remote sensing

MDPI

Article

SC-Net: Structural Constrained Contrastive Learning for Landslide Extraction Toward Power Transmission Corridor Safety Monitoring

Wei Song [1], Shilian Liu [1], Shun Wu [2], Cheng Liao [3], Zongyuan Wu [4], Shiming Li [1,*], Xiaobin Zheng [2] and Yanping Duan [1]

[1] College of Surveying and Geo-Informatics, North China University of Water Resources and Electric Power, Zhengzhou 450045, China
[2] State Grid Jibei Electric Power Co., Ltd., Beijing 100053, China
[3] State Key Laboratory of Intelligent Geotechnics and Tunnelling (FSDI), Xi'an 710043, China
[4] School of Civil Engineering and Transportation, North China University of Water Resources and Electric Power, Zhengzhou 450045, China
* Correspondence: lishiming@ncwu.edu.cn

Highlights

What are the main findings?

- The proposed SC-Net effectively captures both structural and semantic characteristics of landslides, achieving accurate and spatially consistent extraction results in complex, vegetation-covered scenes.
- The integration of object-level consistency learning with pixel-level supervision improves feature discriminability and yields superior performance across UAV and satellite datasets.

What are the implications of the main findings?

- The proposed framework provides a reliable and scalable solution for landslide detection from multi-source remote sensing imagery, supporting large-area and cross-platform applications.
- The improved accuracy and robustness of landslide delineation offer practical value for infrastructure safety monitoring, particularly in power transmission corridor inspection and risk assessment.

Abstract

Landslides are among the most common and destructive geological hazards and pose a significant threat to the long-term stability of infrastructure systems. In particular, long-distance power transmission corridors often traverse mountainous and forested regions, where landslides can endanger tower foundations and transmission line safety. Such landslides predominantly occur in sloped forested areas, where dense vegetation causes severe occlusion that blurs landslide boundaries and creates strong visual similarity with surrounding land covers. Consequently, accurate and efficient landslide identification from remote sensing imagery remains a significant challenge. To address these challenges, we propose a structural constrained contrastive learning network (SC-Net) for reliable landslide extraction from remote sensing images. First, a multi-structural feature extraction module is designed to capture landslide-specific geometric characteristics. These features are further enhanced by fusing multi-scale semantic representations extracted from a pretrained backbone network through an attention-based adaptive feature fusion module. Additionally, a mask-constrained object-level contrastive learning strategy is introduced

Check for updates

Academic Editor: Gino Dardanelli

Received: 27 May 2026
Revised: 28 June 2026
Accepted: 29 June 2026
Published: 6 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.

to enforce global structural consistency at the landslide object-level, thereby improving the discriminability between landslide and non-landslide regions. Extensive experiments conducted on the publicly available CAS landslide dataset demonstrate the effectiveness of the proposed method. The proposed SC-Net achieves IoU scores of 89.89% and 79.76% on the CAS-UAV and CAS-SAT datasets, respectively, outperforming the best-performing baseline by 2.09% and 0.46%. The proposed method provides an effective solution for large-scale landslide monitoring and demonstrates potential for applications in power transmission corridor inspection and infrastructure safety assessment.

Keywords: landslide extraction; remote sensing imagery; multi-structural feature enhancement; object-level contrastive optimization

1. Introduction

Landslides are among the most frequent and destructive geological hazards, characterized by high occurrence rates and strong suddenness, and have caused severe casualties and economic losses worldwide [1]. In recent years, with the rapid expansion of long-distance power transmission networks, a large number of transmission lines and towers have been deployed across mountainous and densely vegetated regions. In such environments, landslides pose a critical threat to the structural stability of tower foundations and the safe operation of transmission corridors. Moreover, because landslides typically occur in rugged terrain with significant gradients, acquiring accurate post-disaster information rapidly is often difficult, posing substantial challenges for both emergency response and infrastructure safety management [2].

Achieving accurate and reliable landslide extraction in complex natural environments remains highly challenging [3,4]. This is especially critical for long-distance power transmission corridors, which traverse narrow mountainous and canyon areas, where landslides near towers or along lines directly endanger safety [5]. On the one hand, landslide regions exhibit considerable variations in scale, appearance, and spatial distribution, and are frequently affected by vegetation occlusion, resulting in blurred boundaries. In transmission corridors, the elongated right-of-way and rugged terrain further aggravate multi-scale landslide detection and incomplete boundary observations. On the other hand, landslides often show strong visual similarity to bare soil or farmland, leading to frequent misclassification. For power infrastructure, such confusion directly threatens tower foundations and lines, making reliable discrimination an even more pressing need.

Driven by advances in data-driven learning paradigms, particularly deep convolutional neural networks (DCNNs), automatic interpretation of remote sensing imagery has achieved remarkable success. Fully Convolutional Networks [6] and their variants have significantly improved the efficiency and flexibility of landslide detection by learning discriminative features from large-scale annotated datasets. Most existing methods adopt encoder–decoder architectures to extract and fuse multi-scale semantic features for landslide segmentation. Recent developments, including attention mechanisms, Transformer-based architectures, and visual foundation models, have further enhanced recognition performance [7,8]. Nevertheless, several critical challenges remain insufficiently addressed:

(1) Insufficient exploitation of landslide-specific fine structural features. Landslides typically occur on sloped terrains, and their surfaces often exhibit distinctive fine-grained structural textures formed during mass movement processes. These structural patterns provide more robust cues for distinguishing landslides from visually similar objects such as cultivated land or bare soil. However, most existing methods primarily

rely on pretrained encoders to extract multi-scale semantic features or introduce digital elevation models (DEMs) to enhance performance [9,10]. In practice, high-quality DEMs that are spatially and temporally aligned with optical imagery are often difficult to obtain, limiting the applicability of such approaches.

(2) Inadequate exploration of global feature consistency among landslide objects. In remote sensing imagery, landslides often exhibit distinguishable feature characteristics from surrounding backgrounds, especially in complex mountainous areas crossed by power transmission lines. Current methods are predominantly optimized under pixel-level supervision [11], which neglects the latent global similarity shared among different landslide instances while overlooking their inherent distinction from background objects. Explicitly modeling object-level global consistency among landslide regions has the potential to enhance feature discriminability and improve the robustness and generalization capability of segmentation models.

In summary, although existing methods have achieved substantial progress in landslide extraction, most approaches either focus primarily on pixel-level semantic learning or rely on auxiliary data and large-scale annotations. The intrinsic structural characteristics of landslides and their object-level global consistency remain insufficiently explored. These limitations motivate the development of a unified framework that simultaneously models landslide-specific structural features and exploits object-level contrastive constraints to achieve reliable and robust landslide extraction.

To address the above limitations, we propose a structural constrained contrastive learning network for reliable landslide extraction from remote sensing images. Specifically, a multi-structural feature extraction module is designed to capture landslide-specific structural characteristics, which are further enhanced through adaptive fusion with multi-scale semantic features extracted from the original images using attention mechanisms. Moreover, an object-level contrastive learning strategy guided by landslide masks is introduced to explicitly exploit global structural constraints, thereby strengthening feature separability between landslide and non-landslide regions to reduce false alarms along transmission line corridors.

The main contributions of this work can be summarized as follows:

(1) A multi-branch landslide extraction architecture is proposed. The proposed SC-Net simultaneously extracts multi-scale semantic features from original images and landslide-specific structural features through dedicated branches, and integrates additional prior supervisory information, resulting in improved landslide extraction performance.

(2) An attention-based adaptive feature fusion module is designed. The proposed module effectively optimizes and fuses multi-scale semantic and structural features, improving the integrity and consistency of landslide detection results across varying spatial scales.

(3) An object-level contrastive learning strategy is introduced. By jointly optimizing object-level contrastive loss and pixel-level segmentation loss, the proposed method learns latent global consistency features of landslides, thereby enhancing model robustness and transferability to facilitate more reliable identification of landslide threats along long-distance transmission networks.

The remainder of this paper is organized as follows. Section 2 reviews related work on landslide detection and deep learning-based remote sensing interpretation. Section 3 presents the proposed SC-Net in detail. Section 4 reports comprehensive experimental evaluations and analyzes the results. Finally, Section 5 concludes the paper and discusses future research directions.

2. Related Works

With the rapid development of deep learning, Convolutional Neural Network (CNN)-based methods have become the mainstream approach for automatic landslide extraction from remote sensing imagery, enabling efficient and accurate identification of landslide regions [2,5]. To cope with large-scale variations and complex background interference commonly observed in landslide scenes, many studies have incorporated attention mechanisms [12,13] into CNN architectures to enhance feature discrimination. For instance, ref. [14] proposed an attention-enhanced CNN for landslide detection in high-resolution optical satellite imagery, improving robustness under heterogeneous surface conditions. Similarly, ref. [15] improved the U-Net [16] architecture by introducing a RegNet-based encoder with attention modules, achieving competitive accuracy with reduced model parameters and computational cost without relying on multi-temporal or geological auxiliary data. To further enhance model generalization across different regions and imaging conditions, ref. [17] proposed a cross-domain landslide extraction framework that integrates pixel-level masking with a morphological information module to generate pseudo-labels for the target domain, thereby alleviating domain shift effects.

Balancing detection accuracy and computational efficiency is particularly critical for large-scale mapping and real-time landslide monitoring applications. To this end, ref. [18] proposed BisDeNet, a lightweight framework that replaces the BiSeNet context path with DenseNet to reduce model complexity while preserving effective feature representation. For onboard and real-time scenarios, ref. [19] developed a lightweight landslide detection network deployed on a QK-series satellite, improving in-orbit processing efficiency under complex surface conditions. In addition, ref. [20] introduced Re-Net, a structurally reparameterized U-shaped network that supports multi-scale feature learning during training while substantially reducing parameter count and computational cost during inference, without sacrificing segmentation accuracy. Although these methods demonstrate the feasibility of efficient landslide extraction, they often emphasize computational efficiency at the expense of detailed structural representation.

Improving the completeness and boundary accuracy of landslide segmentation has therefore become a key research focus. To enhance contextual modeling, ref. [7] proposed Attention U-Mamba, which integrates CNN-based local feature extraction with Mamba-based global context modeling, achieving improved segmentation performance with fewer parameters. Similarly, ref. [21] introduced ARU-Net, incorporating attention mechanisms and residual connections into U-Net, together with binary focal loss to address class imbalance and complex landslide samples. In addition, ref. [14] embedded squeeze-and-excitation (SE) modules into an attention U-Net framework to improve landslide extraction performance on Sentinel-2A imagery. To mitigate coarse boundary delineation, ref. [22] optimized landslide segmentation using multi-resolution feature fusion and selective kernel attention to integrate coarse-to-fine features, effectively capturing landslide geometry and spatial distribution. Contrastive learning has also been explored for boundary refinement: ref. [23] proposed a contrastive learning based framework employing residual blocks, channel attention, and a contrastive Dice loss to improve boundary precision, while ref. [24] introduced AMU-Net, which integrates attention and multi-scale modules with a shifted window strategy to enlarge the receptive field and reduce boundary errors. Further improvements in boundary preservation across different segmentation architectures were reported in [25].

Due to the complex shapes, textures, and spectral characteristics of landslides, distinguishing landslide regions from surrounding land-cover types remains challenging. To enhance feature discriminability, ref. [26] proposed AST-UNet, which incorporates a channel and spatial interaction module into SwinUNet to improve segmentation complete-

ness. Ref. [27] introduced FFS-Net, which jointly extracts texture and shape features from optical imagery and terrain features from DEM data using transposed convolution and multi-scale channel attention. To exploit complementary information from different data sources, multi-modal fusion strategies have also been widely investigated. For example, ref. [28] proposed a dual-channel framework that combines EfficientNetB7 with spatial attention for optical imagery and a depthwise separable CNN with Transformer modules for DEM features. While integrating terrain and spectral information can improve landslide discrimination in complex environments, the performance of such methods strongly depends on the availability and quality of DEM data, which is often difficult to obtain in practice. In addition, ref. [29] developed LandslideNet for accurate landslide segmentation from single-temporal optical images, achieving high-precision identification without auxiliary data. Ref. [30] further integrated generative AI with pixel-level segmentation to enhance image quality and localization accuracy, while ref. [31] proposed GMNet, which exploits global contextual information and fuses low-level multi-scale features to improve discrimination between landslides and non-landslide objects.

Despite these advances, supervised learning-based methods generally rely on large-scale, high-quality annotated datasets, resulting in high labeling costs and limited generalization capability. To alleviate this dependence, ref. [32] proposed a weakly supervised approach that combines class activation maps with CycleGAN, enabling pixel-level landslide segmentation using only image-level labels. To address data scarcity, ref. [33] introduced a StyleGAN2 Transformer framework that generates high-quality synthetic landslide samples, improving the segmentation accuracy of CNN- and Transformer-based models by more than 10%. Domain adaptation techniques have also been explored to enhance cross-domain robustness. For instance, ref. [34] proposed a progressive domain adaptation framework that integrates near-to-far adaptation, adversarial learning, and prototype consistency. Alternatively, ref. [35] introduced a lightweight universal adapter that enables transferable landslide mapping with minimal parameter updates. Ref. [36] proposed DS-Net, which dynamically aligns local details with global context while incorporating domain knowledge through a multi-attention prior module. Model robustness was further enhanced in [11] through virtual adversarial training and probability-constrained pixel-level contrastive learning.

More recently, large-scale pre-trained foundation models have been introduced for landslide extraction to further improve cross-domain generalization. Ref. [37] adopted a dual-branch encoder architecture to extract optical and terrain features, leveraging the strong representation capability of the Segment Anything Model (SAM). Building on this, ref. [38] proposed TransLandSeg, which freezes SAM's image encoder and trains only lightweight adaptive modules, achieving effective landslide segmentation with low training cost. Ref. [39] further integrated SAM's image encoder with a cross-feature fusion decoder and a shallow feature extractor, producing high-precision landslide maps without requiring user prompts.

Considering the importance of subtle structural differences in landslide regions, recent studies have increasingly focused on enhancing high-frequency and object-level representations. To address the neglect of high-frequency information, ref. [40] incorporated guided attention mechanisms into skip connections to preserve structural details. Since pixel-level optimization often overlooks object-level characteristics, ref. [41] combined ResU-Net with rule-based object-based image analysis, integrating pixel-wise probabilities with object-level rules. Similarly, refs. [42,43] leveraged object-level priors through bounding box prompts and instance-aware architectures, improving boundary delineation under limited training samples. More recently, ref. [44] proposed an iterative classification-segmentation

framework with object-level contrastive learning, enhancing both boundary representation and global feature consistency.

3. Methods

3.1. Overview of the SC-Net

Figure 1 illustrates the overall architecture of the proposed SC-Net. The framework mainly consists of three components: (1) multi-structural feature extraction from remote sensing images, (2) multi-scale fusion of semantic and structural features, and (3) global feature optimization guided by landslide object annotations. By fully exploiting the subtle structural characteristics unique to landslides in remote sensing imagery and exploring global feature consistency at the object level, SC-Net enhances the discriminability between landslides and complex background regions, thereby enabling more reliable landslide extraction.

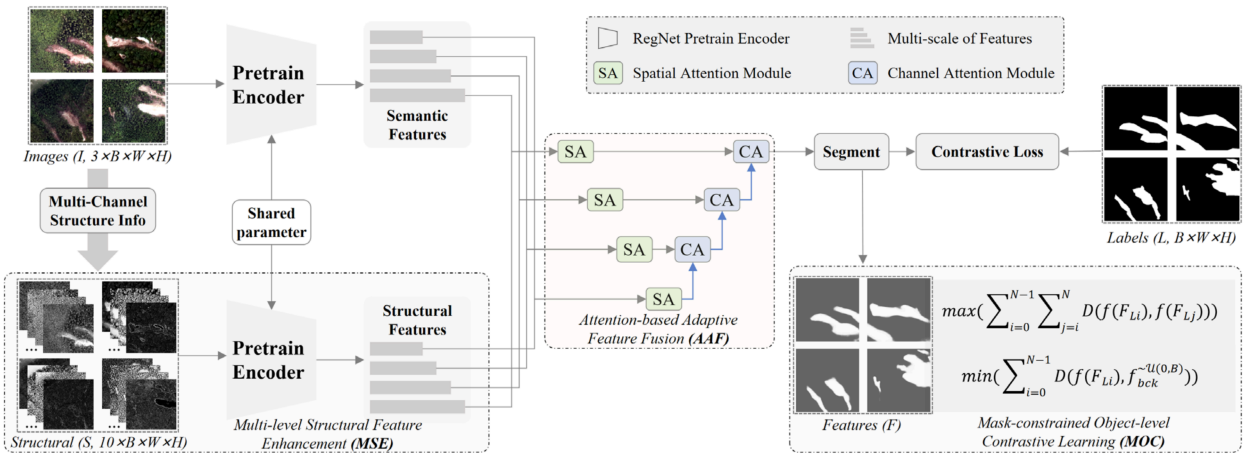


Figure 1. Overview of the proposed SC-Net.

Specifically, SC-Net employs a parameter-shared, pre-trained encoder to extract multi-scale semantic and structural features from the original images and their corresponding multi-channel structural representations. To further strengthen landslide-specific details, an attention-based adaptive fusion module is introduced to effectively integrate semantic and structural features. Moreover, a landslide mask-based object-level contrastive learning optimization strategy is designed to exploit potential object-level global constraints, promoting intra-class feature consistency among landslide objects and inter-class feature separability between landslide and non-landslide regions.

3.2. Multi-Structural Feature Enhancement

To address the challenges of blurred landslide boundaries and their visual similarity to other land-cover types, such as bare soil and plowed farmland, existing methods commonly improve detection reliability by fusing optical imagery with DEM data. However, DEM data that are temporally and spatially consistent with remote sensing images are often difficult to acquire, especially for long-distance power transmission corridors spanning hundreds of kilometers, where high-resolution, up-to-date DEMs are rarely available or prohibitively expensive, limiting the practical applicability of DEM-aided methods in real-world landslide monitoring.

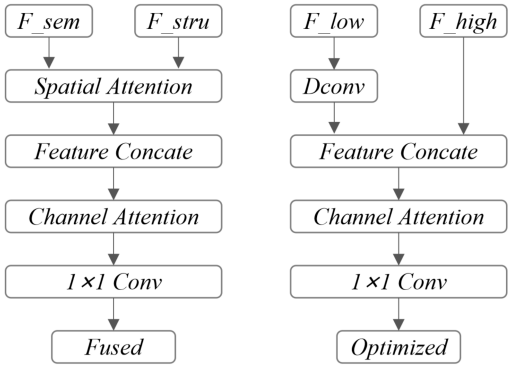
To alleviate this limitation, we design a multi-structural feature extraction module that enhances landslide discriminability by exploiting fine-grained structural characteristics, thereby improving model reliability without requiring any auxiliary data. Specifically, multiple image feature operators are employed to construct multi-channel structural rep-

resentations. These operators include histogram equalization (Hist_eq), denoised raw image (Denoised_raw), denoised after equalization (Denoised_eq), adaptive Gaussian thresholding (Adaptive_th), Sobel gradient (Sobel_grad), Gaussian low-pass filtering (Gaussian_low), Gaussian high-pass filtering (Gaussian_high), Laplacian sharpening convolution (Conv_sharpen), Hough transform-based edge detection (Hough_edge), and Laplacian zero-crossing edge detection (Laplacian_edge). Each operator is designed to extract complementary structural cues from the input image, as summarized in Table 1. They capture multi-scale, multi-directional edge, corner, and texture features that are sensitive to landslide boundaries and surface roughness. The resulting multi-channel structural representations are stacked to form a multi-channel structural S, $S \in \mathbb{R}^{W \times H \times 10}$. Subsequently, to let the subsequent convolutional layers learn optimal fusion, a multi-structural feature encoding module is developed. The stacked structural tensor is first transformed into a compact representation S1, $S \in \mathbb{R}^{W \times H \times 3}$ using convolutional layers with 3×3 and 1×1 kernels. A parameter-shared, pre-trained encoder is then employed to extract multi-scale structural features with higher spatial and channel dimensions. The semantic features extracted from the original images and the structural features derived from the multi-channel representations are generated at four scales and share identical spatial resolutions.

Table 1. Ten structural prior descriptors and their functional characteristics.

Operator	Description
Hist_eq	Enhances global contrast by flattening and spreading the intensity histogram.
Denoised_raw	Removes noise while preserving fine textures using non-local means on the original image.
Denoised_eq	Reduces amplified noise after histogram equalization using a larger search window.
Adaptive_th	Converts local image regions to binary using Gaussian-weighted adaptive thresholding.
Sobel_grad	Computes the first-order derivative magnitude to detect strong edges and intensity changes.
Gaussian_low	Retains global structure and illumination by suppressing high-frequency details with Gaussian blur.
Gaussian_high	Extracts fine textures and sharp edges as the residual between the original and blurred images.
Conv_sharpen	Enhances edge contrast by applying a Laplacian-based sharpening convolution kernel.
Hough_edge	Detects parametric geometric shapes such as lines and circles through Hough space voting.
Laplacian_edge	Identifies zero-crossings of the second derivative for precise, isotropic edge localization.

During feature fusion, a spatial attention (SA) block is first applied to model long-range spatial dependencies, followed by a channel attention (CA) block to achieve bottom-up adaptive feature weighting and fusion across different scales, as illustrated in Figure 2. The introduction of multi-structural features helps preserve fine-grained spatial structural characteristics of landslide regions while maintaining a lightweight parameter configuration, thereby providing potential support for reliable landslide recognition.



(a) SA block (b) CA block

Figure 2. Detailed architecture of the attention-based adaptive fusion (AAF) module, including the cascaded SA and CA blocks.

3.3. Mask Constraint Contrastive Learning Optimization

Most existing landslide recognition methods optimize models by minimizing the pixel-wise discrepancy between predicted landslide regions and ground-truth labels, typically using the binary cross-entropy (BCE) loss for pixel-level supervision. Although effective for local classification, this paradigm neglects the global feature consistency constraints that naturally exist among landslide objects as a specific type of geological hazard, making the learned representations vulnerable to background noise in complex environments. Consequently, pixel-wise supervision alone cannot simultaneously enforce feature compactness across different landslide objects and feature contrast between landslides and the background, both of which are essential for stable recognition in geologically complex scenes. In addition, by incorporating image structural features, landslides can be further differentiated from similar land cover types (e.g., farmland). These observations motivate extending the contrastive loss function to the object level, where global feature descriptors help mitigate noise interference in complex environments.

To address this limitation, we introduce an object-level contrastive learning loss built upon the global feature similarity among landslide instances, which is integrated with the conventional BCE loss. This design explicitly exploits the latent global structural consistency shared by different landslide objects, suppresses interference from confusing backgrounds (e.g., bare soil and farmland), and improves the stability and discriminability of the learned feature representations.

Specifically, for each training batch, let the ground-truth landslide labels be denoted as $L \in \mathbb{R}^{w \times h}$, the predicted probability maps as $P \in \mathbb{R}^{w \times h}$, and the corresponding deep feature maps as $F \in \mathbb{R}^{w \times h \times c}$, where w and h denote the spatial resolution and c is the number of feature channels. The batch size is B. Based on the landslide masks l_n extracted from L, the corresponding landslide instance features f_n are cropped from F, where $n \in \mathrm{N}$ indexes all landslide instances in the batch. Similarly, for each sample in the batch, the background features are extracted from the corresponding feature map using the complement of the landslide mask.

For each landslide and background instance, global average pooling is applied to obtain an instance-level feature embedding $\mathrm{pos_n}$ and f^b_{bck}, respectively. The global average pooling is calculated as shown in Equation (1), where x and y denote the spatial coordinate indices of pixels falling within the landslide regions in the ground-truth annotations of the training samples. For all landslide instances within a batch, the pairwise cosine similarity is computed between their embeddings to measure the global feature consistency between different landslide objects, denoted as d. To further enforce feature discrimination between landslide and background features, we additionally compute the

distance between each landslide embedding and randomly selected background embeddings within the batch. The resulting object-level contrastive loss $\mathcal{L}_{\text{con}}$ is formulated as shown in Equations (1) and (2), encouraging features of different landslide instances to be highly consistent in the embedding space while simultaneously pulling landslide features away from background features.

$$pos_n = \frac{1}{\sum_{x,y} l_n(x,y)} \sum_{x,y} l_n(x,y) \cdot f_n(x,y) \tag{1}$$

$$\mathcal{L}_{con} = (1 - mean(\sum_{i=n}^{N-1} \sum_{j=i}^{N} d(pos_i, pos_j))) + mean(\sum_{i=n}^{N-1} d(pos_i, f_{bck}^{\sim \mathcal{U}(0,B)})) \tag{2}$$

$$\mathcal{L}_{bce} = mean(\sum_{b=0}^{B} [-(l_b * log(p_b) + (1 - l_b) * log(1 - p_b))]) \tag{3}$$

$$\mathcal{L} = \mathcal{L}_{con} + \mathcal{L}_{bce} \tag{4}$$

Finally, the overall training objective $\mathcal{L}$ is defined as the combination of the contrastive loss and the binary cross-entropy loss $\mathcal{L}_{\text{bce}}$ as expressed in Equations (3) and (4). By jointly enforcing pixel-level classification accuracy and object-level global feature consistency, the proposed loss function effectively reduces the influence of complex background noise and leads to more robust and discriminative landslide representations.

3.4. Evaluation Metrics and Hyperparameters

Since the proposed landslide extraction method follows a classical binary semantic segmentation paradigm, standard pixel-level evaluation metrics are adopted for quantitative performance assessment. Specifically, true positives (TP), false positives (FP), and false negatives (FN) are computed by comparing the predicted results with the ground truth, where TP denotes pixels correctly classified as landslides, FP represents background pixels incorrectly labeled as landslides, and FN corresponds to landslide pixels misclassified as background. Based on these quantities, Precision, Recall, Intersection over Union (IoU), and F1-score are calculated according to Equations (5)–(8) to comprehensively evaluate segmentation accuracy.

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$IoU = \frac{TP}{TP + FP + FN} \tag{7}$$

$$F1\_Score = \frac{2 * TP}{2 * TP + FP + FN} \tag{8}$$

All experiments were implemented using PyTorch 1.13.1 and conducted on a single NVIDIA RTX A6000 GPU. Model optimization was performed using stochastic gradient descent with an initial learning rate of 0.01. During training, a RegNet64 model pretrained on ImageNet was adopted as the backbone, and extensive data augmentation strategies, including random flipping, rotation, and scaling, were applied to improve model robustness and generalization performance. The detailed configuration parameters and source code are publicly available at https://github.com/LiaoChengCRCC/SC-Net (accessed on 27 June 2026).

4. Experiments and Analysis

4.1. Datasets Description

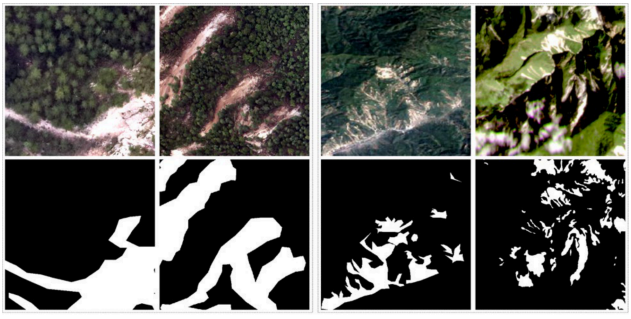
Due to the absence of publicly available landslide datasets tailored for long-distance power transmission corridor applications, the CAS dataset [45] is therefore adopted as a representative benchmark for evaluating model performance. The CAS landslide extraction dataset, including the UAV and AST subdataset, is a large-scale, multisensor benchmark dataset for deep learning-based landslide identification developed by the Artificial Intelligence Group at the Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, comprising 20,865 georeferenced RGB image patches of 512 × 512 pixels acquired from satellite (e.g., Sentinel-2 and Landsat) and unmanned aerial vehicle (UAV) imagery across nine geographically diverse regions with corresponding pixel-wise annotations and rigorous quality control procedures, explicitly addressing challenges from cloud occlusion, seam artifacts, and label misalignment, to support reliable training and evaluation under diverse terrain and environmental conditions. The data cover the following regions: Tiburon Peninsula, Hokkaido Iburi-Tobu, Palu, Lombok, Mengdong Township, Jiuzhai Valley, Longxi River, Luding, and Wenchuan.

Importantly, the landslides contained in the CAS dataset are primarily earthquake-induced and rainfall-induced landslides, which constitute common and representative types of landslide disasters in mountainous regions. These landslides exhibit substantial variability in morphology, scale, vegetation recovery stage, and background complexity, reflecting diverse post-event surface conditions under different triggering mechanisms. Therefore, the dataset provides a comprehensive and challenging benchmark for evaluating landslide detection models and offers a reasonable approximation of the environmental conditions encountered in long-distance power transmission corridor scenarios.

Further detailed comparisons of these datasets are shown in Table 2. Due to the diversity of image acquisition sources, spatial resolutions vary substantially across the dataset, which poses additional challenges for robust and generalized landslide extraction. Sample examples from these datasets are shown in Figure 3. In our experiments, following the same data split strategy as the original CAS dataset [45], we randomly divided the CAS-UAV and CAS-SAT datasets into training and testing sets with a ratio of 7:3, respectively.

Table 2. Detailed comparison of the CAS-UAV and CAS-SAT datasets.

Datasets	Source	Pixel_Size	Resolution	Tiles
CAS-UAV	UAV	512	0.2~1 m	13,443
CAS-SAT	Satellite	512	0.5~5 m	7422



(a) CAS-UAV (b) CAS-SAT

Figure 3. Sample examples from the CAS-UAV and CAS-SAT landslide datasets, respectively.

4.2. Performance Comparison of Recent Related Methods

To validate the effectiveness of the proposed method, comparative experiments were conducted on the CAS-UAV and CAS-SAT datasets. There are two classical and widely

adopted semantic segmentation models, U-Net and DeepLabV3+ [44], were selected as comparison methods, based on the pretrained ResNet50 backbone. In addition, we compared our method with recent state-of-the-art landslide extraction approaches specifically designed for remote sensing on the CAS dataset, including MFFENet [43] and SegFormer [45]. We note that although numerous newer methods have been reported on the CAS landslide dataset, most of them use private data splits and have not released their source code, making it infeasible to reproduce their results under the same experimental setup. Therefore, the performance of these latest methods is quoted directly from their original publications, and the corresponding results are reported in Table 3 (CAS-UAV) and Table 4 (CAS-SAT). The pre-training backbone used for each method is specified in parentheses after the corresponding method name (e.g., ResNet50).

Table 3. Comparison of experimental results on the CAS-UAV test dataset.

Methods	IoU	F1_Score	Precision	Recall
U-Net (ResNet50)	84.82	91.79	91.78	91.80
DeepLabV3+ (ResNet50)	87.51	93.34	92.76	93.92
MFFENet (2024) [43]	84.40	-	89.33	93.84
Segformer (2025) [45]	87.80	-	-	-
SC-Net (RegNet64)	89.89	94.67	94.82	94.51

Table 4. Comparison of experimental results on the CAS-SAT test dataset.

Methods	IoU	F1_Score	Precision	Recall
U-Net (ResNet50)	75.69	86.17	84.39	88.02
DeepLabV3+ (ResNet50)	77.68	87.44	86.90	87.99
MFFENet (2024) [43]	68.00	-	74.14	89.14
Segformer (2025) [45]	79.30	-	-	-
SC-Net (RegNet64)	79.76	88.74	87.77	89.74

The experimental results show that the proposed method consistently outperforms the classical U-Net and DeepLabV3+ models across all semantic segmentation evaluation metrics on both the high-resolution CAS-UAV dataset and the relatively low-resolution CAS-SAT dataset. Specifically, on CAS-UAV, our method achieves IoU improvements of 5.07% and 2.38% over U-Net and DeepLabV3+, respectively. On CAS-SAT, the corresponding IoU gains over U-Net and DeepLabV3+ are 4.07% and 2.08%, respectively. Furthermore, when excluding the performance differences introduced by pretrained models, the proposed method still attains IoU improvements of 2.87% and 3.76% over U-Net on the CAS-UAV and CAS-SAT datasets, respectively. Detailed experimental analyses are provided in Section 4.3.

More importantly, the proposed method also surpasses recent specialized landslide extraction approaches on both datasets. On the CAS-UAV dataset, both MFFENet and SegFormer adopt advanced Transformer-based architectures and multi-scale feature optimization strategies, such as attention mechanisms, which generally provide stronger robustness than conventional CNN-based models. Notably, MFFENet serves as a benchmark method for the CAS landslide dataset. Nevertheless, the proposed method achieves IoU improvements of 5.49% and 2.09% over MFFENet and SegFormer, respectively. On the CAS-SAT dataset, the proposed method further demonstrates superior performance, outperforming MFFENet and SegFormer by 11.76% and 0.46% in terms of IoU, respectively. These results indicate the effectiveness of the proposed approach across different spatial resolutions and sensing conditions.

For qualitative comparison of landslide extraction performance, Figures 4 and 5 present qualitative visualization results on the CAS-UAV and CAS-SAT datasets, respectively. Each row corresponds to a representative sample, and columns (a)–(e) illustrate the original image, the prediction results of U-Net, DeepLabV3+, the proposed method, and the ground-truth, respectively. As observed from the visual comparisons, U-Net frequently generates fragmented and discontinuous landslide regions, often missing small-scale landslides near the edges of sliding areas. DeepLabV3+, although achieving relatively smoother predictions, still exhibits obvious under-segmentation in regions with vegetation occlusion, where landslides are mistakenly merged with surrounding bare soil or shadows. In contrast, the proposed method tends to produce more complete landslide regions and more accurate boundary delineation, further demonstrating its advantage in complex landslide scenarios. Specifically, in areas with dense vegetation cover, where landslide boundaries are often blurred or partially occluded, our method maintains relatively stable detection performance, whereas the U-Net and DeepLabV3+ frequently produce fragmented or missing predictions. Note that the compared baselines employ varying pre-trained backbones, which may affect the comparison; we address this in Section 4.3 via ablation studies on the U-Net with controlled backbone configurations.

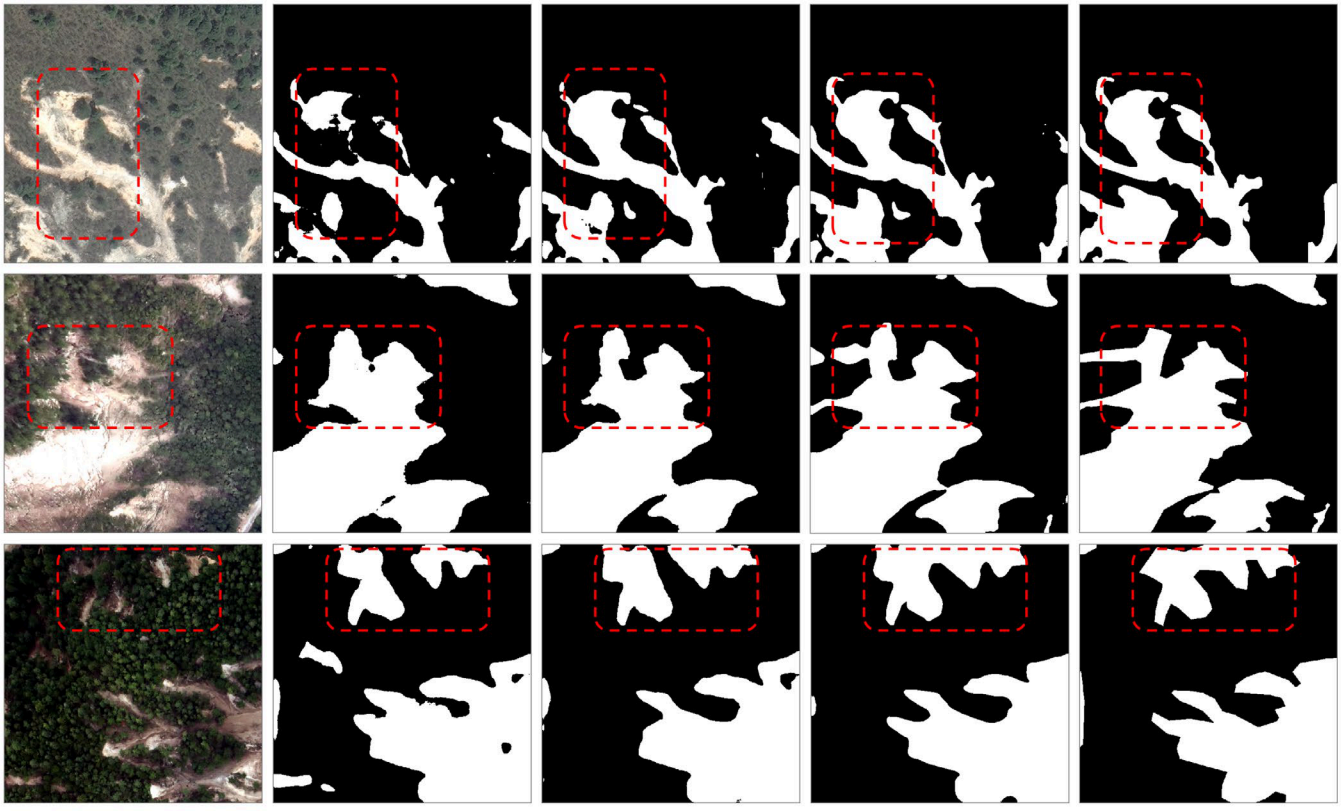


Figure 4. Qualitative comparison of landslide extraction results on the CAS-UAV test dataset. From left to right: (**a**) original images, (**b**) U-Net, (**c**) DeepLabV3+, (**d**) the proposed SC-Net, and (**e**) ground-truth labels.

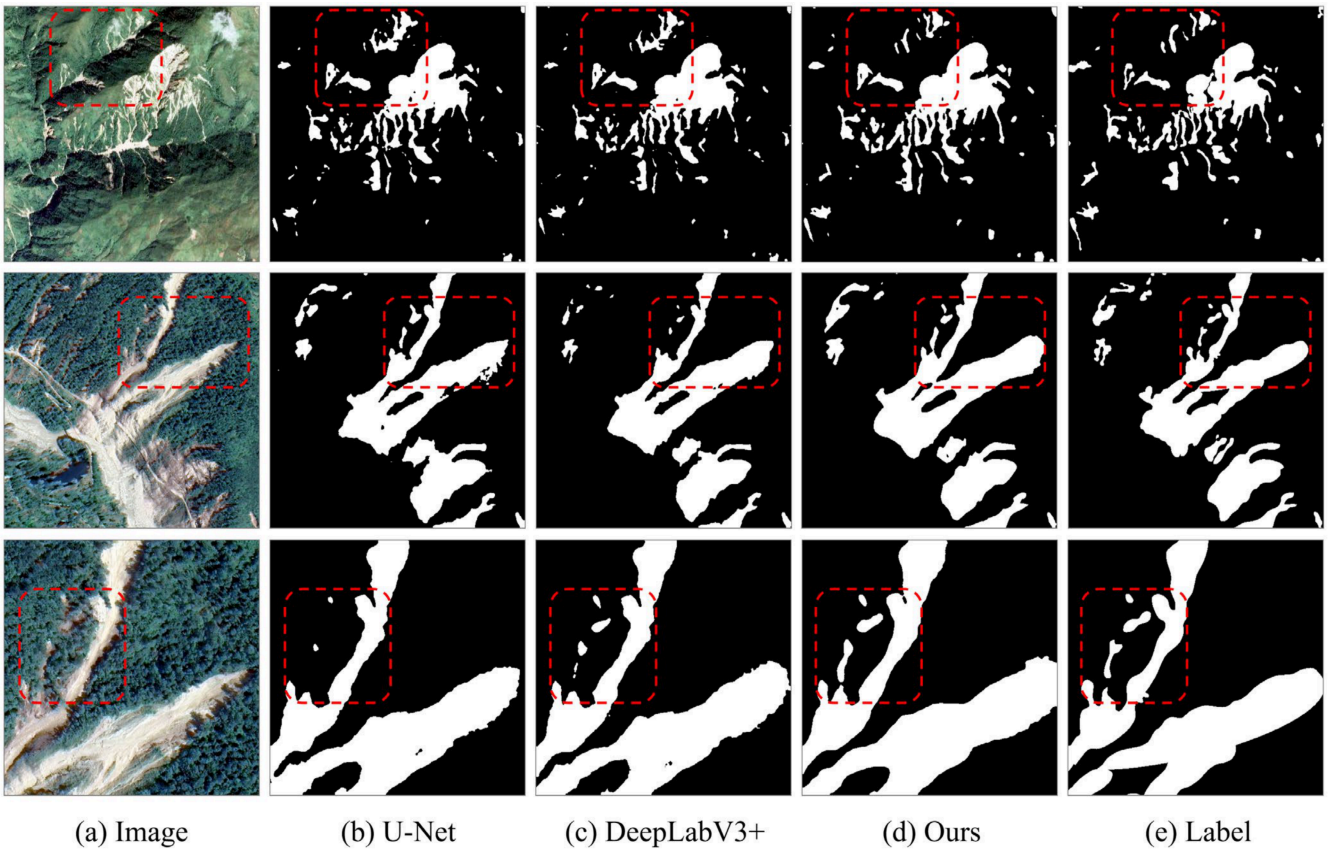


Figure 5. Qualitative comparison of landslide extraction results on the CAS-SAT test dataset. From left to right: (**a**) original images, (**b**) U-Net, (**c**) DeepLabV3+, (**d**) the proposed SC-Net, and (**e**) ground-truth labels.

To further validate the reliability of our proposed method for landslide hazard monitoring along power transmission corridors, we present qualitative identification results on representative samples that explicitly contain power transmission equipment (e.g., towers and nearby corridors) from the CAS-UAV and CAS-SAT datasets, as shown in Figures 6 and 7, respectively. In each figure, every row represents a representative sample, and the columns from left to right show: (a) original optical images, (b) landslide detection heatmaps generated by our SC-Net, (c) the final binary segmentation results, and (d) ground-truth landslide labels.

The visual comparisons reveal several important observations. First, although the spectral signatures of power transmission tower foundations and surrounding disturbed soils often exhibit strong similarity to those of landslide scars, particularly in mountainous environments where bare soil and shadowed areas are abundant, our method correctly distinguishes between them in most cases. As shown in the heatmap (column b), landslide regions exhibit high-confidence responses (bright white), while tower bases, access roads, and other man-made structures receive relatively low confidence scores, despite their visual resemblance. Second, beyond tower-related confusion, the proposed framework demonstrates enhanced ability to suppress false positives on other common distractors along transmission corridors, such as plowed farmland, bare soil patches, and unpaved road surfaces. These land-cover types often result in false-positive predictions, yet our approach produces fewer false-positive responses. Third, the qualitative results further demonstrate the robustness of the proposed framework under challenging conditions, including blurred landslide boundaries caused by vegetation occlusion, significant scale variations, and the presence of spectrally similar background features.

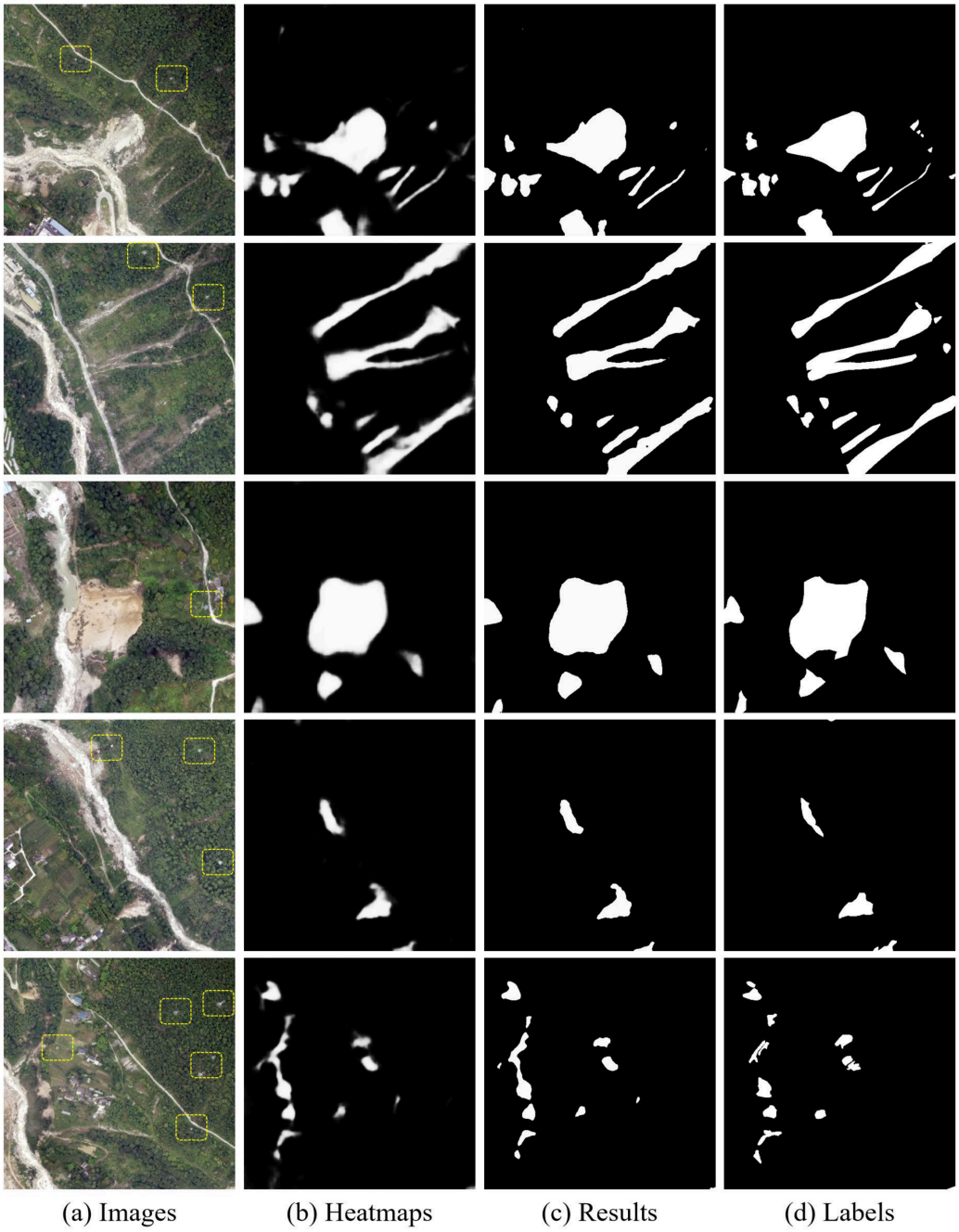


Figure 6. Qualitative results of landslide identification in power transmission scenarios on the CAS-UAV test dataset.

Nevertheless, some limitations remain. In cases where landslides are relatively small or partially obscured by dense vegetation, the model does not always delineate complete landslide boundaries. Additionally, when non-landslide regions share highly similar textures and colors with landslides, the model occasionally generates false positives, albeit less frequently than the baseline methods. Although such cases were observed only occasionally in the evaluated examples, they highlight opportunities for further improvement. Overall, these observations indicate that the proposed method has promising application potential for identifying landslides along power transmission corridors.

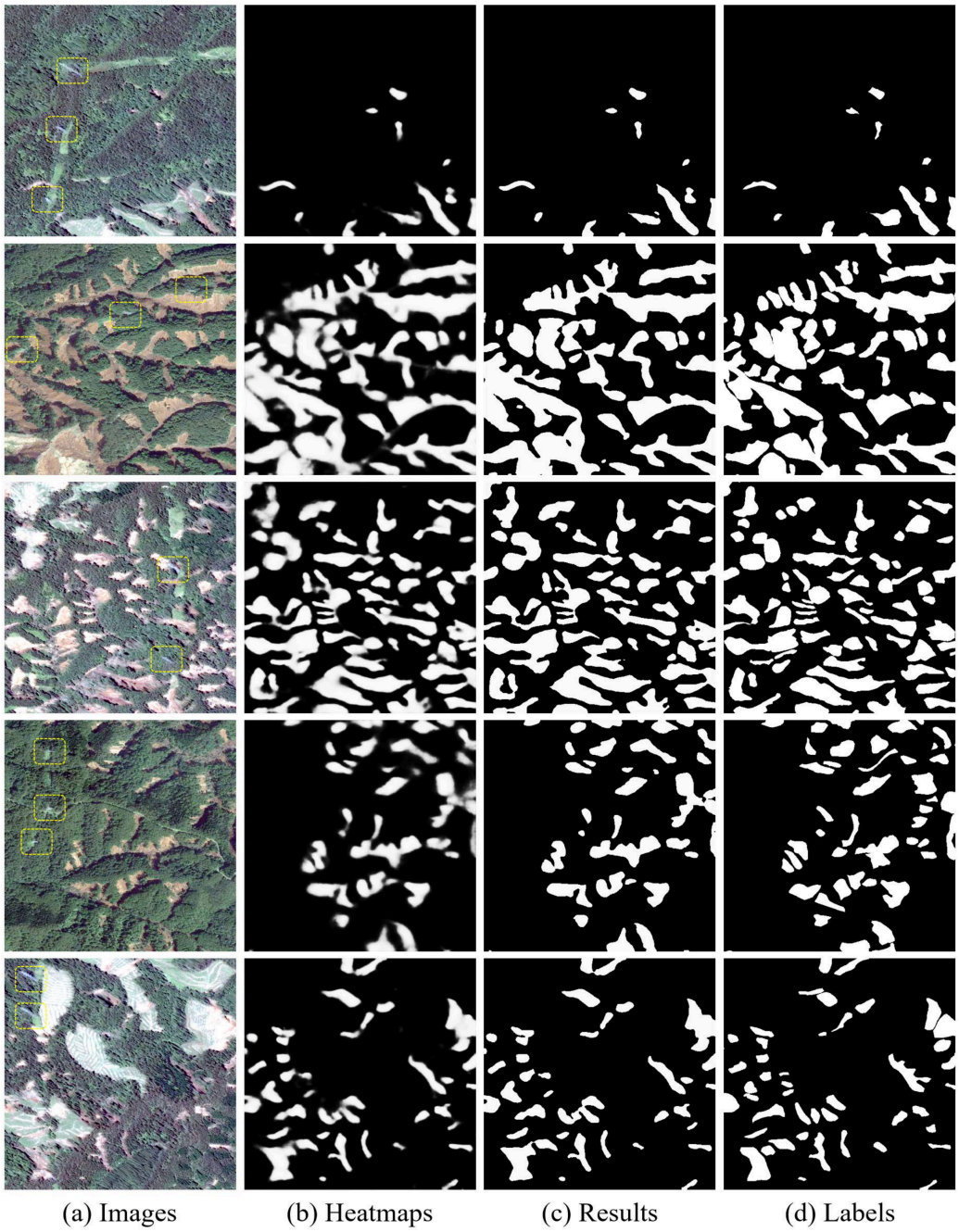


Figure 7. Qualitative results of landslide identification in power transmission scenarios on the CAS-SAT test dataset.

4.3. Ablation Experiments

The proposed SC-Net is built upon the U-Net architecture. As the baseline, a RegNet-pretrained backbone is introduced into U-Net and trained under binary cross-entropy (BCE) loss supervision for landslide extraction. Based on this baseline, different modules are progressively incorporated, including a data augmentation strategy (DA), a multi-level structural feature enhancement module (MSE), an attention-based adaptive feature fusion module (AAF), and a mask-constrained object-level contrastive learning strategy (MOC). The quantitative ablation results on the CAS-UAV and CAS-SAT datasets are reported in Tables 5 and 6, respectively.

Table 5. Results of the ablation experiment using the CAS-UAV dataset. It evaluated the data augmentation (DA) strategy, multi-level structural feature enhancement (MSE) module, attention-based adaptive feature fusion (AAF) module, and mask-constrained object-level contrastive learning (MOC) strategy, respectively.

Ablation Module	DA	MSE	AAF	MOC	F1 (%)	IoU (%)
Baseline (U-Net, RegNet64)					93.06	87.02
+DA	√				93.68	88.11
+MSE	√	√			94.14	88.93
+AAF	√	√	√		94.29	89.20
+MOC (Ours)	√	√	√	√	94.67	89.89

Table 6. Results of the ablation experiment using the CAS-SAT dataset. It evaluated the data augmentation (DA) strategy, multi-level structural feature enhancement (MSE) module, attention-based adaptive feature fusion (AAF) module, and mask-constrained object-level contrastive learning (MOC) strategy, respectively.

Ablation Module	DA	MSE	AAF	MOC	F1 (%)	IoU (%)
Baseline (U-Net, RegNet64)					86.37	76.00
+DA	√				87.10	77.14
+MSE	√	√			87.75	78.17
+AAF	√	√	√		88.40	79.21
+MOC (Ours)	√	√	√	√	88.74	79.76

Experimental results indicate that, compared with the conventional U-Net architecture pretrained on ResNet-50, our baseline model—which adopts a more powerful RegNetY-64 encoder pretrained on ImageNet—achieves IoU improvements of 2.20% and 0.31% on the CAS-UAV and CAS-SAT datasets, respectively. Building upon this stronger baseline, each proposed module is shown to contribute positively to landslide segmentation, collectively yielding an overall IoU improvement of 2.87% on CAS-UAV and 3.76% on CAS-SAT over the optimized baseline.

Specifically, the data augmentation strategy enhances sample diversity during training, yielding IoU gains over the baseline of 1.09% and 1.14% on the two datasets. Furthermore, the incorporation of multi-level structural features, which explicitly encode the spatial characteristics of landslide objects, brings additional IoU improvements of 0.82% and 1.03%, respectively. Furthermore, the attention-based adaptive feature fusion module consistently enhances recognition accuracy, achieving IoU improvements of 0.27% on CAS-UAV and 1.04% on CAS-SAT. Notably, the performance gain is more pronounced on the lower-resolution CAS-SAT dataset, indicating that this module effectively optimizes the fusion of structural and semantic features under challenging imaging conditions. Finally, incorporating the proposed global feature contrastive optimization strategy further boosts the recognition performance, resulting in additional IoU improvements of 0.69% and 0.55% on the CAS-UAV and CAS-SAT datasets, respectively.

The consistent improvement across all evaluation metrics in the ablation experiments confirms the effectiveness of each component in SC-Net. To provide a more intuitive comparison of the impact of different modules, Figures 8 and 9 present qualitative visualization results on the CAS-UAV and CAS-SAT datasets, respectively. In these figures, subfigures (a)–(h) correspond to the original image, the baseline prediction, the results obtained by progressively introducing DA, MSE, AAF, and MOC, the ground-truth landslide annotations, and the feature activation heatmaps produced by the proposed method. The visual comparisons further illustrate how each module incrementally enhances landslide completeness and boundary delineation.

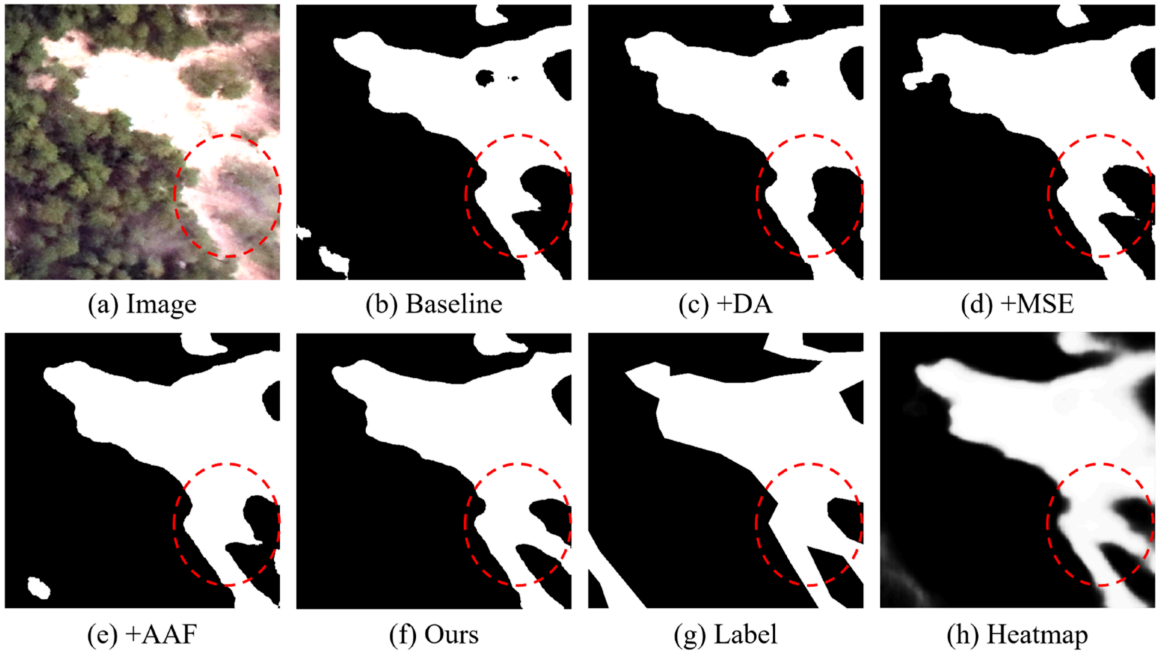


Figure 8. Comparison of the proposed method with each introduced module on the CAS-UAV dataset.

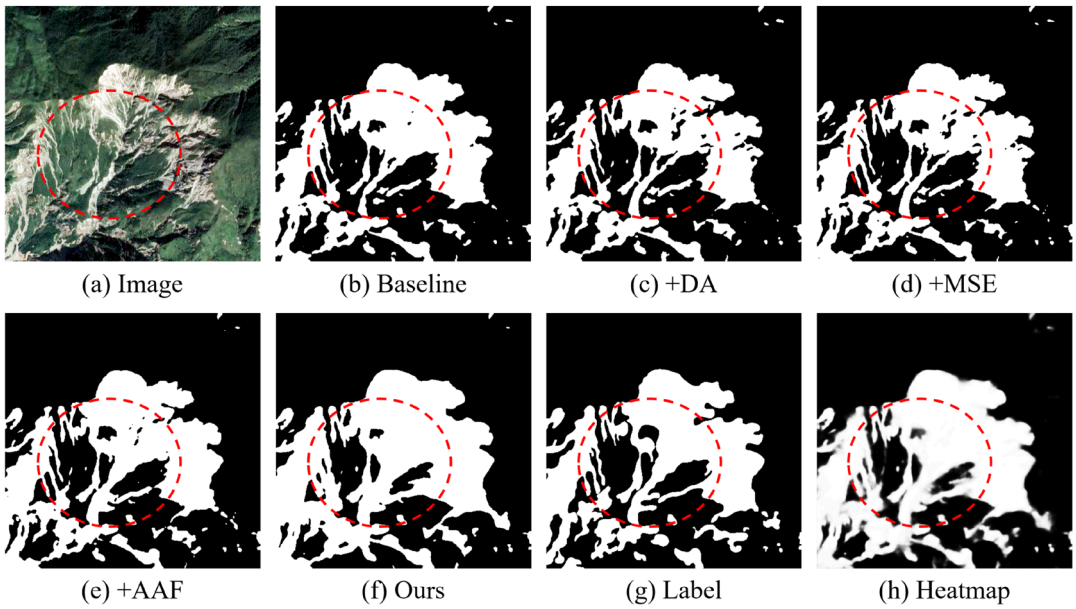


Figure 9. Comparison of the proposed method with each introduced module on the CAS-SAT dataset.

4.4. Efficiency Analyses

In addition to segmentation accuracy, computational efficiency is a critical factor for practical deployment, especially for large-scale landslide monitoring along power transmission corridors. To evaluate the model complexity and inference efficiency of the proposed SC-Net, we compared it with several representative methods, including U-Net, DeepLabV3+, and SegFormer. For fair comparison, all models were tested under the same hardware environment, and their trainable parameters and floating-point operations (FLOPs) were calculated with a unified input size of $512 \times 512 \times 3$. Since the source code of MFFENet is not publicly available, its detailed parameter information could not be obtained and thus is excluded from this quantitative comparison.

Table 7 reports the parameter count and FLOPs of each model. Among all compared methods, the proposed SC-Net achieves the smallest parameter size (approximately

30.03 M), while its computational complexity is about 60.07 G FLOPs. These results indicate that SC-Net achieves a favorable balance between model complexity and segmentation accuracy. Beyond theoretical complexity, we further evaluated the practical inference speed of SC-Net on a real-world test set consisting of 2200 images. Two scenarios were considered: (1) pure model inference, excluding the time for data I/O and structural feature preprocessing; and (2) end-to-end pipeline, including image loading, the construction of ten multi-channel structural descriptors (as described in Section 3.2), model forward pass, and result saving.

Table 7. Comparison of the efficiency of recent works.

Methods	Parameters ($\times 10^6$)	FLOPs ($\times 10^9$)
U-Net	43.93	54.82
DeepLabV3+	227.21	166.84
Segformer	47.20	62.49
SC-Net (Ours)	30.03	60.07

The results show that the pure model inference achieves a high speed of approximately 2550 frames per second (FPS), indicating the low inference latency of the proposed network. When the complete preprocessing pipeline is taken into account, the overall processing rate remains at about 45 FPS, which may be potentially suitable for near-real-time landslide hazard assessment in many practical scenarios. Furthermore, given the notable runtime difference between the full evaluation pipeline and pure inference, we conducted a detailed breakdown of the time spent on structural feature generation and data I/O operations. The experimental results are summarized in Table 8, with a batch size of 50. Here, "Time per batch" denotes the average inference time per batch, and "Percentage" indicates the proportion of each component within the total prediction time. All measurements were performed on a workstation equipped with an AMD Ryzen 9 9900X 12-core CPU and an NVIDIA GPU with 24 GB of memory. The results show that data I/O dominates the inference stage, with writing predictions to disk accounting for 48.89% of the time, multi-structural descriptor generation for 28.6%, and batch data reading for 20.58%.

Table 8. Time consumption distribution for each processing stage.

Processing Stage	Time per Batch (s)	Percentage (%)
Read image from disk	0.211	20.58
Generate structural descriptor	0.293	28.60
Model inference	0.018	1.78
Sigmoid-based binarization	0.002	0.15
Write the result to disk	0.501	48.89

While the structural descriptor computation introduces some overhead, the resulting improvements in accuracy and the overall throughput of 45 FPS justify this trade-off for practical UAV-based monitoring. These results suggest that the proposed framework may support efficient large-scale landslide monitoring applications.

5. Discussion

Our proposed SC-Net integrates a multi-structural feature extraction module with an attention-based adaptive fusion strategy and introduces a mask-constrained object-level contrastive learning loss. Experimental results on the CAS-UAV and CAS-SAT datasets demonstrate that SC-Net outperforms both classical segmentation models and recent state-of-the-art landslide extraction methods. Notably, in several qualitative examples that

contain power transmission towers and their surrounding corridors, our method shows promising capability in distinguishing landslides from spectrally similar features such as bare tower foundations and access roads, while also maintaining relatively clean boundary delineation under vegetation occlusion. Ablation studies further confirm the effectiveness of each module and their complementary contributions.

Despite these promising results, we acknowledge certain limitations in the current implementation. First, the object-level contrastive loss employs uniform random sampling of negative background features within each batch, as dedicated hard-negative mining lies beyond the scope of this study. Although this random sampling strategy, aided by multi-level structural patterns, effectively separates landslides from common distractors and thus yields satisfactory general background suppression, it lacks a mechanism to prioritize the most challenging negatives that share strong visual similarities with landslides. Second, the scarcity of publicly available datasets specifically designed for power-transmission scenarios constitutes another limitation. Owing to this constraint, we resorted to large-scale public datasets that provide diverse landslide samples and exhibit considerable environmental resemblance to power-transmission corridors. While we have performed qualitative comparisons on a limited subset of samples containing power-transmission facilities, quantitative validation in real-world power-transmission environments has not yet been conducted. Consequently, the present experimental results offer preliminary evidence of effectiveness but remain insufficient for a comprehensive assessment in the target application context.

In future work, we will focus on three aspects. First, hard negative mining strategies will be explored to better distinguish challenging confusers, thereby enhancing the discriminative power of the contrastive loss. Second, we will construct a dedicated dataset specifically for power-transmission scenarios and conduct more extensive quantitative evaluations in real-world environments, so as to verify the generalization capability of the proposed method under practical application conditions. Third, we note that the current multi-level structural generation module incurs relatively high computational overhead, which is largely attributed to the concurrent use of multiple structural feature types. To mitigate this issue, we will investigate selective pruning of structural feature types to identify potentially redundant components and reduce computational overhead while maintaining competitive accuracy.

6. Conclusions

This research proposes a structural constrained contrastive learning network for accurate and robust landslide extraction from remote sensing imagery, with a particular focus on long-distance power transmission corridors where landslide hazards may threaten tower foundations. By jointly addressing complex background interference and blurred boundaries, SC-Net improves landslide representation under diverse imaging conditions. Specifically, a multi-level structural feature extraction module with an attention-based adaptive fusion strategy strengthens landslide-specific geometric and semantic representations, while a mask-constrained object-level contrastive learning loss enforces global structural consistency among landslide objects.

Extensive comparative and ablation experiments on the CAS-UAV and CAS-SAT datasets show that SC-Net achieves higher quantitative performance than the compared methods across multiple metrics. The proposed method shows promise for large-scale recognition of geological hazards along power transmission corridors. In addition, SC-Net has the potential to process newly acquired satellite or UAV imagery to localize landslide occurrences near towers and lines, potentially supporting rapid emergency response and condition assessment. Future work will focus on hard negative mining, selective pruning

of structural features, and validation on dedicated power-transmission-corridor datasets to further assess its practical applicability.

Author Contributions: Conceptualization, W.S. and S.L. (Shiming Li); methodology, W.S., S.L. (Shiming Li) and S.L. (Shilian Liu); validation, S.W., C.L. and S.L. (Shilian Liu); formal analysis, Z.W.; investigation, Y.D. and S.W.; resources, X.Z.; data curation, S.W. and X.Z.; writing—original draft preparation, S.L. (Shilian Liu) and W.S.; writing—review and editing, C.L.; visualization, S.L. (Shilian Liu); supervision, S.L. (Shiming Li); project administration, S.W. and S.L. (Shiming Li); funding acquisition, S.L. (Shiming Li). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, No. 2024YFC3210800 and the Science Foundation of Henan Province (General Program), No. 262300421762, No. 252300423427.

Data Availability Statement: The data supporting the findings of this study are publicly available at Zenodo: https://zenodo.org/records/10294997 (accessed on 27 June 2026).

Acknowledgments: The authors would like to thank the editor and reviewers for their contributions to the paper.

Conflicts of Interest: Authors Shun Wu and Xiaobin Zheng were employed by the State Grid Jibei Electric Power Co., Ltd. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DEM	Digital elevation model
CNN	Convolutional Neural Network
CA	Channel attention module
SA	Spatial attention module
AAF	Attention-based adaptive fusion module
BCE	Binary cross-entropy
TP	True positives
FP	False positives
FN	False negatives
IoU	Intersection over union
DA	Data augmentation strategy
MSE	Multi-level structural feature enhancement module
MOC	Mask-constrained object-level contrastive learning strategy
FPS	Frames per second

References

1. Mohan, A.; Singh, A.K.; Kumar, B.; Dwivedi, R. Review on remote sensing methods for landslide detection using machine and deep learning. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e3998. [CrossRef]
2. Yang, Z.; Qi, W.; Xu, C.; Shao, X. Exploring deep learning for landslide mapping: A comprehensive review. *China Geol.* **2024**, *7*, 330–350. [CrossRef]
3. Liu, S.; Yin, K.; Zhou, C.; Gui, L.; Liang, X.; Lin, W.; Zhao, B. Susceptibility Assessment for Landslide Initiated along Power Transmission Lines. *Remote Sens.* **2021**, *13*, 5068. [CrossRef]
4. Chen, X.; Li, W.; Hsu, C.-Y.; Arundel, S.T.; Higman, B. Harnessing Geospatial Artificial Intelligence and Deep Learning for Landslide Inventory Mapping: Advances, Challenges, and Emerging Directions. *Remote Sens.* **2025**, *17*, 1856. [CrossRef]
5. Ghorani, R.; Fattaheian-Dehkordi, S.; Farrokhi, M.; Fotuhi-Firuzabad, M.; Lehtonen, M. Modeling and Quantification of Power System Resilience to Natural Hazards: A Case of Landslide. *IEEE Access* **2021**, *9*, 80300–80309. [CrossRef]
6. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2015; pp. 3431–3440.

7. Fu, Y.; Zhong, H.; Fang, C. Attention U-Mamba: A Simple and Efficient Method for Landslide Segmentation. *IEEE Geosci. Remote Sens. Lett.* **2025**, *22*, 3002705. [CrossRef]
8. Hou, C.; Yu, J.; Ge, D.; Yang, L.; Xi, L.; Pang, Y.; Wen, Y. A Transfer Learning Approach for Landslide Semantic Segmentation Based on Visual Foundation Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 11561–11572. [CrossRef]
9. Ji, S.; Yu, D.; Shen, C.; Li, W.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [CrossRef]
10. Qu, Y.; Xing, H.; Sun, L.; Shi, X.; Huang, J.; Ao, Z.; Chang, Z.; Li, J. Integrating sentinel-2a imagery, DEM data, and spectral feature analysis for landslide detection via fully convolutional networks. *Landslides* **2025**, *22*, 335–352. [CrossRef]
11. Lv, J.; Zhang, R.; Wu, R.; Bao, X.; Liu, G. Landslide detection based on pixel-level contrastive learning for semi-supervised semantic segmentation in wide areas. *Landslides* **2025**, *22*, 1087–1105. [CrossRef]
12. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In *Proceedings of the European Conference on Computer Vision (ECCV)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–19.
13. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018; pp. 7132–7141.
14. Chen, H.; He, Y.; Zhang, L.; Yao, S.; Yang, W.; Fang, Y.; Liu, Y.; Gao, B. A landslide extraction method of channel attention mechanism U-Net network based on Sentinel-2A remote sensing images. *Int. J. Digit. Earth* **2023**, *16*, 552–577. [CrossRef]
15. Niu, C.; Gao, O.; Lu, W.; Liu, W.; Lai, T. Reg-SA–UNet++: A Lightweight Landslide Detection Network Based on Single-Temporal Images Captured Postlandslide. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9746–9759. [CrossRef]
16. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. [CrossRef]
17. Chen, J.; Liu, J.; Zeng, X.; Zhou, S.; Sun, G.; Rao, S.; Guo, Y.; Zhu, J. A Cross-Domain Landslide Extraction Method Utilizing Image Masking and Morphological Information Enhancement. *Remote Sens.* **2025**, *17*, 1464. [CrossRef]
18. Chen, T.; Gao, X.; Liu, G.; Wang, C.; Zhao, Z.; Dou, J.; Niu, R.; Plaza, A.J. BisDeNet: A New Lightweight Deep Learning-Based Framework for Efficient Landslide Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 3648–3663. [CrossRef]
19. Fu, Y.; Li, W.; Fan, S.; Jiang, Y.; Bai, H. CAL-Net: Conditional Attention Lightweight Network for In-Orbit Landslide Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4408515. [CrossRef]
20. Zhang, R.; Zhu, W.; Li, Z.; Zhang, B.; Chen, B. Re-Net: Multibranch Network with Structural Reparameterization for Landslide Detection in Optical Imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 2828–2837. [CrossRef]
21. Hussaine, S.M.; Mu, L.; Lu, Y.; Hussain, S.S. Landslide Image Segmentation with Attention Residual U-Net: A Hybrid Deep Learning Model. *Procedia Comput. Sci.* **2025**, *258*, 2029–2039. [CrossRef]
22. Ling, Y.; Wang, Y.; Lin, Y.; On Chan, T.; Awange, J. Optimized Landslide Segmentation from Satellite Imagery Based on Multiresolution Fusion and Attention Mechanism. *IEEE Geosci. Remote Sens. Lett.* **2024**, *21*, 6015505. [CrossRef]
23. Li, P.; Wang, Y.; Xu, G.; Wang, L. LandslideCL: Towards robust landslide analysis guided by contrastive learning. *Landslides* **2023**, *20*, 461–474. [CrossRef]
24. Wei, R.; Ye, C.; Sui, T.; Zhang, H.; Ge, Y.; Li, Y. A feature enhancement framework for landslide detection. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *124*, 103521. [CrossRef]
25. Yang, Y.; Miao, Z.; Li, X.; Zhang, H.; Chen, S. LSL-SS-Net: Level set loss-guided semantic segmentation networks for landslide extraction. *GIScience Remote Sens.* **2024**, *61*, 2416727. [CrossRef]
26. Liu, B.; Wang, W.; Wu, Y.; Gao, X. Attention Swin Transformer UNet for Landslide Segmentation in Remotely Sensed Images. *Remote Sens.* **2024**, *16*, 4464. [CrossRef]
27. Liu, X.; Peng, Y.; Lu, Z.; Li, W.; Yu, J.; Ge, D.; Xiang, W. Feature-Fusion Segmentation Network for Landslide Detection Using High-Resolution Remote Sensing Images and Digital Elevation Model Data. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4500314. [CrossRef]
28. Wang, J.; Zhang, Q.; Xie, H.; Chen, Y.; Sun, R. Enhanced Dual-Channel Model-Based with Improved Unet++ Network for Landslide Monitoring and Region Extraction in Remote Sensing Images. *Remote Sens.* **2024**, *16*, 2990. [CrossRef]
29. Zhu, X.; Zhang, Z.; He, Y.; Wang, W.; Yang, S.; Hou, Y. LandslideNet: A landslide semantic segmentation network based on single-temporal optical remote sensing images. *Adv. Space Res.* **2024**, *74*, 4616–4638. [CrossRef]
30. Ha, Y.-S.; Nguyen, H.-H.-D.; Nguyen, T.-N.; Pham, M.-V. UAV imagery-based landslide detection in challenging environment using pixel segmentation and generative AI approach. *Model. Earth Syst. Environ.* **2025**, *11*, 342. [CrossRef]
31. Wang, K.; He, D.; Sun, Q.; Yi, L.; Yuan, X.; Wang, Y. A novel network for semantic segmentation of landslide areas in remote sensing images with multi-branch and multi-scale fusion. *Appl. Soft Comput.* **2024**, *158*, 111542. [CrossRef]
32. Zhou, Y.; Wang, H.; Yang, R.; Yao, G.; Xu, Q.; Zhang, X. A Novel Weakly Supervised Remote Sensing Landslide Semantic Segmentation Method: Combining CAM and cycleGAN Algorithms. *Remote Sens.* **2022**, *14*, 3650. [CrossRef]

33. Feng, X.; Du, J.; Wu, M.; Chai, B.; Miao, F.; Wang, Y. Potential of synthetic images in landslide segmentation in data-poor scenario: A framework combining GAN and transformer models. *Landslides* **2024**, *21*, 2211–2226. [CrossRef]
34. Zhang, X.; Yu, W.; Pun, M.-O.; Shi, W. Cross-domain landslide mapping from large-scale remote sensing images using prototype-guided domain-aware progressive representation learning. *ISPRS J. Photogramm. Remote Sens.* **2023**, *197*, 1–17. [CrossRef]
35. Wei, R.; Li, Y.; Li, Y.; Zhang, B.; Wang, J.; Wu, C.; Yao, S.; Ye, C. A universal adapter in segmentation models for transferable landslide mapping. *ISPRS J. Photogramm. Remote Sens.* **2024**, *218*, 446–465. [CrossRef]
36. Wang, X.; Zhong, D.; Liu, C.; Song, X.; Xu, L.; Deng, Y.; Li, S. DS Net: A Dual-Coded Segmentation Network Leveraging Large Model Prior Knowledge for Intelligent Landslide Extraction. *Remote Sens.* **2025**, *17*, 1912. [CrossRef]
37. Yang, C.; Zhu, Y.; Zhang, J.; Wei, X.; Zhu, H.; Zhu, Z. A feature fusion method on landslide identification in remote sensing with Segment Anything Model. *Landslides* **2025**, *22*, 471–483. [CrossRef]
38. Xi, L.; Yu, J.; Ge, D.; Pang, Y.; Zhou, P.; Hou, C.; Li, Y.; Chen, Y.; Dong, Y. SAM-CFFNet: SAM-Based Cross-Feature Fusion Network for Intelligent Identification of Landslides. *Remote Sens.* **2024**, *16*, 2334. [CrossRef]
39. Yan, K.; Shen, F.; Li, Z. Enhancing Landslide Segmentation with Guide Attention Mechanism and Fast Fourier Transformer. In *Proceedings of the Advanced Intelligent Computing Technology and Applications*; Huang, D.-S., Zhang, C., Guo, J., Eds.; Springer Nature: Singapore, 2024; pp. 296–307. [CrossRef]
40. Ghorbanzadeh, O.; Shahabi, H.; Crivellari, A.; Homayouni, S.; Blaschke, T.; Ghamisi, P. Landslide detection using deep learning and object-based image analysis. *Landslides* **2022**, *19*, 929–939. [CrossRef]
41. Li, H.; He, Y.; Xu, Q.; Deng, J.; Li, W.; Wei, Y. Detection and segmentation of loess landslides via satellite images: A two-phase framework. *Landslides* **2022**, *19*, 673–686. [CrossRef]
42. Lu, Z.; Peng, Y.; Li, W.; Yu, J.; Ge, D.; Han, L.; Xiang, W. An Iterative Classification and Semantic Segmentation Network for Old Landslide Detection Using High-Resolution Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4408813. [CrossRef]
43. Xu, Y.; Ouyang, C.; Xu, Q.; Wang, D.; Zhao, B.; Luo, Y. CAS Landslide Dataset: A Large-Scale and Multisensor Dataset for Deep Learning-Based Landslide Detection. *Sci. Data* **2024**, *11*, 12. [CrossRef] [PubMed]
44. 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 (ECCV)*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 801–818.
45. Syed, H.M.; Oghaz, M.M.; Saheer, L.B. Semantic Segmentation for Landslide Detection Using Segformer. In *Proceedings of the Artificial Intelligence XLI*; Bramer, M., Stahl, F., Eds.; Springer Nature: Cham, Switzerland, 2025; pp. 35–45. [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.