TerrainGraphNet: Terrain-Constrained Graph Reasoning for Landslide Segmentation

Shahid Shafi Dar[1,⋆], Pranjal Pandey[2], and Nagendra Kumar[1]

[1] Indian Institute of Technology Indore, Indore, India
[2] Indian Institute of Information Technology Bhagalpur, Bihar, India
phd2201201004@iiti.ac.in, pranjal.230103027@iiitbh.ac.in, nagendra@iiti.ac.in,

Abstract. Accurate landslide segmentation from high-resolution remote sensing imagery remains challenging due to weak spectral contrast and strong dependence on terrain structure. Unlike appearance-defined objects, landslides follow geomorphological patterns governed by slope continuity and elevation discontinuities. However, most existing segmentation methods treat landslides as generic semantic regions and incorporate Digital Elevation Models (DEM) through simple feature fusion without explicitly enforcing terrain-consistent spatial reasoning. To address this limitation, we reformulate landslide segmentation as a terrain-conditioned structured prediction problem and propose TerrainGraphNet, a terrain-aware graph reasoning framework that embeds geomorphological constraints into representation learning and spatial propagation. The proposed method introduces terrain-modulated feature interaction, where elevation structure adaptively regulates visual representations, and terrain-aware graph construction, where spatial connectivity is defined jointly by feature similarity and slope continuity. This formulation can be interpreted as learning a terrain-weighted smoothness prior that encourages predictions to respect geomorphological coherence. Extensive experiments on three benchmark datasets demonstrate consistent improvements over strong CNN and transformer baselines. In addition to higher IoU and F1 scores, TerrainGraphNet significantly improves boundary accuracy and topological consistency, highlighting the importance of terrain-constrained reasoning for reliable landslide delineation.

Keywords: Landslide Segmentation, Terrain-Aware Learning, Graph Neural Networks, Graph Attention Networks, Remote Sensing

1 Introduction

Landslides are among the most destructive natural hazards, causing severe human and economic losses worldwide [12, 31, 35]. Accurate identification of landslide regions from remote sensing imagery is therefore essential for disaster response, hazard assessment, and environmental monitoring [1, 7, 10]. Recent advances in deep learning have significantly improved semantic segmentation for

⋆ Corresponding author.

Earth observation tasks [6, 8, 24]. Early approaches relied on convolutional encoder–decoder architectures such as U-Net [36] and its extensions [47], while later work introduced multi-scale context modeling through atrous convolution networks [5] and hybrid transformer–CNN architectures [3]. Transformer-based segmentation models such as SegFormer [42] further improve global context modeling and have shown strong performance on high-resolution imagery.

For landslide detection, several specialized architectures have been proposed that incorporate multiscale attention, contextual aggregation, and efficient sequence modeling to capture complex terrain patterns [13, 15, 20, 27, 39, 49]. These approaches substantially improve landslide mapping accuracy from high-resolution satellite imagery. Despite these advances, landslide segmentation remains challenging because landslides are geomorphological phenomena rather than appearance-defined objects. Their spatial extent is strongly governed by terrain structure: landslide regions typically follow slope directions, remain spatially connected along terrain gradients, and align with elevation discontinuities. Therefore, purely appearance-driven segmentation models often produce fragmented predictions or boundary leakage in visually ambiguous regions.

To mitigate this limitation, several studies incorporate Digital Elevation Models (DEM) as auxiliary inputs to provide terrain-related information. Multimodal frameworks combine optical imagery with elevation data to improve landslide detection performance [28], and recent remote sensing foundation models explore unified representations across heterogeneous earth observation modalities [11, 19]. However, most existing methods integrate DEM through simple feature concatenation or uniform fusion, treating elevation as an additional appearance channel rather than a structural constraint guiding spatial reasoning.

Graph-based learning offers a natural mechanism for modeling long-range spatial dependencies in segmentation tasks. Graph neural networks propagate information across spatially distant regions and capture non-local relationships beyond convolutional receptive fields [17, 26]. Recent advances in remote sensing representation learning also explore state-space models and foundation models to improve large-scale scene understanding [4, 32, 44]. Nevertheless, these approaches typically construct spatial relationships based solely on feature similarity or proximity.

In this work, we instead formulate landslide segmentation as a terrain-conditioned structured prediction problem. Predictions should propagate along terrain-consistent regions while respecting elevation discontinuities. To realize this principle, we propose **TerrainGraphNet**, a terrain-aware graph reasoning framework that embeds geomorphological constraints into both representation learning and spatial propagation. The model introduces terrain-modulated feature interaction to adaptively fuse visual and elevation features, followed by terrain-aware graph construction where spatial connectivity depends jointly on feature similarity and slope continuity derived from DEM data. Graph attention layers then propagate information along terrain-consistent regions while suppressing diffusion across geomorphological boundaries. Extensive experiments on multiple benchmark datasets demonstrate that TerrainGraphNet consistently improves landslide segmentation performance. In addition to higher IoU and F1 scores, the proposed method significantly improves boundary accuracy and structural consistency, highlighting the importance of terrain-aware reasoning for reliable landslide segmentation. The main contributions of this work are as follows:

- We formulate landslide segmentation as a terrain-conditioned structured prediction problem that explicitly incorporates geomorphological constraints.
- We propose terrain-modulated feature interaction, where elevation structure adaptively regulates visual representations during multimodal fusion.
- We introduce terrain-aware graph construction and topology-aligned graph propagation to enforce spatial coherence along slope-consistent regions.

2 Related Work

2.1 Deep Learning for Landslide Segmentation

Deep learning has significantly improved landslide mapping from remote sensing imagery. Early studies mainly relied on convolutional encoder–decoder architectures such as U-Net [36]. Subsequent works introduced stronger contextual modeling through graph-based and hierarchical segmentation frameworks [17,26] as well as transformer architectures that capture long-range dependencies and global scene context [38]. Recent representation learning approaches further explore large-scale pretrained models and efficient sequence modeling architectures for remote sensing imagery [19,32]. Despite these advances, most existing methods treat landslides as generic semantic regions defined primarily by appearance features, which may lead to fragmented predictions in heterogeneous terrain. In contrast, our approach incorporates terrain structure to enforce spatial coherence during segmentation.

2.2 Multi-Modal Remote Sensing Fusion

Multi-modal learning has been widely used to improve geospatial segmentation by combining optical imagery with auxiliary data such as Digital Elevation Models (DEMs) [2, 9, 14, 30, 33, 37]. Several studies integrate elevation information through channel concatenation or multi-branch architectures [11, 41]. More recent approaches explore multimodal representation learning to jointly process heterogeneous data sources [11]. However, most existing methods treat DEMs as additional input features rather than structural constraints guiding spatial reasoning. Our approach instead uses terrain information to define terrain-aware connectivity during graph propagation.

2.3 Graph-Based and Topology-Aware Segmentation

Graph-based learning has been explored to capture long-range spatial dependencies in segmentation tasks by propagating information across pixels or regions

[17, 22, 26, 34, 40, 45, 46]. Semi-supervised frameworks further improve segmentation by combining pixel-level predictions with region-level representations [48]. In addition, recent works investigate topology-aware modeling and boundary-aware architectures for better structural consistency [18, 32].
Nevertheless, most existing approaches construct connectivity purely based on spatial proximity or feature similarity. In contrast, our proposed approach constructs graphs by incorporating terrain information derived from elevation data, allowing information to propagate along terrain-consistent regions while preserving geomorphological boundaries [4, 25, 44].

3 Methodology

Problem Definition: Landslides are terrain-driven processes whose spatial extent is governed by elevation geometry and slope continuity. Pixels belonging to the same landslide often follow coherent slope directions and remain connected across large spatial distances, even when local appearance varies due to vegetation or illumination. Conventional segmentation networks treat the task as pixel-wise classification primarily driven by visual appearance. While effective for generic semantic objects, such models may produce fragmented or topologically inconsistent predictions in complex terrain.

We instead formulate landslide segmentation as a *terrain-conditioned structured prediction problem*. The key principle of TerrainGraphNet is that predictions should propagate along terrain-consistent regions while respecting elevation discontinuities.

Let $\mathcal{D} = \{(I_i, D_i, Y_i)\}_{i=1}^{N}$ denote the dataset, where I_i is an RGB image, D_i is the aligned DEM, and Y_i is the corresponding landslide mask. The goal is to learn a mapping $f_\theta : (I, D) \rightarrow \hat{Y}$ that integrates visual evidence with terrain structure to produce spatially coherent predictions. To operationalize this principle, our proposed method, TerrainGraphNet, as illustrated in Figure 1 performs three steps: (i) joint visual–terrain encoding, (ii) terrain-modulated feature interaction, and (iii) terrain-aware graph construction and propagation.

3.1 Joint Visual–Terrain Encoding

We first learn a joint embedding that captures both visual semantics and terrain structure. To achieve this, the network employs two parallel encoders: a visual encoder that extracts appearance features from RGB imagery and a terrain encoder that captures elevation-driven structure from DEM data.

Visual Encoding: Remote sensing imagery contains information at multiple spatial scales. To capture both global context and fine-grained terrain details, we employ a dual-stream visual encoder. Global semantic features are extracted using a pretrained Vision Transformer backbone (SAM) [23]. Given a high-resolution input image I^{high}, the vision encoder produces:

$$F_{\text{g}} = E_{\text{SAM}}(I^{\text{high}}), \tag{1}$$

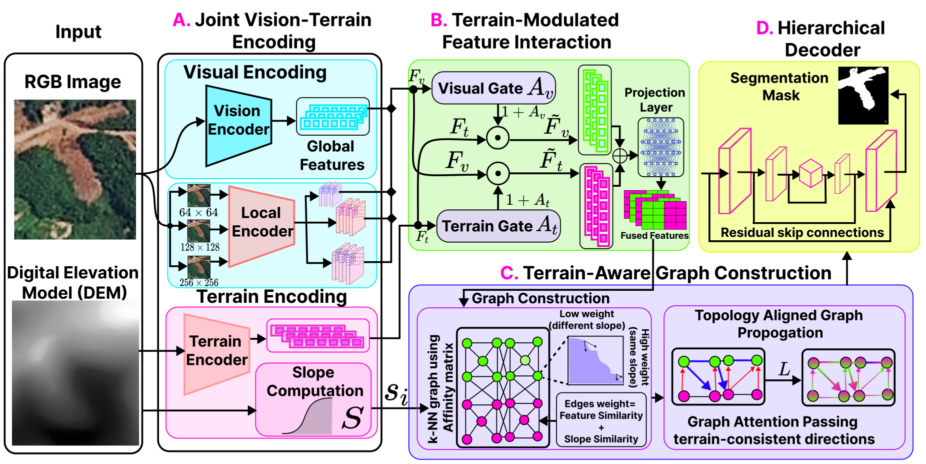


Fig. 1: Architecture of TerrainGraphNet for landslide segmentation. RGB images and DEM data are first encoded to obtain complementary visual and terrain representations. A terrain-modulated interaction module fuses these features through adaptive gating. The fused features are then used to construct a terrain-aware graph where topology-aligned attention propagates information along terrain-consistent directions. Finally, a hierarchical decoder with residual skip connections generates the final segmentation mask.

where $F_{\text{g}} \in \mathbb{R}^{H' \times W' \times C}$ denotes the global feature representation. Although the transformer encoder captures long-range contextual information, it may miss fine-scale texture variations that are important for landslide boundaries. To complement these features, we employ a lightweight multi-scale convolutional encoder on a medium-resolution image I^{mid}. The image is processed at three spatial scales, $64, 128, 256$, to capture local terrain patterns at different receptive fields. The resulting multi-scale features are fused to obtain:

$$F_{\text{l}} = E_{\text{local}}(I^{\text{mid}}). \tag{2}$$

Finally, the global and local features are projected into a common feature space and fused as:

$$F_{\text{v}} = \phi([F_{\text{g}}, F_{\text{l}}]), \tag{3}$$

where $\phi(\cdot)$ denotes a learnable 1×1 projection and $[\cdot]$ denotes channel concatenation.

Terrain Encoding: Elevation information is processed through a parallel encoder:

$$F_{\text{t}} = E_{\text{terrain}}(D) \tag{4}$$

However, raw elevation alone does not directly indicate instability. Landslides correlate strongly with local slope magnitude. We therefore compute slope explicitly from the DEM:

$$S(x, y) = \sqrt{\left(\frac{\partial D}{\partial x}\right)^2 + \left(\frac{\partial D}{\partial y}\right)^2} \tag{5}$$

Here, $S(x, y)$ measures the gradient magnitude of elevation and highlights terrain discontinuities and steep regions, which are critical for geomorphological reasoning. All feature maps are spatially aligned to resolution $H' \times W'$ to enable joint processing in subsequent stages.

3.2 Terrain-Modulated Feature Interaction

A naive strategy is to concatenate visual and terrain features and learn their interaction through convolutional layers. However, such uniform fusion assumes that both modalities contribute equally at all spatial locations. In practice, terrain information is more informative in steep regions, whereas visual texture dominates in flat or homogeneous areas. We therefore introduce terrain-modulated feature interaction to adaptively regulate cross-modal influence. We compute terrain-driven and visual-driven gating maps as follows:

$$A_{\mathrm{t}} = \sigma(W_{\mathrm{t}} F_{\mathrm{t}}), \qquad A_{\mathrm{v}} = \sigma(W_{\mathrm{v}} F_{\mathrm{v}}) \tag{6}$$

where W_{t} and W_{v} are learnable 1×1 projections and $\sigma(\cdot)$ denotes the sigmoid function. The gating map A_{t} indicates where terrain structure should influence visual features, while A_{v} indicates where visual features should refine terrain features. Because the sigmoid constrains values to $[0, 1]$, the gates behave as soft spatial attention masks. We then modulate the features multiplicatively:

$$\tilde{F}_{\mathrm{v}} = F_{\mathrm{v}} \odot (1 + A_{\mathrm{t}}), \qquad \tilde{F}_{\mathrm{t}} = F_{\mathrm{t}} \odot (1 + A_{\mathrm{v}}) \tag{7}$$

where $\odot$ denotes element-wise multiplication. The additive bias of 1 ensures that features are enhanced rather than suppressed entirely. Intuitively, in steep regions where slope magnitude is high, A_{t} increases the contribution of terrain features to the visual embedding. Conversely, in visually distinctive areas aligned with terrain structure, A_{v} reinforces elevation features. The final fused representation is obtained by projecting the concatenated modulated features:

$$F = \psi([\tilde{F}_{\mathrm{v}}, \tilde{F}_{\mathrm{t}}]) \tag{8}$$

where $\psi(\cdot)$ is a learnable projection layer. This process produces a terrain-conditioned embedding in which appearance features are adaptively regulated by elevation structure. The resulting representation forms the basis for topology-aware graph propagation in the next stage.

3.3 Terrain-Aware Graph Construction and Propagation

The terrain-modulated embedding F captures local interactions between appearance and elevation. However, landslides frequently extend across large spatial extents and may appear fragmented due to vegetation, shadows, or imaging artifacts. Purely convolutional processing remains local and therefore cannot explicitly enforce long-range structural continuity. To propagate information across spatially distant but terrain-consistent regions, we construct a sparse graph over spatial locations. Each spatial position is treated as a node whose feature representation encodes both visual and terrain information.

Let $f_i \in \mathbb{R}^C$ denote the fused feature vector at spatial location i, and let s_i denote the slope magnitude at that location derived from the DEM. The slope term provides an explicit measure of terrain continuity and elevation change.

We define the affinity between two locations i and j as:

$$w_{ij} = \exp\left(-\alpha \|f_i - f_j\|_2^2 - (1-\alpha)\|s_i - s_j\|_2^2\right) \tag{9}$$

where $\alpha \in [0, 1]$ balances similarity in the learned embedding space and similarity in slope structure. The above affinity formulation has a simple interpretation. Two nodes are assigned a strong connection when they have similar visual-terrain features and similar slope values, indicating that they are likely to belong to the same terrain structure. In contrast, large differences in slope reduce the affinity, preventing information from propagating across natural terrain boundaries such as ridges and valleys.

For each node i, we select the top-k neighbors based on affinity scores to construct the graph. As a result, a sparse graph is created that links regions with similar terrain characteristics, even when they are not close to each other.

Topology-Aligned Graph Propagation: After constructing the terrain-aware graph, node features are updated through graph attention. Each node updates its features using information from its neighbors, allowing features to spread across similar terrain while avoiding natural terrain boundaries. Let the initial node representation be $h_i^{(0)} = f_i$. At each layer l, the information from neighboring nodes is captured as shown in the Eq. 10:

$$h_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(l)} W^{(l)} h_j^{(l)} \tag{10}$$

where $W^{(l)}$ is a learnable linear transformation. The attention weights $\alpha_{ij}^{(l)}$ determine the relative importance of each neighbor:

$$\alpha_{ij}^{(l)} = \frac{\exp\left(\mathrm{LeakyReLU}(\mathbf{a}^{(l)\top}[W^{(l)} h_i^{(l)} \| W^{(l)} h_j^{(l)}])\right)}{\sum_{k \in \mathcal{N}(i)} \exp\left(\mathrm{LeakyReLU}(\mathbf{a}^{(l)\top}[W^{(l)} h_i^{(l)} \| W^{(l)} h_k^{(l)}])\right)} \tag{11}$$

where $\mathbf{a}^{(l)}$ is a learnable vector and $\|$ denotes concatenation. The attention weights determine how much information each neighboring node contributes during feature aggregation. Since the graph is built using terrain-aware affinity (Eq. 9), nodes with similar terrain are more likely to exchange information, whereas connections across sharp elevation changes receive lower weights.
After L layers, the updated node features are obtained through residual integration, $F' = F + h^{(L)}$. This residual connection preserves local information while incorporating terrain-aligned global context. As a result, the model promotes connected predictions within the same geomorphological structure, reduces fragmentation, and better preserves topological consistency.

3.4 Decoding and Optimization

The terrain-aware representation F' is decoded into a full-resolution mask using a hierarchical decoder. The decoder progressively upsamples the feature map while integrating encoder information through skip connections. This process results in the final feature map (F_{dec}). The final segmentation logits $\hat{Y} = \sigma(\text{Conv}_{1\times1}(F_{\text{dec}}))$ are produced by a 1×1 convolution applied to the decoded feature map F_{dec}.

Training Objective: The network is trained using a compound segmentation loss that combines Dice loss and binary cross-entropy (BCE):

$$\mathcal{L}_{\text{main}} = 0.7\,\mathcal{L}_{\text{Dice}} + 0.3\,\mathcal{L}_{\text{BCE}} \tag{12}$$

Dice loss improves region overlap, while BCE provides stable pixel-wise supervision. To improve optimization stability and encourage multi-scale consistency, auxiliary supervision is applied at intermediate decoder stages. The final training objective is:

$$\mathcal{L} = \mathcal{L}_{\text{main}} + \lambda_{\text{aux}}\mathcal{L}_{\text{aux}} \tag{13}$$

Here, $\mathcal{L}_{\text{aux}}$ denotes the auxiliary loss applied to intermediate decoder outputs using the same Dice formulation, and $\lambda_{\text{aux}} = 0.4$ is a weighting coefficient that balances the contribution of auxiliary supervision during training.

4 Experiments

4.1 Datasets and Evaluation Metrics

We evaluate our proposed method, TerrainGraphNet, on three landslide segmentation datasets: Landslide4Sense [16], Bijie [21], and Luding [29]. These datasets provide high-resolution remote sensing imagery with pixel-level landslide annotations and represent diverse terrain conditions and geographic regions. Detailed dataset specifications are provided in the supplementary material.

Evaluation Metrics: We evaluate segmentation performance using standard metrics, such as Precision (P), Recall (R), F1-score (F1), Intersection over Union (IoU), mean Intersection over Union (mIoU), and Overall Accuracy (OA). Some of these metrics may overlook the thin, fragmented, or highly irregular boundaries in landslides. For this, we additionally report several boundary and geometric metrics, such as Connectivity Error (CE), Betti Error (BE), the 95th-percentile Hausdorff Distance (HD95), and Boundary F1-score (BF1). These metrics assess how well the predicted boundaries and overall shape of a landslide match the ground truth. These metrics provide an exhaustive evaluation of landslide segmentation quality.

4.2 Performance Comparison on the Bijie Dataset

On the Bijie dataset, TerrainGraphNet outperforms existing methods across all the evaluation metrics, as reported in Table 1. Conventional encoder–decoder architectures, such as U-Net [36] and U-Net++ [47], provide strong baselines but often fail to capture fine-scale landslide structures, resulting in lower IoU scores. Transformer-based models such as TransUNet [3] and SegFormer [42] improve global contextual modeling, yet their boundary delineation remains less accurate. More recent approaches, such as SegMamba [39], CalandNet [20], OLSNet [27], and FFS-Net [28], further enhance segmentation through multi-scale feature aggregation and attention mechanisms. Among them, FFS-Net is the strongest baseline, reporting an IoU of 77.19 and a BF1 score of 52.22. In comparison, TerrainGraphNet obtains the best overall results, improving the IoU to 79.08, mIoU to 88.26, and F1-score to 88.32. It also records the lowest connectivity error (CE = 0.0172) while achieving the highest boundary F1-score of 57.31. These findings indicate that terrain-aware constraints into graph reasoning enables more accurate feature interactions, leading to better segmentation masks and substantially better boundary localization.

Table 1: Performance comparison on the Bijie Dataset. Bold values denote the best results. ↓ indicates lower is better and ↑ indicates higher is better.

Model	**Standard Metrics**						**Geometric & Boundary Metrics**			
	P	**R**	**F1**	**OA**	**IoU**	**mIoU**	**BE ↓**	**CE ↓**	**HD95 ↓**	**BF1 ↑**
U-Net [36]	84.64	63.51	72.57	92.18	56.94	76.20	2.9500	0.8540	27.84	35.12
U-Net++ [47]	82.04	83.26	82.64	96.50	70.42	83.30	3.8233	1.3103	21.26	44.00
DeepLab V3+ [5]	81.35	73.81	77.40	95.68	63.13	79.23	0.6422	0.2974	25.06	36.94
TransUNet [3]	85.46	79.00	82.11	96.56	69.64	82.95	0.6120	0.1534	22.45	42.65
SegFormer [42]	84.71	75.87	80.05	96.21	66.74	81.32	0.3060	0.0431	19.89	43.47
SegMamba [39]	79.61	71.51	75.34	95.30	60.44	77.69	1.1039	0.3853	25.31	33.57
CalandNet [20]	78.75	78.42	78.59	95.72	64.73	80.04	0.2759	0.1164	19.89	40.42
OLSNet [27]	84.77	78.10	81.30	96.39	68.49	82.28	0.5584	0.0736	26.73	42.17
LandSegNet [49]	79.01	84.81	81.81	96.21	69.22	82.54	4.4372	1.0390	21.99	41.75
Swin-MA [15]	84.13	80.12	82.08	96.48	69.60	82.89	**0.0606**	0.0216	18.92	42.80
SCGC-Net [13]	85.37	79.17	82.15	96.56	69.71	82.99	0.5200	0.1600	23.46	43.01
FFS-Net [28]	87.64	86.61	87.12	97.37	77.19	87.15	0.0862	0.0474	15.06	52.22
Ours	**88.84**	**87.80**	**88.32**	**97.67**	**79.08**	**88.26**	0.0776	**0.0172**	**14.42**	**57.31**

4.3 Performance Comparison on the Landslide4Sense Dataset

Landslide4Sense dataset contains more complex terrain conditions and smaller landslide regions, which makes the segmentation task more challenging. Table 2 presents the results on the Landslide4Sense dataset. We observe a similar trend to that of the Bijie dataset. U-Net and U-Net++ achieve relatively low IoU scores due to limited contextual modeling capability. Recent landslide-specific methods, such as OLSNet and FFS-Net, further improve performance by leveraging multi-scale feature fusion and different variations of the attention mechanisms. In particular, FFS-Net achieves strong results with an IoU of 53.84 and highest BF1 of 65.46. TerrainGraphNet achieves the highest IoU of 54.20 and mIoU of 76.39,

Table 2: Performance comparison on the Landslide4Sense dataset.

Model	Standard Metrics						Geometric & Boundary Metrics			
	P	**R**	**F1**	**IoU**	**mIoU**	**OA**	**BE ↓**	**CE ↓**	**HD95 ↓**	**BF1 ↑**
U-Net [36]	52.50	45.81	48.93	32.39	65.15	97.93	3.5775	1.3947	106.88	48.58
U-Net++ [47]	48.39	64.57	55.32	38.24	67.98	97.74	2.3041	0.3640	92.6525	52.60
DeepLab V3+ [5]	59.38	54.43	56.80	39.36	68.92	97.34	2.0807	0.1974	96.5478	56.65
TransUNet [3]	57.69	63.43	60.42	43.29	70.68	98.09	2.7217	0.4291	96.8033	52.01
SegFormer [42]	62.13	63.43	62.77	45.74	71.97	98.22	2.5768	0.2914	94.7195	54.01
SegMamba [39]	62.17	63.40	62.78	45.75	71.97	98.22	2.4325	0.6982	94.6435	57.60
CalandNet [20]	67.38	64.02	65.65	48.87	73.63	98.41	2.2002	**0.1291**	101.0078	52.90
OLSNet [27]	**75.63**	63.45	69.01	52.68	75.66	**98.65**	1.4609	0.1975	79.6516	63.44
LandsegNet [49]	71.84	64.75	68.11	51.64	75.09	98.57	6.8051	2.1106	82.3716	61.40
Swin-MA [15]	66.69	65.27	65.97	49.22	73.80	98.41	3.1554	0.2204	122.3186	44.85
SCGC-Net [13]	67.16	64.95	66.03	49.29	73.84	98.42	4.9962	0.5882	141.1206	42.04
FFS-Net [28]	71.87	68.23	70.00	53.84	76.22	98.44	1.5667	0.1772	74.4361	**65.46**
Ours	70.61	**69.99**	**70.30**	**54.20**	**76.39**	98.59	**1.0562**	0.3169	**74.4099**	63.49

while some of the boundary metrics, such as BE (1.0562) and HD95 (74.4099), are lower. The other metrics, such as CE (0.3169) and BF1 (63.49), are moderate on this dataset. These results indicate the complex nature of the Landslide4sense dataset. We observe that incorporating terrain-aware representations and graph reasoning improves the robustness of segmentation in complex terrain.

4.4 Performance Comparison on the Luding Dataset

In the Luding dataset, ground-truth DEMs are not available. Instead, the terrain branch uses depth maps estimated by Depth Anything V2 [43] to derive the slope map, while the same terrain-aware graph affinity is applied without modifying the network architecture. We present the quantitative results on the Luding dataset, as reported in Table 3. Despite relying only on estimated depth, TerrainGraphNet achieves the highest IoU (78.92), mIoU (87.18), and F1-score (88.22) which shows improved region-level segmentation performance. Also, TerrainGraphNet achieves the lowest geometric metrics (CE = 0.2400, BE = 0.3400,

Table 3: Quantitative comparison on the Luding dataset.

Model	Standard Metrics						Geometric & Boundary Metrics			
	P	**R**	**F1**	**IoU**	**mIoU**	**OA**	**BE ↓**	**CE ↓**	**HD95 ↓**	**BF1 ↑**
U-Net [36]	80.34	75.37	77.78	63.63	79.62	95.92	1.0400	0.8200	22.93	36.31
U-Net++ [47]	82.72	81.98	82.35	70.00	81.70	94.28	0.5500	0.4100	28.27	43.06
DeepLab V3+ [5]	83.23	82.94	83.08	71.06	82.35	94.50	0.3900	0.3100	29.54	35.56
TransUNet [3]	79.59	82.09	80.82	67.82	80.25	93.66	0.9900	0.7900	37.31	38.40
SegFormer [42]	81.52	83.73	82.61	70.37	81.86	94.26	0.3500	0.3300	25.21	39.34
SegMamba [39]	83.17	83.81	83.49	71.66	82.70	94.60	0.4600	0.3800	27.99	41.11
CalandNet [20]	82.64	77.93	80.22	66.97	79.90	93.74	0.6800	0.5400	39.06	36.99
OLSNet [27]	86.41	78.20	82.10	69.63	81.63	94.45	0.4705	0.4525	25.78	30.90
LandSegNet [49]	79.02	83.59	81.24	68.40	80.56	93.71	1.2120	0.6775	32.06	37.86
Swin-MA [15]	77.25	86.73	81.72	69.09	80.86	93.68	0.4325	0.4265	29.38	26.83
SCGC-Net [13]	79.39	85.05	82.12	69.67	81.33	93.97	1.6540	1.1190	28.70	43.24
FFS-Net [28]	86.00	86.85	86.42	76.09	84.73	94.03	0.9500	0.6400	25.69	50.04
Ours	**87.06**	**89.40**	**88.22**	**78.92**	**87.18**	**96.11**	**0.3400**	**0.2400**	**21.94**	**52.03**

and HD95 = 21.94) and the highest boundary accuracy (BF1 = 52.03). These results show that terrain-aware feature fusion and graph-based reasoning improve the representation of terrain structures, leading to better boundary delineation and segmentation accuracy. Please refer to supplementary material for the implementation of the Depth Anything V2 model.

4.5 Ablation Study

Table 4 reports the contribution of each component in TerrainGraphNet. Starting with the backbone model, which is a Segment Anything Model (SAM) encoder, and adding the local encoder yields a modest improvement by strengthening visual feature representations. We observe an improved segmentation performance when we incorporate additional components such as a terrain encoder with terrain feature interaction, indicating that terrain features provide complementary information beyond appearance alone. Adding the hierarchical decoder yields additional gains by improving multi-scale feature aggregation. The largest improvement is observed after introducing the terrain-aware graph reasoning module, which increases the IoU from 74.80 to 79.08 and the F1-score from 85.58 to 88.32. Overall, the progressive improvements across the ablation study show that each component contributes to the final performance. At the same time, terrain-aware graph reasoning plays the most significant role in improving structural consistency and boundary localization.

Table 4: Ablation study of TerrainGraphNet. A checkmark indicates that the corresponding module is enabled.

Component Configuration						Performance Metrics				
BB	**LE**	**TE**	**TMFI**	**HD**	**TAGR**	**P (%)**	**R (%)**	**F1 (%)**	**IoU (%)**	**mIoU (%)**
✓	✗	✗	✗	✗	✗	80.93	83.65	82.30	69.92	81.34
✓	✓	✗	✗	✗	✗	84.65	80.41	82.48	70.20	82.94
✓	✓	✓	✓	✗	✗	88.14	79.75	83.75	72.04	83.72
✓	✓	✓	✓	✓	✗	85.78	85.39	85.58	74.80	85.82
✓	✓	✓	✓	✓	✓	**88.84**	**87.80**	**88.32**	**79.08**	**88.26**

BB: Backbone Encoder; **LE**: Local Encoder; **TE**: Terrain Encoder; **TMFI**: Terrain-Modulated Feature Interaction; **HD**: Hierarchical Decoder; **TAGR**: Terrain-Aware Graph Reasoning Module.

4.6 Effect of Feature Fusion and Graph Reasoning Analysis

In Table 5, we compare different feature fusion strategies and evaluate the effect of the Terrain-Modulated Feature Interaction (TMFI) and Terrain-Aware Graph Reasoning (TAGR) module. The complete model contains both modules (TMFI and TAGR) and achieves the best overall performance, with an IoU of 79.08 and a BF1 score of 57.31. By replacing TMFI with element-wise addition or multiplication consistently reduces performance, indicating that these simple fusion operations are less effective at combining visual and terrain information. By removing the TAGR module decreases the IoU from 79.08 to 74.55 and

Table 5: Ablation analysis of multimodal fusion strategies and the terrain-aware graph reasoning (TAGR) module. TMFI denotes the proposed terrain-modulated feature interaction used for visual–terrain fusion.

Experiment Setup	IoU ↑	mIoU ↑	F1 ↑	BE ↓	HD95 ↓	BF1 ↑
TMFI + TAGR (Full Model)	79.08	88.26	88.32	0.0776	14.42	57.31
Add + TAGR	77.32	87.28	87.21	0.0563	15.32	56.08
Multiply + TAGR	76.21	86.63	86.50	0.1082	15.53	54.74
Concat + (w/o TAGR)	74.55	85.62	85.42	0.1515	18.20	51.69

TMFI: Terrain-Modulated Feature Interaction; TAGR: Terrain-Aware Graph Reasoning module.

increases the HD95 from 14.42 to 18.20, resulting in less accurate boundaries. These results show that both terrain feature interaction and terrain-aware graph reasoning contribute to the final performance, with TAGR providing the largest improvement.

4.7 Qualitative Comparison

Figure 2 presents qualitative comparisons of landslide segmentation results across six representative test samples. Compared with existing methods, TerrainGraph-Net produces predictions that better align with the ground-truth boundaries while reducing spurious false-positive regions. Competing approaches often ex-

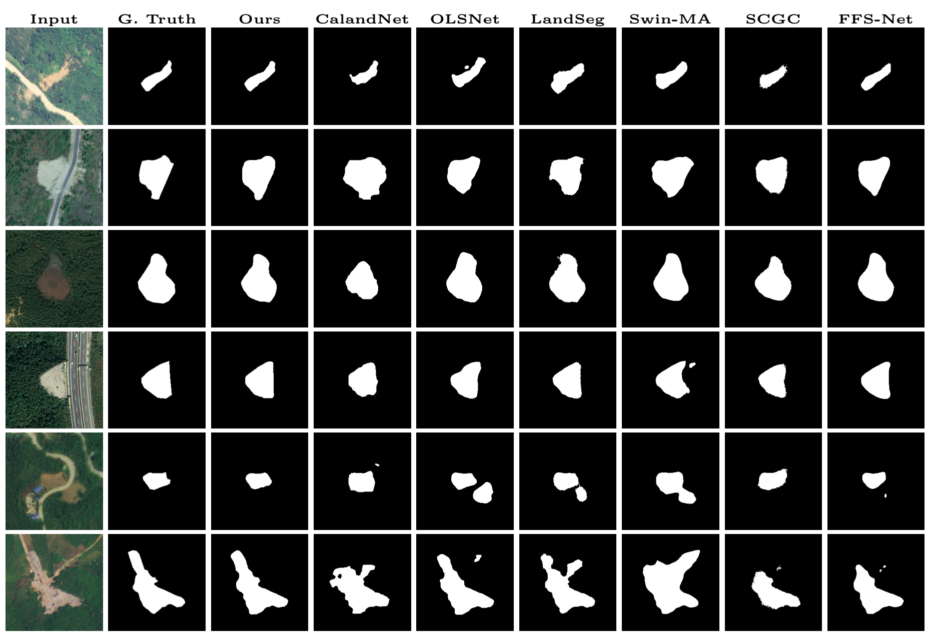
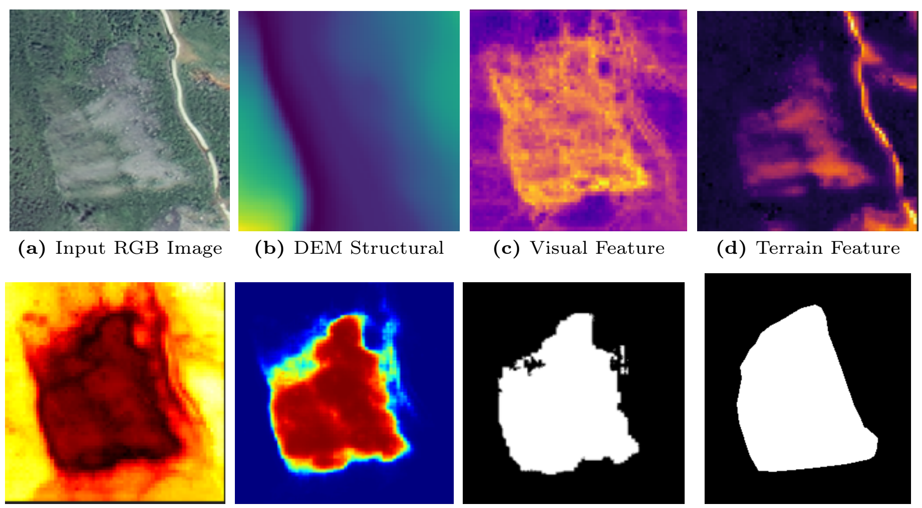


Fig. 2: Qualitative visual comparison of landslide segmentation methods.

hibit fragmented masks or boundary leakage, particularly in complex terrain areas. In contrast, the proposed model maintains more coherent and structurally consistent segmentation, highlighting the effectiveness of the terrain-aware fusion and graph reasoning modules.

4.8 Analysis of Intermediate Feature Representations

We show the intermediate feature representations learned at different stages of TerrainGraphNet in Figure 3. The inputs are an RGB image and a DEM, capturing appearance and terrain information, respectively. The early feature maps reveal texture and elevation patterns that are useful for identifying landslide regions. After fusing the two modalities in the VT Fusion stage, we get a shared representation with stronger structural features. Before decoding, the features become more discriminative by incorporating broader contextual information. The Terrain-Graph module further organizes these representations by strengthening terrain-consistent responses and suppressing noisy activations, resulting in a more accurate final segmentation mask.



(a) Input RGB Image **(b)** DEM Structural **(c)** Visual Feature **(d)** Terrain Feature

(e) VT Fusion **(f)** Terrain-Graph **(g)** Predicted Mask **(h)** Ground Truth Mask

Fig. 3: Stage-wise feature progression in TerrainGraphNet.

4.9 Sensitivity Analysis of α

Table 6 evaluates the effect of the balancing parameter α used in Eq. (9). The best performance is obtained with $\alpha = 0.5$, indicating that equally weighting feature and slope similarity yields the most effective graph construction. Performance varies only slightly across different values, demonstrating that the proposed framework is robust to the choice of α.

Table 6: Sensitivity analysis of α.

α	P	R	F1	IoU	mIoU	↓BE	↓CE	↓HD95	↑BF1
0.00	87.70	86.28	86.99	77.55	86.97	0.0788	0.0175	14.58	56.67
0.25	87.38	86.73	87.06	77.66	87.02	0.0778	0.0174	14.48	56.96
0.50	**88.84**	**87.80**	**88.32**	**79.08**	**88.26**	**0.0776**	**0.0172**	**14.42**	**57.31**
0.75	88.69	87.20	87.94	78.47	87.92	0.0783	0.0482	14.98	56.95
1.00	87.81	86.76	87.28	78.02	87.23	0.0817	0.0348	15.43	56.74

4.10 Effect of Terrain Descriptors

We compare different terrain descriptors for graph construction, as shown in Table 7. We observe that aspect and curvature improve some individual metrics, but they do not consistently outperform slope across all evaluation metrics. Therefore, we use slope throughout the paper as an effective and computationally efficient terrain descriptor.

Table 7: Effect of different terrain descriptors on the Bijie dataset.

Descriptor	P	R	F1	IoU	mIoU	BE ↓	CE ↓	HD95 ↓	BF1 ↑
No Slope ($\alpha = 1$)	87.81	86.76	87.28	78.02	87.23	0.0817	0.0348	15.43	56.74
Slope Only ($\alpha = 0$)	87.70	86.28	86.99	77.55	86.97	0.0788	0.0175	14.58	56.67
Aspect	87.90	88.25	88.07	78.69	88.03	0.0939	0.0171	13.99	**57.74**
Curvature	86.30	**89.48**	87.86	78.35	87.82	0.0819	0.0182	15.01	56.58
Slope + Aspect	89.12	86.91	88.00	78.58	87.99	0.0817	0.0214	**13.36**	57.58
Slope + Curvature	88.22	88.06	**88.14**	**78.79**	**88.09**	0.0866	0.0272	13.56	57.55
Slope + Aspect + Curvature	**89.14**	86.47	87.79	78.23	87.80	**0.0766**	**0.0161**	14.57	57.23

4.11 Complexity Analysis

We summarize the computational cost of each module in Table 8. Most of the computation is attributed to the frozen SAM encoder, whereas the proposed Terrain-Aware Graph Reasoning (TAGR) module introduces only 0.54 GFLOPs and 3.61 ms of inference time. Despite its modest overhead, TAGR provides the largest performance gain in the ablation study, improving IoU from 74.80 to 79.08. Additional complexity comparisons are provided in the supplementary material.

Table 8: Computational complexity of different components in TerrainGraphNet.

Module	Params (M)	GFLOPs	Time (ms)
RGB Encoder	0.52	13.33	2.48
DEM Encoder	0.51	13.11	2.46
TMFI	0.26	1.08	0.31
TAGR	0.13	0.54	3.61
Decoder	1.24	23.09	6.36

4.12 Backbone-Agnostic Validation

To evaluate whether the proposed modules generalize across backbones, we replace SAM with ResNet-50 while keeping the remaining architecture unchanged.

As shown in Table 9, the proposed framework consistently improves both segmentation and boundary-aware metrics, indicating that the performance gains are not tied to a specific visual backbone. Negative sign in BE, CE, and HD95 is an improvement over the pre-trained backbone.

Table 9: Backbone-agnostic validation on the Bijie dataset.

Method	P	R	F1	IoU	mIoU	BE↓	CE↓	HD95↓	BF1↑
ResNet-50	72.70	86.61	79.05	65.36	80.16	0.4060	0.0712	24.57	44.57
ResNet-50 + TerrainGraphNet	**83.65**	86.34	**84.97**	**73.87**	**85.26**	**0.1514**	**0.0307**	**17.89**	**51.23**
Gain	+10.95	-0.27	+5.92	+8.51	+5.10	-0.255	-0.041	-6.68	+6.66

4.13 Impact of Graph Connectivity and Layer Depth

We examine the effect of varying the number of neighbors (K) and the number of graph attention layers (L) in the terrain-aware graph module. These parameters determine the extent of spatial interaction and message propagation across terrain-aware features. Table 10 summarizes the results. The configuration with

Table 10: Effect of terrain-aware graph configurations with different neighborhood sizes (K) and graph layer depths (L).

Configuration	P	R	F1	IoU	mIoU
$K = 4, L = 2$	88.84	**87.80**	**88.32**	**79.08**	**88.26**
$K = 4, L = 1$	88.95	87.39	88.16	78.83	88.13
$K = 8, L = 2$	88.99	86.86	87.91	78.43	87.91
$K = 16, L = 3$	**90.85**	83.67	87.11	77.16	87.23

$K = 4$ and $L = 2$ achieves the best overall performance, obtaining an F1 score of 88.32 and an IoU of 79.08. Increasing either the number of neighbors or the graph depth does not lead to further improvement. In particular, large connectivity ($K = 16$) leads to a noticeable performance degradation, indicating that excessive neighborhood aggregation may introduce irrelevant contextual information and dilute terrain-specific representations. These results suggest that moderate graph connectivity with shallow propagation depth provides the most effective balance between contextual modeling and feature discrimination.

5 Conclusion

Terrain plays a key role in landslide mapping, yet it is often treated as an auxiliary input in existing segmentation methods. In this work, we introduced TerrainGraphNet, which incorporates terrain information into both feature learning and graph-based reasoning to better capture the spatial structure of landslide regions. Results on three benchmark datasets demonstrate the effectiveness of this design for landslide segmentation. Future work will investigate richer geospatial information and more efficient terrain-aware reasoning to improve performance across diverse landscapes.

References

1. Cai, W., Huang, W., Tian, L., Huang, C., Yan, J.: Multiscale global attention for abnormal geological hazard segmentation. IEEE Sensors Journal **24**(10), 16961–16971 (2024). https://doi.org/10.1109/JSEN.2024.3383233
2. Chang, H., Bi, H., Li, F., Xu, C., Chanussot, J., Hong, D.: Deep symmetric fusion transformer for multimodal remote sensing data classification. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–15 (2024). https://doi.org/10.1109/TGRS.2024.3476975
3. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M.P., Zhang, S., Xing, L., Lu, L., Yuille, A., Zhou, Y.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis **97**, 103280 (Oct 2024). https://doi.org/10.1016/j.media.2024.103280
4. Chen, K., Liu, C., Chen, B., Zhang, J., Zou, Z., Shi, Z.: Rsrefseg 2: Decoupling referring remote sensing image segmentation with foundation models. IEEE Transactions on Geoscience and Remote Sensing **64**, 1–20 (2026). https://doi.org/10.1109/tgrs.2025.3647535
5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part VII. p. 833–851. Springer-Verlag, Berlin, Heidelberg (2018). https://doi.org/10.1007/978-3-030-01234-2_49
6. Chen, M., Zhang, J., Li, C., Guo, W., Fang, T., Zhao, Y., Ge, X., Xu, B., Hu, H., Zhu, Q.: Ls-dsperceiver: A dual-domain feature encoding and scale-aware perceiver for robust multiscenario landslide extraction. IEEE Transactions on Geoscience and Remote Sensing **64**, 1–19 (2026). https://doi.org/10.1109/TGRS.2026.3687990
7. Chen, X., Zhao, C., Liu, X., Zhang, S., Xi, J., Khan, B.A.: An embedding swin transformer model for automatic slow-moving landslide detection based on insar products. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–15 (2024). https://doi.org/10.1109/TGRS.2024.3470325
8. Chen, Y., Ming, D., Yu, J., Xu, L., Ma, Y., Li, Y., Ling, X., Zhu, Y.: Susceptibility-guided landslide detection using fully convolutional neural network. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **16**, 998–1018 (2023). https://doi.org/10.1109/JSTARS.2022.3233043
9. Chen, Y., Zhao, M., Bruzzone, L.: A novel approach to incomplete multimodal learning for remote sensing data fusion. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–14 (2024). https://doi.org/10.1109/TGRS.2024.3387837
10. Chendeb El Rai, M., Darweesh, M., Beya Far, A., Gawanmeh, A.: Multi-source attention-based fusion for segmentation of natural disasters. IEEE Geoscience and Remote Sensing Letters **21**, 1–5 (2024). https://doi.org/10.1109/LGRS.2024.3414421
11. Do, M.K., Han, K., Lai, P., Phan, K.T., Xiang, W.: Robsense: A robust multimodal foundation model for remote sensing with static, temporal, and incomplete data adaptability. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 7427–7436. IEEE (Jun 2025). https://doi.org/10.1109/cvpr52734.2025.00696
12. Dong, A., Dou, J., Li, C., Chen, Z., Ji, J., Xing, K., Zhang, J., Daud, H.: Accelerating cross-scene co-seismic landslide detection through progressive transfer learning and lightweight deep learning strategies. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–13 (2024). https://doi.org/10.1109/TGRS.2024.3424680

13. Fan, Y., Ma, P., Hu, Q., Liu, G., Guo, Z., Tang, Y., Wu, F., Zhang, H.: Scgc-net: Spatial context-guided calibration network for multisource rsi landslides detection. IEEE Transactions on Geoscience and Remote Sensing **63**, 1–17 (2025). https://doi.org/10.1109/tgrs.2025.3528750
14. Feng, H., Hu, Q., Zhao, P., Wang, S., Ai, M., Zheng, D., Liu, T.: Ftransdeeplab: Multimodal fusion transformer-based deeplabv3+ for remote sensing semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing **63**, 1–18 (2025). https://doi.org/10.1109/TGRS.2025.3553478
15. Gao, M., Chen, F., Wang, L., Zhao, H., Yu, B.: Swin transformer-based multiscale attention model for landslide extraction from large-scale area. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–14 (2024). https://doi.org/10.1109/tgrs.2024.3477910
16. Ghorbanzadeh, O., Xu, Y., Ghamisi, P., Kopp, M., Kreil, D.: Landslide4sense: Reference benchmark data and deep learning models for landslide detection. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–17 (2022). https://doi.org/10.1109/tgrs.2022.3215209
17. He, T., Chen, J.: Dc-gcn: A lightweight graph convolutional network integrating local and global context for semantic segmentation. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **18**, 28283–28299 (2025). https://doi.org/10.1109/jstars.2025.3626006
18. He, Z., Weng, X., Fang, K., Li, Y., Ruan, Y., Feng, H.: Brsmamba: Boundary-aware mamba for forest and shrub segmentation from diverse satellite imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **19**, 4607–4628 (2026). https://doi.org/10.1109/jstars.2025.3650425
19. Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J.A., Chanussot, J.: Spectralgpt: Spectral remote sensing foundation model. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(8), 5227–5244 (Aug 2024). https://doi.org/10.1109/tpami.2024.3362475
20. Hu, W., Sun, G., Zeng, X., Tong, B., Wang, Z., Wu, X., Song, P.: Hierarchical cross attention achieves pixel precise landslide segmentation in submeter optical imagery. Scientific Reports **15**(1) (Jul 2025). https://doi.org/10.1038/s41598-025-08695-8
21. 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 **17**(6), 1337–1352 (Feb 2020). https://doi.org/10.1007/s10346-020-01353-2
22. Kang, X., Hong, Y., Duan, P., Li, S.: Fusion of hierarchical class graphs for remote sensing semantic segmentation. Information Fusion **109**, 102409 (2024). https://doi.org/10.1016/j.inffus.2024.102409
23. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollár, P., Girshick, R.: Segment anything. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 3992–4003 (2023). https://doi.org/10.1109/ICCV51070.2023.00371
24. Li, H., Wang, J., Wu, S., Wang, Y., Gao, Y., Nie, G.: Ep-unet++: Advancing landslide monitoring through enhanced remote sensing semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing **64**, 1–16 (2026). https://doi.org/10.1109/TGRS.2026.3682705
25. Li, K., Liu, R., Cao, X., Bai, X., Zhou, F., Meng, D., Wang, Z.: Segearth-ov: Towards training-free open-vocabulary segmentation for remote sensing im-

ages. In: 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). p. 10545–10556. IEEE (Jun 2025). https://doi.org/10.1109/cvpr52734.2025.00986

26. Liang, J., Wei, G.: Hieraasgsegnet: Hierarchical context fusion for semantic segmentation via adaptive superpixel graph reasoning. IEEE Access **13**, 186449–186464 (2025). https://doi.org/10.1109/access.2025.3626744
27. 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 Geoscience and Remote Sensing Letters **21**, 1–5 (2024). https://doi.org/10.1109/lgrs.2024.3454319
28. 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 Transactions on Geoscience and Remote Sensing **61**, 1–14 (2023). https://doi.org/10.1109/tgrs.2022.3233637
29. Liu, X.: The 2022 luding earthquake landslide shape dataset (2023). https://doi.org/10.5281/zenodo.7860906, https://doi.org/10.5281/zenodo.7860906
30. Liu, Y., Gao, K., Wang, H., Yang, Z., Wang, P., Ji, S., Huang, Y., Zhu, Z., Zhao, X.: A transformer-based multi-modal fusion network for semantic segmentation of high-resolution remote sensing imagery. International Journal of Applied Earth Observation and Geoinformation **133**, 104083 (2024). https://doi.org/10.1016/j.jag.2024.104083
31. 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 Transactions on Geoscience and Remote Sensing **61**, 1–13 (2023). https://doi.org/10.1109/TGRS.2023.3313586
32. Ma, X., Zhang, X., Pun, M.O.: Rs3mamba: Visual state space model for remote sensing image semantic segmentation. IEEE Geoscience and Remote Sensing Letters **21**, 1–5 (2024). https://doi.org/10.1109/lgrs.2024.3414293
33. Ma, X., Zhang, X., Pun, M.O., Liu, M.: A multilevel multimodal fusion transformer for remote sensing semantic segmentation. IEEE Transactions on Geoscience and Remote Sensing **62**, 1–15 (2024). https://doi.org/10.1109/TGRS.2024.3373033
34. Nie, J., Huang, L., Zheng, C., Lv, X., Wang, R.: Cross-scale graph interaction network for semantic segmentation of remote sensing images. ACM Trans. Multimedia Comput. Commun. Appl. **19**(6) (May 2023). https://doi.org/10.1145/3558770
35. 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing **15**, 9746–9759 (2022). https://doi.org/10.1109/JSTARS.2022.3219897
36. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
37. Roy, S.K., Deria, A., Hong, D., Rasti, B., Plaza, A., Chanussot, J.: Multimodal fusion transformer for remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–20 (2023). https://doi.org/10.1109/TGRS.2023.3286826
38. Sun, Y., Yan, J., Qian, J., Xu, C., Yang, J., Luo, L.: Dual-perspective united transformer for object segmentation in optical remote sensing images. In: Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence. p.

1909–1917. IJCAI-2025, International Joint Conferences on Artificial Intelligence Organization (Sep 2025). https://doi.org/10.24963/ijcai.2025/213
39. Tang, X., Lu, Z., Fan, X., Yan, X., Yuan, X., Li, D., Li, H., Li, H., Raj Meena, S., Novellino, A., Nava, L., Catani, F.: Mamba for landslide detection: A lightweight model for mapping landslides with very high-resolution images. IEEE Transactions on Geoscience and Remote Sensing **63**, 1–17 (2025). https://doi.org/10.1109/tgrs.2025.3598446
40. Tang, Y., Hu, X., Ke, T., Zhang, M.: Semantic segmentation of high-resolution remote sensing imagery via an end-to-end graph attention network with superpixel embedding. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **18**, 7236–7252 (2025). https://doi.org/10.1109/JSTARS.2025.3542255
41. Verma, S., Lindseth, F., Kiss, G.: Segdesicnet: Lightweight semantic segmentation in remote sensing with geo-coordinate embeddings for domain adaptation. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). p. 9093–9104. IEEE (Feb 2025). https://doi.org/10.1109/wacv61041.2025.00881
42. Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J.M., Luo, P.: Segformer: Simple and efficient design for semantic segmentation with transformers. In: Neural Information Processing Systems (NeurIPS) (2021), https://dl.acm.org/doi/10.5555/3540261.3541185
43. Yang, L., Kang, B., Huang, Z., Zhao, Z., Xu, X., Feng, J., Zhao, H.: Depth anything v2. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. NIPS '24, Curran Associates Inc., Red Hook, NY, USA (2024), https://dl.acm.org/doi/10.5555/3737916.3738604
44. Zheng, Z., Ermon, S., Kim, D., Zhang, L., Zhong, Y.: Changen2: Multi-temporal remote sensing generative change foundation model. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(2), 725–741 (Feb 2025). https://doi.org/10.1109/tpami.2024.3475824
45. Zhou, G., Chen, W., Gui, Q., Li, X., Wang, L.: Split depth-wise separable graph-convolution network for road extraction in complex environments from high-resolution remote-sensing images. IEEE Transactions on Geoscience and Remote Sensing **60**, 1–15 (2022). https://doi.org/10.1109/TGRS.2021.3128033
46. Zhou, W., Fan, X., Yan, W., Shan, S., Jiang, Q., Hwang, J.N.: Graph attention guidance network with knowledge distillation for semantic segmentation of remote sensing images. IEEE Transactions on Geoscience and Remote Sensing **61**, 1–15 (2023). https://doi.org/10.1109/TGRS.2023.3311480
47. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. pp. 3–11. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_1
48. Zhu, X., Zhang, X., Zhang, T., Fang, C., Tang, X., Jiao, L.: Regionmatch: Pixel-region collaboration for semi-supervised semantic segmentation in remote sensing images. In: Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence. p. 2530–2538. IJCAI-2025, International Joint Conferences on Artificial Intelligence Organization (Sep 2025). https://doi.org/10.24963/ijcai.2025/282
49. Şener, A., Ergen, B.: Landslidesegnet: an effective deep learning network for landslide segmentation using remote sensing imagery. Earth Science Informatics **17**(5), 3963–3977 (Aug 2024). https://doi.org/10.1007/s12145-024-01434-z