

Article

LHEM-MSegNet: A Landslide Hazard Extraction Model Based on Multi-Source Data Fusion

Fukang Shen ¹ , Weibin Li ^{1,*} , Tianyi Zhang ¹, Xuan Sun ¹, Zhixiong Han ² and Xiaolei Yang ²

¹ School of Artificial Intelligence, Xidian University, Xi'an 710071, China; 25171214073@stu.xidian.edu.cn (F.S.); 22171110624@xidian.edu.cn (T.Z.); 24171214029@xidian.edu.cn (X.S.)

² Key Laboratory of Coal Resources Exploration and Comprehensive Utilization, Ministry of Natural Resources, Xi'an 710071, China; smdzhzx0625@163.com (Z.H.); yangxiaolei2003@163.com (X.Y.)

* Correspondence: weibinli@xidian.edu.cn

Highlights

What are the main findings?

- Multi-source fusion of optical imagery and terrain information improves the accuracy and reliability of loess landslide hazard extraction by providing complementary spectral, textural, and geomorphological cues.
- LHEM-MSegNet achieves the best overall performance on the LYPLH dataset, with an mIoU of 89.29%, and outperforms LHEM-SegNet and other baseline models in complex terrain scenarios.
- The ablation experiments show that DEM-derived terrain factors, especially slope and aspect, help improve segmentation stability and reduce false and missed detections.
- The supplementary experiment on Landslide4Sense indicates that the proposed framework has potential applicability under different data distributions, although further validation is still needed for broader regions and different landslide types.

What are the implications of the main findings?

- Terrain-related information provides complementary geomorphological cues for improving landslide hazard extraction in loess areas.
- The proposed multisource fusion framework can support more reliable remote sensing interpretation and geohazard monitoring in complex terrain environments.

Abstract

Accurate detection of loess landslide hazards is critical for disaster prevention, yet remains challenging due to complex terrain, spectral similarity to background regions, and the lack of specialized datasets. To address these challenges, this paper proposes LHEM-MSegNet, a multisource landslide hazard extraction model based on Transformer and attention fusion. A loess landslide hazard region segmentation dataset for Linyou County (LYPLH), Shaanxi Province, China, is constructed using a transfer-learning strategy based on the Segment Anything Model (SAM), followed by iterative manual correction and data augmentation to improve annotation quality. LHEM-MSegNet integrates Transformer blocks with spatial and channel attention modules to enhance feature representation, and adopts a multi-pathway architecture for multisource fusion. Specifically, optical remote sensing imagery is combined with geomorphological information, including DEM, slope, and aspect maps, to improve landslide hazard extraction. Experimental results on the LYPLH dataset demonstrate that the proposed LHEM-MSegNet achieves an mIoU of 89.29% by integrating optical imagery with geomorphological information. Ablation studies first demonstrate the effectiveness of the proposed model structure and further confirm the significant contribution of terrain-related features to landslide hazard extraction.



Academic Editor: Fumio Yamazaki

Received: 4 June 2026

Revised: 7 July 2026

Accepted: 14 July 2026

Published: 22 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](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

In addition, the supplementary evaluation on the external Landslide4Sense benchmark shows that the proposed method maintains competitive performance under different data distributions. The proposed approach provides an effective solution for early loess landslide hazard identification and supports disaster mitigation in the Loess Plateau region.

Keywords: loess landslides; hazard point extraction; multi-source data; remote sensing

1. Introduction

With the intensification of global climate change and increasing human activities, geological hazards-particularly landslides-have become a major global concern. In China, landslides account for approximately half of all geological hazards and represent the most frequent and destructive type among them [1]. The formation mechanisms of landslides are highly complex, involving a combination of geological conditions, human activities, and natural factors. In recent years, numerous landslide events have resulted in significant casualties and economic losses. For example, the Shuicheng landslide in Guizhou Province in 2019 resulted in 43 fatalities and direct economic losses of approximately 190 million Chinese yuan(CNY) [2]. Compared with rock or soil landslides in other regions, loess landslides exhibit unique geomorphic characteristics, such as weak structural stability, strong erosion by rainfall, and gradual deformation processes. These properties often result in ambiguous landslide boundaries, weak spectral contrast with surrounding terrain, and complex surface textures, making accurate identification from optical remote sensing imagery particularly challenging. Consequently, developing robust remote sensing based methods for extracting loess landslide hazard regions remains a critical yet unresolved problem for disaster prevention and mitigation.

Remote sensing imagery provides spatially continuous observations over large areas and has become an important data source for landslide investigation, hazard monitoring, and disaster assessment. Driven by the rapid development of computer vision and deep learning techniques, remote sensing image interpretation has gradually shifted from manual visual interpretation and rule-based feature extraction to automatic feature learning. Deep learning methods have shown strong representation capability in various optical remote sensing extraction tasks, such as water body segmentation and land-cover mapping [3,4]. In the field of landslide interpretation, they have also been widely applied to landslide detection from remote sensing imagery, including image classification [5], object detection [6], and semantic segmentation [7]. Image classification methods can determine whether a candidate image patch contains landslides, but they usually provide only coarse category level results and cannot accurately locate landslide boundaries. Object detection methods further improve spatial localization by predicting bounding boxes, but they are still unable to delineate irregular landslide contours and accurately estimate landslide areas. In contrast, semantic segmentation can achieve pixel level landslide mapping, which makes it more suitable for landslide boundary extraction, area calculation, and subsequent hazard assessment.

In recent years, semantic segmentation networks have been widely introduced into landslide extraction, including models derived from U-Net, DeepLab series models, attention-based convolutional networks, and Transformer-based architectures [8–10]. Networks with encoder-decoder structures can effectively capture local texture and edge information, while attention mechanisms help suppress irrelevant background features and enhance landslide-related responses. Transformer-based models further improve the representation of long-range spatial dependencies and global contextual relationships,

which is beneficial for identifying landslide bodies with irregular shapes and complex backgrounds. These studies have greatly advanced pixel-level landslide mapping. However, most existing methods are still mainly developed for general landslide extraction or post-event landslide inventory mapping, and their applicability to loess landslide hazard regions remains insufficiently investigated.

Loess landslides have distinctive geological and geomorphological characteristics compared with landslides in rock-dominated mountainous regions. Loess is characterized by weak cementation, high collapsibility, well-developed vertical joints, and strong sensitivity to rainfall infiltration and human engineering activities [11]. In loess hilly and gully regions, landslide hazard regions are often distributed along gully margins, steep slopes, road cuts, and slope toe areas. Their surface features may be affected by erosion, vegetation disturbance, agricultural terraces, and exposed loess surfaces [12,13]. As a result, loess landslide hazard regions often show gradual deformation traces, fragmented surface textures, ambiguous boundaries, and weak spectral contrast with the surrounding terrain. These characteristics make them difficult to distinguish from non-landslide bare slopes, erosion gullies, construction-disturbed surfaces, and other visually similar backgrounds in optical remote sensing images.

Previous studies have explored loess landslide identification using field investigation, expert visual interpretation, high-resolution optical imagery, terrain factors, machine learning, and deep learning methods [6,14,15]. Traditional loess landslide mapping mainly relies on field surveys, geomorphic interpretation, and object-based image analysis, which can provide reliable inventory information but are usually time-consuming and difficult to apply over large areas. With the development of remote sensing and intelligent interpretation techniques, spectral, textural, and topographic features have been combined with machine-learning classifiers to improve the automatic identification of loess landslides. More recently, deep learning methods have been introduced into loess landslide detection. For example, Ju et al. [6] applied object detection algorithms to loess landslide detection in northwest China, demonstrating the potential of deep learning for automatic loess landslide interpretation.

However, many existing studies focus on landslide points, landslide bodies, or historical landslide events with relatively clear visual evidence, whereas pixel-level extraction of loess landslide hazard regions remains more challenging. Compared with occurred landslide bodies, loess landslide hazard regions may not always have complete scar boundaries, fresh exposed surfaces, or strong spectral differences. In addition, road cuts, bare loess slopes, erosion gullies, construction-disturbed areas, and agricultural terraces may show visual characteristics similar to loess landslide hazard regions in optical remote sensing images. Therefore, relying only on optical spectral and textural features may easily lead to confusion between loess landslide hazard regions and visually similar non-landslide surfaces, indicating the need for pixel-level segmentation methods that can integrate both optical features and terrain-related geomorphological information.

Another limitation of existing studies lies in the insufficient use of geological and geomorphological information. Although optical spectral and textural features are important for identifying landslide-related surface characteristics, relying only on optical imagery may easily lead to confusion between loess landslide hazard regions and visually similar non-landslide surfaces, such as bare loess slopes, erosion gullies, road cuts, construction-disturbed areas, and agricultural terraces. This problem is particularly prominent in loess landform regions because landslide occurrence is closely associated with terrain factors such as elevation, slope, and aspect [14]. Elevation provides information on the regional terrain background, slope reflects terrain steepness and potential gravitational instability, and aspect is related to slope orientation, illumination, erosion, and moisture conditions.

Although some studies have attempted to combine optical imagery with DEM data or other topographic information, the contribution of derived terrain factors, such as slope and aspect, has not yet been fully investigated for loess landslide hazard region extraction [15]. Therefore, an effective segmentation framework for loess landslide hazards should integrate optical features with terrain-related geomorphological cues to improve the reliability of hazard region extraction.

To address the aforementioned challenges, this study constructs the LYPLH dataset based on the NXPLH dataset by leveraging the Segment Anything Model (SAM) [16], and proposes a loess landslide hazard point segmentation method termed LHEM-SegNet. By integrating a Transformer architecture with spatial attention (SA) and channel attention modules (CAM), the proposed method enhances the model's ability to focus on landslide-prone regions and improves segmentation accuracy. Furthermore, by incorporating auxiliary geological and geomorphological data, an extended multi-source fusion model, namely LHEM-MSegNet, is developed to further enhance the segmentation performance of landslide areas.

The main contributions of this study are summarized as follows:

1. A loess landslide hazard region segmentation dataset for Linyou County, Shaanxi Province, China, termed LYPLH, is constructed to support model training and evaluation.
2. A Transformer-based segmentation framework with spatial and channel attention modules is proposed to enhance the extraction of landslide-related features from optical remote sensing images.
3. A multi-source fusion segmentation model, LHEM-MSegNet, is developed to integrate optical imagery with terrain-derived geomorphological information, thereby improving the discrimination of loess landslide hazard regions with weak spectral contrast and ambiguous boundaries.

2. Study Area and Data

The LYPLH dataset was constructed within Linyou County, Baoji City, Shaanxi Province, China. Linyou County is located on the southern margin of the Loess Plateau, spanning longitudes from 107°19' E to 108°00' E and latitudes from 34°33' N to 34°58' N. The region is predominantly characterized by loess hills and gully landforms, with an average elevation ranging from approximately 1000 to 1500 m, making it a representative area of the Loess Plateau [17]. In this study, Linyou County provides the administrative and geomorphological context for the dataset, while the actual image coverage used for dataset construction and model evaluation is the GF-2 satellite image acquired on 22 July 2023. As shown in Figure 1, this GF-2 image coverage is located within Linyou County and mainly covers the Yonglong mining area and its surrounding region.

The raw data were subjected to a series of preprocessing steps, including radiometric correction, orthorectification, image fusion, and spatial cropping, to construct the initial dataset. The landslide hazard inventory used for constructing the LYPLH dataset was derived from registered landslide hazard-point records of Linyou County, which provide coordinate information and attribute descriptions of known loess landslide hazard points. These records were acquired in the same period as the GF-2 imagery and were used as auxiliary spatial references for locating candidate hazard regions during annotation.

The GF-2 imagery acquired on 22 July 2023 was used as the image basis for annotation. During label construction, the registered hazard-point coordinates were overlaid with the GF-2 imagery to identify candidate regions, and the final pixel-level masks were manually interpreted and refined according to the visible landslide-related geomorphic features in the GF-2 imagery. Therefore, the ground-truth labels and the satellite imagery correspond

to the same period, and the final masks represent the landslide hazard regions visible in the GF-2 imagery at the time of image acquisition. The inventory records were not directly converted into ground-truth masks; instead, they were used to guide the annotation process and improve the spatial reliability of label construction.

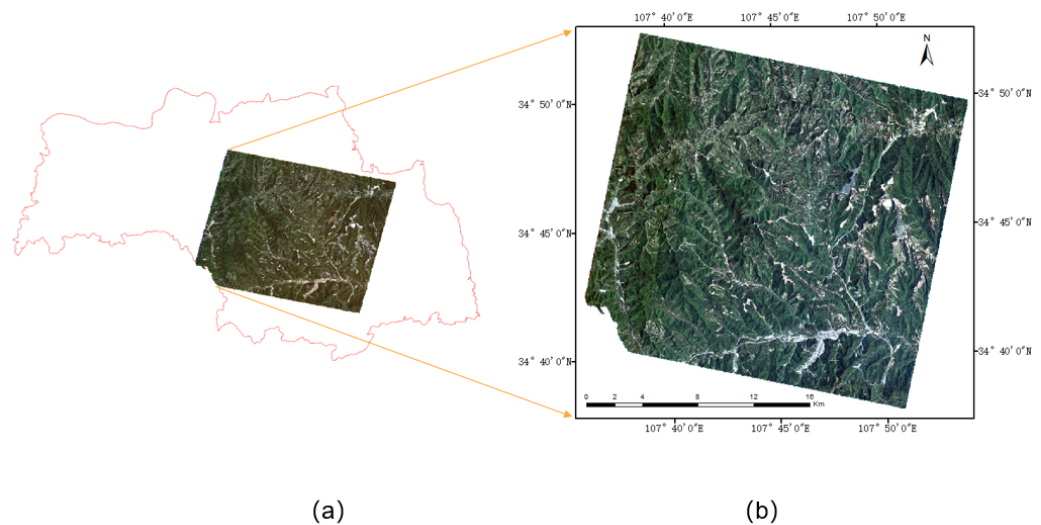


Figure 1. Location of the study area. (a) Administrative boundary of Linyou County, Shaanxi Province, China, with the complete GF-2 image coverage used in this study indicated. (b) Enlarged view of the GF-2 satellite image covering the Yonglong mining area and its surrounding region within Linyou County. The satellite image in (b) represents the complete image coverage used for dataset construction and model evaluation, rather than the entire administrative area of Linyou County. The orange connection lines show the spatial correspondence between the county-scale location map and the GF-2 image coverage.

Although loess landslide hazard-point segmentation cannot directly adopt SAM for semi-automatic annotation due to ambiguous boundaries and weak visual contrast in certain regions, SAM's general segmentation capability can still be leveraged to reduce annotation effort. Given the geographic proximity and similar topographic characteristics between Linyou County and Ningxia Province, a SAM-assisted labeling strategy [18] is employed to construct the LYPLH. It should be emphasized that SAM was used only as an auxiliary tool to generate initial candidate masks, rather than as the source of the ground-truth labels.

The LYPLH dataset production is organized into three phases: SAM fine-tuning with the NXPLH dataset [19], assisted manual annotation, and visual-interpretation-based boundary refinement. In the first phase, SAM is fine-tuned using the NXPLH dataset, which contains 315 potential loess landslide hazard points and 1039 landslide-free samples from Ningxia Province. Foreground and background prompts are interactively provided to correct segmentation results, and the model is retrained for six iterations as annotated samples accumulate.

In the second phase, the fine-tuned SAM model is applied to remote-sensing images of Linyou County. After five rounds of adjustment and retraining, the generated masks were manually refined based on the registered landslide hazard-point locations to correct misclassification errors, yielding a preliminary image-mask dataset, as shown in Figure 2a,b.

In the third phase, Labelme is used to manually refine mask boundaries with reference to the original GF-2 imagery, registered landslide hazard-point information, and visual interpretation results, as illustrated in Figure 2c. Ambiguous regions and obviously incorrect candidate masks were removed or adjusted to improve annotation reliability. After filtering

redundant negative samples, the final LYPLH dataset consists of 2294 images with a size of 256×256 pixels, including 475 positive samples. The relatively small number of positive samples is mainly because individual hazard points often span areas larger than a single image patch. The LYPLH dataset focuses on potential loess landslide hazard regions in Linyou County and is constructed as a binary semantic segmentation dataset, in which the target class represents loess landslide hazard regions and the background class represents non landslide surfaces.

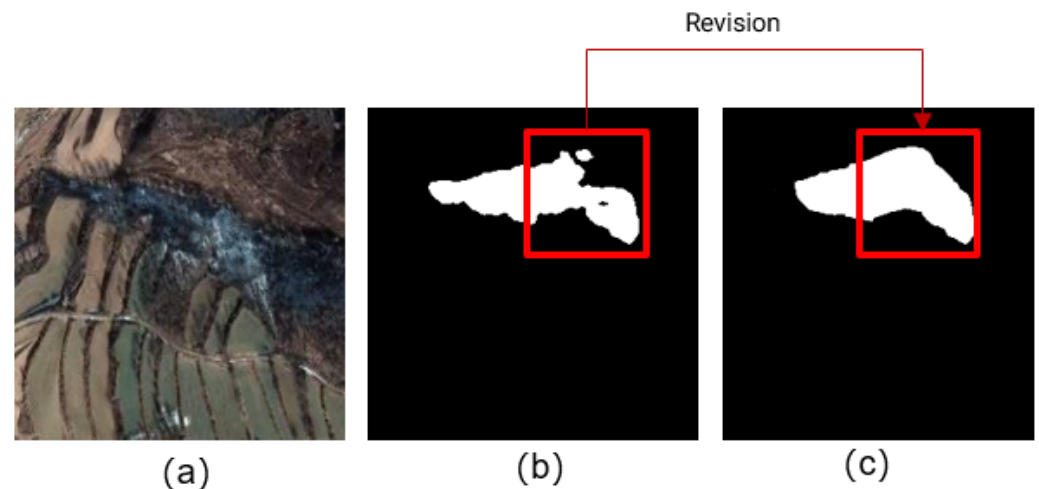


Figure 2. Example of LYPLH dataset samples, including (a) original remote-sensing images, (b) mask predictions generated by the fine-tuned SAM model, and (c) refined annotation masks after manual correction.

In this study, the term loess landslide hazard region refers to the area associated with registered landslide hazard-point information and visually identifiable landslide-related geomorphic features in the GF-2 imagery. It should not be interpreted as a newly predicted unstable slope solely inferred from optical imagery. Existing landslides generally refer to landslide bodies or scars that have already occurred and can be observed in remote sensing images, whereas landslide-prone areas usually refer to susceptibility zones inferred from conditioning factors without explicit boundary labels. The LYPLH dataset is designed for pixel-level extraction of visible loess landslide hazard regions associated with registered hazard points, rather than for predicting future landslide occurrence based on deformation information.

To further evaluate the applicability and robustness of the proposed method under different data distributions, the public Landslide4Sense dataset was used as an external benchmark. It contains image patches collected from four landslide-affected regions, including Iburi, Kodagu, Gorkha, and Taiwan. The dataset consists of three subsets, including 3799 training patches, 245 validation patches, and 800 test patches. Each image patch has a size of 128×128 pixels and contains 14 bands, including Sentinel-2 multispectral bands B1–B12, slope data derived from ALOS PALSAR as B13, and DEM data derived from ALOS PALSAR as B14 [20]. The ground-truth masks used in this experiment are the official binary labels provided by the Landslide4Sense dataset, in which landslide and non landslide pixels are encoded as two classes. These masks were not generated or modified in this study, but were directly used to ensure a fair comparison with existing benchmark settings.

Considering that the proposed framework in this study mainly focuses on the fusion of optical images and terrain-related information, only the Sentinel-2 RGB composite and terrain factors were used in the external generalization experiment. As shown in Figure 3, bands B4, B3, and B2 were selected to construct the RGB image, while B13 and B14 were

used as the slope and DEM inputs, respectively. Since Landslide4Sense does not provide aspect maps, the external experiment was conducted using RGB, slope, and DEM data only. This experiment was used as a supplementary validation to evaluate whether the proposed method can maintain stable performance under different data distributions.

It should be noted that the masks shown in Figure 3 follow the official binary labeling criterion of the Landslide4Sense dataset. According to the dataset description, each pixel is labeled as either landslide or non landslide based on multiple data sources and manual annotation [20]. In this benchmark, pixels included in the official mask are treated as landslide, whereas all unlabeled pixels are treated as non landslide background. In addition, Figure 3 only presents a Sentinel-2 RGB composite using B4/B3/B2 for visualization, while the original Landslide4Sense data contain multiple spectral bands and terrain-related information. Therefore, some visually disturbed, bare, or debris-flow-like areas in the RGB composite may appear different from the official binary mask because the RGB visualization cannot fully represent all spectral and auxiliary information used in the dataset construction. This does not mean that the masks were modified in this study; rather, the official labels were directly used to maintain consistency with the public benchmark.

Such annotation uncertainty may influence both model training and evaluation. During training, incomplete or ambiguous labels may cause the model to learn relatively conservative landslide boundaries and weaken its response to visually similar disturbed regions. During evaluation, if the model extracts visually landslide-like areas that are not included in the official mask, these pixels are counted as false positives, which may reduce Precision, IoU, Dice/F1, and related quantitative metrics. Therefore, the Landslide4Sense experiment is used in this study as a supplementary external benchmark, and the quantitative results should be interpreted together with the possible uncertainty of the official annotations.

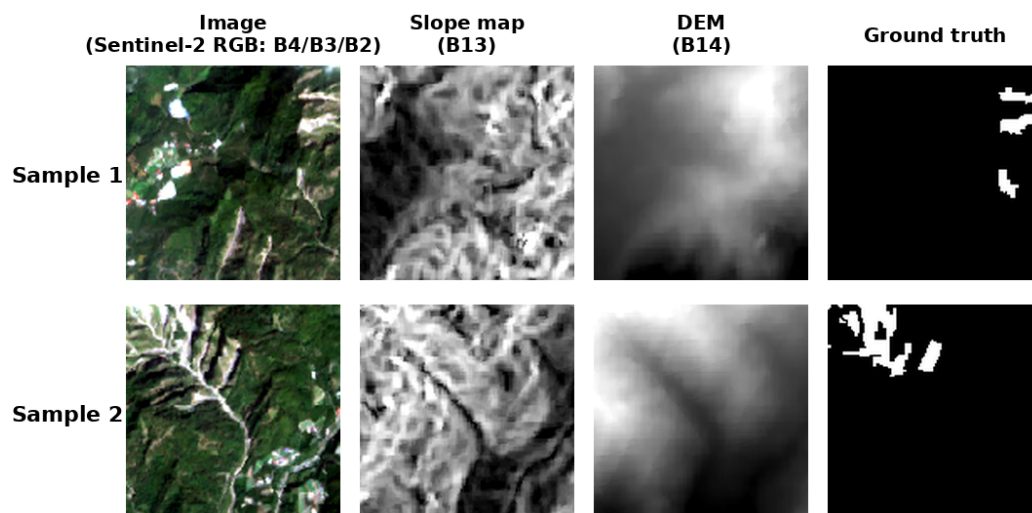


Figure 3. Visualization of the Landslide4Sense training data used in the generalization experiment. The columns show the Sentinel-2 RGB composite using B4/B3/B2, the slope map from B13, the DEM from B14, and the corresponding ground truth mask.

3. Methods

This study proposes a Transformer-based attention fusion network, termed LHEM-SegNet, which is further extended to a multi-source architecture, namely LHEM-MSegNet, as illustrated in Figure 4.

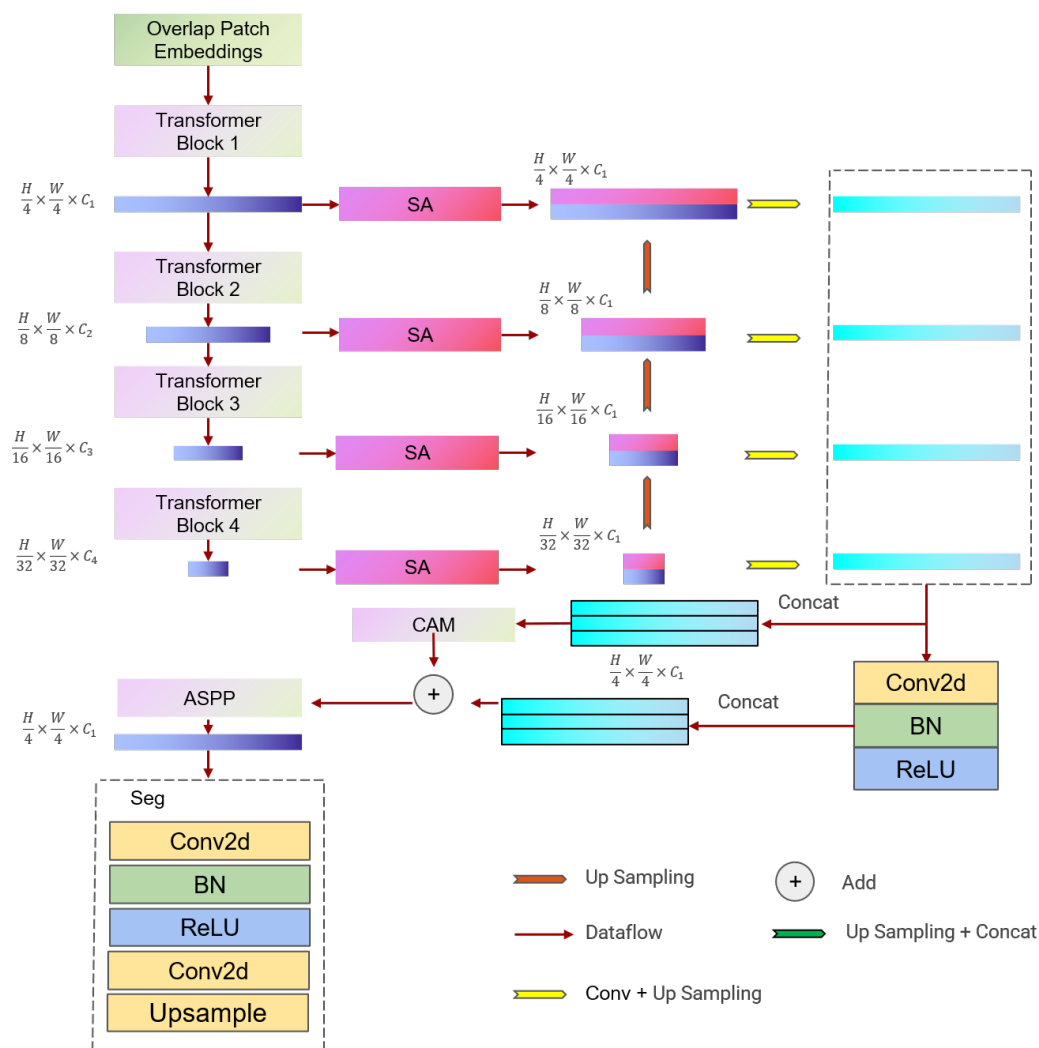


Figure 4. Architecture of the proposed LHEM-SegNet.

3.1. Transformer-Based Attention-Fusion Loess Landslide Hazard Extraction Network

As shown in Figure 4, the Transformer-based landslide hazard point extraction model consists of multiple layers of Transformer blocks [21], SA, CAM, and ASPP [22]. Loess landslide hazards exhibit complex spatial distributions and pronounced spectral and textural differences from the surrounding terrain over time. These characteristics pose challenges for conventional convolutional or single-scale attention-based segmentation models.

To address these challenges, the spatial attention module (SA) dynamically adjusts the importance of different spatial locations within feature maps, enabling the model to focus on landslide hazard areas while suppressing background interference. This mechanism allows the network to better accommodate feature variations under diverse conditions and improves robustness in complex scenarios. In the proposed architecture, SA is deployed after each Transformer block to progressively refine feature maps at multiple levels, thereby enhancing spatial emphasis on hazard regions.

For an input feature map $F \in \mathbb{R}^{H \times W \times C}$, the SA module first applies average pooling and max pooling along the channel dimension to obtain two spatial descriptors, denoted as $F_{avg}^s \in \mathbb{R}^{H \times W \times 1}$ and $F_{max}^s \in \mathbb{R}^{H \times W \times 1}$, respectively. These two descriptors are then concatenated along the channel dimension and passed through a convolution layer to

generate a spatial attention map. Finally, a sigmoid activation function is used to normalize the attention weights. The calculation process of the SA module is defined as follows:

$$SA_s(F) = \sigma \left(f^{7 \times 7} ([AvgPool(F); MaxPool(F)]) \right), \quad (1)$$

where $SA_s(F)$ denotes the generated spatial attention map, σ represents the sigmoid activation function, $f^{7 \times 7}$ denotes a convolution operation with a kernel size of 7×7 , and $[\cdot; \cdot]$ indicates concatenation along the channel dimension. The refined feature map can be further expressed as:

$$F' = SA_s(F) \otimes F, \quad (2)$$

Here, $\otimes$ denotes element-wise multiplication with broadcasting. This spatial weighting operation enables the SA module to emphasize landslide-related regions while suppressing irrelevant background responses, thereby improving the model's ability to identify landslide hazard areas in complex remote sensing scenes.

In parallel, the channel attention module (CAM) enhances segmentation performance by assigning higher weights to informative feature channels, thereby strengthening feature representation and improving landslide identification accuracy. Different from SA, which focuses on spatial locations, CAM emphasizes the importance of different feature channels. Since high-level features contain richer semantic information related to landslide morphology, vegetation disturbance, bare soil exposure, and background interference, CAM is introduced only in the later stages of the network to selectively recalibrate high-level channel responses. This design enhances the extraction of critical semantic information while avoiding unnecessary computational overhead in shallow layers.

Given an input feature map $X \in \mathbb{R}^{H \times W \times C}$, CAM first compresses the spatial information of each channel using global average pooling and global max pooling. This operation produces two channel-wise descriptors, $X_{avg}^c \in \mathbb{R}^{1 \times 1 \times C}$ and $X_{max}^c \in \mathbb{R}^{1 \times 1 \times C}$. The two descriptors are then fed into a shared multilayer perceptron to learn channel dependencies. After that, the two outputs are summed and activated by a sigmoid function to obtain the channel attention weights. The CAM operation can be written as:

$$CAM_c(X) = \sigma(MLP(AvgPool(X)) + MLP(MaxPool(X))), \quad (3)$$

where $CAM_c(X)$ denotes the channel attention map, σ represents the sigmoid activation function, $AvgPool(\cdot)$ and $MaxPool(\cdot)$ indicate global average pooling and global max pooling along the spatial dimensions, respectively, and $MLP(\cdot)$ denotes the shared multilayer perceptron. The generated channel attention map is subsequently multiplied by the original feature map to obtain an enhanced channel representation. Through this recalibration process, CAM increases the responses of channels that contain discriminative semantic information related to landslides and reduces the influence of less informative or noisy channels, thereby improving the ability of the model to distinguish landslide hazard regions from complex background objects.

Furthermore, the atrous spatial pyramid pooling (ASPP) module enhances the representation of landslide hazard regions at different scales through multi-scale feature fusion [22]. After ASPP, the output feature map is fed into the Seg module to generate the final pixel-wise segmentation result. Specifically, given the ASPP output feature map $F_{ASPP} \in \mathbb{R}^{B \times 128 \times H/4 \times W/4}$, the Seg module first applies a 3×3 convolution followed by batch normalization and ReLU activation to reduce the channel dimension from 128 to 64. Then, a 1×1 convolution with C output channels is used as the pixel-wise classification layer to generate segmentation logits $Y' \in \mathbb{R}^{B \times C \times H/4 \times W/4}$, where C denotes the number of semantic classes. In this study, $C = 2$, corresponding to background and loess landslide

hazard region. Finally, the logits are upsampled to the original image resolution $H \times W$ using bilinear interpolation, producing the final output $Y \in \mathbb{R}^{B \times C \times H \times W}$. Through these operations, the network ultimately achieves pixel-level semantic segmentation of loess landslide hazards. The Seg module does not contain additional skip connections or attention mechanisms; these operations are completed in the preceding multi-scale feature fusion, SA/CAM, and ASPP modules.

3.2. Multi-Pathway Loess Landslide Hazard Extraction Model Research

Due to the characteristics of fissure development, high porosity, and strong susceptibility to water-induced softening, landslides occur frequently in loess areas. The occurrence of loess landslides is closely associated with complex geological and geomorphological conditions, among which topographic factors such as slope gradient and aspect play an important role in controlling slope stability. Steep slopes are more likely to be affected by gravitational forces, while slope aspect influences solar radiation, precipitation distribution, soil moisture, and vegetation coverage, thereby indirectly affecting the stability of loess slopes. However, optical remote-sensing (RS) images mainly reflect surface spectral and texture information, and their ability to characterize micro-topographic variations is limited. As a result, areas with similar spectral features but different topographic conditions may be misclassified, leading to missed or false detections of potential landslide hazards. Therefore, it is necessary to integrate topographic and geological information with optical RS images to improve the reliability of loess landslide hazard identification.

A Digital Elevation Model (DEM) provides continuous surface elevation information and serves as an important data source for terrain analysis. Based on DEM data, key topographic parameters such as slope and aspect can be derived, which are directly or indirectly related to surface runoff, erosion processes, soil water redistribution, and slope stability [23]. As shown in Figure 5, the DEM derived topographic factors reveal the complex geomorphological conditions of the study area. The DEM shows pronounced elevation variations and well-developed valley incision, indicating strong terrain relief in the loess region. The slope map presents clear spatial differences in terrain steepness. In particular, medium and high slope zones are mainly distributed along valley margins, ridge flanks, and strongly incised gullies, where gravitational driving forces and surface erosion are generally more active. The aspect map further reflects the directional heterogeneity of slopes, which may influence local solar radiation, precipitation redistribution, evaporation, vegetation growth, and soil moisture conditions, thereby indirectly affecting the stability of loess slopes [11,24,25].

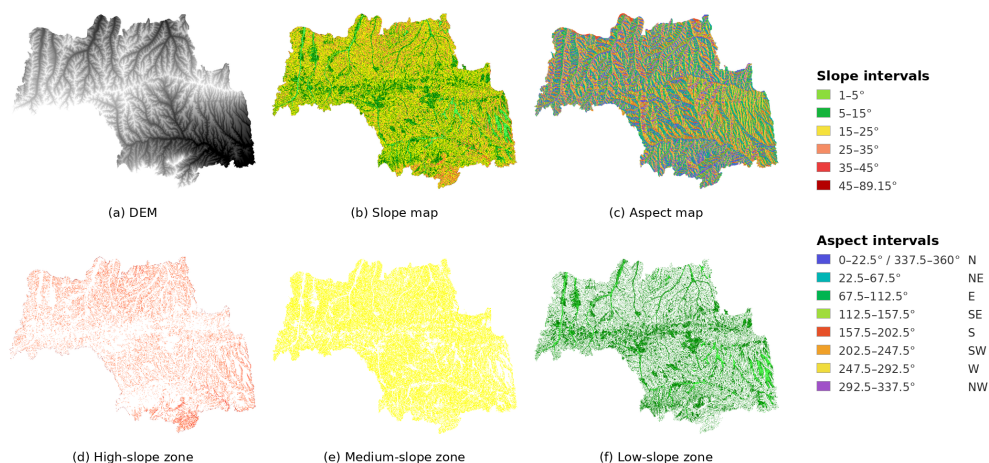


Figure 5. DEM-derived topographic factors for loess landslide hazard extraction.

Therefore, DEM, slope, and aspect can provide mechanism related topographic information for loess landslide hazard extraction. In this study, these three factors were selected as representative terrain inputs because they are physically interpretable, relatively stable, and can directly describe elevation conditions, terrain steepness, and slope orientation, which are closely related to the development of loess landslide hazards. Optical RS images mainly describe surface cover characteristics, while DEM based terrain factors provide complementary topographic constraints. The combination of these complementary data sources enables more comprehensive characterization of landslide prone environments and helps distinguish areas with similar spectral or textural features but different geomorphological conditions. For example, bare soil, gullies, and exposed slopes may show similar spectral responses in optical images, but their terrain conditions can differ significantly in terms of elevation, slope, and aspect.

Based on this, this study designs a multi-channel network for extracting landslide disaster points in loess areas, as shown in Figure 6. The proposed network integrates optical RS images with geological information such as DEM, slope gradient maps, and aspect maps. The model adopts a dual-channel structure, extracting features from both the RS images and topographic information separately, and achieving attention enhancement at the spatial and channel levels through Transformer, SA and CAM. The fused features are subjected to multi-scale modeling by ASPP, and the pixel-level landslide disaster point extraction is completed by the segmentation head. This method effectively improves the accuracy and stability of landslide disaster point identification in loess areas.

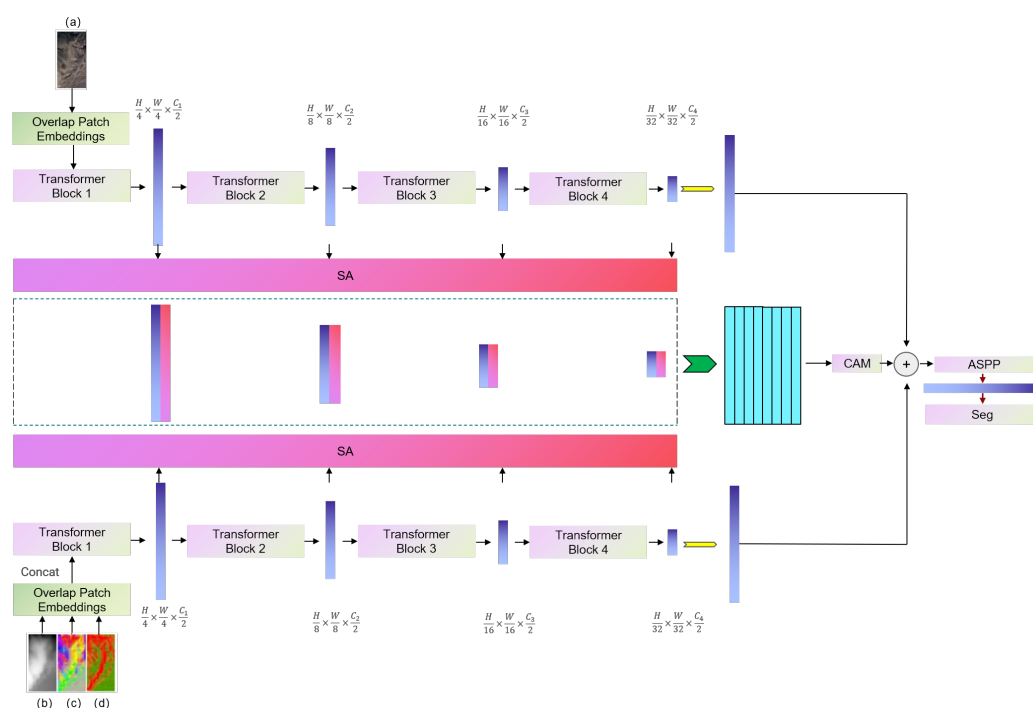


Figure 6. Architecture of the proposed multi-pathway loess landslide hazard extraction network, LHEM-MSegNet. (a) denotes the optical remote sensing image input to the upper branch. (b), (c), and (d) denote the terrain-related inputs, corresponding to DEM, aspect, and slope, respectively. In the multisource framework, the lower branch is used to encode terrain information, which can be provided either as a DEM map alone or as the combination of aspect and slope maps. The extracted optical and terrain features are fused in a stage-wise manner and then passed to the subsequent CAM, ASPP, and Seg modules for final prediction.

3.3. Experimental Setup

To verify the performance of the models, we designed multiple sets of comparative experiments on the LYPLH dataset. As shown in Table 1, the experimental design includes optical-image-based model comparison, structural ablation of the proposed network, multisource fusion comparison, and multisource data ablation. Furthermore, a supplementary evaluation was conducted on the public Landslide4Sense dataset to assess the applicability and robustness of the proposed method under different data distributions.

Table 1. Design of Comparative and Ablation Experiments.

Group	Model	Input
Group A	LHEM-SegNet	Optical Images
	DeepLabV3+	Optical Images
	U-Net	Optical Images
	SegFormer	Optical Images
	SETR	Optical Images
Group B	Transformer baseline	Optical Images
	Transformer + SA	Optical Images
	Transformer + ASPP	Optical Images
	Transformer + CAM	Optical Images
	Transformer + SA + ASPP + CAM, namely LHEM-SegNet	Optical Images
Group C	LHEM-SegNet	Optical Images + DEM
	LHEM-MSegNet	Optical Images + DEM
Group D	LHEM-MSegNet	Optical Images + DEM
	LHEM-MSegNet	Optical Images + Slope Map
	LHEM-MSegNet	Optical Images + Aspect Map
	LHEM-MSegNet	Optical Images + Slope Map + Aspect Map

Firstly, LHEM-SegNet was compared with mainstream segmentation models such as SegFormer [26], SETR [27], DeepLabV3+ [28], and U-Net [29]. The training was conducted using only optical RS images, and all models were evaluated under the same experimental conditions. To ensure a balanced distribution of landslide and non-landslide samples, the dataset was not directly divided as a whole. Instead, the positive samples containing landslide hazard points and the negative samples without landslides were separately divided into training, validation, and testing subsets with a ratio of 6:2:2, and then merged to form the final datasets. This stratified splitting strategy ensured that the proportions of positive and negative samples remained consistent across the training, validation, and testing sets.

To alleviate overfitting caused by the limited number of positive landslide samples and to improve the generalization ability of the model, data augmentation [30–33] was applied to the training set. The augmentation operations included random horizontal flipping with a probability of 0.5, random vertical flipping with a probability of 0.5, random rotation by 90°, 180°, or 270° with a probability of 0.5, random small-angle rotation within the range of −15° to +15° with a probability of 0.5, random scaling from 0.75 to 1.25 with a probability of 0.5, random cropping or padding back to the predefined image size with a probability of 0.5, and random brightness, contrast, and saturation perturbations with an intensity of 0.2 and a probability of 0.5. The validation and testing sets were kept unchanged to ensure a fair and objective evaluation. Specifically, the experiments were carried out on two NVIDIA RTX 4090 GPUs, with an AdamW optimizer, a batch size of 16, and 200 epochs, as shown in Table 2.

Table 2. Experimental environment and training parameters.

Name	Parameter
Hardware conditions	NVIDIA RTX 4090 × 2
Experimental environment	PyCharm 2024.3 + Python 3.8 + CUDA 12.4
Epoch	200
Batch size	16
Optimizer	AdamW
Learning rate	1×10^{-5}
Learning rate decay	CosineAnnealingLR
Loss function	Focal Loss + Cross Entropy Loss

Second, to verify the effectiveness of the proposed network structure for landslide hazard extraction, a structural ablation experiment was conducted under the optical-only input setting. Specifically, the Transformer baseline, Transformer + SA, Transformer + ASPP, Transformer + CAM, and the complete LHEM-SegNet were compared under identical experimental conditions. This experiment was designed to evaluate the individual contributions of spatial attention, channel attention, and multi-scale context aggregation to the extraction of landslide hazard regions.

Third, under identical experimental settings, LHEM-SegNet is compared with its DEM-integrated counterpart, LHEM-MSegNet, to investigate the contribution of multi-source terrain information. In addition, a series of ablation experiments are conducted based on LHEM-MSegNet by progressively incorporating DEM, slope, aspect, and their combinations, in order to analyze the impact of different geomorphological factors on model performance. Finally, a qualitative visual comparison is performed to evaluate the extraction results of different models across several representative scenarios.

This study employed the mean Intersection over Union (mIoU), accuracy (Acc), Dice Similarity Coefficient (Dice/F1), and Kappa Coefficient (Kappa) as evaluation metrics to quantitatively assess the performance of different landslide hazard extraction models. Kappa was retained as a complementary metric because it measures the agreement between the predicted results and the reference labels after considering chance agreement, which helps evaluate the overall consistency of binary segmentation results under imbalanced landslide and background classes. For binary semantic segmentation, the Dice coefficient is mathematically equivalent to the F1-score calculated from precision and recall; therefore, Dice/F1 is reported instead of listing F1-score as a separate metric to avoid redundancy. In addition, Precision_{LS} and Recall_{LS} were further reported in the main comparative experiment to evaluate false extractions and missed detections of the landslide hazard region class. The specific formulas are shown in Equations (4)–(9).

$$Acc = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}, \quad (4)$$

$$Kappa = \frac{P_o - P_e}{1 - P_e}, \quad (5)$$

$$Dice = \frac{2T_P}{2T_P + F_P + F_N}, \quad (6)$$

$$mIoU = \frac{1}{C} \sum_{i=1}^C \frac{T_{P_i}}{T_{P_i} + F_{P_i} + F_{N_i}}, \quad (7)$$

$$Precision_{LS} = \frac{TP_{LS}}{TP_{LS} + FP_{LS}}, \quad (8)$$

$$Recall_LS = \frac{TP_{LS}}{TP_{LS} + FN_{LS}}, \quad (9)$$

where T_P , T_N , F_P , and F_N represent true positives, true negatives, false positives, and false negatives, respectively. C denotes the number of classes. In Equation (5), P_o represents the observed agreement, which reflects the proportion of correctly classified samples, while P_e denotes the chance agreement, indicating the expected agreement under random classification. The Kappa coefficient measures the consistency between the predicted results and the ground truth. For the landslide hazard region class, TP_{LS} represents the number of pixels correctly predicted as landslide hazard regions, FP_{LS} represents the number of background pixels incorrectly predicted as landslide hazard regions, and FN_{LS} represents the number of landslide hazard pixels incorrectly predicted as background. Therefore, Precision_LS and Recall_LS specifically evaluate the reliability and completeness of the extracted landslide hazard regions, respectively.

4. Results and Discussion

Table 3 presents the results of the first group of comparative experiments. A higher mIoU value indicates better hazard region extraction, and a Kappa coefficient closer to 1 suggests stronger consistency between the predicted results and the ground truth. Pre_LS and Rec_LS further evaluate the reliability and completeness of the extracted landslide hazard regions, respectively. Overall, LHEM-SegNet achieves the best performance across all evaluation metrics, with an mIoU of 87.70%, accuracy of 94.20%, Dice/F1 of 93.40%, Kappa coefficient of 0.869, Pre_LS of 85.97%, and Rec_LS of 86.74%. These results demonstrate its effectiveness in segmenting loess landslide hazard regions. In particular, the highest Pre_LS indicates that LHEM-SegNet produces fewer false extractions of landslide hazard regions, while the highest Rec_LS indicates that it also reduces missed detections compared with the baseline models. However, since only optical remote sensing imagery is utilized, the performance gains over baseline networks remain limited. For example, compared with DeepLabV3+, the improvements in mIoU and Kappa coefficient are only 0.15% and 0.003, respectively, while the improvements in Pre_LS and Rec_LS are 6.54% and 2.48%, respectively. This indicates that LHEM-SegNet improves the reliability and completeness of landslide hazard region extraction to some extent, but also highlights the necessity of integrating geological and geomorphological information to further enhance landslide extraction performance.

Table 3. Assessment Table of Results for the First Group of Comparative Experiments. Pre-LS and Rec-LS denote Precision and Recall calculated for the landslide hazard region class, respectively.

Model	mIoU (%)	Acc (%)	Dice/F1 (%)	Kappa	Pre-LS (%)	Rec-LS (%)
LHEM-SegNet	87.70	94.20	93.40	0.869	85.97	86.74
DeepLabV3+	87.55	93.20	92.30	0.866	79.43	84.26
U-Net	86.83	93.71	92.90	0.858	82.36	83.12
SegFormer	87.31	93.01	92.16	0.863	80.24	80.76
SETR	86.45	93.15	91.61	0.853	76.00	78.99

To further evaluate the contribution of different structural components in the optical-only LHEM-SegNet, a structural ablation experiment was conducted, as shown in Table 4. Compared with the baseline model without SA, CAM, and ASPP, the introduction of SA significantly improves the performance, increasing the mIoU from 58.22% to 80.00%, the Acc from 78.27% to 92.12%, and the Dice from 72.30% to 88.52%. This indicates that spatial attention plays an important role in enhancing the localization ability of landslide-related regions. When only ASPP is introduced, the model also achieves a clear improvement,

with an mIoU of 77.08%, suggesting that multi-scale contextual information is beneficial for extracting landslide regions with different sizes and complex morphologies.

In contrast, the CAM-only variant shows limited improvement compared with the baseline, with an mIoU of 58.26%. This suggests that channel attention alone is insufficient for accurate landslide extraction because it mainly recalibrates channel-wise feature responses and lacks explicit spatial localization capability. However, when SA, CAM, and ASPP are jointly integrated, the model achieves the best performance, with an mIoU of 87.70%, an Acc of 94.20%, a Dice of 93.40%, and a Kappa coefficient of 0.869. These results demonstrate that the three modules are complementary: SA enhances spatial feature localization, CAM improves channel-wise feature selection, and ASPP captures multi-scale contextual information. Their joint integration effectively improves the feature representation capability of the optical-only LHEM-SegNet.

Table 4. Structural Ablation Results of the Optical-Only LHEM-SegNet. The symbols “✓” and “×” indicate that the corresponding module is used and not used, respectively.

ID	SA	CAM	ASPP	mIoU (%)	Acc (%)	Dice (%)	Kappa
A	×	×	×	58.22	78.27	72.30	0.451
B	✓	×	×	80.00	92.12	88.52	0.770
C	×	×	✓	77.08	90.19	86.64	0.733
D	×	✓	×	58.26	78.52	72.27	0.450
E	✓	✓	✓	87.70	94.20	93.40	0.869

As shown in Table 5, when DEM data are incorporated, LHEM-MSegNet consistently outperforms the baseline LHEM-SegNet, particularly in terms of mIoU and Dice coefficient, demonstrating the effectiveness of multi-source feature fusion for loess landslide hazard-point detection. LHEM-SegNet adopts a single-input structure, in which the DEM data and optical images are combined through channel-wise addition before being fed into the network. In contrast, LHEM-MSegNet processes the optical image and DEM image through separate pathways, allowing the model to extract spectral-spatial features and terrain-related features independently before feature fusion. This design enables more effective utilization of complementary information from optical and DEM data.

Table 5. Evaluation Table of Results for the Second Group of Comparative Experiments.

Model	mIoU (%)	Acc (%)	Dice (%)	Kappa
LHEM-SegNet	88.25	94.21	93.29	0.873
LHEM-MSegNet	88.64	94.54	94.43	0.876

Although model accuracy is essential for evaluating landslide hazard extraction methods, computational cost should also be considered when assessing their practical applicability. Therefore, the number of parameters, FLOPs, training time per epoch, and inference time per image are reported in Table 6. The evaluation settings are consistent with those used in the comparative experiments.

As shown in Table 6, different models show obvious differences in model size and computational cost. LHEM-MSegNet has 6.87 M parameters and 38.80 G FLOPs, which are lower than those of the optical-only LHEM-SegNet. This is mainly because each branch in the multi-source model uses half of the channel width of the corresponding single-source encoder, and the channel dimension is restored after feature concatenation. Therefore, the multi-source design does not simply increase the model size.

Table 6. Comparison of model complexity and computational efficiency.

Model	Input	Params (M)	FLOPs (G)	Train (s/Epoch)	Inference (ms/Image)
LHEM-MSegNet	Optical images + DEM	6.87	38.80	41.28	18.98
LHEM-SegNet	Optical images	12.34	41.02	29.81	7.54
DeepLabV3+	Optical images	2.54	158.06	40.90	8.15
SegFormer	Optical images	3.71	15.55	17.26	8.51
SETR	Optical images	6.96	22.79	12.78	7.63
U-Net	Optical images	7.70	281.71	51.36	15.70

Compared with DeepLabV3+ and U-Net, LHEM-MSegNet has a clear advantage in computational complexity. Although DeepLabV3+ has only 2.54 M parameters, its FLOPs reach 158.06 G, indicating that a small number of parameters does not necessarily correspond to low computational cost. This may be related to its multi-scale atrous convolution and decoder operations on relatively large feature maps. U-Net has 7.70 M parameters but the highest FLOPs of 281.71 G, mainly because its symmetric encoder-decoder structure and high-resolution skip connections require intensive feature reconstruction operations. In comparison, LHEM-MSegNet achieves lower FLOPs than both models while maintaining stronger landslide hazard extraction performance.

SegFormer and SETR show lower FLOPs and shorter inference time than LHEM-MSegNet, indicating that they are computationally efficient baseline models. However, their accuracy is lower than that of the proposed model in the comparative experiments, suggesting that lower computational cost alone is insufficient for reliable loess landslide hazard region extraction. LHEM-MSegNet introduces terrain information and multi-source feature fusion, which improves the representation of landslide-related geomorphological features under complex loess terrain conditions.

It should be noted that LHEM-MSegNet requires a longer inference time than the optical-only models. This is mainly due to the additional multi-source input processing, feature fusion, and memory access operations. In addition, FLOPs mainly reflect theoretical arithmetic operations, while actual inference time is also affected by data loading, branch parallelism, feature concatenation, and hardware execution efficiency. Nevertheless, the average inference time of LHEM-MSegNet is 18.98 ms per image, which remains acceptable for tile-based landslide hazard region extraction. Overall, these results indicate that LHEM-MSegNet improves extraction accuracy while maintaining a controllable computational cost, achieving a reasonable trade-off between segmentation performance and computational efficiency.

To further assess the impact of geological and geographical information on loess landslide hazard point segmentation, Group D was designed to evaluate the contribution of DEM, slope, and aspect data. Since slope and aspect are derived from DEM data, the ablation experiments were designed based on these DEM-related data combinations rather than treating them as entirely independent data sources. The corresponding results obtained using the proposed multi-source model LHEM-MSegNet are summarized in Table 7.

Table 7. Results of the Third Group of Ablation Experiments.

Data Type	mIoU (%)	Acc (%)	Dice (%)	Kappa
Optical + DEM	88.64	94.54	94.43	0.876
Optical + Slope Map	88.51	94.98	93.84	0.877
Optical + Aspect Map	88.91	94.72	94.09	0.882
Optical + Slope Map + Aspect Map	89.29	94.59	94.28	0.885

The experimental results presented in Table 7 demonstrate that different input data combinations yield distinct model performance levels, as detailed below:

- The Optical + DEM combination achieves an mIoU of 88.64%, an Acc of 94.54%, a Dice coefficient of 94.43%, and a Kappa coefficient of 0.876. The relatively high Dice value indicates that DEM information is effective in improving the spatial overlap between the predicted landslide regions and the reference labels. This is because DEM provides elevation information and reflects the overall terrain relief of the study area, which can complement the surface spectral and texture features provided by optical RS images. Therefore, the combination of optical images and DEM is useful for representing both surface cover and terrain variations in landslide hazard point extraction.
- The Optical + Slope Map combination obtains an mIoU of 88.51%, an Acc of 94.98%, a Dice coefficient of 93.84%, and a Kappa coefficient of 0.877. Although its mIoU is slightly lower than that of the Optical + Aspect Map combination, it achieves the highest Acc among the terrain based combinations, indicating strong overall pixel classification performance. The Slope Map reflects terrain steepness and is closely related to gravitational driving forces, surface runoff, and erosion processes. Therefore, slope information provides important topographic constraints for identifying landslide prone areas, especially in medium and high slope regions.
- The Optical + Aspect Map combination achieves an mIoU of 88.91%, an Acc of 94.72%, a Dice coefficient of 94.09%, and a Kappa coefficient of 0.882. Compared with the DEM and Slope Map combinations, the Aspect Map combination obtains a higher mIoU and Kappa coefficient, indicating that aspect information contributes to improving the overall segmentation performance and result consistency. The Aspect Map describes slope orientation and may indirectly affect slope stability by influencing solar radiation, vegetation development, evaporation, and soil moisture distribution. Therefore, aspect information provides complementary environmental information for distinguishing landslide prone areas from background regions with similar spectral characteristics.
- The Optical + Slope Map + Aspect Map combination achieves the best mIoU of 89.29% and the highest Kappa coefficient of 0.885, with an Acc of 94.59% and a Dice coefficient of 94.28%. Although its Acc and Dice are slightly lower than the best values obtained by the Optical + Slope Map and Optical + DEM combinations, respectively, this combination shows the most balanced overall performance. The improvement in mIoU and Kappa indicates that jointly using slope and aspect information can enhance the discrimination of landslide hazard areas and improve the consistency of the extraction results. This is because the Slope Map provides information on terrain steepness and potential gravitational instability, while the Aspect Map describes slope orientation and related environmental differences. Their combination allows the model to incorporate complementary terrain constraints into optical image interpretation, thereby improving the accuracy and stability of landslide hazard point extraction.

Overall, the multisource ablation results show that each data source contributes to the final prediction from a different perspective. Optical remote sensing images provide the basic spectral, textural, and boundary information for identifying visible loess landslide hazard regions. DEM introduces elevation and terrain-background information, which helps the model perceive the overall topographic setting of the study area. Slope provides terrain steepness information related to gravitational instability and erosion processes, while aspect provides slope-orientation information associated with illumination, vegetation, evaporation, and moisture differences. These complementary cues explain why the combination of optical imagery with terrain-related information improves the accuracy and stability of loess landslide hazard region extraction.

To qualitatively compare the experimental results, a visual analysis was conducted for several typical scenarios, as shown in Figure 7. These scenarios include large steep slopes, small slopes, complex terrain, plains, and small residential areas. Each column represents one typical scenario and presents the extraction results of different models. The first row shows the original remote sensing images, and the second row shows the corresponding labeled images. The subsequent rows show the extraction results of LHEM-MSegNet, LHEM-SegNet + DEM, LHEM-SegNet, DeepLabV3+, SegFormer, UNet, and SETR, respectively. In each extraction result, the white regions indicate correctly extracted landslide areas that overlap with the labels, the red regions represent false extractions, and the blue regions indicate missed detections.

As illustrated in Figure 7, LHEM-MSegNet achieves the best overall extraction performance, exhibiting the fewest false and missed detections across diverse terrain conditions. The model performs consistently well in both regular and complex areas, with residual errors primarily concentrated along landslide hazard point boundaries.

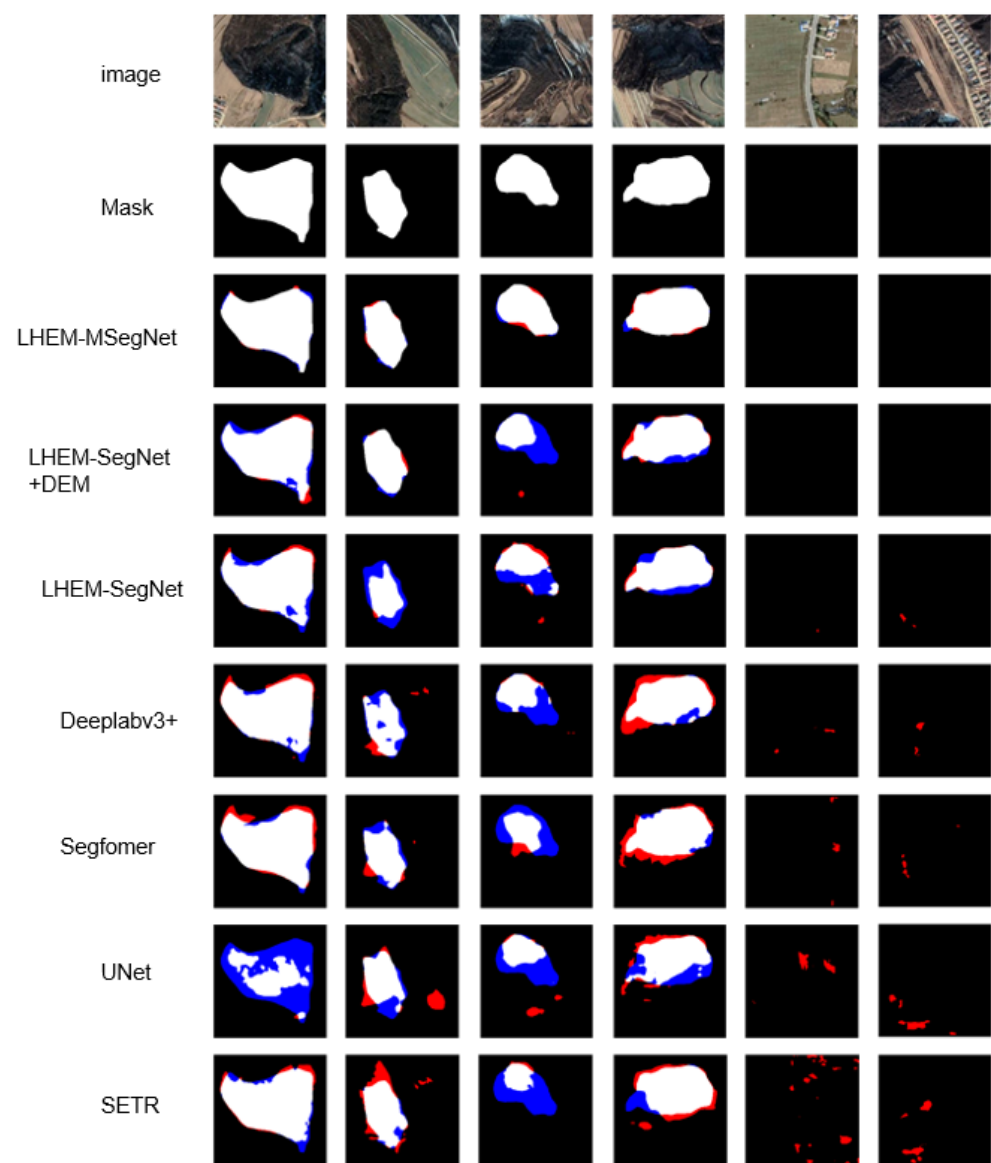


Figure 7. Qualitative comparison of landslide hazard extraction results under typical scenarios. White regions represent correctly extracted landslide hazard regions overlapping with the labels, red regions represent false extractions, and blue regions represent missed detections.

In contrast, LHEM-SegNet without DEM information shows degraded performance in steep and complex terrains, indicating that the absence of elevation data limits its ability to effectively characterize terrain features. After incorporating DEM data, LHEM-SegNet + DEM demonstrates a substantial improvement in extraction accuracy and ranks second only to LHEM-MSegNet, underscoring the effectiveness of elevation information in landslide hazard point segmentation.

Among the comparative methods, DeepLabv3+ and SegFormer achieve moderate performance but exhibit increased false and missed detections in complex regions, while UNet and SETR perform the weakest, particularly in challenging scenes. Notably, only LHEM-MSegNet and LHEM-SegNet + DEM produce no false detections in non-landslide areas, highlighting their superior robustness and reliability for loess landslide hazard point extraction.

In addition, Figure 7 also provides a qualitative analysis of confusing non landslide surfaces. In optical remote sensing images, plains, exposed soils, road cuts, human disturbed areas near residential regions, and other bare surfaces may show visual characteristics similar to landslide hazard regions. These confusing areas are reflected by the red regions in the extraction results, which represent false extractions. Compared with the other models, LHEM-MSegNet produces fewer false extractions in these scenarios. This improvement indicates that the integration of DEM information and attention enhanced feature representation can provide complementary terrain and contextual information, thereby helping the model distinguish landslide hazard regions from visually similar bare surfaces.

As shown in Table 8, the external generalization experiment was conducted on the Landslide4Sense dataset to evaluate the robustness of different models under different data distributions. When only the Sentinel-2 RGB composite, namely B4/B3/B2, was used as input, LHEM-SegNet achieved an mIoU of 63.48%, an Acc of 97.19%, a Dice coefficient of 72.13%, and a Kappa coefficient of 0.443. Compared with the strongest RGB baseline in terms of mIoU and Kappa, SETR, LHEM-SegNet improved the mIoU, Acc, Dice, and Kappa by 0.14%, 0.70%, 0.24%, and 0.003, respectively. This indicates that the proposed optical-image-based network maintains competitive landslide extraction performance on the external dataset.

After incorporating DEM and slope information, LHEM-MSegNet achieved the best overall performance, with an mIoU of 65.12%, an Acc of 97.49%, a Dice coefficient of 74.05%, and a Kappa coefficient of 0.482. Compared with LHEM-SegNet, the multisource version improved the mIoU by 1.64%, Acc by 0.30%, Dice by 1.92%, and Kappa by 0.039. These improvements demonstrate that terrain-related information provides complementary geomorphological cues for landslide hazard extraction and helps the proposed method maintain more robust performance on an external benchmark with different data distributions.

It should be noted that the improvement on Landslide4Sense is relatively moderate. This may be partly related to the reduced spectral input used in this experiment. The original Landslide4Sense dataset contains 14 bands, including multiple Sentinel-2 multispectral bands and terrain-related data. However, to maintain consistency with the input setting of the proposed framework, this experiment used only the B4/B3/B2 RGB composite together with DEM and slope information. As a result, some spectral differences that may be useful for distinguishing landslide surfaces, bare soil, vegetation disturbance, and other background areas cannot be fully represented by the RGB composite alone. This limited spectral representation may increase the visual similarity between landslide and non-landslide disturbed surfaces, leading to more difficult boundary discrimination and relatively conservative model predictions.

Table 8. External dataset generalization comparison on Landslide4Sense.

Model	Input	mIoU (%)	Acc (%)	Dice (%)	Kappa
LHEM-MSegNet	Sentinel-2 RGB composite (B4/B3/B2) + DEM + Slope	65.12	97.49	74.05	0.482
LHEM-SegNet	Sentinel-2 RGB composite (B4/B3/B2)	63.48	97.19	72.13	0.443
DeepLabV3+	Sentinel-2 RGB composite (B4/B3/B2)	62.10	95.94	71.13	0.434
U-Net	Sentinel-2 RGB composite (B4/B3/B2)	62.97	96.29	71.56	0.422
SegFormer	Sentinel-2 RGB composite (B4/B3/B2)	61.71	95.75	70.24	0.407
SETR	Sentinel-2 RGB composite (B4/B3/B2)	63.34	96.49	71.89	0.440

In addition, the samples shown in Figure 3 are also visualized using the B4/B3/B2 RGB composite. Therefore, some areas that appear disturbed or debris-flow-like in the RGB visualization may not fully reflect the multispectral and terrain information contained in the original dataset. Such visualization limitation may make the official binary masks appear partially inconsistent with the RGB image, but it does not indicate that the masks were modified in this study. The official masks were directly used for training and evaluation to maintain consistency with the public benchmark.

Therefore, although the performance gain on Landslide4Sense is not as large as that observed in the internal LYPLH experiments, LHEM-MSegNet still achieves the best overall performance. This result indicates that the proposed multisource fusion strategy has a certain degree of transferability, and that DEM and slope information can provide useful terrain constraints when spectral information is limited. Nevertheless, broader validation in more loess regions and other geological and geomorphological settings is still needed in future work.

5. Conclusions

This study focuses on improving the extraction of loess landslide hazard points from remote-sensing images. To alleviate the scarcity of loess landslide hazard datasets, the LYPLH dataset was constructed with reference to the NXPLH dataset, and data augmentation and preprocessing were applied to improve data diversity. Based on this dataset, a Transformer-based LHEM-SegNet model was developed by incorporating spatial and channel attention modules to enhance feature representation for landslide hazard-point segmentation. To make better use of multi-source geo-information, including optical remote-sensing images, DEM data, slope-gradient maps, and slope-aspect maps, LHEM-MSegNet was further developed to integrate spectral, spatial, and terrain-related features for loess landslide hazard-point segmentation.

The experimental results on the LYPLH dataset indicate that LHEM-MSegNet achieves better overall performance than the compared models in terms of the evaluated metrics. These results suggest that DEM-related terrain information and multi-source feature fusion can contribute to loess landslide hazard-point segmentation. In addition, the computational comparison shows that LHEM-MSegNet maintains a controllable model complexity, with fewer parameters and lower FLOPs than the optical-only LHEM-SegNet. This indicates that the performance improvement is not achieved by simply increasing model size or computational burden. To further evaluate the generalization ability of the proposed method for landslide detection under different geomorphological environments, an additional experiment was conducted on the public Landslide4Sense dataset. The experimental results show that LHEM-SegNet achieves better overall performance than other RGB-based baseline models under the same input conditions, while the multisource LHEM-MSegNet further improves the mIoU, Acc, Dice, and Kappa after incorporating DEM and slope information. This indicates that the proposed method maintains competitive performance on

the external Landslide4Sense benchmark, and that terrain-related information can provide effective complementary cues for landslide hazard extraction under different data distributions. However, due to differences in sensor type, landslide characteristics, and regional geomorphological conditions, the improvement on Landslide4Sense is relatively moderate. Therefore, the generalization ability of the proposed method in more loess regions and other complex geomorphological environments still needs to be further validated in future work.

Several limitations should be acknowledged in this study. First, although manual correction and visual interpretation were performed during dataset construction, potential annotation bias may still exist in the LYPLH dataset, particularly because SAM-assisted pre-annotation was used in the initial labeling stage. Second, although a supplementary evaluation was conducted on the public Landslide4Sense dataset to assess the applicability of the proposed method on an external benchmark, this experiment was still limited by differences in sensor type, available data sources, landslide characteristics, and regional geomorphological conditions. In particular, Landslide4Sense does not provide aspect maps, which prevents a fully consistent multisource input configuration with the LYPLH dataset. In addition, the visual samples shown in this study are based on the Sentinel-2 B4/B3/B2 RGB composite, whereas the original Landslide4Sense dataset contains multiple spectral bands and terrain-related information. Therefore, some visually disturbed or bare areas in the RGB composite may not be fully consistent with the official binary landslide masks, because RGB visualization cannot represent all information contained in the original multisource dataset. The current study focuses on loess landslide hazard regions, and the applicability of the proposed method to landslide hazards in other geological and geomorphological settings still requires further validation. Therefore, although the external experiment indicates that the proposed method has a certain degree of transferability, its robustness in more loess regions and other complex geomorphological environments still needs further validation.

Third, the introduction of attention modules and multisource branches improves feature representation but also increases the structural complexity of the model, which may affect computational efficiency in large scale practical applications. It should also be noted that other DEM derived geomorphological factors, such as curvature, roughness, terrain position index, and topographic wetness index, were not systematically evaluated in this study. These factors may provide additional information on slope morphology, local surface variation, relative terrain position, and potential water accumulation conditions, and may further improve landslide extraction performance in some complex geomorphological environments. However, these derivatives may also be correlated with DEM, slope, and aspect, and their reliability can be affected by DEM resolution, noise, and calculation scale. Introducing too many terrain channels may increase model complexity and the risk of overfitting, especially when the number of training samples is limited. Therefore, this study focused on DEM, slope, and aspect as representative and interpretable terrain factors. Future work will focus on expanding the dataset, improving annotation consistency, evaluating additional DEM derived geomorphological factors, conducting broader cross regional and cross dataset validation, and further optimizing the balance between segmentation accuracy and model complexity.

Author Contributions: Conceptualization, F.S., T.Z. and W.L.; methodology, F.S. and T.Z.; validation, X.S. and T.Z.; formal analysis, F.S.; investigation, W.L.; resources, W.L.; data curation, Z.H. and X.Y.; writing—original draft preparation, F.S.; writing—review and editing, F.S., T.Z. and W.L.; visualization, F.S.; supervision, W.L.; funding acquisition, W.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Xi'an Key Industrial Chain Technology R&D Projects, grant numbers 25ZDLYB00006 and 25ZDLYB00015; Shaanxi Provincial Innovation Capability Support Program, grant number 2025ZC-WTXM-36; the Shaanxi Provincial Qinchuangyuan "Scientists + Engineers" Project, grant number 2025QCY-KXJ-016; and the Scientific Research Projects of Shaanxi Coalfield Geology Group, grant numbers SMDZ-2023CX-14 and SMDZ-2025ZD-11.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Liu, J.; Chen, C.; Fu, Y.; Wang, C.; Li, T.; Xiao, R.; Liu, Y. Mechanism of rainfall-induced shallow landslide and stability prediction model. *Hydrogeol. Eng. Geol.* **2024**, *51*, 183–191.
2. Zhang, L.; Liao, M.; Dong, J.; Xu, Q.; Gong, J. Early detection of landslide hazards in mountainous areas of west China using time series SAR interferometry—a case study of Danba, Sichuan. *Geomat. Inf. Sci. Wuhan Univ.* **2018**, *43*, 2039–2049.
3. Zhang, T.; Ren, Y.; Li, W.; Qin, C.; Jiao, L.; Su, H. CSW-SAM: A cross-scale algorithm for very-high-resolution water body segmentation based on Segment Anything Model 2. *ISPRS J. Photogramm. Remote Sens.* **2025**, *228*, 208–227. [\[CrossRef\]](#)
4. Zhang, T.; Ji, W.; Li, W.; Qin, C.; Wang, T.; Ren, Y.; Fang, Y.; Han, Z.; Jiao, L. EDWNet: A Novel Encoder–Decoder Architecture Network for Water Body Extraction from Optical Images. *Remote. Sens.* **2024**, *16*, 4275. [\[CrossRef\]](#)
5. Zhang, J.; Cui, Q.; Ma, X. Deep evidential remote sensing landslide image classification with a new divergence, multiscale saliency and an improved three-branched fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2024**, *17*, 3799–3820. [\[CrossRef\]](#)
6. Ju, Y.; Xu, Q.; Jin, S.; Li, W.; Su, Y.; Dong, X.; Guo, Q. Loess landslide detection using object detection algorithms in northwest China. *Remote. Sens.* **2022**, *14*, 1182. [\[CrossRef\]](#)
7. Zhou, N.; Hong, J.; Cui, W.; Wu, S.; Zhang, Z. A multiscale attention segment network-based semantic segmentation model for landslide remote sensing images. *Remote. Sens.* **2024**, *16*, 1712. [\[CrossRef\]](#)
8. Wang, Z.; Sun, T.; Hu, K.; Zhang, Y.; Yu, X.; Li, Y. A deep learning semantic segmentation method for landslide scene based on transformer architecture. *Sustainability* **2022**, *14*, 16311. [\[CrossRef\]](#)
9. Li, Z.-H.; Shi, A.-C.; Xiao, H.-X.; Niu, Z.-H.; Jiang, N.; Li, H.-B.; Hu, Y.-X. Robust landslide recognition using UAV datasets: A case study in Baihetan Reservoir. *Remote. Sens.* **2024**, *16*, 2558. [\[CrossRef\]](#)
10. Fu, Z.; Wang, F.; Zhong, J.; Catani, F.; Dou, J.; You, Q.; Zhang, B. A CNN–Transformer hybrid network for efficient cross-region landslide detection by transfer learning. *Landslides* **2026**, 1–17. [\[CrossRef\]](#)
11. Wei, L.; Zeng, Z.; Yan, J. Factors affecting the stability of loess landslides: A review. *Appl. Sci.* **2024**, *14*, 2735. [\[CrossRef\]](#)
12. Zhang, M.-S.; Liu, J. Controlling factors of loess landslides in western China. *Environ. Earth Sci.* **2010**, *59*, 1671–1680.
13. Zhuang, J.; Peng, J.; Wang, G.; Iqbal, J.; Wang, Y.; Li, W. Distribution and characteristics of landslide in Loess Plateau: A case study in Shaanxi province. *Eng. Geol.* **2018**, *236*, 89–96. [\[CrossRef\]](#)
14. Shao, X.; Ma, S.; Xu, C.; Xu, Y. Insight into the characteristics and triggers of loess landslides during the 2013 heavy rainfall event in the Tianshui area, China. *Remote. Sens.* **2023**, *15*, 4304. [\[CrossRef\]](#)
15. Zhao, C.; Li, L.; Wang, Y.; Li, X.; Xu, C.; Song, Y.; Ren, D.; Xiao, C. D2FLS-Net: Dual-Stage DEM-guided Fusion Transformer for landslide segmentation. *PLoS ONE* **2025**, *20*, e0337412. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–3 October 2023; pp. 4015–4026.
17. Wang, J.J.; Liang, Y.; Zhang, H.P.; Wu, Y.; Lin, X. A loess landslide induced by excavation and rainfall. *Landslides* **2014**, *11*, 141–152.
18. Wang, C.; Li, D.; Wang, S.; Zhang, C.; Wang, Y.; Liu, Y.; Yang, G. Sammed: A medical image annotation framework based on large vision model. *arXiv* **2023**, arXiv:2307.05617.
19. Wang, R.; Wei, H.; Kong, Y.; Zhang, P.; Wang, X.; Li, W. A Potential Landslide Hazard Benchmark Dataset for Geological Disaster Detection of Remote Sensing. In Proceedings of the IGARSS 2024—2024 IEEE International Geoscience and Remote Sensing Symposium; IEEE: Piscataway, NJ, USA, 2024; pp. 7410–7414.
20. Ghorbanzadeh, O.; Xu, Y.; Ghamisi, P.; Kopp, M.; Kreil, D. Landslide4sense: Reference benchmark data and deep learning models for landslide detection. *arXiv* **2022**, arXiv:2206.00515.
21. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 5998–6008.
22. 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.* **2017**, *40*, 834–848. [\[CrossRef\]](#) [\[PubMed\]](#)

23. Saleem, N.; Huq, M.E.; Twumasi, N.Y.D.; Javed, A.; Sajjad, A. Parameters derived from and/or used with digital elevation models (DEMs) for landslide susceptibility mapping and landslide risk assessment: A review. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 545. [[CrossRef](#)]
24. Bilir, T.E.; Fung, I.; Dawson, T.E. Slope-aspect induced climate differences influence how water is exchanged between the land and atmosphere. *J. Geophys. Res. Biogeosciences* **2021**, *126*, e2020JG006027. [[CrossRef](#)]
25. Yang, J.; El-Kassaby, Y.A.; Guan, W. The effect of slope aspect on vegetation attributes in a mountainous dry valley, Southwest China. *Sci. Rep.* **2020**, *10*, 16465. [[CrossRef](#)] [[PubMed](#)]
26. Xie, E.; Wang, W.; Yu, Z.; Anandkumar, A.; Alvarez, J.M.; Luo, P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 12077–12090.
27. Zheng, S.; Lu, J.; Zhao, H.; Zhu, X.; Luo, Z.; Wang, Y.; Fu, Y.; Feng, J.; Xiang, T.; Torr, P.H.; et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Nashville, TN, USA, 20–25 June 2021; pp. 6881–6890.
28. Liu, Y.; Bai, X.; Wang, J.; Li, G.; Li, J.; Lv, Z. Image semantic segmentation approach based on DeepLabV3 plus network with an attention mechanism. *Eng. Appl. Artif. Intell.* **2024**, *127*, 107260. [[CrossRef](#)]
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
30. Song, Y.; Hao, L.; Li, W. Landslide detection using deep learning on remotely sensed images. *Appl. Comput. Geosci.* **2025**, *27*, 100278. [[CrossRef](#)]
31. Bazi, Y.; Bashmal, L.; Rahhal, M.M.A.; Dayil, R.A.; Ajlan, N.A. Vision transformers for remote sensing image classification. *Remote. Sens.* **2021**, *13*, 516. [[CrossRef](#)]
32. Zhang, Y.; Zhang, S.; Li, Y.; Zhang, Y. Coarse-to-fine satellite images change detection framework via boundary-aware attentive network. *Sensors* **2020**, *20*, 6735. [[CrossRef](#)] [[PubMed](#)]
33. Shi, B.; Guo, H.; Sun, Y.; Long, J.; Yang, L.; Zhou, Y.; Jiao, J.; Zhou, J.; He, Y.; Li, H. BiFusion-LDSeg: A Latent Diffusion Framework with Bi-Directional Attention Fusion for Landslide Segmentation in Satellite Imagery. *Remote. Sens.* **2026**, *18*, 719. [[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.