

Synergistic fusion network for landslide segmentation: the vital complementarity of geological information to remote sensing imagery

Guangle Yao, Chuyi Zhou, Haoran Zhou, Shaoze Ye, Yi Yang, Xiaodong Zhang & Honghui Wang

To cite this article: Guangle Yao, Chuyi Zhou, Haoran Zhou, Shaoze Ye, Yi Yang, Xiaodong Zhang & Honghui Wang (2025) Synergistic fusion network for landslide segmentation: the vital complementarity of geological information to remote sensing imagery, International Journal of Parallel, Emergent and Distributed Systems, 40:3, 282-295, DOI: [10.1080/17445760.2024.2352881](https://doi.org/10.1080/17445760.2024.2352881)

To link to this article: <https://doi.org/10.1080/17445760.2024.2352881>



Published online: 17 May 2024.



Submit your article to this journal [↗](#)



Article views: 169



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)



Synergistic fusion network for landslide segmentation: the vital complementarity of geological information to remote sensing imagery

Guangle Yao^{a,b,c}, Chuyi Zhou^c, Haoran Zhou^{d,e}, Shaoze Ye^{d,e}, Yi Yang^a, Xiaodong Zhang^c and Honghui Wang^{b,c}

^aChinese Academy of Surveying and Mapping, Beijing, People's Republic of China; ^bState Key Laboratory of Geohazard Prevention and Geoenvironment Protection (Chengdu University of Technology), Chengdu, Sichuan, People's Republic of China; ^cCollege of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu, Sichuan, People's Republic of China; ^dKey Laboratory of Natural Disaster Monitoring, Early Warning and Assessment of Jiangxi Province (Jiangxi Normal University), Nanchang, Jiangxi, People's Republic of China; ^eShenzhen Sensing Data Technology Co., Ltd., Shenzhen, Guangdong, People's Republic of China

ABSTRACT

In the contemporary technological environment, access to a multitude of data sources, such as digital elevation models (DEM), slope data, and other geological information, has become essential for augmenting remote sensing imagery in landslide analysis. These datasets enhance landslide identification capabilities, but existing methods often fail to fully leverage their unique characteristics due to oversimplified fusion strategies. To overcome this challenge, we introduce the Synergistic Fusion Network (SFNet), which is specifically designed to utilize the complementary nature of geological data to significantly improve the accuracy and efficiency of landslide segmentation. SFNet incorporates Multi-Source Fusion Modules (MSFM) and Reverse Feature Refinement Modules (RFRM), through an innovative network architecture, enhancing remote sensing imagery with integrated geological information. In our experiments, SFNet outperforms existing mainstream networks, demonstrating superior F1 scores, thereby showcasing its ability to effectively segment landslide regions with high precision. This advancement highlights the importance of considering the unique contributions of each data source, especially the complementary effects of geological information.

ARTICLE HISTORY

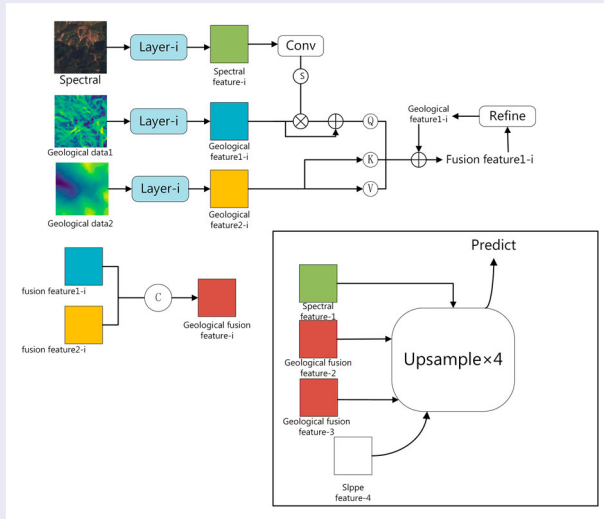
Received 7 April 2024

Accepted 4 May 2024

KEYWORDS

Landslide segmentation; geological information; remote sensing imagery; multisource fusion; attention mechanism

CONTACT Guangle Yao  guangle.yao@cdut.edu.cn  Chinese Academy of Surveying and Mapping, Beijing, People's Republic of China; State Key Laboratory of Geohazard Prevention and Geoenvironment Protection (Chengdu University of Technology), Chengdu, Sichuan, 610059, People's Republic of China; College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu, Sichuan 610059, People's Republic of China



1. Introduction

Landslides, known for their devastating consequences, require prompt and accurate identification for efficient disaster management. Historically, manual analysis of remote sensing imagery was the primary method for landslide identification [1]. This approach, while insightful, was labour-intensive, time-consuming, and prone to human error. Furthermore, manual interpretation lacked consistency and scalability, limiting its practical application for large-scale landslide identification projects.

With the advent of deep learning, there has been a monumental shift in landslide identification towards automated and data-driven approaches. Deep learning-based semantic segmentation networks have emerged as powerful tools for automatically delineating landslide-affected areas from remote sensing imagery. FCN [2] was one of the pioneering deep learning models for semantic segmentation. It uses convolutional layers for feature extraction and upsampling layers to generate dense segmentation maps, enabling end-to-end learning for semantic segmentation tasks. Subsequently, many excellent networks have been proposed for segmentation tasks, and applying these advancements to landslide segmentation has consistently shown promising performance [3–7].

However, the variability in morphology and local blurriness in landslide imagery present significant obstacles for precise segmentation. Current algorithms, heavily reliant on imagery, are hitting precision plateaus, with marginal gains from network refinement alone. Recognizing that effective landslide segmentation depends on a multitude of factors, not just imagery, underscores the importance of complementary information provided by various data types, such as geological information. Geological data introduce additional, comprehensive features of landslides, reducing data ambiguity and improving segmentation accuracy. This approach has sparked an increased interest in exploiting geological information comprehensively for more effective landslide segmentation.

Each data type highlights distinct aspects, making it crucial to efficiently and effectively combine their strengths. Direct concatenation of remote sensing imagery and geological data often fails to fully exploit the complementary information from different sources, as it can introduce redundancy or noise. In recent years, a surge of research has focused on leveraging geological information for remote sensing imagery segmentation tasks, with various approaches proposed. For instance, [8] compared a one-layer CNN with deeper architectures, such as residual networks, and employed layer

stacking and feature-level fusion techniques for improved landslide segmentation. This study [9] introduced a method that fused spectral and DEM data during preprocessing, employing convolutional neural networks and channel attention mechanisms. In another study [10], they combined geological features like terrain morphology, elevation characteristics, and river distribution with spectral information. By concatenating these features and using dot product operations to assign weights to multispectral data, they developed an automatic fusion technique. Zhang et al. [11] proposed a dual-encoder approach, where they separately extracted features from spectral and elevation data, then concatenated them and fed them into a convolutional neural network to learn complementary features. However, these methods do not fully capitalize on the geological information, failing to extract its full potential.

To address this limitation, we present a network that achieves synergistic fusion of multispectral, slope, and DEM data, effectively leveraging the complementary information provided by geological data. Our approach includes innovative fusion mechanisms, reverse feature refinement modules to enhance the geological feature extraction branch and skip connections between encoders and decoders to facilitate effective fusion of diverse information. These design elements enable us to maximize the benefits of each data source, ultimately resulting in more accurate landslide segmentation.

The main contributions of this paper are:

- (1) Our research explores the significant role of geological data in landslide segmentation and emphasizes the crucial role of geological data in complementing remote sensing imagery for landslide segmentation. This emphasis sets the stage for further investigation into fully leveraging the complementary effects of geological information to enhance remote sensing imagery for more accurate and comprehensive landslide analysis.
- (2) We propose Synergistic Fusion Network (SFNet), an innovative network for landslide segmentation crafted to address the underutilization of geological information. SFNet leverages skip connections and our designed Multi-source Fusion Module, which maximizes the utilization of the complementary information provided by geological data. Moreover, our novel Reverse Feature Refinement Module integrates multi-source information into the intermediate feature maps of the geological branch, guiding the network to learn more comprehensive and informative feature representations.

2. Proposed synergistic fusion network

2.1. The overall framework of the network

The overall architecture of our proposed SFNet, depicted in Figure 1, is an encoder-decoder design. The encoder part of our network mainly consists of a multi-branch feature extraction module (MFEM) that comprises three distinct deep feature extraction networks, a channel attention module (CAM) that enhances feature representation by emphasizing important channels, multisource fusion modules (MSFM) integrate the features from different data sources, achieving seamless fusion among them, and reverse feature refinement modules (RFRM) to refine the pre-fusion geological feature maps with multispectral features or comprehensive multisource information obtained after fusion. The following sections provide detailed descriptions of each module mentioned above. For the decoder, we adopt the decoder structure proposed by Res-Unet [5], utilizing residual blocks in each upsampling section to alleviate the issues of semantic and gradient vanishing.

For the skip connections between the encoder and decoder, in the deep features of the encoder, we concatenate the features of two geological information types refined by MSFM with the corresponding layers of the decoder. As for the shallow features of the encoder, we concatenate the multispectral features processed by CAM into the corresponding layers of the decoder. The deep features refined by MSFM integrate multispectral information, providing rich semantic context for landslide segmentation, while still preserving their unique geological characteristics to a certain extent. This

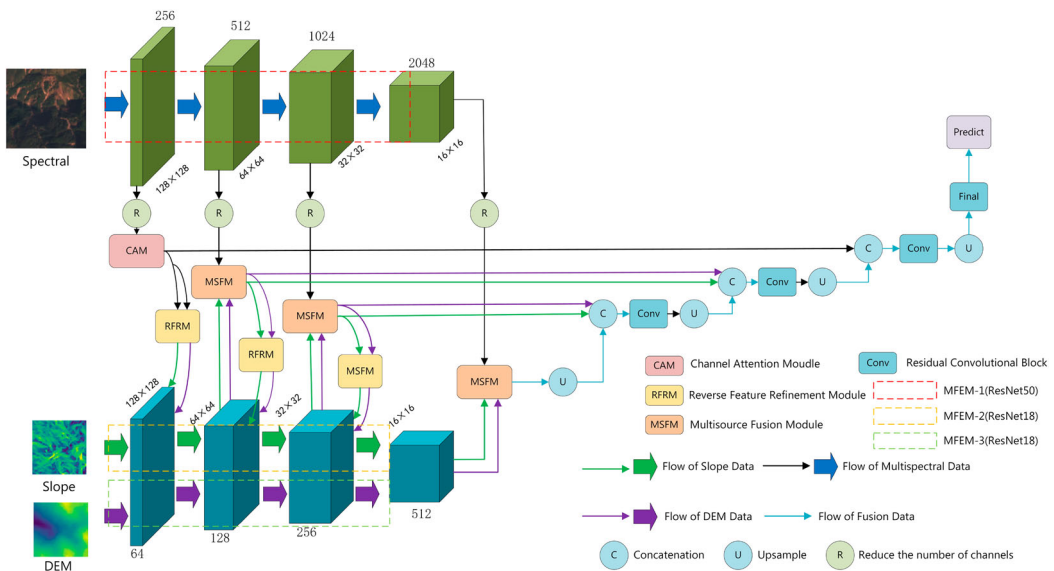


Figure 1. Overall framework of proposed synergistic fusion network.

preservation of their unique geological characteristics is crucial. The overall terrain characteristics derived from the deep features of geological information are pivotal for landslide segmentation, as landslide occurrences often correlate with changes in terrain morphology. Additionally, geological data exhibit strong geographical relevance and stability, remaining unaffected by external factors such as weather and lighting conditions, thus enhancing the model's adaptability across diverse regions and scenarios. Conversely, the integration of shallow features from multispectral data is crucial, as multispectral features are more responsive to local surface characteristics, accurately capturing fine-grained surface details and providing precise location cues essential for accurate landslide localization. Through this integration strategy, we harness the strengths of each data source, effectively combining complementary data to enable more comprehensive and accurate landslide segmentation.

2.2. Multibranch feature extraction module

Multispectral remote sensing imagery possess detailed features like colour, texture, and shape, which are crucial for landslide segmentation. These features are complex and require careful extraction. Consequently, a more powerful deep feature extraction network is necessary for effective extraction. Conversely, geological information offers more straightforward and easily accessible information. Directly incorporating geological information into a feature extraction network of the same depth as multispectral imagery would lead to redundant parameters, increased training costs, and a higher risk of severe overfitting, ultimately degrading the network's practical performance.

Therefore, as illustrated in Figure 1, we adopt ResNet50 [12] as the backbone network for extracting features from multispectral imagery. For the two types of geological information, slope data and DEM, we utilize a lighter network ResNet18, to extract features separately. Since the input is a high-resolution 128×128 remote sensing imagery, excessive downsampling might lead to the loss of crucial detailed information. Therefore, we removed the second downsampling layer from each of the three ResNet network branches. Consequently, the output of the final encoding part is reduced to 1/16 of the original input size.

2.3. Channel attention module

The multispectral data contains a wealth of spectral information, yet not all of it is essential for landslide segmentation. Some information may have minimal impact or even cause interference in the

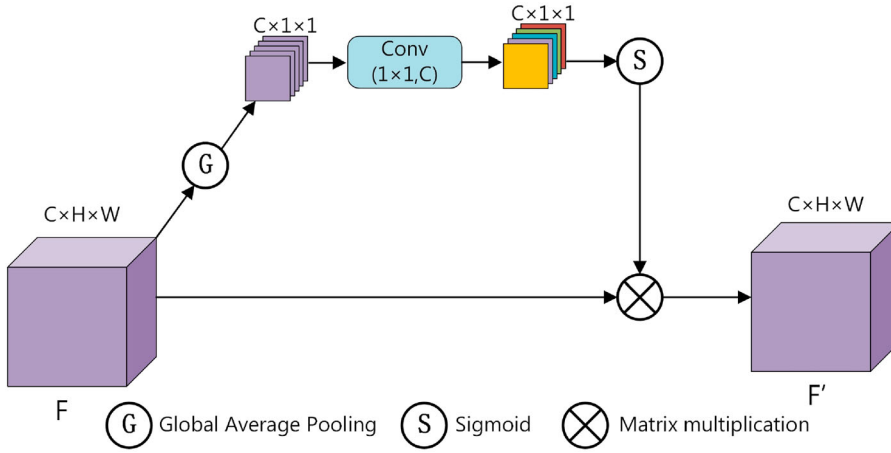


Figure 2. Channel attention module.

segmentation process. Furthermore, multispectral data can be affected by factors like atmospheric conditions and cloud cover, rendering certain bands unreliable. As a result, there is a significant amount of redundant information in multispectral data. To selectively extract the most important parts of multispectral features and geological features for landslide segmentation, we utilize a channel attention mechanism.

Our channel attention module is similar to the Squeeze-and-Excitation (SE) [13] block. However, distinct from the SE block, our module preserves the original dimensionality during channel weighting. This choice is made to prevent potential information loss or degradation that could result from dimensionality reduction via fully connected layers. Even though dimensionality reduction followed by dimensionality restoration may theoretically enhance channel representation, the loss of information during the reduction process could result in incomplete or inaccurate information recovery.

As shown in Figure 2, given the input features $F \in \mathbb{R}^{C \times H \times W}$, we first perform global average pooling $\in \mathbb{R}^{C \times 1 \times 1}$. Then, we adopt a 1×1 convolutional layer to capture the importance of each channel, and the output is passed through a sigmoid function to obtain the weights for each channel. These weights accurately capture the contribution of each channel to the overall feature representation. By assigning these weights to the input features, we obtain the final feature map $F' \in \mathbb{R}^{C \times H \times W}$ of the weighted feature representation.

The above process can be summarized as follows:

$$F' = \sigma(\text{Conv}_C^{1 \times 1}(\text{GAP}(F))) \odot F \quad (1)$$

where GAP represents Global Average Pooling; σ represents the sigmoid function; $\text{Conv}_C^{1 \times 1}$ represents a 2D convolution operation with output channels of C and a kernel size of 1×1 ; $\odot$ represents element-wise multiplication.

2.4. Multi-source fusion module

Due to insufficient feature extraction and significant noise in the shallow feature maps of geological information, we opt not to perform multi-source data fusion in the first layer of the encoder in order to fully leverage the fine-grained information provided by multispectral data.

The MSFM module is composed of two instances of the structure shown in Figure 3. As illustrated in Figure 1, the module receives inputs of multispectral features, slope features, and DEM feature maps. These feature maps are processed by each of the two identical structures sequentially. However, the entry points for the slope and DEM features are different in each structure, resulting in two distinct

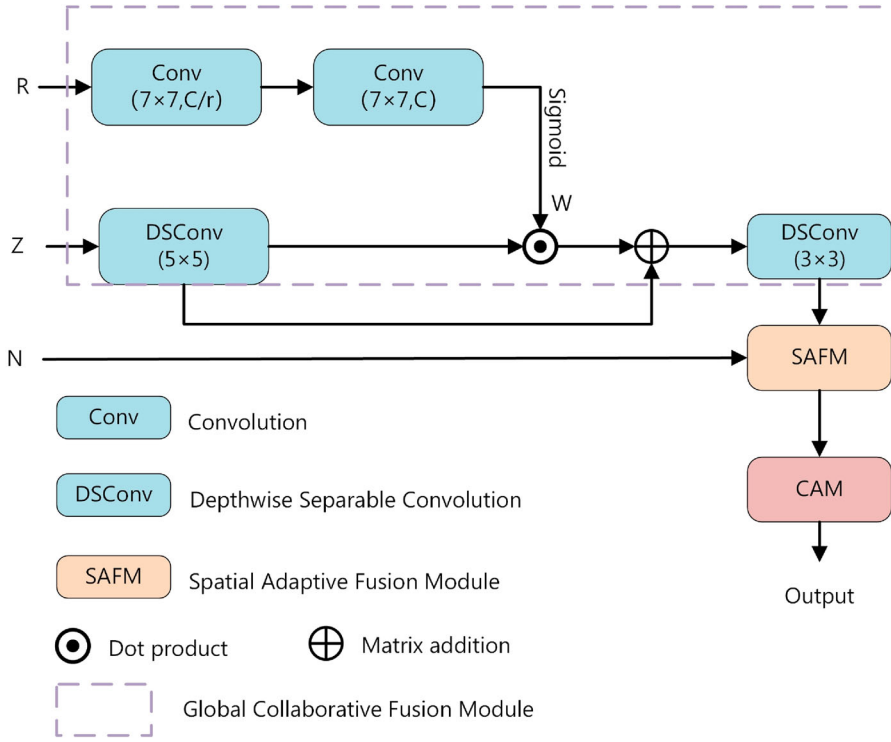


Figure 3. Multi-source fusion structure.

outputs. Each of the two outputs represents a feature map that has undergone refinement through additional data processing, aggregating multisource information while still retaining its unique geological characteristics. As shown in Figure 3, the module first takes the multispectral information as input R , the slope information as input Z , and the DEM information as input N . Next, the slope information is fused with the multispectral information through the global collaborative fusion module, integrating multispectral features into the slope information. Then, through the spatial adaptive fusion module, global information from the DEM is extracted. Subsequently, the multispectral information is taken as input R , the DEM information as input Z , and the slope information as input N to refine the DEM features using the multispectral and slope information.

2.4.1. Global collaborative fusion module

Each pixel value in the geological feature map reflects the corresponding surface's slope degree or height; its information is limited to surface geological characteristics. Multispectral information is more sensitive to changes in specific surface features. By integrating multispectral information into the geological feature map, the geological extraction branch learns from multiple types of data. This enables the geological feature branch to focus on complex landslides with smaller slopes or less distinct geological characteristics.

For the fusion of multispectral and geological information, we propose a global collaborative fusion module. As illustrated in Figure 3, we employ a spatial weighted fusion strategy. In this strategy, when refining geological features using multispectral features, the multispectral features serve as input $R \in \mathbb{R}^{C \times H \times W}$, while the geological features serve as input $Z \in \mathbb{R}^{C \times H \times W}$. To extract spatial weights, R is first processed through a convolutional layer with a kernel size of 7 to reduce channel numbers, thereby decreasing computational complexity. Subsequently, the output undergoes another convolutional layer with a kernel size of 7 to increase channel numbers, maintaining consistency in channel numbers.

Finally, the output passes through a sigmoid function to obtain the weight $w \in \mathbb{R}^{C \times H \times W}$. This weight fully reflects the global contextual information of the multispectral data.

The process of extracting spatial weight is as follows:

$$w = \sigma(\text{Conv}_C^{7 \times 7}(\text{Conv}_{C_1}^{7 \times 7}(R))) \quad (2)$$

where $C_1 = C/r$; $\text{Conv}_{C_1}^{7 \times 7}$ represents a 2D convolution operation with output channels of C_1 and a kernel size of 7×7 ; σ represents the sigmoid function.

Next, we perform element-wise multiplication and matrix addition between the obtained weight w and feature map Z , followed by further feature transformation through depth-wise separable convolution, resulting in the fused output $\in \mathbb{R}^{C \times H \times W}$.

The entire fusion process is as follows:

$$Z' = \text{Conv}((w \cdot \text{Conv}(Z)) + Z) \quad (3)$$

where Conv represents Depthwise separable convolution.

2.4.2. Spatial adaptive fusion module

For the fusion of geological features, we introduce a spatial adaptive fusion module. This module computes spatial dependencies between any two positions in slope and DEM features, allowing one type of geological data to leverage the global context provided by the other. Based on feature similarity, it integrates additional geological data with varying weights at each position. Consequently, this module adjusts fusion weights adaptively for distinct features. This enhances the integration of valuable information for landslide segmentation while minimizing the inclusion of irrelevant data. The following section elaborates on this module in detail.

As illustrated in Figure 4, Given input features $X \in \mathbb{R}^{C \times H \times W}$ and features $Y \in \mathbb{R}^{C \times H \times W}$, we first pass each of X and Y through a convolutional layer to reduce the number of channels to the same size, resulting in $Q \in \mathbb{R}^{C_1 \times H \times W}$ and $K \in \mathbb{R}^{C_1 \times H \times W}$. Then, we pass Y through another convolutional layer to maintain the same number of channels, generating $V \in \mathbb{R}^{C \times H \times W}$. The process of obtaining Q, K, V is as follows:

$$Q = \text{Conv}_{C_1}^{1 \times 1}(X) \quad (4)$$

$$K = \text{Conv}_{C_1}^{1 \times 1}(Y) \quad (5)$$

$$V = \text{Conv}_C^{1 \times 1}(Y) \quad (6)$$

where $\text{Conv}_{C_1}^{1 \times 1}$ represents 2D convolution with 1×1 kernel for channel reduction; $\text{Conv}_C^{1 \times 1}$ represents 1×1 convolution that preserves channels.

Then, we reshape $Q \in \mathbb{R}^{C_1 \times N}$ and transpose it to $\mathbb{R}^{N \times C_1}$, where $N = H \times W$, represents the total number of pixels. Similarly, we reshape $K \in \mathbb{R}^{C_1 \times N}$. Next, we perform matrix multiplication between Q and K , followed by applying the softmax function to the result, obtaining $W \in \mathbb{R}^{N \times N}$. W represents the spatial dependency between each position in the one type of geological features and each position in another type of geological features.

The entire process of obtaining W is as follows:

$$W = \text{softmax}(\text{Trans}(\text{Fla}(Q)) \times \text{Fla}(K)) \quad (7)$$

where softmax represents the softmax function; Trans represents transpose; Fla represents flatten; $\times$ represents matrix multiplication.

For the weight W , we transpose it to $\mathbb{R}^{N \times N}$ and perform matrix multiplication with V , which has been reshaped to $\mathbb{R}^{C \times N}$. Then, we reshape the resulting matrix to $\mathbb{R}^{C \times H \times W}$. The process to obtain the final output is represented by the following equation:

$$\text{Output} = (\text{Fla}(V) \times \text{Trans}(W)) + X \quad (8)$$

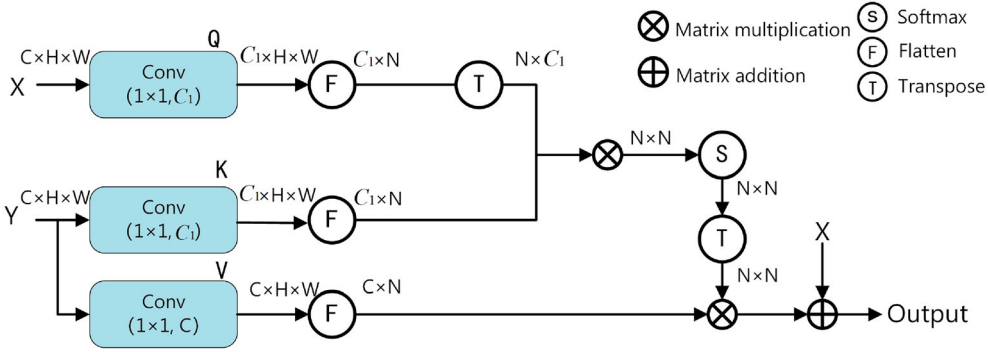


Figure 4. Spatial adaptive fusion module.

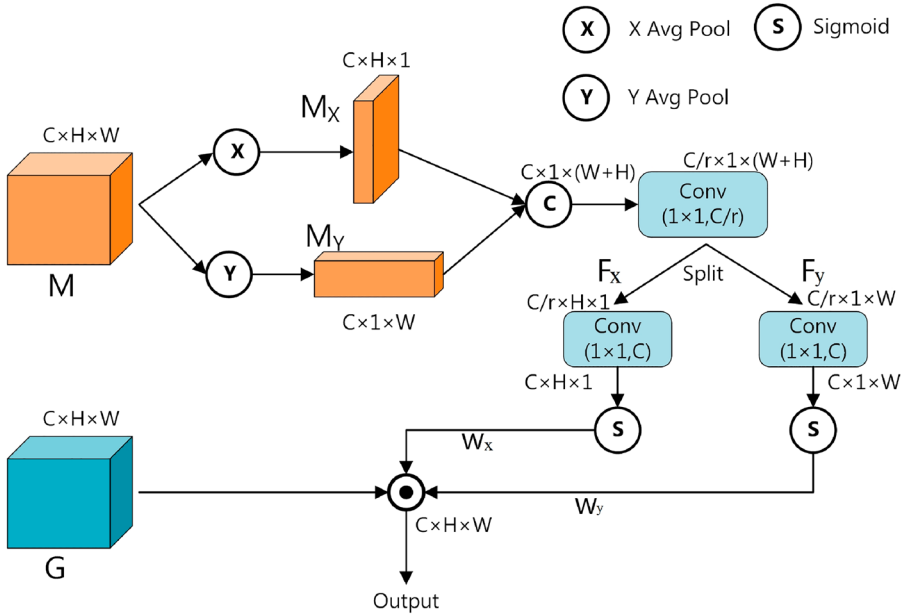


Figure 5. Reverse feature refinement structure.

2.5. Reverse feature refinement module

As shown in Figure 1, each layer, except for the final one, includes a reverse feature refinement module, which consists of two instances of the structure depicted in Figure 5. Each structure is used for the reverse refinement of a different type of geological feature map within that layer. For the multispectral features or those refined after passing through the MSFM module, we reverse-propagate them to the corresponding layers of geological features or the original features. The fused features integrate information from multispectral data, DEM, and slope data. Through the reverse refinement module, these features are fed back to the intermediate layers of the network, allowing the network to learn more discriminative and informative feature representations, thereby enhancing its ability to perceive and recognize surface features. Moreover, this process guides the network to learn the relationships between surface features more efficiently, accelerating convergence and improving training effectiveness. Our reverse feature refinement module is illustrated in Figure 5.

Table 1. Hardware and software information.

Experiential environment	Specific information
GPU	Nvidia Geforce RTX 3070
CPU	Intel(R) Core(TM) i7-10700
Memory	32 GB
Pytorch	2.1.1

We adopt a coordinate attention mechanism [14], which captures not only inter-channel information but also direction-aware and position-sensitive details, effectively extracting key information from the feature maps. Next, we will provide a detailed explanation.

First, the multispectral feature map or the fused feature map is taken as input $M \in \mathbb{R}^{C \times H \times W}$, and average pooling operations are separately applied along the X and Y directions to obtain two output feature maps. These two feature maps retain precise positional information in their respective directions and capture long-distance spatial dependencies in the other direction. For the output of the c th channel with height h or width w after specific spatial direction pooling, it is as follows:

$$M_X(c, h) = \frac{1}{W} \sum_{i=1}^W M(c, h, i) \quad (9)$$

$$M_Y(c, w) = \frac{1}{H} \sum_{i=1}^H M(c, i, w) \quad (10)$$

where X represents output after average pooling operation along the X-direction; Y represents output after average pooling operation along the Y-direction.

Next, the two feature maps are concatenated and fed into a 1×1 convolutional layer to extract channel relationships across different spatial positions. The output $\in \mathbb{R}^{C/r \times 1 \times (W+H)}$ of the convolution is split back into the original two feature maps. The above operations can be expressed by the following equation:

$$F_X, F_Y = \text{Split}(\text{Conv}_{C/r}^{1 \times 1}([M_X, M_Y])) \quad (11)$$

where $[]$ represents concatenation.

Each of these outputs $F_X \in \mathbb{R}^{C/r \times H \times 1}$, $F_Y \in \mathbb{R}^{C/r \times 1 \times W}$ then enters a separate 1×1 convolutional layer to restore the channel numbers. After passing through a sigmoid function, the resulting two spatial direction weights $w_X \in \mathbb{R}^{C \times H \times 1}$, $w_Y \in \mathbb{R}^{C \times 1 \times W}$ are obtained. Finally, we perform element-wise multiplication between the geological feature map $G \in \mathbb{R}^{C \times H \times W}$ and the two weight tensors broadcasted to the same shape. The final refined geological feature map $O \in \mathbb{R}^{C \times H \times W}$ is obtained after the element-wise multiplication. The final process can be described by the following equation:

$$w_X = \sigma(\text{Conv}_C^{1 \times 1}(F_X)) \quad (12)$$

$$w_Y = \sigma(\text{Conv}_C^{1 \times 1}(F_Y)) \quad (13)$$

$$O(c, i, j) = G(c, i, j) \odot w_X(c, i) \odot w_Y(c, j) \quad (14)$$

where σ represents sigmoid function; O represents Output; $\odot$ represents element-wise multiplication.

3. Experiments and results analysis

3.1. Experimental setup

Our experiments are based on the PyTorch, with details of the experimental setup provided in Table 1. The relevant hyperparameters of our proposed network are presented in Table 2.

Table 2. Model parameter.

Parameter name	Parameter value	Note
Weight decay	0.0005	Weight decay coefficient
lr	0.001	Initial learning rate
Batch size	16	Number of training batches
Epochs	100	Training epochs

We adopt Landslide4Sense [15] as the dataset, as depicted in Figure 6. The Landslide4Sense dataset is divided into a training set consisting of 3799 image patches and a test set consisting of 800 image patches. Our model is trained on the entire training set for 100 epochs and evaluated on the entire test set. Each image patch in the dataset consists of 14 channels. The first 12 channels contain multispectral data from Sentinel-2, the 13th channel contains slope information from Sentinel-2, and the 14th channel contains DEM information from ALOS PALSAR. All image patches are of size 128 x 128 pixels. We selected a sample and visualized its RGB image and all bands along with the corresponding mask, as showed in Figure 7.

We adopted the cross-entropy loss function and utilized the Adam optimizer. The formula for the cross-entropy loss function is as follows:

$$\mathcal{L} = \sum_{i=1} y_i \log(\hat{y}_i) \quad (15)$$

where $\mathcal{L}$ represents loss; y_i represents the ground truth label of the i th pixel; $\hat{y}_i$ represents the predicted value for the i th pixel.

In this study, we employed warm-up training, where the learning rate linearly increases from 0 to the initial learning rate during the first epoch. Subsequently, the learning rate gradually decreases over the following epochs until it reaches 0 at the end of the final epoch. The variation of the learning rate during training is illustrated in Figure 8.

3.2. Evaluation metrics

In this section, we employ the F1 score as the evaluation metric to assess the performance of our proposed landslide segmentation model. The F1 score, a balanced measure of precision and recall, is particularly suitable for imbalanced datasets where one class may significantly outnumber the other. It takes into account both false positives and false negatives, which are crucial for evaluating the model's performance in such scenarios.

The F1 score is calculated using the following formula:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (16)$$

Precision is the ratio of true positive (TP) predictions to the total predicted positives, and it measures the accuracy of positive predictions. It is calculated as follows:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (17)$$

Recall is the ratio of true positive (TP) predictions to the total actual positives, and it measures the ability of the model to correctly identify positive samples. It is calculated as follows:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (18)$$

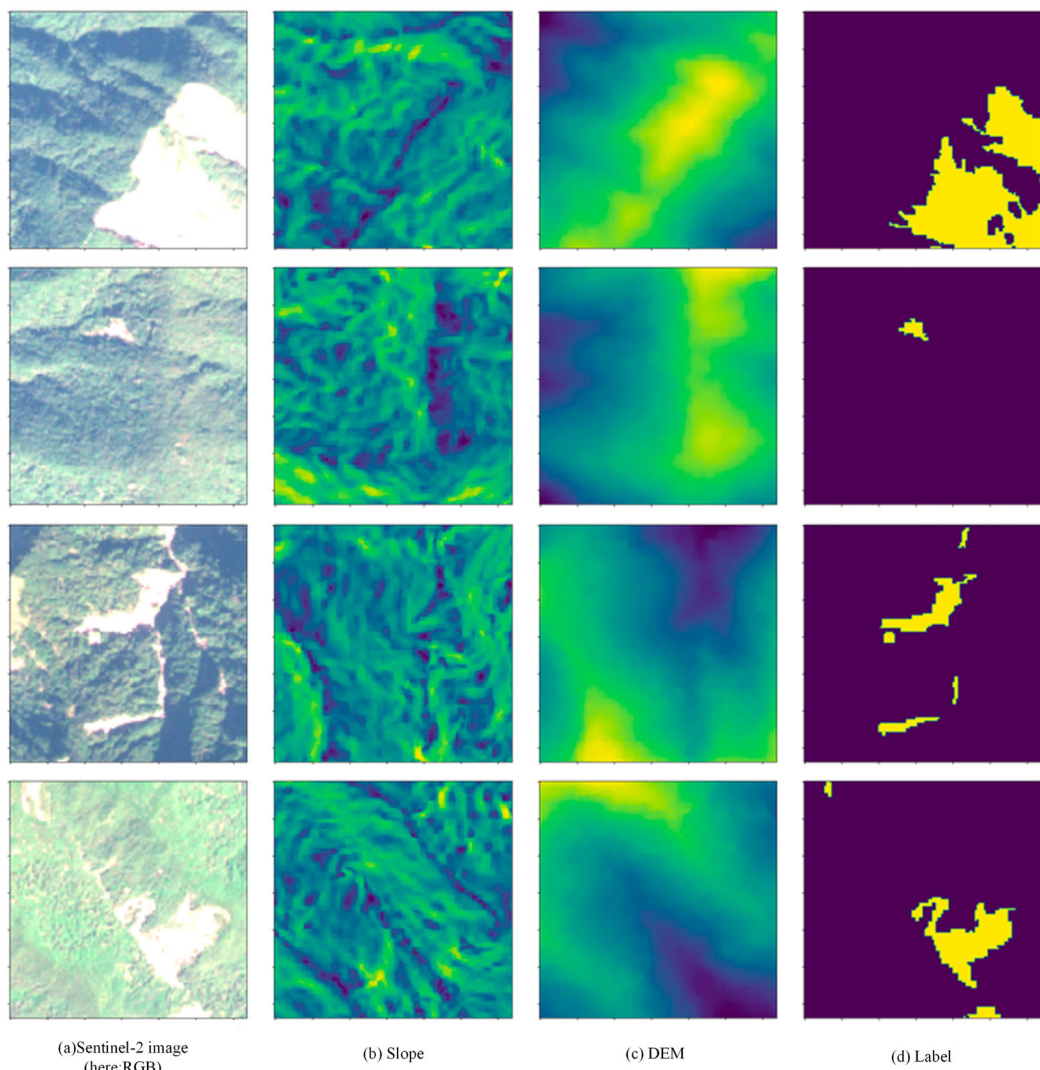


Figure 6. Landslide4Sense dataset [15].

3.3. Experimental results and analyzes

Based on the established experimental setup, dataset, and evaluation metrics, we will compare our approach with state-of-the-art models and conduct ablation experiments to validate the effectiveness of our proposed modules. We compared our model with several state-of-the-art semantic segmentation models, and the experimental results are presented in Table 3. The comparative experiments clearly show that our proposed model achieves significant performance enhancement in landslide segmentation tasks. Our model outperforms state-of-the-art semantic segmentation models in both accuracy and generalization ability.

In the ablation study, we will systematically assess the model's performance by sequentially removing or disabling each proposed component and evaluating its effect on segmentation outcomes. We use a single branch input for multispectral information, with a feature extraction network based on ResNet50 and a decoder derived from ResUNet as the baseline. The experimental findings are detailed in Table 4. In Table 4, the Multi-Branch Feature Extraction Module is denoted as MFEM, the Global

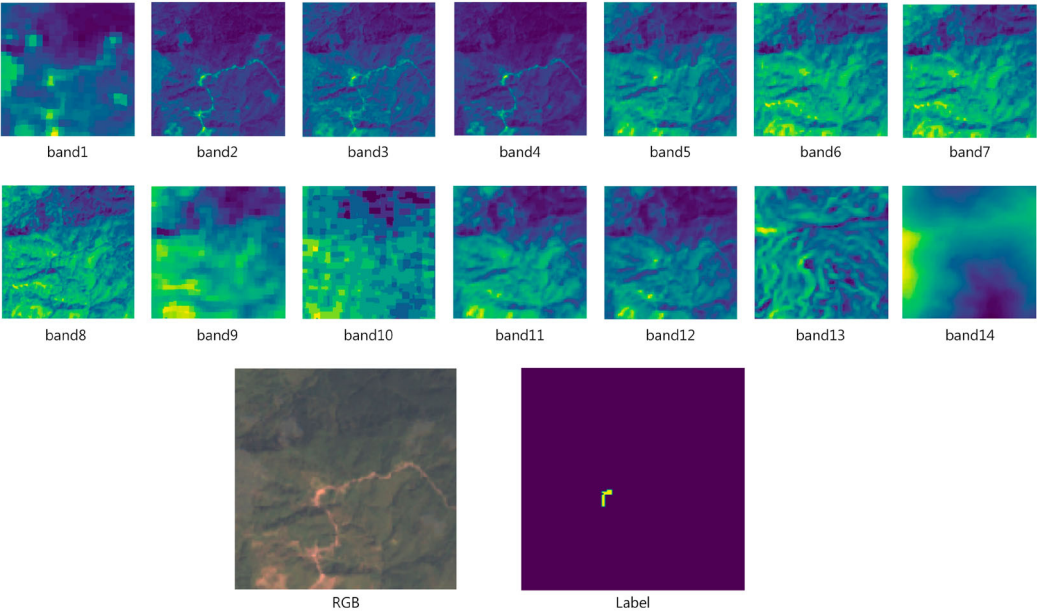


Figure 7. Visualization.

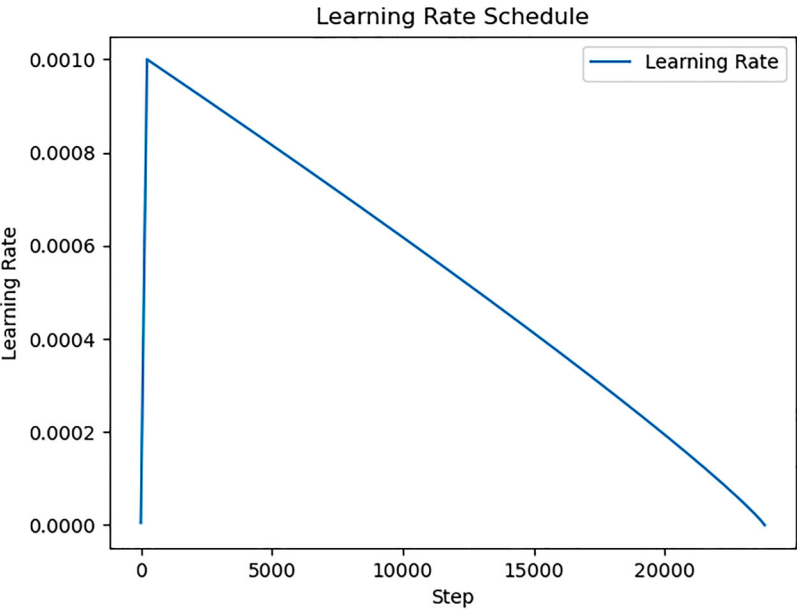


Figure 8. The variation of the learning rate.

Collaborative Fusion Module is denoted as GCFM, the Spatial Adaptive Fusion Module is denoted as SAFM, and the Reverse Feature Refinement Module is denoted as RFRM. The results of the ablation experiments show that each module in our approach positively contributes to landslide segmentation performance. Specifically, the GCFM effectively integrates multispectral information into geological features when employing multi-branch networks of varying depths. The SAFM dynamically adjusts fusion weights based on different scenes, enhancing the fusion of geological features. Furthermore,

Table 3. Comparison of various model experiments.

Model	F1 Score
FCN [2]	60.69
Unet [3]	60.33
deeplabV3 [4]	62.88
ResUnet [5]	61.86
ResUnet++ [16]	63.59
HRnet [17]	61.34
Segformer [7]	58.20
SYnet(ours)	68.61

Table 4. Results of ablation experiments.

Method	F1 Score
baseline(12 channels)	60.12
baseline(14 channels)	62.96
baseline+MFEM+GCFM	64.54
baseline+MFEM+SAFM	64.89
baseline+MFEM+GCFM+SAFM	65.93
baseline+MFEM+GCFM+RFRM	66.40
baseline+MFEM+SAFM+RFRM	66.77
baseline+MFEM+GCFM+SAFM+RFRM(Our final model)	68.61

the RFRM improves the representational capacity of geological features, facilitating more precise localization and segmentation of landslide areas.

4. Conclusion

In this study, we proposed a novel landslide segmentation network aimed at leveraging the complementary role of geological information for enhanced landslide segmentation. Different-depth multi-branch networks were employed to extract features from various data sources. By introducing a module to integrate multispectral information into geological features, we tackled the limitations of geological features. We also proposed a module for adaptive fusion to efficiently integrate geological features. Additionally, the introduced RFRM empowered the geological feature network to learn stronger feature representations, unconstrained by a single data source. Moreover, by utilizing skip connection structures, we incorporated rich semantic features provided by geological information. Furthermore, by fusing shallow features from multispectral data, we further enriched the feature representation. This enabled us to achieve precise localization of landslides by accurately reflecting detailed surface characteristics and providing precise location information. Through experiments on the Landslide4Sense dataset, we demonstrated the effectiveness of our modules. Comparisons with various state-of-the-art networks further validated the superior accuracy of our network in landslide segmentation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was Funded by Sichuan Science and Technology Program (MZGC20230081) and also funded by State Key Laboratory of Geo-Information Engineering and Key Laboratory of Surveying and Mapping Science and Geospatial Information Technology of MNR, CASM (2023-05-14).

References

- [1] Tsai F, Hwang J-H, Chen L-C, et al. Post-disaster assessment of landslides in southern Taiwan after 2009 Typhoon Morakot using remote sensing and spatial analysis. *Nat Hazards Earth Syst Sci.* **2010**;10(10):2179–2190. doi: [10.5194/nhess-10-2179-2010](https://doi.org/10.5194/nhess-10-2179-2010)
- [2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Boston (MA), USA; 2015. p. 3431–3440.
- [3] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III* 18. Springer; 2015. p. 234–241.
- [4] Chen L-C, Zhu Y, Papandreou G, et al. Encoder–decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)*; Munich, Germany; 2018. p. 801–818.
- [5] Zhang Z, Liu Q, Wang Y. Road extraction by deep residual u-net. *IEEE Geosci Remote Sens Lett.* **2018**;15(5):749–753. doi: [10.1109/LGRS.2018.2802944](https://doi.org/10.1109/LGRS.2018.2802944)
- [6] Fu J, Liu J, Tian H, et al. Dual attention network for scene segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; Long Beach (CA), USA; 2019. p. 3146–3154.
- [7] Xie E, Wang W, Yu Z, et al. Segformer: simple and efficient design for semantic segmentation with transformers. *Adv Neural Inf Process Syst.* **2021**;34:12077–12090.
- [8] Sameen MI, Pradhan B. Landslide detection using residual networks and the fusion of spectral and topographic information. *IEEE Access.* **2019**;7:114363–114373. doi: [10.1109/Access.6287639](https://doi.org/10.1109/Access.6287639)
- [9] Xin L-b, Han L, Li L-z. Landslide intelligent recognition based on multi-source data fusion. *J Earth Sci Environ.* **2023**;45:920–928.
- [10] Jin Y, Li X, Zhu S, et al. Accurate landslide identification by multisource data fusion analysis with improved feature extraction backbone network. *Geom Nat Hazards Risk.* **2022**;13(1):2313–2332. doi: [10.1080/19475705.2022.2116357](https://doi.org/10.1080/19475705.2022.2116357)
- [11] Zhang WK, Liu WJ, Sun X, et al. Multi-source features adaptation fusion network for semantic segmentation in high-resolution remote sensing images. *J Image Graph.* **2022**;27:2516–2526.
- [12] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Las Vegas (NV), USA; 2016. p. 770–778.
- [13] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; Salt Lake City (UT), USA; 2018. p. 7132–7141.
- [14] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2021. p. 13713–13722.
- [15] Ghorbanzadeh O, Xu Y, Ghamisi P, et al. Landslide4sense: reference benchmark data and deep learning models for landslide detection; 2022. Available from: [arXiv preprint arXiv:2206.00515](https://arxiv.org/abs/2206.00515).
- [16] Jha D, Smedsrud PH, Riegler MA, et al. Resunet++: an advanced architecture for medical image segmentation. In: *2019 IEEE International Symposium on Multimedia (ISM)*. IEEE; 2019. p. 225–2255.
- [17] Wang J, Sun K, Cheng T, et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans Pattern Anal Mach Intell.* **2020**;43(10):3349–3364. doi: [10.1109/TPAMI.2020.2983686](https://doi.org/10.1109/TPAMI.2020.2983686)