remote sensing

MDPI

Article

A UAV–TLS Point-Cloud Fusion Framework for Three-Dimensional Characterization of Mining-Induced Surface Cracks

Weiwei Zhou [1], Youfeng Zou [1,*], Huabin Chai [1], Lailiang Cai [1], Weibing Du [1], Jibiao Hu [2], Miaomiao Ma [1] and Mengnan Li [1]

[1] School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China; 112404010003@home.hpu.edu.cn (W.Z.); chaihb@hpu.edu.cn (H.C.); cll@hpu.edu.cn (L.C.); dwb@hpu.edu.cn (W.D.); mamiaomiao@home.hpu.edu.cn (M.M.); 112504010016@home.hpu.edu.cn (M.L.)
[2] College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China; gqt2300204005@student.cumtb.edu.cn
* Correspondence: zouyf@hpu.edu.cn

Highlights

What are the main findings?

- A UAV–TLS fusion framework was developed to convert mining-induced surface cracks from flat image-based detections into measurable three-dimensional spatial objects.
- Semantic crack point clouds were constructed by linking DRA-UNet-derived crack semantics with TLS geometry through skeleton vectorization and KDTree-based spatial matching.

What are the implications of the main findings?

- The framework bridges the gap between crack recognition and true 3D geometric characterization under real topographic conditions.
- The proposed workflow enables more reliable crack measurement and provides a practical tool for mining-induced ground damage monitoring and ecological restoration assessment.

Abstract

Mining-induced surface cracks are direct indicators of ground damage caused by underground coal extraction, yet their accurate three-dimensional (3D) characterization remains challenging because most unmanned aerial vehicle (UAV)-based studies are limited to two-dimensional (2D) detection and planar parameter extraction. This study proposes a UAV–terrestrial laser scanning (TLS) fusion framework for measurable 3D characterization of mining-induced surface cracks under real topographic conditions. High-resolution UAV orthomosaics were used to extract crack semantics with a dual residual-attention U-Net (DRA-UNet). The resulting crack masks were skeletonized and vectorized, and a k-dimensional tree (KDTree)-based spatial matching strategy was applied to link crack centerlines with TLS point clouds, thereby constructing semantic crack point clouds. The framework was evaluated at the Huojitu Mine, Shenmu City, China, using approximately 5000 crack samples. DRA-UNet achieved precision, recall, F1-score, and mean intersection over union values of 85.13%, 77.84%, 81.32%, and 70.26%, respectively. Compared with conventional 2D length measurements, the proposed 3D estimation reduced the mean relative error from 4.49% to 2.64%. Width validation at 24 field points yielded a mean relative error of 4.47%. These results show that UAV–TLS fusion can bridge the gap between

Check for updates

Academic Editor: Guoqing Zhou

Received: 9 June 2026
Revised: 2 July 2026
Accepted: 15 July 2026
Published: 21 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

planar crack detection and true 3D geometric characterization, providing a practical tool for mining-induced ground damage monitoring.

Keywords: mining-induced surface cracks; UAV remote sensing; TLS; DRA-UNet; 3D characterization; point cloud fusion

1. Introduction

Mining-induced surface cracks are among the most direct manifestations of ground damage caused by underground coal extraction [1–5]. Their development can disrupt surface continuity and land structural stability and may further trigger secondary environmental problems, including soil erosion, vegetation degradation, and localized ground instability. Accurate identification of mining-induced surface cracks and quantitative characterization of their geometric properties are therefore essential for hazard monitoring, ecological restoration, and surface stability assessment in mining areas [1,2].

Conventional crack investigation mainly relies on field surveys, tape measurements, total station observations, Global Navigation Satellite System (GNSS) measurements, and visual interpretation of remote sensing imagery. Although these methods can provide reliable crack information at local scales, they are generally labor-intensive, spatially discontinuous, and strongly dependent on human interpretation. In mining-affected areas with extensive crack networks and complex terrain, conventional approaches are often insufficient for large-scale, high-precision, and high-frequency monitoring [2,5].

Recent advances in UAV remote sensing and deep learning have substantially improved the automatic detection of mining-induced surface cracks [6–9]. Semantic segmentation methods based on high-resolution UAV orthomosaics enable pixel-level crack extraction and support the derivation of 2D geometric parameters, including crack length, width, area, density, and spatial distribution. However, most existing studies still treat cracks as 2D image objects, with results mainly represented by crack masks, skeletons, or planar geometric descriptors. In mining areas, surface deformation is commonly accompanied by subsidence basins, terrain undulations, and localized collapses. As a result, the actual propagation path of a crack may deviate from its planar projection. Measurements derived solely from UAV orthomosaics therefore cannot fully represent crack geometry under real topographic conditions, which limits further analysis of the relationship between crack development and mining-induced surface deformation.

Point-cloud data provide new opportunities for 3D crack characterization [10–13]. TLS and photogrammetric point clouds can capture detailed 3D coordinates and elevation information, enabling the representation of surface structures, terrain undulations, and micro-topographic variations. Nevertheless, direct extraction of mining-induced surface cracks from point clouds remains challenging [12,13]. Vegetation cover, shadows, loose surface materials, surface roughness, point-density variations, occlusions, and noise may lead to unstable representations of fine crack boundaries. Therefore, rather than replacing image-based crack detection with point-cloud-based extraction, a more feasible strategy is to integrate crack semantic information derived from UAV imagery with the geometric information provided by TLS point clouds. Such integration can extend crack representation from 2D image objects to 3D spatial objects [10,11].

Although multi-source remote sensing data fusion has received increasing attention, studies on mining-induced surface cracks have predominantly focused on 2D detection and planar parameter extraction. Research on constructing semantic 3D crack objects and quantifying crack geometric parameters under real topographic conditions remains

limited [11,12]. Establishing a spatial linkage between crack semantic information and 3D terrain information is therefore critical for transforming crack analysis from 2D detection to 3D quantitative characterization.

To address these limitations, this study proposes a UAV–TLS fusion framework for 3D characterization of mining-induced surface cracks [13,14]. First, crack semantic information is extracted from UAV orthomosaics using DRA-UNet [11,12,15–17]. Crack centerlines are then generated through connected-component analysis, skeleton extraction, and vectorization [18]. Subsequently, KDTree-based neighborhood matching is employed to map 2D crack objects into 3D point cloud space, thereby constructing semantic crack point clouds [19,20]. Finally, 3D geometric parameters, including crack length and width, are estimated by integrating planar crack information with TLS-derived elevation information.

Different from previous studies that mainly focused on 2D crack detection, planar geometric parameter extraction, or direct crack extraction from point clouds, this study emphasizes the transformation of image-derived crack semantics into measurable 3D spatial objects. Specifically, UAV orthomosaics are used to provide reliable crack semantic boundaries and centerlines, while TLS point clouds provide elevation and terrain morphology information. By linking these two data sources through skeleton vectorization and KDTree-based spatial matching, the proposed framework enables the construction of semantic crack point clouds and the calculation of 3D crack geometric parameters under real topographic conditions.

The main contributions of this study are threefold. First, a UAV–TLS fusion framework is developed to extend the representation of mining-induced surface cracks from 2D image objects to 3D spatial objects. Second, a spatial association strategy between crack centerlines and TLS point clouds is established to construct semantic crack point clouds. Third, an automated workflow for estimating 3D crack length and width is developed to support quantitative geometric characterization under real topographic conditions.

2. Study Area and Data

2.1. Study Area Overview

The study area is located in the mining-affected region of the Huojitu Coalfield, Daliuta Town, Shenmu City, Shaanxi Province, China. It lies in the transitional zone between the loess plateau hills and gullies and the Mu Us Desert. The region has a semi-arid to arid continental climate, with annual precipitation of approximately 250–440 mm, mainly concentrated from July to September, and annual potential evaporation of 1800–2500 mm. The terrain is dominated by loess-covered hills and is widely covered by mobile and semi-fixed sand. Loose surface materials, low erosion resistance, and fragile ecological conditions make the area sensitive to mining-induced disturbance [21].

Affected by underground coal extraction, the study area exhibits surface subsidence, tensile cracks, and localized deformation. The cracks show considerable variability in length, width, continuity, and spatial distribution, providing representative surface damage features for method validation. In addition, the open terrain and well-developed crack morphology allow the acquisition of high-resolution UAV orthomosaics and TLS point clouds over the same area. Therefore, this site is suitable for evaluating 2D crack identification and 3D crack characterization methods. The location of the study area is shown in Figure 1.

2.2. UAV Imagery and TLS Point-Cloud Data

High-resolution UAV orthomosaics and TLS point clouds were used as the primary data sources in this study. UAV orthomosaics provide centimeter-level spatial resolution and rich texture information, allowing surface features such as cracks, bare soil, roads, and vegetation to be clearly distinguished. They were therefore used as the basis for crack

semantic segmentation, boundary extraction, and skeleton analysis. In contrast, TLS point clouds record the 3D morphology of the ground surface and can represent local subsidence, terrain undulations, and micro-topographic variations around surface cracks [22]. These data provide essential elevation information for 3D crack representation and geometric parameter estimation.

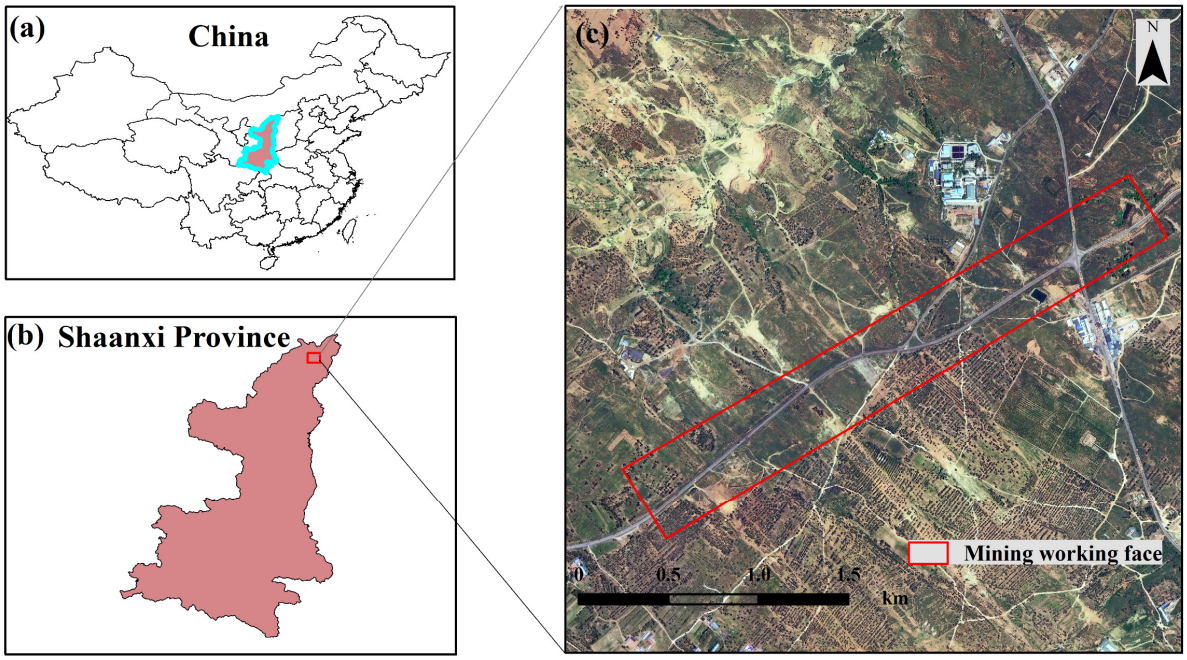


Figure 1. Location of the study area. (**a**) Location of Shaanxi Province within China; (**b**) location of the study area in Shaanxi Province; (**c**) mining panel and local orthomosaic image.

To ensure spatial consistency among the crack objects extracted from UAV orthomosaics, crack centerline vectors, and TLS point clouds, all datasets were transformed into the CGCS2000/3-degree Gauss–Kruger CM 111E projected coordinate system (EPSG:4546; units: m). This coordinate system is based on the CGCS2000 datum and uses a 3-degree Gauss–Kruger projection with a central meridian of 111°E, a false easting of 500,000 m, a false northing of 0 m, a scale factor of 1.0, and a latitude of origin of 0°. The spatial co-registration accuracy between the UAV orthomosaic and TLS point clouds was evaluated using common check points distributed across the study area. The horizontal root mean square error (RMSE) of the UAV-TLS co-registration was 0.03 m, indicating that the two datasets had sufficient planimetric consistency for subsequent crack centerline-to-point-cloud matching and 3D geometric parameter calculation. In the proposed framework, UAV orthomosaics provide crack semantic boundaries and planar spatial distributions, whereas TLS point clouds supply the corresponding 3D geometric information at crack locations. The complementary characteristics of these two data sources form the basis for semantic crack point-cloud construction and subsequent 3D geometric characterization. The acquisition parameters of the UAV orthomosaics and TLS point clouds are summarized in Tables 1 and 2, respectively.

It should be noted that the flight altitude of approximately 1190 m refers to the absolute UAV flight elevation above mean sea level rather than the relative flight height above the ground surface. Therefore, this value should not be directly used to interpret the ground sampling distance. The generated orthomosaic had a ground sampling distance of 2.76 cm/pixel after photogrammetric processing.

Table 1. Acquisition parameters of the UAV imagery.

Item	Description
Data type	UAV orthomosaic imagery
UAV platform	DJI Matrice 300 RTK
Image type	True-color visible imagery
Image size	4056 × 3040 pixels
Flight altitude	Approximately 1190 m above mean sea level
Focal length	40–60 mm
Exposure time	1/640 s
Ground sampling distance	2.76 cm/pixel
Data processing	Image matching, aerial triangulation, orthorectification, and image mosaicking
Main application	Crack extraction, binary mask generation, and skeleton extraction

Table 2. Acquisition parameters of the TLS point clouds.

Item	Description
Data type	TLS point clouds
Scanner	RIEGL VZ-1000 terrestrial laser scanner
Maximum range	1400 m
Minimum range	2.5 m
Range accuracy	8 mm
Range repeatability	5 mm
Scanning field of view	360°
Angular resolution	0.0005°
Data processing	Point cloud registration, denoising, cropping, and spatial reference verification
Main application	3D surface characterization, crack point-cloud matching, and estimation of crack length and elevation variations

2.3. Sample Annotation and Point-Cloud Preprocessing

The crack sample dataset was constructed from UAV orthomosaics covering representative crack-development areas within the same mining-affected area of the Huojitu Coalfield, Daliuta Town, Shenmu City, Shaanxi Province, China. The UAV orthomosaics were divided into image patches of 256 × 256 pixels using a sliding-window strategy with 20% overlap between adjacent patches. This overlap reduces the risk of crack truncation at patch boundaries and preserves the continuity of crack features. Crack regions in each patch were manually annotated using Labelme [23], and the annotations were subsequently converted into binary masks, with crack pixels assigned a value of 255 and background pixels assigned a value of 0.

The dataset included both crack-containing patches and background-dominated patches, including bare soil, roads, vegetation, and shadowed areas, to improve the model's ability to distinguish cracks from non-crack surface features. Approximately 5000 annotated image patches were established and then randomly divided at the patch level into training, validation, and testing subsets at a ratio of 8:1:1. The partitioning was performed after sliding-window cropping, rather than at the orthomosaic or acquisition-area level. Therefore, adjacent overlapping patches originating from the same orthomosaic may exist among the training, validation, and testing subsets, which may lead to optimistic estimates of segmentation performance compared with a fully spatially independent evaluation. Accordingly, the reported segmentation results mainly reflect patch-level crack extraction performance within the same study area, while cross-area generalization should be further evaluated using independent orthomosaics or datasets from other mining areas.

Prior to analysis, the raw TLS point clouds underwent preprocessing, including registration, denoising, cropping, and spatial consistency verification. Point clouds acquired

from different scanning stations were first registered into a unified spatial reference system, and registration accuracy was assessed using overlapping regions between adjacent scans. Statistical filtering combined with manual inspection was then applied to remove outliers, floating points, and abnormal points near scan boundaries, thereby reducing noise influence on elevation assignment and 3D crack-length estimation. Finally, the point clouds were clipped according to the study-area boundary and moderately downsampled while preserving major terrain characteristics. This procedure reduced redundant points and improved the computational efficiency of subsequent neighborhood-search operations.

3. Methodology

3.1. Overall Workflow

A UAV–TLS fusion framework was developed for the 2D detection and 3D characterization of mining-induced surface cracks. The framework integrates UAV orthomosaics and TLS point clouds to provide a complete workflow from crack extraction to 3D geometric parameter estimation. The workflow consists of four main stages: data preparation, 2D crack extraction, 3D spatial mapping, and geometric parameter estimation.

During the data-preparation stage, UAV orthomosaics and TLS point clouds are preprocessed to produce spatially consistent imagery and terrain point clouds within a unified coordinate reference system. Crack semantic information is then extracted from the UAV orthomosaics using a semantic segmentation model. The resulting binary crack masks are further processed through post-processing, connected-component analysis, skeleton extraction, and vectorization to generate crack centerlines with real-world coordinates.

In the 3D spatial-mapping stage, the spatial relationship between crack centerlines and TLS point clouds is established. A KDTree-based neighborhood search identifies point-cloud subsets corresponding to crack locations and assigns crack semantic attributes to the points. This process constructs semantic crack point clouds and extends crack representation from 2D image objects to 3D spatial objects.

Finally, geometric parameters are derived by integrating binary crack masks, crack centerlines, and semantic crack point clouds. 3D crack length and width are subsequently calculated, enabling quantitative characterization of crack geometry under real topographic conditions. The overall workflow of the proposed method is illustrated in Figure 2.

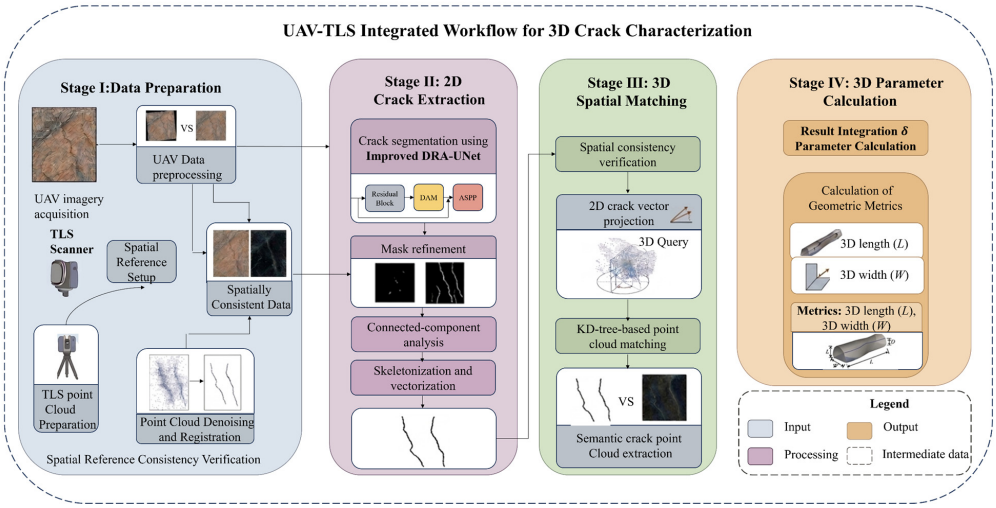


Figure 2. Overall workflow of the UAV–TLS fusion framework for mining-induced surface crack extraction, 3D spatial mapping, and geometric parameter estimation. The arrows indicate the processing sequence, and the different colors distinguish the main workflow stages and data-processing modules.

3.2. Crack Semantic Segmentation Based on DRA-UNet

Mining-induced surface cracks are typically characterized by elongated geometries, discontinuous boundaries, large-scale variations, and strong background interference. These characteristics make accurate crack extraction challenging for conventional semantic segmentation models, which may produce missed detections of narrow cracks, blurred crack boundaries, and false segmentation in complex surface environments. To address these challenges, DRA-UNet, proposed in our previous study, was adopted for crack semantic segmentation in this work.

DRA-UNet is based on the encoder–decoder architecture of U-Net and integrates a residual network (RN), a dual attention module (DAM), and an atrous spatial pyramid pooling (ASPP) module [24–27]. In this study, DRA-UNet consistently refers to the U-Net-based encoder–decoder segmentation network enhanced by residual learning, dual attention, and ASPP-based multi-scale atrous convolution. The RN was introduced by embedding residual blocks with identity shortcut connections. These shortcut connections allow shallow feature representations to be propagated to deeper layers, which can mitigate network degradation and gradient attenuation during deep feature extraction, rather than guaranteeing lossless information transmission. The DAM was designed to assign adaptive weights to spatial and channel features so that crack-related responses could be emphasized while background responses could be relatively reduced. The ASPP module was used to aggregate multi-scale contextual information through parallel atrous convolutions with different dilation rates, thereby supporting the representation of cracks with varying widths and spatial patterns. These descriptions represent the design motivations of the modules, and their practical contribution was further evaluated through the ablation experiment presented in Section 4.1.

By combining residual learning, attention mechanisms, and multi-scale feature aggregation, DRA-UNet was designed to improve the discrimination of crack features under complex mining-area conditions. It was also intended to support the extraction of fine crack structures and boundary details, thereby providing semantic information for subsequent skeleton extraction and 3D geometric characterization. The architecture of DRA-UNet is illustrated in Figure 3. In the decoder, the ASPP output from the bottleneck layer serves as the primary input for progressive upsampling. The skip connections from the encoder are used as auxiliary high-resolution features to compensate for spatial detail loss during downsampling. At each decoder stage, the upsampled decoder feature map is fused with the corresponding encoder feature map at the same spatial resolution by channel-wise concatenation, followed by convolutional refinement. Therefore, the skip connections shown in Figure 3 indicate scale-matched feature fusion rather than all-to-all routing from encoder stages to decoder stages.

RN blocks were embedded in both the encoder and decoder to facilitate information propagation between shallow texture features and deep semantic features, which may help alleviate gradient vanishing and network degradation during training. DAM was introduced to assign adaptive weights to spatial and channel features, with the aim of increasing the network's attention to crack-related regions. In the encoding and decoding stages, DAM was designed to support crack-feature representation and high-resolution feature refinement under complex background conditions.

The ASPP module was incorporated into the bottleneck layer to aggregate multi-scale contextual information through parallel atrous convolutions with different dilation rates. This design was intended to support the representation of narrow cracks, wide cracks, and branched crack structures. The integration of RN, DAM, and ASPP was designed to support feature propagation, attention-based feature weighting, and multi-scale contextual representation under complex mining-area conditions.

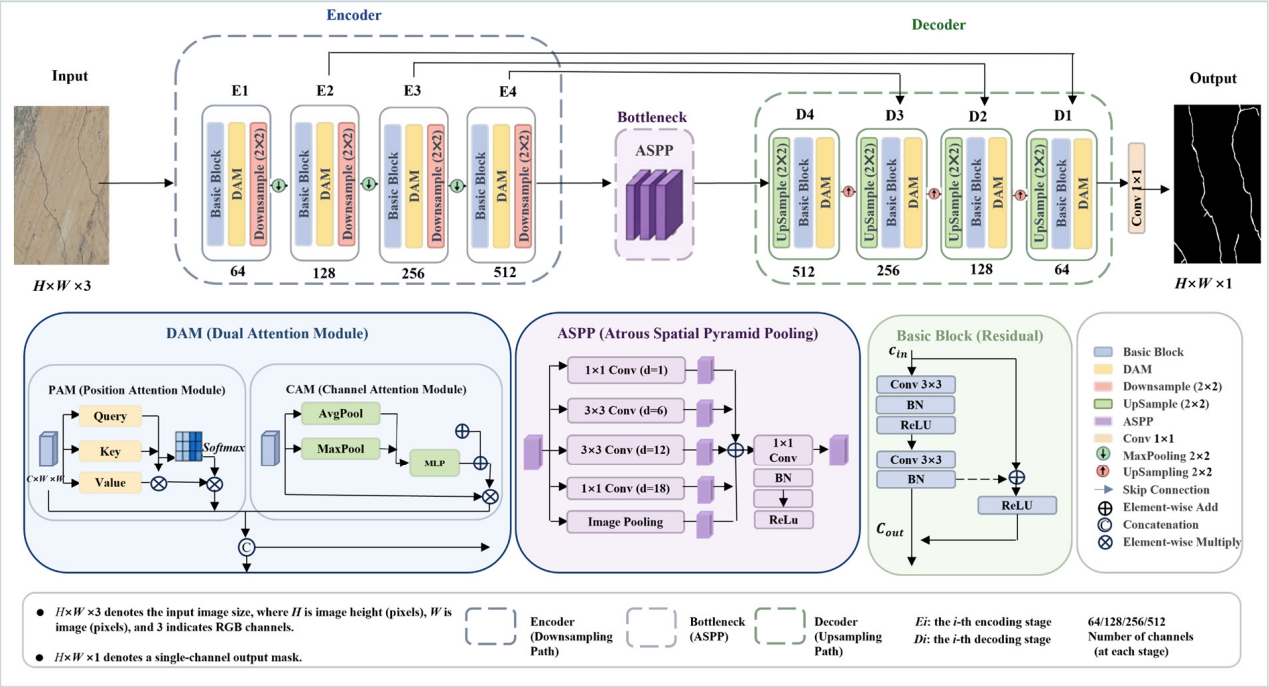


Figure 3. Architecture of DRA-UNet and its key components for mining-induced surface crack segmentation. The ASPP output serves as the primary input to the decoder, while encoder skip features provide auxiliary high-resolution information. At each decoder stage, the upsampled decoder feature map is fused with the corresponding encoder feature map at the same spatial resolution by channel-wise concatenation followed by convolutional refinement.

3.3. Crack Skeleton Extraction and Vectorization

After crack semantic segmentation, connected-component analysis, skeleton extraction, and vectorization were performed on the binary crack masks. Skeletonization reduces crack regions with finite widths to single-pixel-wide centerlines while preserving their topological structure and spatial connectivity. It should be noted that skeletonization is used to extract centerline representations from binary crack masks, rather than to reconstruct crack boundaries. The crack boundaries used for width estimation are extracted separately from the binary masks. This process removes redundant boundary information and improves the efficiency of subsequent crack-length estimation and point-cloud spatial matching. The binary crack mask is defined as:

$$M(\mathrm{x}, \mathrm{y}) \in \{0, 255\} \tag{1}$$

where $M(x, y) = 255$ represents crack pixels and $M(x, y) = 0$ represents background pixels. To avoid interference among adjacent crack regions, connected-component analysis was first used to separate and label individual crack objects. Independent crack units were identified using the eight-neighborhood connectivity criterion.

A topology-preserving skeletonization algorithm was then applied to each crack object. During this process, boundary pixels were iteratively removed while crack connectivity and the main topological structure were retained. The resulting single-pixel-wide crack skeleton can be expressed as:

$$S = Skeleton(M) \tag{2}$$

where S denotes the set of crack skeleton pixels. The extracted skeleton retains the primary propagation direction and spatial structural characteristics of each crack, providing a

compact representation of its centerline geometry. Therefore, it can be used to describe the crack centerline and serves as the basis for subsequent geometric analysis.

After skeleton extraction, crack centerlines were generated according to the neighborhood connectivity among skeleton pixels and converted into vector line objects. The vectorized centerlines preserve the principal propagation direction and topological relationships of crack structures while providing an efficient geometric representation for spatial analysis. More importantly, they serve as a geometric link between 2D crack semantic information and 3D point cloud space, enabling the subsequent construction of semantic crack point clouds and 3D geometric characterization.

3.4. Spatial Matching Between Crack Centerlines and Point Clouds

To extend crack representation from 2D image objects to 3D spatial objects, a spatial matching procedure was established between the vectorized crack centerlines and TLS point clouds. After transforming UAV orthomosaics, crack vector data, and TLS point clouds into the CGCS2000/3-degree Gauss–Kruger CM 111E projected coordinate system (EPSG:4546; units: m), neighborhood matching could be directly performed based on their spatial relationships. Let the set of crack centerline points be defined as:

$$C = \{ c_i(x_i, y_i) \mid i = 1, 2, \ldots, n \} \tag{3}$$

where $c_i(x_i, y_i)$ denotes the i-th centerline point and n is the total number of centerline points. The TLS point cloud is represented as:

$$P = \{ p_j(x_j, y_j, z_j) \mid j = 1, 2, \ldots, m \} \tag{4}$$

where $p_j(x_j, y_j, z_j)$ denotes the j-th point in the TLS point cloud, and m is the total number of points.

To improve the efficiency of spatial queries in large-scale point clouds, a KDTree structure was used to construct a spatial index for TLS points and facilitate neighborhood searches around crack centerline points, as illustrated in Figure 4 [28]. For each centerline point c_i, neighboring point-cloud points within a search radius r were identified according to:

$$N(c_i) = \{ p_j \in P \mid d_{xy}(c_i, p_j) \leq r \} \tag{5}$$

where $N(c_i)$ denotes the neighborhood point-cloud set associated with the centerline point c_i, and $d_{xy}(c_i, p_j)$ represents the horizontal Euclidean distance between the centerline point c_i and the TLS point p_j in the x-y plane. Specifically, $d_{xy}(c_i, p_j) = [(x_j - x_i)^2 + (y_j - y_i)^2]^{1/2}$, and r denotes the search radius. Although each TLS point contains a three-dimensional coordinate $p_j = (x_j, y_j, z_j)$, the elevation coordinate z_j was not used in the KDTree neighborhood search. This is because the purpose of this step is to establish the planar spatial association between image-derived crack centerlines and TLS points. The elevation coordinate z_j of the matched TLS points was retained and used in the subsequent 3D geometric parameter calculation. The KDTree search radius r is a key parameter controlling the spatial association between crack centerline points and TLS point-cloud points. In this study, r was set to 0.15 m by considering the UAV-TLS co-registration accuracy, UAV image ground sampling distance, TLS point density, and the typical crack-width scale in the study area. This radius is larger than the horizontal co-registration RMSE of 0.03 m and can tolerate minor spatial misalignment and local point-cloud sampling gaps, while limiting the inclusion of unrelated ground points around the crack. If r is too small, semantic crack point clouds may become discontinuous because some crack-related TLS points cannot be matched. Conversely, if r is too large, non-crack ground points may be included, which may reduce the geometric precision of subsequent crack length and width estimation. Therefore,

r = 0.15 m was adopted as a compromise between matching completeness and geometric reliability for semantic crack point-cloud construction.

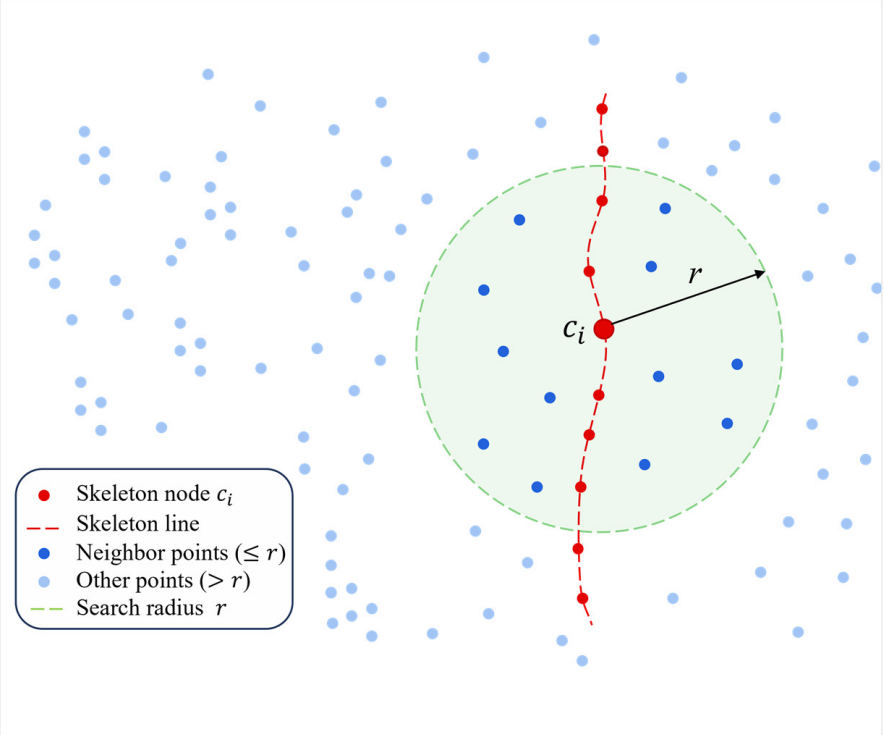


Figure 4. KDTree-based neighborhood search for constructing semantic crack point clouds.

Neighborhood queries were performed independently for all sampled centerline points. The semantic crack point cloud was constructed as the union of all matched TLS points. If the same TLS point was included in the neighborhoods of multiple adjacent centerline points, it was retained only once according to its point index to avoid duplicate counting. When point-level association was required, the TLS point was assigned to the nearest centerline point based on the minimum horizontal distance dxy. Therefore, the matching procedure can be regarded as a point-wise approximation of a buffer search along the continuous crack centerline. The matched TLS points were then assigned the corresponding crack ID, producing a subset of the TLS point cloud with crack semantic attributes, referred to as the semantic crack point cloud.

3.5. 3D Geometric Parameter Calculation

After skeleton extraction and point cloud matching, 3D crack objects with spatial semantic attributes can be constructed for quantitative analysis of crack geometry. Compared with conventional 2D plane-projection methods, this approach integrates crack centerlines, boundary structures, and point cloud elevation data to produce a realistic 3D representation of crack morphology, capturing curvature, branching, and topographic variations more accurately.

3.5.1. 3D Length Calculation

To construct the 3D skeleton of a crack, each 2D skeleton point is assigned an elevation value. To reduce the influence of local outliers and point cloud noise, the median elevation of neighboring point cloud points is used for each skeleton point:

$$z_i = median(\{ z_j \mid p_j \in N(c_i) \}) \quad (6)$$

where z_i denotes the elevation of skeleton point c_i, and z_j represents the elevations of neighboring point cloud points. Using the median rather than a single nearest neighbor mitigates the effect of local anomalies and scanning noise, enhancing the stability and reliability of 3D skeleton construction.

Combining the planar coordinates of skeleton points with the assigned elevations yields the 3D skeleton point set:

$$C_{3D} = \{ c_i(x_i, y_i, z_i) \mid i = 1, 2, \ldots, n \} \tag{7}$$

where (x_i, y_i, z_i) denotes the 3D coordinates of the i-th skeleton point, and n is the total number of skeleton points.

Cracks typically exhibit curvature, branching, and irregular skeleton point distributions. Directly connecting skeleton points in sequence may introduce local jump errors, especially when skeleton points are unordered or locally discontinuous. Therefore, a minimum spanning tree (MST) is constructed to establish spatial connections between skeleton points. Compared with sequential point connection, MST provides a distance-based graph structure for connecting skeleton points without requiring a predefined point order, thereby reducing artificial long-distance jumps caused by unordered point sequences and improving the stability of length calculation for complex crack skeletons. The sum of all edge lengths in the MST defines the 3D crack length:

$$L_{3D} = \sum_{(i,j) \in E_{MST}} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \tag{8}$$

where E_{MST} represents the edge set of the MST, and (x_i, y_i, z_i) and (x_j, y_j, z_j) denote the 3D coordinates of adjacent skeleton points. $L_{3\mathrm{D}}$ is the resulting 3D crack length.

It should be noted that the MST is used here as a practical graph-based approximation for organizing unordered crack skeleton points into a connected structure. The input skeleton points are derived from extracted crack regions and semantic crack point clouds, and therefore are already constrained by the detected crack morphology rather than being arbitrary scattered points. Compared with directly connecting skeleton points according to their storage order, the MST can reduce artificial local jump errors caused by unordered point sequences and provide an acyclic connected representation for curved or locally branching cracks. However, the MST minimizes the total edge length only and does not explicitly enforce curvature continuity, local orientation consistency, or physical crack topology. Therefore, shortcut connections may still occur in noisy skeletons, dense crack networks, or areas where spatially adjacent branches are physically disconnected.

3.5.2. 3D Crack Width Calculation

Crack width reflects the opening degree and spatial development of mining-induced surface cracks. Because these cracks commonly exhibit curvature, irregular shapes, and non-uniform width variations, conventional fixed-direction or buffer-based measurement methods are prone to errors caused by local bending and boundary noise. To address this issue, a skeleton-guided boundary point-pairing method was adopted to estimate crack width adaptively. In this study, crack-width estimation follows a "2D boundary pairing and 3D distance calculation" strategy. Boundary pairing is performed in the horizontal image plane because crack edges are extracted from UAV orthomosaic-based semantic masks, whereas the final width calculation incorporates TLS-derived elevation information after the opposite boundary points have been identified.

In this method, crack skeleton points are used as local measurement centers. For the i-th skeleton point, the local tangent direction is first calculated based on neighboring skeleton points:

$$t_i = (t_x, t_y) \tag{9}$$

A local normal direction perpendicular to the crack direction is then established:

$$n_i = (-t_y, t_x) \tag{10}$$

where t_i and n_i denote the local tangent and normal vectors at the current skeleton point, respectively, and t_x and t_y are the components of the tangent vector along the x and y axes.

Let the set of boundary points along the crack be:

$$B = \left\{ b_j(x_j, y_j, z_j) \mid j = 1, 2, \ldots m \right\} \tag{11}$$

To improve boundary-point search efficiency, a KDTree spatial index is constructed for the boundary points, and a local search region is established around each skeleton point. For each candidate boundary point, its projections along the local normal and tangent directions are calculated as follows:

$$u = \Delta x n_x + \Delta y n_y \tag{12}$$

$$v = \Delta x t_x + \Delta y t_y \tag{13}$$

where u is the projection distance along the local normal direction, v is the projection distance along the local tangent direction, Δx and Δy are the coordinate differences between the candidate point and the skeleton point in the x and y directions, respectively, and n_x and n_y are the components of the normal vector. The tangent projection v is used to constrain candidate points close to the local normal cross-section. Here, the local tangent and normal directions are defined in the x-y plane of the common projected coordinate system. The purpose of this 2D projection-based filtering is to ensure that the paired boundary points are located on opposite sides of the crack and close to the same local cross-section. Elevation is not used in this pairing step to avoid unstable pairing caused by point-cloud noise or local terrain fluctuations.

Based on the sign of the normal projection, candidate boundary points are classified into left (B_L) and right (B_R) boundary sets:

$$\begin{cases} B_L = \left\{ b_j \in B \mid u < 0 \right\} \\ B_R = \left\{ b_j \in B \mid u > 0 \right\} \end{cases} \tag{14}$$

Only points satisfying $\mid v \mid$ below a predefined threshold are retained to reduce the influence of crack curvature and local noise on width measurements. After the opposite boundary points are determined in the 2D plane, their corresponding elevation values are assigned from the TLS point cloud. The 3D Euclidean distance between the paired boundary points is then calculated as the terrain-aware edge-to-edge crack width at the current skeleton point:

$$W_i = \sqrt{(x_r - x_l)^2 + (y_r - y_l)^2 + (z_r - z_l)^2} \tag{15}$$

where (x_l, y_l, z_l) and (x_r, y_r, z_r) denote the coordinates of the B_L and B_R points. To reduce the influence of anomalous points and mismatches, upper and lower width constraints are imposed, and unreasonable width values are discarded. All valid width measurements are then aggregated for statistical analysis.

Compared with traditional fixed-direction width measurement methods, the proposed approach adaptively adjusts the measurement direction according to local crack

morphology. It can accommodate curvature, branching, and non-uniform width variations in mining-induced surface cracks, thereby improving the stability and spatial fidelity of crack-width estimation. It should be noted that when a large vertical offset exists between the two crack edges or when the local terrain slope is very steep, the calculated 3D distance may include part of the terrain-relief effect and may therefore differ from the horizontal crack opening. In the present study area, the vertical offset between paired crack edges was generally limited, and the field validation results showed acceptable errors. For areas with strong vertical displacement or complex micro-topography, local terrain-plane projection or cross-sectional correction should be considered in future work.

4. Results

4.1. Crack Semantic Segmentation Results

The proposed DRA-UNet model was implemented using the PyTorch framework (version 2.0.1) and trained on a workstation equipped with GPU acceleration. During training, the AdamW optimizer was employed to update network parameters, with an initial learning rate of 1×10^{-4} and a batch size of 8. All input images were uniformly resized to 256×256 pixels, and the maximum number of training epochs was set to 180.

A hybrid loss function was adopted as the optimization objective. It consisted of Focal Tversky Loss, BCEWithLogitsLoss, and Edge Consistency Loss, with corresponding weight coefficients of 0.6, 0.3, and 0.1, respectively. Focal Tversky Loss was used to reduce the influence of the strong class imbalance between crack and background pixels, BCEWithLogitsLoss was used to constrain pixel-level crack/background classification, and Edge Consistency Loss was introduced to provide boundary-related supervision. It should be noted that the hybrid loss was designed to support crack-region recognition and boundary continuity, while the independent contribution of each loss term was not separately evaluated in this study.

During model training, data augmentation strategies were introduced to increase the diversity of training samples. Specifically, random brightness and contrast adjustment was applied with a probability of 0.5, with brightness and contrast factors randomly sampled from 0.8 to 1.2. Random horizontal flipping and random vertical flipping were each applied with a probability of 0.5. Random rotations of 90° or 180° were also applied with a probability of 0.5. These augmentation operations were intended to improve the robustness of the model under complex mining-area background conditions.

Segmentation performance was evaluated using four metrics: precision, recall, F1-score, and MIoU. To assess the effectiveness of DRA-UNet, comparative experiments were conducted using several representative semantic segmentation models, including DeepLabV3+, SegNet, PSPNet, SegFormer, and FastSCNN. All models were trained and evaluated using the same dataset partition and experimental settings to ensure a fair comparison.

As shown in Figure 5, all models were able to identify the major crack regions when cracks exhibited clear morphological characteristics and strong contrast against the background. However, substantial differences were observed in areas containing narrow cracks, low-contrast cracks, and complex surface textures. In these regions, several models exhibited crack discontinuities, local omissions, and false detections, leading to reduced connectivity and incomplete boundary delineation. In contrast, DRA-UNet preserved the overall crack structure and spatial topology more effectively. The model demonstrated superior capability in maintaining the continuity of narrow branch cracks and suppressing background interference, particularly in areas with heterogeneous textures and complex surface conditions.

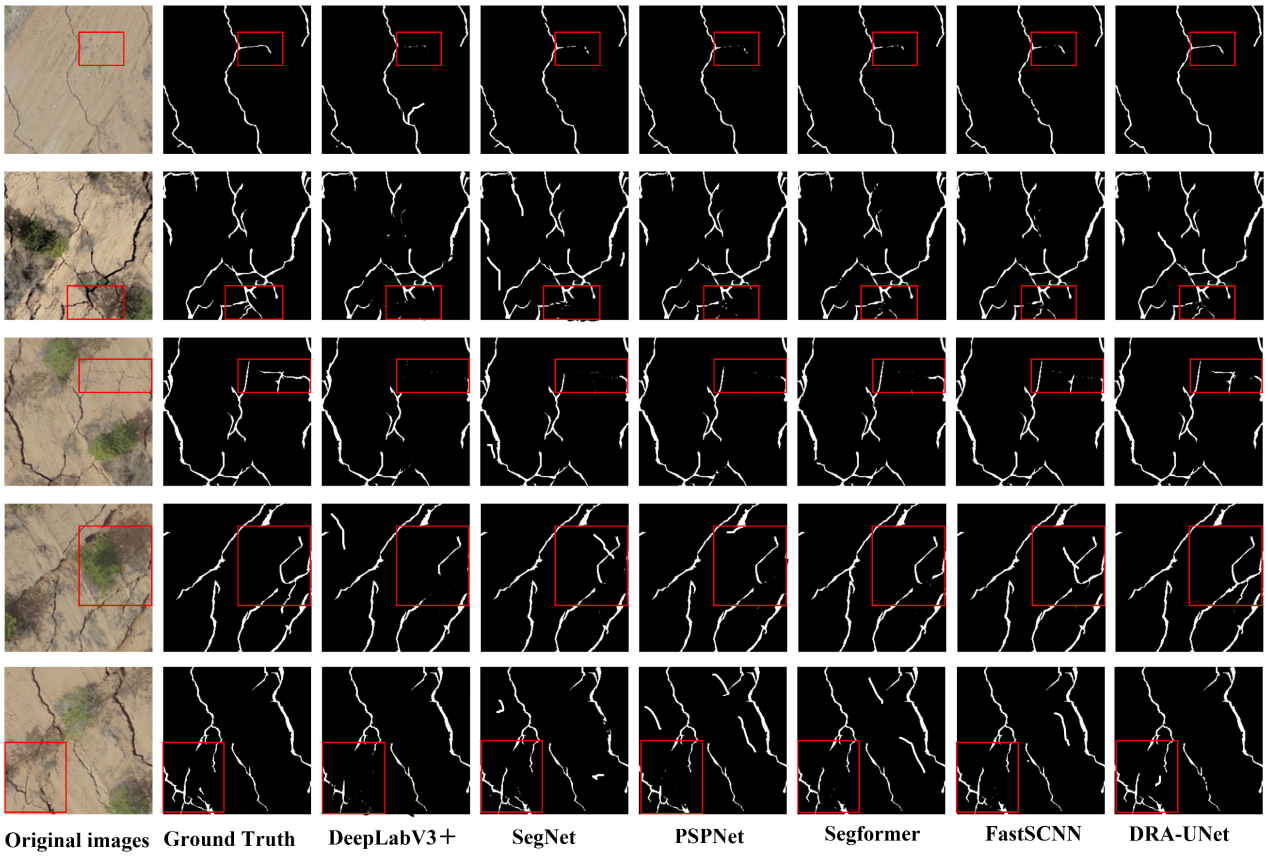


Figure 5. Visual comparison of crack semantic segmentation results obtained by different deep learning models. The red boxes highlight representative local regions used to compare crack extraction details among different models.

As shown in Table 3, DRA-UNet achieved the highest performance across all four evaluation metrics, with a precision of 85.13%, recall of 77.84%, F1-score of 81.32%, and MIoU of 70.26%. Compared with the baseline models, DRA-UNet showed better overall segmentation performance in the test set, especially in terms of the balance between precision and recall. However, these results alone cannot fully determine the specific contribution mechanism of each architectural component. Therefore, an ablation experiment was further conducted under the same dataset partition and training settings. Starting from the U-Net baseline, RN, DAM, and ASPP were introduced individually and in combination to assess their effects on crack segmentation performance. The results are summarized in Table 4.

Table 3. Quantitative comparison of crack semantic segmentation performance among different models.

Model	Pr (%)	Re (%)	F1 (%)	MIoU (%)
DeepLabV3+	84.85	68.12	75.57	65.42
SegNet	76.42	69.37	72.72	64.11
PSPNet	83.71	72.08	77.46	68.33
Segformer	80.56	76.92	78.70	69.18
FastSCNN	78.25	74.36	76.25	67.42
DRA-UNet	85.13	77.84	81.32	70.26

As shown in Table 4, the U-Net baseline achieved a precision of 79.55%, recall of 73.07%, F1-score of 76.18%, and MIoU of 66.16%. Introducing RN, DAM, and ASPP individually improved the F1-score to 78.43%, 76.78%, and 76.83%, respectively. Among the single-module variants, RN produced the largest improvement in F1-score. When multiple

modules were combined, the segmentation performance further improved. The full DRA-UNet achieved the best overall performance, with a precision of 85.13%, recall of 77.84%, F1-score of 81.32%, and MIoU of 70.26%. Compared with the U-Net baseline, DRA-UNet improved precision, recall, F1-score, and MIoU by 5.58, 4.77, 5.14, and 4.10 percentage points, respectively. These results suggest that the combined use of RN, DAM, and ASPP is associated with improved crack segmentation performance in the present experimental setting. However, the exact contribution mechanism of each module and the possible interaction among modules were not further analyzed through feature visualization or more detailed controlled experiments. Therefore, the functional roles of RN, DAM, and ASPP should be interpreted as design motivations and empirical observations rather than definitive causal explanations. It should also be noted that the ablation experiment mainly evaluates the architectural components, while the independent contribution of the hybrid loss function to segmentation performance was not separately quantified in this study.

Table 4. Ablation study of the main components in DRA-UNet.

Model	Pr (%)	Re (%)	F1 (%)	MIoU (%)
U-Net	79.55	73.07	76.18	66.16
U-Net + RN	82.49	74.61	78.43	67.12
U-Net + DAM	80.60	73.20	76.78	66.31
U-Net + ASPP	80.81	73.08	76.83	67.14
U-Net + RN + DAM	82.93	75.72	79.20	68.49
U-Net + RN + ASPP	83.01	75.17	78.97	68.21
U-Net + DAM + ASPP	81.49	73.22	77.24	67.70
DRA-UNet	85.13	77.84	81.32	70.26

4.2. 3D Mapping and Semantic Point Cloud Extraction

To validate the effectiveness of mapping 2D crack semantic information into 3D point cloud space, a representative area of crack development was selected for visualization. This area includes both elongated primary cracks and localized short or branching cracks, providing a comprehensive view of crack spatial morphology under complex surface conditions. The results of crack binarization, skeleton extraction, skeleton-to-point-cloud mapping, and semantic crack point cloud construction are shown in Figure 6.

Figure 6a,b illustrate the conversion of crack segmentation results into skeletonized centerlines. The binarized crack masks effectively preserve the main crack structures, while skeletonization reduces crack regions to single-pixel-wide centerlines. This process removes redundant boundary information while maintaining topological structure and spatial connectivity. Compared with direct use of crack boundaries for spatial analysis, skeletonized centerlines provide a more stable geometric representation of crack propagation paths and a robust basis for point-cloud matching.

Figure 6c,d show the mapping of crack information from 2D image space to 3D point cloud space. The crack skeletons and TLS point clouds exhibit strong spatial correspondence, and the mapped centerlines accurately represent crack locations under real surface conditions. The constructed semantic crack point clouds retain the planar distribution of cracks while incorporating elevation information and 3D geometric attributes, thereby transforming cracks from 2D image objects into 3D spatial objects. These results demonstrate that integrating image-based semantic information with point cloud geometry overcomes the limitation of UAV orthomosaics lacking elevation information and provides a reliable data basis for subsequent analysis of 3D crack length, width, and elevation variation. Although independent ground-truth data for skeleton extraction and 2D–3D point-cloud matching were not available, the intermediate results were checked through visual overlay and spatial consistency analysis. The extracted skeletons were located within

the segmented crack regions and preserved the main propagation directions of the cracks. After KDTree-based matching, the mapped crack centerlines showed good spatial correspondence with the TLS point clouds, indicating that the semantic crack information was successfully transferred from the 2D image plane to the 3D point-cloud space. In addition, the UAV–TLS co-registration RMSE of 0.03 m and the adopted search radius of 0.15 m provided a practical tolerance for minor spatial misalignment during matching.

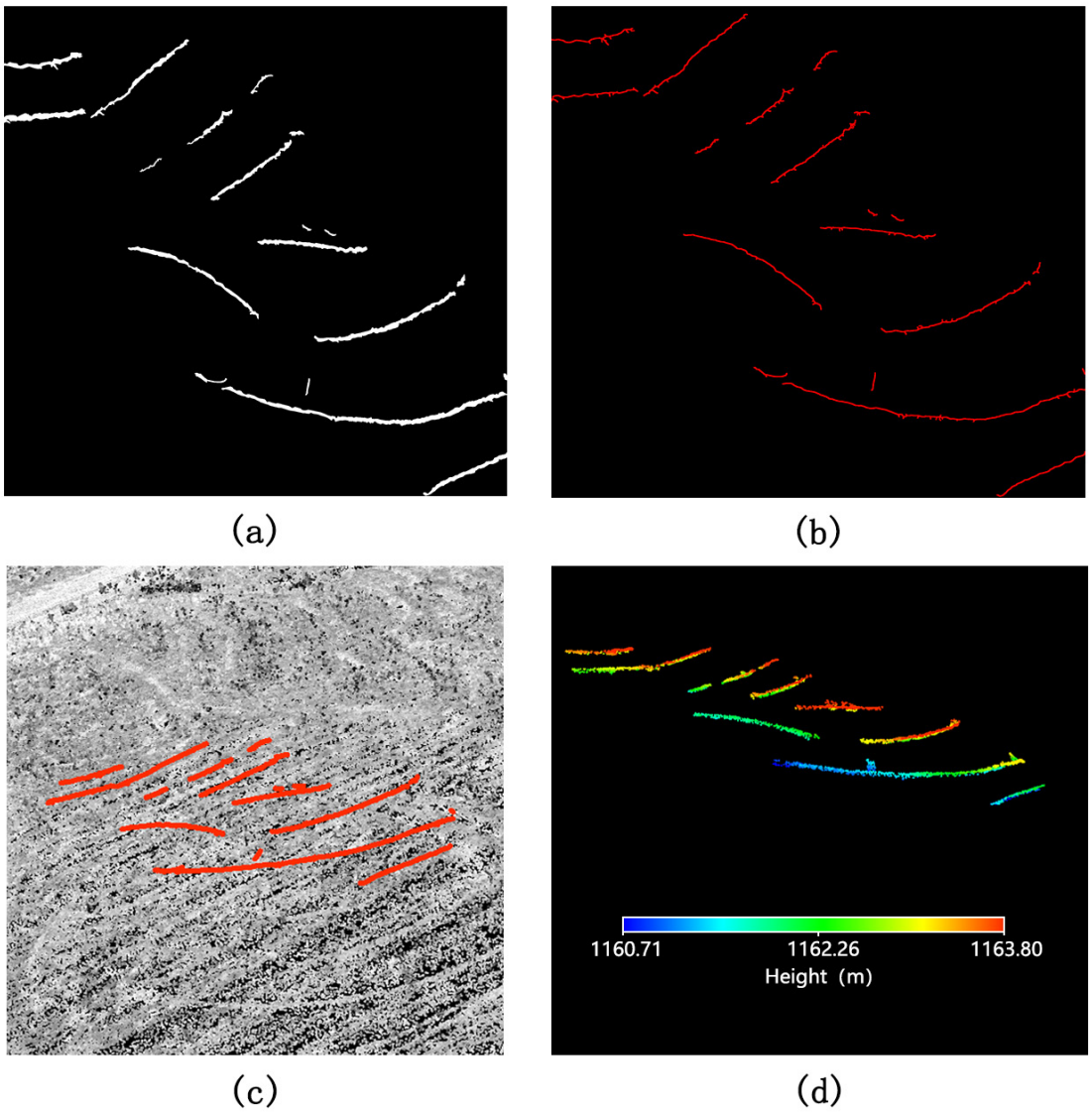


Figure 6. Results of crack extraction and 3D spatial mapping. (**a**) Binarized crack segmentation; (**b**) extracted crack skeleton; (**c**) mapping of crack skeleton onto TLS point cloud; (**d**) semantic crack point cloud after KDTree spatial matching. The colors in the point cloud indicate elevation variations, and the red lines/points represent crack-related features.

4.3. 3D Geometric Parameter Estimation and Validation

4.3.1. 3D Length Estimation and Validation

Figure 7 presents the semantic crack point clouds of representative cracks and their corresponding labels in the study area. The color variations indicate elevation differences within the crack point clouds, enabling visualization of crack propagation patterns under real topographic conditions. Overall, the identified cracks show heterogeneous spatial distributions, with clear variations in length, continuity, and propagation direction.

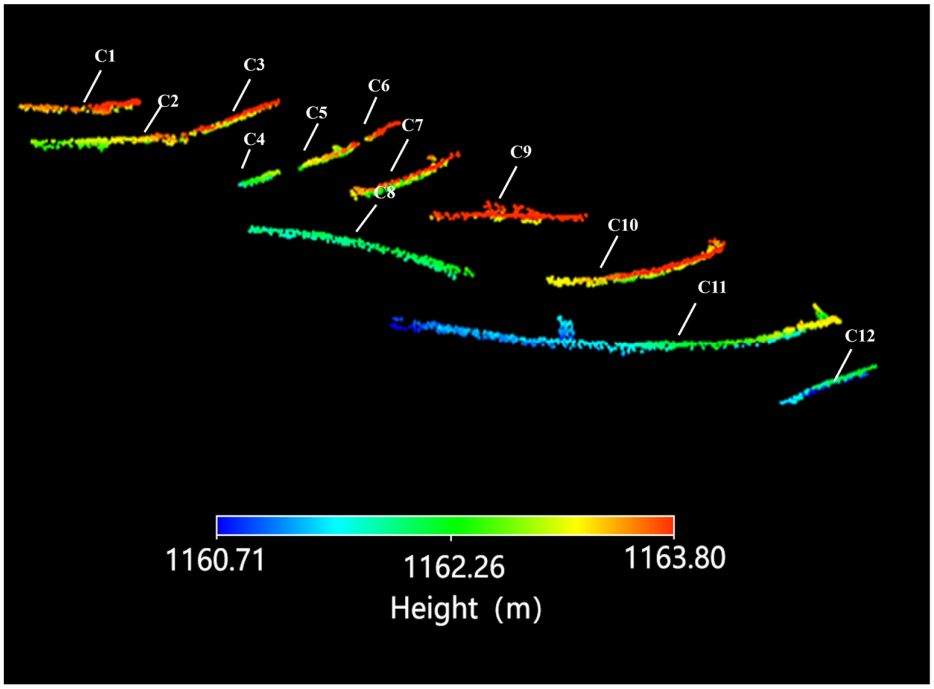


Figure 7. 3D semantic crack point clouds of representative cracks and their corresponding labels.

Among the representative cracks, C11 is the most continuous primary crack and has the largest spatial extent, indicating a well-developed propagation pattern across the mining-affected area. Cracks C10, C3, and C8 are medium-to-long cracks with relatively large spatial coverage, whereas C4 and C6 are short cracks mainly associated with local secondary fractures or crack termination zones. These results demonstrate that the semantic crack point cloud representation can effectively capture the spatial variability and geometric characteristics of mining-induced surface cracks under complex terrain conditions.

To evaluate the reliability of the proposed 3D crack characterization method, the twelve representative cracks shown in Figure 7 were selected for field validation. These cracks were selected to cover different crack lengths, continuity levels, propagation directions, and spatial locations within the study area, including both long primary cracks and short secondary cracks. Field-measured crack lengths were obtained manually along the visible crack traces using a tape measure. For each representative crack, the two endpoints were identified according to the continuous visible crack trace observed in the field and the corresponding UAV orthomosaic. The tape measure was placed along the main crack path as closely as possible to follow local curvature and branching characteristics. Each crack length was measured repeatedly, and the average value was used as the reference length for validation. The 2D lengths derived from crack skeletons and the 3D lengths calculated from semantic crack point clouds were then compared with the field-measured lengths. Table 5 summarizes the 2D length, 3D length, field-measured length, absolute error, relative error, RMSE, and R^2 for the representative cracks.

As shown in Table 5, the 2D length estimates were consistently smaller than the corresponding field-measured lengths, with a mean absolute error of 0.70 m, an RMSE of 0.81 m, an R^2 of 0.993, and a mean relative error of 4.49%. Here, the 2D length estimation can be regarded as a 2D-only image-based baseline, because the crack lengths were derived from UAV-extracted crack skeletons without incorporating TLS-derived elevation information.

In contrast, the proposed UAV-TLS fusion method incorporates both the planar positions of crack centerlines and the elevation information from TLS point clouds. The resulting 3D lengths show better agreement with field measurements, with a mean absolute error of 0.41 m, an RMSE of 0.48 m, an R^2 of 0.998, and a mean relative error of 2.64%.

Compared with the 2D-only baseline, the proposed 3D method reduced the mean absolute error from 0.70 m to 0.41 m and the RMSE from 0.81 m to 0.48 m. For the longest primary crack (C11), the absolute error was 1.07 m, corresponding to a relative error of 2.51%. Shorter cracks, C4 and C6, exhibited relative errors of 3.17% and 3.06%, respectively.

Table 5. Comparison of 2D Length, 3D Length, and Field-Measured Length of Representative Mining-Induced Surface Cracks.

Crack ID	Measured Length (m)	2D Length (m)	3D Length (m)	2D Absolute Error (m)	2D Relative Error (%)	3D Absolute Error (m)	3D Relative Error (%)
C1	12.08	11.55	12.39	0.53	4.39	0.31	2.57
C2	13.81	13.20	14.16	0.61	4.42	0.35	2.53
C3	20.69	19.76	21.23	0.93	4.50	0.54	2.61
C4	3.47	3.30	3.58	0.17	4.90	0.11	3.17
C5	9.27	8.85	9.51	0.42	4.53	0.24	2.59
C6	4.25	4.04	4.38	0.21	4.94	0.13	3.06
C7	17.29	16.55	17.70	0.74	4.28	0.41	2.37
C8	18.25	17.44	18.73	0.81	4.44	0.48	2.63
C9	14.31	13.69	14.66	0.62	4.33	0.35	2.45
C10	23.12	22.08	23.75	1.04	4.50	0.63	2.72
C11	42.63	40.82	43.70	1.81	4.25	1.07	2.51
C12	11.12	10.63	11.40	0.49	4.41	0.28	2.52
Mean	—	—	—	0.70	4.49	0.41	2.64
RMSE	—	—	—	0.81	—	0.48	—
R^2	—	0.993	0.998	—	—	—	—

4.3.2. Crack Width Analysis and Validation

To evaluate the accuracy of the proposed method in estimating crack widths, four representative cracks (C13–C16) were selected to represent different width ranges and local boundary conditions. Six measurement points (P1–P6) were arranged along each crack, as shown in Figure 8. Field-measured widths were obtained using a tape measure and compared with the calculated 3D widths obtained by the proposed framework. In this study, the calculated width refers to the terrain-aware edge-to-edge crack width derived from 2D boundary pairing and TLS-based 3D distance calculation, rather than a purely 2D image-based width. For field validation, the six measurement points were arranged from the crack head to the crack end to cover local width variations. At each measurement point, the crack width was measured approximately perpendicular to the local crack direction between the two visible crack edges. The measurement positions were identified based on field observations and the corresponding UAV orthomosaic to ensure consistency with the automatically extracted width locations. Small discrepancies may still arise from irregular crack edges, loose surface materials, local boundary fragmentation, and slight positioning differences between field measurement points and automatically extracted cross-sections.

Figure 9a compares the field-measured crack widths with those estimated by the proposed method at all measurement points. Overall, the estimated widths agree well with the field measurements and capture width variations along the crack traces. The method performs consistently across cracks with different width ranges and effectively characterizes local width variability. Among the investigated cracks, C13 exhibits the largest widths, ranging from 9.4 to 13.5 cm, whereas C16 has smaller widths, ranging from 5.7 to 8.7 cm. The estimated widths also closely follow the peaks and troughs observed in the field measurements, indicating that the proposed method can capture the spatial variability of crack width under real surface conditions.

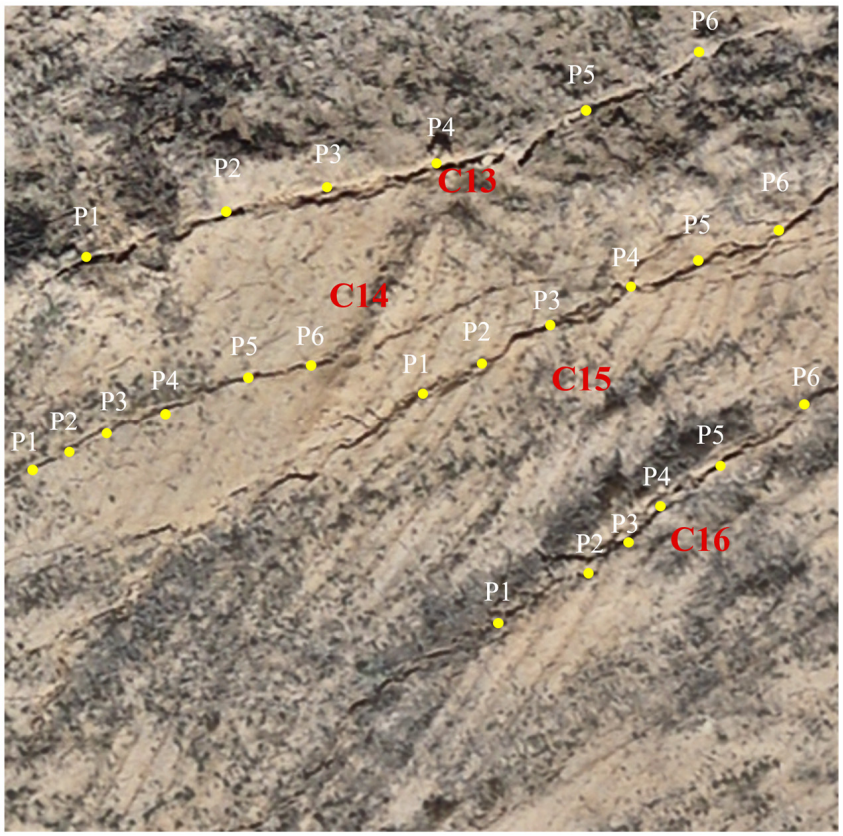


Figure 8. Representative cracks used for width validation and the locations of measurement points (P1–P6).

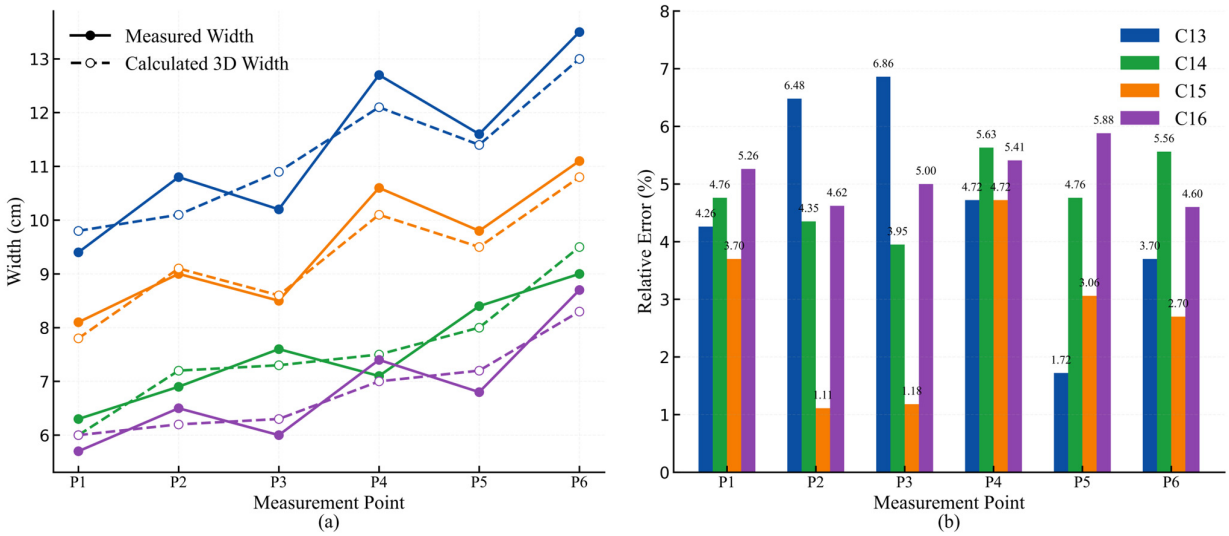


Figure 9. Validation of calculated 3D crack width results. (**a**) Comparison between field-measured widths and calculated 3D widths; (**b**) analysis of relative errors. Different colors represent different representative cracks used for width validation.

The relative errors of the estimated crack widths ranged from 1.11% to 6.86% across all measurement points. The minimum relative error of 1.11% was observed at P2 of crack C15, whereas the maximum relative error of 6.86% occurred at P3 of crack C13. Although local variations were present among measurement points, the errors remained within a limited range, indicating that the skeleton-guided width estimation method provides consistent

and reliable measurements under different crack conditions. For further comparison, the width-validation results of the four representative cracks are summarized in Table 6.

Table 6. Statistical results of calculated 3D crack-width validation.

Crack ID	Measurement Points	Measured Width Range (cm)	Calculated 3D Width Range (cm)	Mean Relative Error (%)	Max. Relative Error (%)
C13	6	9.4–13.5	9.8–13.0	5.00	6.86
C14	6	6.3–9.0	6.0–9.5	4.83	5.63
C15	6	8.1–11.1	7.8–10.8	2.91	4.72
C16	6	5.7–8.7	6.0–8.3	5.13	5.88
Overall	24	5.7–13.5	6.0–13.0	4.47	6.86

As summarized in Table 6, crack C15 achieved the lowest mean relative error of 2.91%, whereas the mean relative errors for cracks C13, C14, and C16 were 5.00%, 4.83%, and 5.13%, respectively. Across all 24 measurement points, the overall mean relative error was 4.47%, with a maximum relative error of 6.86%.

5. Discussion

The results demonstrate that integrating UAV-derived crack semantics with TLS-derived geometric information provides an effective pathway for three-dimensional characterization of mining-induced surface cracks. Unlike road or structural cracks, mining-induced cracks develop under the combined influence of overburden movement, ground subsidence, terrain undulation, and loose surface materials. These processes often produce curved traces, non-uniform openings, fragmented boundaries, and locally branching structures. Under such conditions, crack objects extracted only from UAV orthomosaics are inevitably constrained by planar projection, which limits their ability to represent true propagation paths and local topographic effects [29].

The improvement in length estimation confirms the necessity of incorporating elevation information into crack geometry analysis. The comparison with the 2D-only image-based baseline further supports this point. The 2D-only method produced a mean relative error of 4.49%, whereas the proposed UAV-TLS fusion method reduced the error to 2.64%. This improvement indicates that TLS-derived elevation information can reduce planar-projection-related underestimation and provide a more realistic representation of crack propagation under actual terrain conditions. Therefore, mining-induced surface cracks should not be treated simply as planar image features, particularly in areas affected by subsidence basins, micro-topographic deformation, and local ground collapse [30].

The construction of semantic crack point clouds is a key step in this transformation. Compared with TLS-only crack extraction, the proposed UAV-TLS fusion framework has clear advantages in mining-affected surface environments. Direct extraction of fine mining-induced cracks from TLS point clouds alone is difficult because narrow crack boundaries can be affected by point-density variations, rough ground surfaces, vegetation, occlusion, and scanning noise [31–33]. In contrast, UAV orthomosaics provide clearer texture and boundary information for crack semantic extraction, while TLS point clouds provide reliable elevation and surface morphology information for 3D characterization. Therefore, the proposed framework avoids relying solely on geometric discontinuities in point clouds and instead transfers image-derived crack semantics into 3D space through spatial matching. Compared with direct boundary-to-point-cloud matching, the centerline-guided KDTree matching strategy adopted in this study provides a more stable spatial link

between 2D crack semantics and TLS geometry, because skeletonized centerlines are less sensitive to fragmented boundaries and local segmentation noise than crack edges.

The width validation results further indicate that the skeleton-guided measurement strategy can capture local variations in crack opening. The overall mean relative error of 4.47% suggests that the method provides stable width estimates across cracks with different size ranges. Remaining errors are mainly related to local boundary fragmentation, shadow interference in UAV orthomosaics, small registration uncertainties between image and point-cloud data, and unavoidable differences between field-measured positions and automatically extracted cross-sections. These factors are particularly important for narrow or discontinuous cracks, where slight boundary deviations can lead to relatively large proportional errors.

From an application perspective, the proposed framework extends mining-induced crack analysis from image-based detection to spatially measurable 3D characterization. Crack length, width, spatial distribution, and elevation variation provide complementary information for evaluating the intensity and pattern of mining-induced ground damage. Therefore, the framework can support mining-area surface damage monitoring, crack development assessment, ecological restoration surveys, and surface stability evaluation. Compared with manual field surveys, the workflow offers a more repeatable and spatially continuous approach for monitoring crack development over mining-affected areas by linking crack morphology with local terrain deformation.

Several limitations should be acknowledged. First, the experiments were conducted in a single mining area using single-period UAV and TLS datasets, and the generalizability of the framework to different geomorphic settings, vegetation covers, mining stages, and point-cloud densities requires further validation. In addition, because the training, validation, and testing subsets were randomly divided at the patch level after sliding-window cropping, partial spatial information leakage may occur when adjacent overlapping patches are assigned to different subsets. Therefore, the segmentation metrics may be optimistic compared with evaluations based on spatially independent orthomosaics or acquisition areas. Second, dense crack networks and highly fragmented boundaries may still affect skeleton extraction, centerline vectorization, and local width measurement. In addition, although the MST provides an efficient way to connect unordered skeleton points, it minimizes the total edge length only and does not explicitly model curvature continuity or local orientation consistency. As a result, shortcut connections may occur in noisy skeletons, dense crack networks, or areas where adjacent crack branches are spatially close. Future work should explore graph-based reconstruction methods that incorporate orientation, curvature, and topological constraints to better preserve the physical crack trajectory. Moreover, in areas with dense vegetation cover, severe crack occlusion, complex terrain, or low point-cloud density, the accuracy of crack segmentation, point-cloud matching, and local width estimation may decrease. Third, the accuracy of 3D characterization depends on the spatial consistency between UAV orthomosaics and TLS point clouds. In this study, the horizontal RMSE of the UAV-TLS co-registration was 0.03 m, indicating good spatial consistency between the two datasets. Nevertheless, residual spatial misalignment may still affect local crack-edge pairing and width estimation, especially for very narrow cracks or areas with complex micro-topography. In addition, the KDTree search radius may influence semantic crack point-cloud construction. Although r = 0.15 m was adopted in this study by considering the UAV-TLS co-registration accuracy, UAV image resolution, TLS point density, and crack scale, this parameter should be adjusted according to image resolution, point-cloud density, co-registration accuracy, terrain complexity, and crack morphology when applying the proposed framework to other study areas. In addition, this study mainly validated crack length and width, while crack depth or vertical displacement was

not independently measured in the field. Therefore, although TLS point clouds provide elevation information and can describe local height variations around cracks, the accuracy of crack-depth estimation was not quantitatively evaluated in this study. Future work should include field-measured depth or cross-sectional profiles to validate additional 3D geometric parameters. Moreover, improved segmentation performance does not automatically guarantee proportional improvement in final geometric measurements, because 3D length and width estimation are also affected by skeleton extraction, boundary quality, UAV-TLS co-registration accuracy, KDTree matching parameters, point-cloud density, and local terrain complexity. Therefore, the relationship between segmentation accuracy and downstream geometric accuracy should be further investigated through uncertainty propagation analysis or controlled comparative experiments. Future work should incorporate multi-temporal datasets from multiple mining sites to investigate crack initiation, propagation, closure, and migration. Furthermore, introducing topological constraints, temporal association strategies, and more robust point-cloud matching methods may further improve the stability and applicability of 3D crack reconstruction.

6. Conclusions

This study proposed a UAV–TLS fusion framework for 3D characterization of mining-induced surface cracks. Unlike conventional studies that mainly focus on 2D crack detection and planar parameter statistics, the proposed framework integrates crack semantic information from UAV orthomosaics with geometric information from TLS point clouds. Through crack segmentation, skeleton extraction, vectorization, and KDTree-based spatial matching, 2D crack objects were mapped into 3D point cloud space to construct semantic crack point clouds. This enables the transformation of mining-induced surface cracks from 2D image objects to 3D spatial objects.

The experimental results demonstrate that DRA-UNet provides reliable crack semantic extraction under complex mining-area conditions, achieving precision, recall, F1-score, and MIoU values of 85.13%, 77.84%, 81.32%, and 70.26%, respectively. Based on the constructed semantic crack point clouds, the proposed framework enables automatic estimation of 3D crack geometric parameters, including crack length and width. Compared with conventional 2D length measurements, which produced a mean relative error of 4.49%, the proposed 3D length estimation reduced the mean relative error to 2.64% by incorporating TLS-derived elevation information. Width validation at 24 field measurement points yielded a mean relative error of 4.47% and a maximum relative error of 6.86%, confirming the reliability of the skeleton-guided width estimation method.

These results indicate that UAV–TLS fusion can effectively overcome the limitations of image-only crack analysis, particularly the lack of elevation information in UAV orthomosaics. By linking crack morphology with terrain geometry, the proposed framework provides a more realistic representation of crack propagation under real topographic conditions and improves the accuracy of geometric parameter estimation. The method offers a practical remote sensing data fusion approach for ground damage monitoring, crack development analysis, ecological restoration assessment, and surface stability evaluation in mining areas.

Future work will focus on multi-temporal UAV and TLS data acquisition to monitor crack initiation, propagation, closure, and migration processes. Additional validation across different mining sites, geomorphic settings, vegetation conditions, and mining stages will further improve the robustness and generalizability of the proposed 3D crack characterization framework.

Author Contributions: Conceptualization, W.Z. and Y.Z.; methodology, W.Z. and Y.Z.; software, W.Z.; validation, W.Z., H.C. and M.M.; formal analysis, W.Z.; investigation, W.Z., H.C., L.C., W.D., J.H.,

M.M. and M.L.; resources, Y.Z. and H.C.; data curation, W.Z., H.C. and M.M.; writing—original draft preparation, W.Z.; writing—review and editing, W.Z., Y.Z. and H.C.; visualization, W.Z.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number U22A20620.

Data Availability Statement: The data presented in this study are available from the corresponding author upon reasonable request. The UAV orthomosaics and TLS point clouds are not publicly available due to restrictions related to mining-area data management.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Fu, Y.; Wu, Y.; Yin, X.; Zhang, Y. Mapping Mining-Induced Ground Fissures and Their Evolution Using UAV Photogrammetry. *Front. Earth Sci.* **2023**, *11*, 1260913. [CrossRef]
2. Zhao, Y.; Sun, B.; Liu, S.; Zhang, C.; He, X.; Xu, D.; Tang, W. Identification of Mining-Induced Ground Fissures Using UAV and Infrared Thermal Imager: Temperature Variation and Fissure Evolution. *ISPRS J. Photogramm. Remote Sens.* **2021**, *180*, 45–64. [CrossRef]
3. Xu, D.; Zhao, Y.; Jiang, Y.; Zhang, C.; Sun, B.; He, X. Using Improved Edge Detection Method to Detect Mining-Induced Ground Fissures Identified by Unmanned Aerial Vehicle Remote Sensing. *Remote Sens.* **2021**, *13*, 3652. [CrossRef]
4. Zhang, F.; Hu, Z.; Fu, Y.; Yang, K.; Wu, Q.; Feng, Z. A New Identification Method for Surface Cracks from UAV Images Based on Machine Learning in Coal Mining Areas. *Remote Sens.* **2020**, *12*, 1571. [CrossRef]
5. Wang, K.; Wei, B.; Zhao, T.; Wu, G.; Zhang, J.; Zhu, L.; Wang, L. An Automated Approach for Mapping Mining-Induced Fissures Using CNNs and UAS Photogrammetry. *Remote Sens.* **2024**, *16*, 2090. [CrossRef]
6. Jiang, X.; Mao, S.; Li, M.; Liu, H.; Zhang, H. MFPA-Net: An Efficient Deep Learning Network for Automatic Ground Fissures Extraction in UAV Images of the Coal Mining Area. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 103039. [CrossRef]
7. Yang, K.; Hu, Z.; Liang, Y.; Fu, Y.; Yuan, D.; Guo, J.; Li, G.; Li, Y. Automated Extraction of Ground Fissures Due to Coal Mining Subsidence Based on UAV Photogrammetry. *Remote Sens.* **2022**, *14*, 1071. [CrossRef]
8. An, J.; Dong, S.; Wang, X.; Li, C.; Zhao, W. Research on UAV Aerial Imagery Detection Algorithm for Mining-Induced Surface Cracks Based on Improved YOLOv10. *Sci. Rep.* **2025**, *15*, 30101. [CrossRef] [PubMed]
9. Zhang, F.; Hu, Z.; Liang, Y.; Li, Q. Evaluation of Surface Crack Development and Soil Damage Based on UAV Images of Coal Mining Areas. *Land* **2023**, *12*, 774. [CrossRef]
10. Riquelme, A.J.; 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. [CrossRef]
11. Bello, S.A.; Yu, S.; Wang, C.; Adam, J.M.; Li, J. Deep Learning on 3D Point Clouds. *Remote Sens.* **2020**, *12*, 1729. [CrossRef]
12. Weinmann, M.; Jutzi, B.; Hinz, S.; Mallet, C. Semantic Point Cloud Interpretation Based on Optimal Neighborhoods, Relevant Features and Efficient Classifiers. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 286–304. [CrossRef]
13. Elamin, A.; El-Rabbany, A. UAV-Based Image and LiDAR Fusion for Pavement Crack Segmentation. *Sensors* **2023**, *23*, 9315. [CrossRef] [PubMed]
14. Jia, Y.; Ahmad, A.B. Quality Assessments of Unmanned Aerial Vehicle (UAV) and Terrestrial Laser Scanning (TLS) Methods in Road Cracks Mapping. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2023**, *XLVIII-4/W6-2022*, 183–188. [CrossRef]
15. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
16. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
17. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder–Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 801–818. [CrossRef]
18. Yu, Y.; Li, J.; Guan, H.; Wang, C. 3D Crack Skeleton Extraction from Mobile LiDAR Point Clouds. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 914–917. [CrossRef]
19. Bentley, J.L. Multidimensional Binary Search Trees Used for Associative Searching. *Commun. ACM* **1975**, *18*, 509–517. [CrossRef]

20. Rusu, R.B.; Cousins, S. 3D Is Here: Point Cloud Library (PCL). In Proceedings of the IEEE International Conference on Robotics and Automation, Shanghai, China, 9–13 May 2011; pp. 1–4. [CrossRef]
21. Wang, L.; Mu, Y.; Zhang, Q.; Zhang, X. Groundwater Use by Plants in a Semi-Arid Coal-Mining Area at the Mu Us Desert Frontier. *Environ. Earth Sci.* **2013**, *69*, 1015–1024. [CrossRef]
22. RIEGL Laser Measurement Systems. RIEGL VZ-1000 V-Line 3D Terrestrial Laser Scanner: Data Sheet. 2015. Available online: https://www.riegl-japan.co.jp/product/pdf_1/DataSheet_VZ-1000_2015-01-22.pdf (accessed on 9 June 2026).
23. Russell, B.C.; Torralba, A.; Murphy, K.P.; Freeman, W.T. LabelMe: A Database and Web-Based Tool for Image Annotation. *Int. J. Comput. Vis.* **2008**, *77*, 157–173. [CrossRef]
24. Zhou, Z.; Rahman Siddiquee, M.M.; Tajbakhsh, N.; Liang, J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2018; Volume 11045, pp. 3–11. [CrossRef] [PubMed]
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Volume 9908, pp. 630–645. [CrossRef]
26. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3146–3154.
27. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 834–848. [CrossRef] [PubMed]
28. Friedman, J.H.; Bentley, J.L.; Finkel, R.A. An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Trans. Math. Softw.* **1977**, *3*, 209–226. [CrossRef]
29. Ren, H.; Zhao, Y.; Xiao, W.; Hu, Z. A Review of UAV Monitoring in Mining Areas: Current Status and Future Perspectives. *Int. J. Coal Sci. Technol.* **2019**, *6*, 320–333. [CrossRef]
30. Tong, X.; Liu, X.; Chen, P.; Liu, S.; Luan, K.; Li, L.; Liu, S.; Liu, X.; Xie, H.; Jin, Y.; et al. Integration of UAV-Based Photogrammetry and Terrestrial Laser Scanning for the Three-Dimensional Mapping and Monitoring of Open-Pit Mine Areas. *Remote Sens.* **2015**, *7*, 6635–6662. [CrossRef]
31. Hackel, T.; Wegner, J.D.; Schindler, K. Fast Semantic Segmentation of 3D Point Clouds with Strongly Varying Density. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III-3*, 177–184. [CrossRef]
32. Feng, Z.; El Issaoui, A.; Lehtomäki, M.; Ingman, M.; Kaartinen, H.; Kukko, A.; Savela, J.; Hyyppä, H.; Hyyppä, J. Pavement Distress Detection Using Terrestrial Laser Scanning Point Clouds: Accuracy Evaluation and Algorithm Comparison. *ISPRS Open J. Photogramm. Remote Sens.* **2022**, *3*, 100010. [CrossRef]
33. Yakar, M.; Ulvi, A.; Yiğit, A.Y.; Alptekin, A. Discontinuity Set Extraction from 3D Point Clouds Obtained by UAV Photogrammetry in a Rockfall Site. *Surv. Rev.* **2023**, *55*, 416–428. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.