Morad, A. & Sattarvand, J. Deep learning-based image segmentation for highwall stability monitoring in open pit mines. *Journal of Engineering Research* **13**, 3595-3608, doi:https://doi.org/10.1016/j.jer.2025.04.002 (2025).

[48] Wang, X., Wang, T. & Li, J. Advanced crack detection and quantification strategy based on CLAHE enhanced DeepLabv3+. *Engineering Applications of Artificial Intelligence* **126**, 106880, doi:https://doi.org/10.1016/j.engappai.2023.106880 (2023).

Acknowledgments: This work was supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. 2022XSCX01).

Author Contributions: Conceptualization, W.Z., D.Z., G.F. and Y.M.; methodology, W.Z., D.Z. and Y.M.; software, W.Z. and Y.M.; validation, W.Z., Y.M. and Z.F.; formal analysis, Z.F.; investigation, W.Z.; resources, D.Z., G.F. and Y.M.; data curation, W.Z.; writing—original draft preparation, W.Z.; writing—review and editing, W.Z., D.Z., G.F., Y.M. and Z.F.; visualization, W.Z. and Z.F.; supervision, D.Z. and Y.M.; project administration, W.Z., D.Z and G.F.; funding acquisition, D.Z. and G.F. All authors have read and agreed to the published version of the manuscript.

Competing interests: The authors declare no competing interests.

Data Availability Statement: The data that support the findings of this study are available from State Key Laboratory for Fine Exploration and Intelligent Development of Coal Resources but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Professor Dongsheng Zhang.

Funding: Supported by the Postgraduate Research & Practice Innovation Program of Jiangsu Province (Grant No. 2022XSCX01).