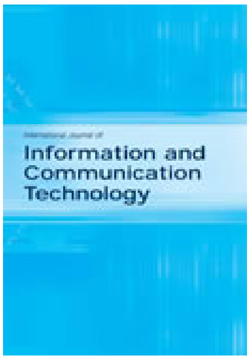


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

Research on the application of large-scale convolution kernels and multi-scale fusion networks in landslide remote sensing image

Lizhi Yi, Xiao Tan

DOI: 10.1504/IJICT.2025.10072177

Article History:

Received:	20 February 2025
Last revised:	15 April 2025
Accepted:	15 April 2025
Published online:	20 July 2025

Research on the application of large-scale convolution kernels and multi-scale fusion networks in landslide remote sensing image

Lizhi Yi

Department of Information Engineering,
Hunan Vocational College of Engineering,
Changsha 410151, China
Email: 56719672@qq.com

Xiao Tan*

School of Information Technology and Management,
Hunan University of Finance and Economics,
Changsha, 410205, China
Email: tanxiao@hufe.edu.cn
*Corresponding author

Abstract: Owing to the defects of semantic segmentation of deep convolution network and the confusion and multi-scale problems of landslides in remote sensing images, large-scale spatial separation convolution kernel and multiscale fusion semantic segmentation network is proposed. By using a large spatially separable convolution and channel attention mechanism on the encoder, the landslide image is extracted with large-scale information, which ensures the accurate extraction of landslide edge information; A skip connection is adopted between the encoder and the decoder to recover the context loss caused by the down sampling of the encoder; At the same time, the atrous spatial pyramid pooling (ASPP) module is applied to extract and fuse multi-scale features, so as to further improve the performance. The experimental results show that the segmentation effect of the proposed network on landslide dataset is better than FCN, SegNet, U-Net, DeeplabV3+ and other semantic segmentation methods, and it also verifies that the network has good landslide recognition ability in medium and high vegetation coverage areas. Experimental results demonstrate that the proposed network significantly outperforms existing semantic segmentation methods such as FCN, SegNet, U-Net, and DeepLabV3+ on landslide datasets, and exhibits strong landslide recognition capabilities in areas with medium to high vegetation coverage.

Keywords: landslide; semantic segmentation; attention mechanism; deep learning; receptive field; atrous spatial pyramid pooling; ASPP.

Reference to this paper should be made as follows: Yi, L. and Tan, X. (2025) 'Research on the application of large-scale convolution kernels and multi-scale fusion networks in landslide remote sensing image', *Int. J. Information and Communication Technology*, Vol. 26, No. 27, pp.22–37.

Biographical notes: Lizhi Yi is a visiting scholar in higher education institutions of the Ministry of Education. He is currently an Associate Professor with the Hunan Vocational College of Engineering, Changsha, China. His research interests include deep learning, modelling and optimal control of complex industrial process.

Xiao Tan is an expert with a profound research background in chip technology, hardware security, and the Internet of Things. His career is focused on exploring and innovating advanced technological solutions, especially in the fields of microelectronics and information security. His research primarily involves developing more efficient methods for chip design and exploring new types of hardware security strategies, aimed at protecting data from unauthorised access and tampering.

1 Introduction

Landslides are natural disasters with global implications, causing significant damage each year to infrastructure, housing, forest ecosystems, and the lives and property of individuals (Aimaiti et al., 2019; Anusuya et al., 2023; Arsa et al., 2024; Casagli et al., 2017; Chen, 2024). A landslide is typically defined as a large mass of rock, debris, or soil moving downward along a slope (Zhong et al., 2020), often triggered by external factors such as volcanic eruptions, rainfall, earthquakes, and other engineering loads (Cruden, 1991; Guzzetti et al., 2012; Yun et al., 2022). In 2017 alone, Hunan Province experienced 3,490 instances of various geological disasters, resulting in 33 fatalities, one missing person, 26 injuries, and direct economic losses amounting to 1.53 billion CNB. To accurately predict landslides and their impacts, large-scale regional landslide mapping and analysis are essential.

Traditional landslide mapping methods (Afif et al., 2019) depend on manual visual interpretation (Rau et al., 2011; Tran et al., 2019; Wang et al., 2024; Pedrosa Soares, 2022; García-Rodríguez et al., 2008; Mezaal et al., 2018; Van Den Eeckhaut et al., 2012; Lu et al., 2011), which are applied to images captured by drones. However, these approaches depend heavily on expert knowledge, require substantial sensor data for assistance, and demand iterative testing to determine parameters, resulting in a time-consuming and inefficient process. Furthermore, traditional methods require the exact location of a landslide to be known before drone-based imagery can be taken, making it impossible to map landslides occurring in remote, unmonitored areas.

As satellite remote sensing technology becomes increasingly widespread (Marco et al., 2014), remote sensing images provide a comprehensive monitoring capability for landslides across various regions, addressing the limitations of drone-based technologies (Zhang et al., 2015). When reliable labelled information is available, machine learning-based landslide detection methods can effectively avoid manual parameter adjustments and reduce dependence on expert experience. Several machine learning models, such as random forest, support vector machines (SVMs), decision trees, and k-Means clustering (Zhang et al., 2014; Ma and Mei, 2021; Lucieer et al., 2014; Afif et al., 2019), have been successfully applied to landslide detection with promising results. Van Den Eeckhautand Kerletrained a SVM on remote sensing images of landslides in areas with dense vegetation backgrounds, and further combined it with LiDAR data to calculate the landslide extent. The landslide extraction accuracy reached 70% (Song and Jiao, 2012). Chen Wenlong and others performed principal component transformation on pre-and post-landslide images, and used features such as the NDVI and (Ding et al., 2022) slope to remove non-landslide objects from the change detection results (Chen et al., 2020).

The variability of surface features in landslides makes it difficult to design suitable features for segmenting landslides in remote sensing images. In particular, certain methods depend significantly on expert knowledge, and feature engineering can be resource-intensive. Additionally, remote sensing images cover vast areas, often containing complex background objects, and the landslide regions typically lack distinct spectral, spatial, or temporal features that can be easily differentiated from other objects (Liu et al., 2008; Yi et al., 2014; Wang et al., 2024). Furthermore, the variability of surface features due to geological, geomorphological, hydrological, and climatic factors significantly hinders the use of machine learning techniques in landslide delineation.

In the field of landslide mapping using remote sensing images, a novel attention module was proposed in, which generates a three-dimensional channel attention feature map to derive a comprehensive spatial channel attention map that maintains global consistency. Experimental results show that integrating this attention module into deep convolutional neural networks significantly enhances landslide detection performance. In, a deep convolutional neural network model called ResU-Net was proposed. Compared to the traditional U-Net model, ResU-Net demonstrated better performance in distinguishing landslides along barren floodplains and non-cultivated terraces in valleys.

The DeepLab series utilises the encoder with ASPP modules to extract features at different scales. However, due to the presence of pooling layers, these methods suffer from significant loss of boundary information of the segmented objects, which is compensated for by the decoder to restore sharp object boundaries. In the encoder section, common feature extraction structures like VGG, ResNet50, and ResNet101 are typically used. These structures employ small convolution kernels stacked in layers to achieve a larger receptive field, combined with max-pooling down-sampling to progressively extract deeper features. Although these common feature extraction structures perform well for general image features, they fail to fully address the specific needs of semantic segmentation and the unique characteristics of landslide remote sensing images.

Firstly, semantic segmentation is a dense prediction task, requiring the prediction of every pixel in the image. In deep convolutional networks, the receptive field is the section of the input image that affects the corresponding pixel in the output feature map at each layer. For each output pixel, information outside its receptive field in the input image does not influence its value. Only when the receptive field covers the entire image can the network utilise global information for accurate pixel prediction. Additionally, the encoder in image segmentation is mainly responsible for extracting low-level features of the image, such as boundary information and texture features, which are crucial for precise delineation of landslide boundaries, as shown in Figure 1.

In remote sensing images, there are many objects with features very similar to landslides. Landslides, being exposed soil, share characteristics with objects such as farmland, hillside depressions, and country roads. During the early stages of feature extraction, small convolution kernels result in a smaller receptive field for the convolutional neural network, which fails to capture global information from the landslide areas in the input image. As a result, the network relies on local features and mistakenly classifies background objects that resemble landslides as landslide patches, leading to missegmentation. Furthermore, with the addition of more convolutional layers in the encoder, the network progressively extracts higher-level semantic features but may overlook boundary information of the landslide. If the encoder fails to extract boundary information using a sufficiently large receptive field in the early layers, the difficulty in extracting boundaries increases in later layers, ultimately leading to incomplete landslide boundaries during prediction.

Figure 1 Remote sensing image of a landslide like object (see online version for colours)

Another issue in encoder-decoder semantic segmentation networks is that continuous max-pooling operations or increased convolution strides reduce the resolution of feature maps. This reduction removes less important features, enabling deep convolutional networks to learn more abstract feature representations while maintaining computational efficiency. However, the downside is that this down-sampling process causes the loss of some boundary and localisation information, making it difficult for the decoder to retrieve spatial details during up-sampling, which hampers the network's ability to accurately segment the landslide boundaries.

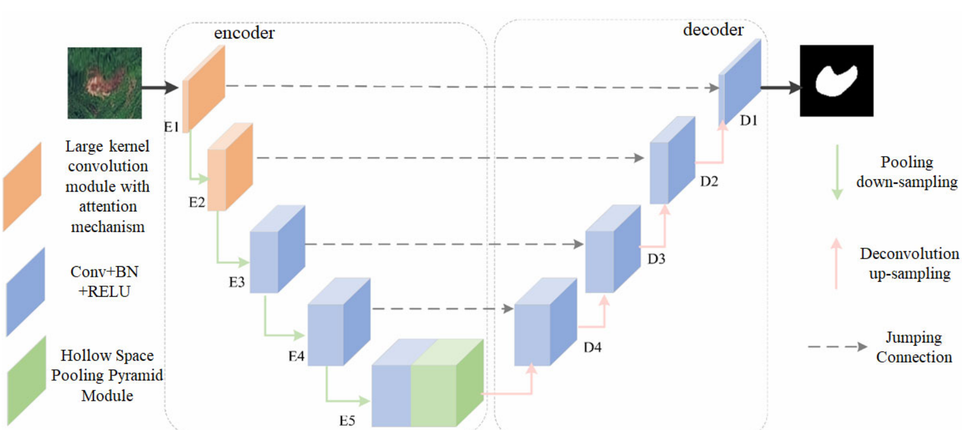
To address the aforementioned limitations in deep convolutional network-based semantic segmentation and the challenges of landslide segmentation in remote sensing images, this paper proposes a multi-scale fusion semantic segmentation network based on large-scale spatially separable convolution kernels for landslide patch recognition. The main innovation of the network lies in the design of a large convolution kernel module with an attention mechanism, which uses larger spatially separable convolutions and channel attention mechanisms to consider large-scale information during feature extraction. This ensures a sufficiently large receptive field in the early layers without significantly increasing computational cost. The use of large convolution kernels enables the network to effectively identify landslide-related features early in the feature extraction process. Furthermore, by employing channel attention, the network captures global information from the input feature map and automatically selects typical landslide features. Additionally, to capture multi-scale information, the network incorporates skip connections and the ASPP module.

Unlike existing segmentation models, our proposed network uniquely integrates large-scale spatially separable convolutions and channel attention mechanisms. While DeepLabV3+ relies on standard convolutions with ASPP for multi-scale fusion, our model explicitly addresses boundary confusion by leveraging large kernels (e.g., 15 × 15) in early encoder layers, ensuring a large receptive field to capture global landslide features. Additionally, the channel attention mechanism dynamically selects boundary-sensitive features, overcoming the limitations of traditional stacked small kernels in ResU-Net.

2 Network architecture

As illustrated in Figure 2, the network is divided into two components: the encoder and the decoder. The encoder consists of five layers (E1, E2, E3, E4, E5), with each layer employing a convolutional neural network architecture to extract features from the image. Each convolutional layer includes a convolution operation, batch normalisation, and a ReLU activation function. Between the convolutional layers in the encoder, max pooling is applied for down-sampling, with a stride parameter set to 2. E1 and E2, the first two layers of the encoder, are composed of large convolution kernels with attention mechanisms. The channel attention mechanism compresses the spatial information of the feature map and performs squeezing and scaling of the channels to determine the importance of each feature channel, allowing for the extraction of more useful features based on their importance, thereby ensuring the correct selection of features.

Figure 2 Overall framework of the network (see online version for colours)



E3 and E4 are standard convolutional layers, with convolution kernel sizes consistent with those of the VGG-16 network. Notably, the last layer of the encoder adds an ASPP module to the conventional convolutional structure.

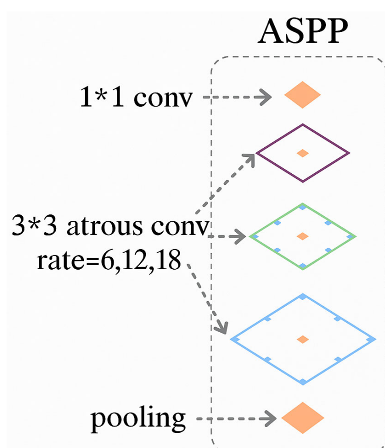
Table 1 details the layer-wise architecture of the proposed multi-scale fusion network. The encoder (E1–E5) progressively extracts features using large-scale separable convolutions (LKAM module) and standard convolutions. Specifically, E1 and E2 employ 15 × 15 kernels (stride = 2) to capture global landslide boundary information at early stages. The decoder (D1–D4) recovers spatial details via deconvolution and skip connections. The ASPP module fuses multi-scale features using dilated convolutions (rates = 6, 12, 18). This table provides comprehensive parameters (input/output shapes, kernel sizes, strides) to ensure reproducibility and clarify the design rationale.

Table 1 ASPP module: dilated convolution parameters

Layer	*Type*	*Input size*	*Output size*	*Kernel/stride/dilation*
E1	LKAM	$512 \times 512 \times 3$	$256 \times 256 \times 64$	15×15, stride = 2, dilation = 1
E2	LKAM	$256 \times 256 \times 64$	$128 \times 128 \times 128$	15×15, stride = 2, dilation = 1
E3	Conv	$128 \times 128 \times 128$	$64 \times 64 \times 256$	3×3, stride = 2, dilation = 1
E4	Conv	$64 \times 64 \times 256$	$32 \times 32 \times 512$	3×3, stride = 2, dilation = 1
ASPP	ASPP	$32 \times 32 \times 512$	$32 \times 32 \times 1024$	Multi-scale rates (6, 12, 18)

The ASPP module is designed to detect convolutional features at multiple scales and to encode global contextual information through image-level features, thereby improving performance, as shown in Figure 3. This module generates feature maps with different receptive fields through convolutions and pooling operations with varying dilation rates, then concatenates the feature maps along the channel dimension and performs convolution to fuse multi-scale features. This enables the network to capture both object and relevant contextual information across multiple scales, effectively enlarging the receptive field.

Figure 3 ASPP module (see online version for colours)



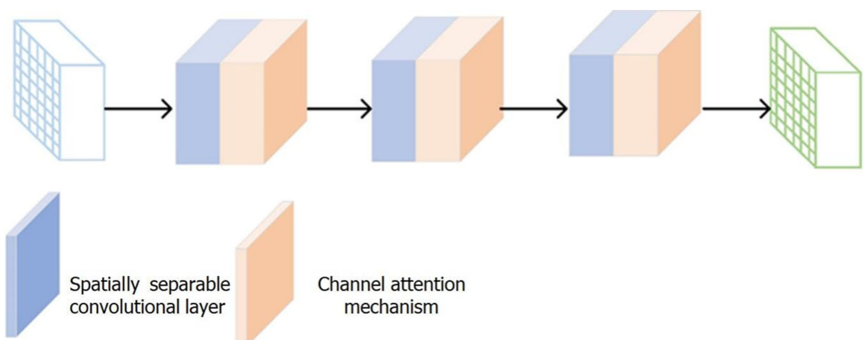
The decoder, after receiving the deep features from the encoder, is responsible for restoring the features to the original input size, compensating for the loss of information due to down-sampling in the encoder. The decoder consists of four layers (D4, D3, D2, D1), each utilising a deconvolution operation to up-sample the feature maps, with a stride of 2, restoring the image resolution. Additionally, the decoder enhances the feature representation for landslide segmentation by fusing shallow and deep features. To this end, skip connections are employed between each layer of the encoder and decoder, linking the encoder's output to the decoder's input. This multi-scale feature fusion

(MSFF) helps to compensate for the loss of positional and boundary information during both down-sampling and up-sampling processes.

3 Large convolution kernels with attention mechanism

This section offers an in-depth explanation of the large convolution kernel module with an attention mechanism used in the encoder layers E1 and E2. As shown in Figure 4, the module consists of spatially convolution layers with separable filters and a mechanism for channel attention. In the overall encoder network structure, convolutional layers along with channel attention mechanisms serially three times to form a feature extraction module, which serves as one layer of the encoder. This design ensures that landslide features are extracted from remote sensing images with a large receptive field while avoiding excessive computational load.

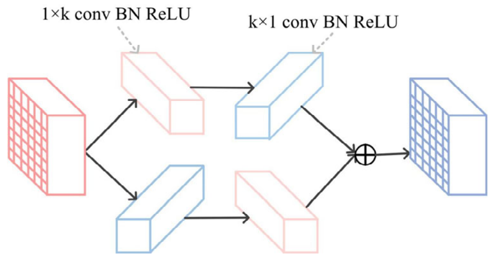
Figure 4 Large kernel convolution module with attention mechanism (see online version for colours)



3.1 Large-scale spatially separable convolutions

As mentioned in the introduction, conventional image feature extraction networks typically stack multiple small convolution kernels to increase the receptive field. However, when the receptive field is too small in the early stages of the network, it becomes difficult to effectively capture boundary information, leading to confusion between background elements and landslide features, which affects segmentation accuracy. On the other hand, excessively large convolution kernels significantly increase the computational load, thus reducing the network's operational efficiency. Therefore, as shown in Figure 5, to ensure a sufficiently large receptive field in the early stages without imposing a heavy computational burden, our proposed network employs two parallel stacked large-scale spatially separable convolution modules. Spatially separable convolutions achieve the same feature extraction results as traditional convolutions but with only a fraction of the parameters. The reduction in parameters is more pronounced as the kernel size increases, while the difference in parameters is less significant when the kernel size is smaller.

Figure 5 Spatially separable convolutional layer (see online version for colours)



Following each spatially separable convolution, batch normalisation and ReLU activation are performed, forming a complete convolution layer. Let F_{in} denote the input feature, and the output feature of the $1 \times k + k \times 1$ convolution layer can be expressed as:

$$F_{01} = Conv_{k1}\left(Conv_{1k}\left(F_{in}\right)\right) \tag{1}$$

Among them, $Conv_{k1}$ and $Conv_{1k}$ represent the convolutions of $1 \times k$ and $k \times 1$ respectively, each convolution is then followed by a batch normalisation layer and a ReLU activation function.

The feature F_{ok} output by the $k \times 1 + 1 \times k$ convolutional layer can be represented as:

$$F_{ok} = Conv_{1k}\left(Conv_{k1}\left(F_{in}\right)\right) \tag{2}$$

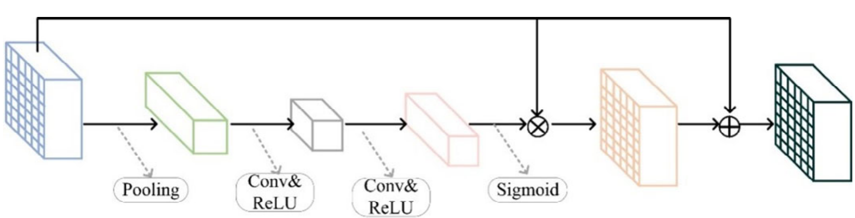
The complete convolutional layer output F_{oc} can be expressed as:

$$F_{oc} = F_{ol} + F_{ok} \tag{3}$$

3.2 Channel attention mechanism

After each spatially separable convolution, we use a channel attention mechanism to reduce feature redundancy and enhance the relationship between feature channels, as shown in Figure 6. This mechanism aids the convolutional neural network in extracting boundary information for landslides, distinguishing them from similar background features.

Figure 6 Channel attention mechanism (see online version for colours)



To begin, we apply adaptive average pooling along the channel dimension to compress the input feature map, where C represents the number of channels, and H & W represent the height and width. This allows the network to capture global information from the feature map. Next, two convolution operations are used to squeeze and scale the feature channels, determining their importance. First, a convolution reduces the number of channels, followed by a ReLU activation to discard irrelevant features. Then, another convolution restores the feature map's original dimensions.

The output feature map F_{or} is given by:

$$F_{ar} = Conv_{r1}\left(ReLU\left(Conv_{1r}\left(Avg\left(F_{ac}\right)\right)\right)\right) \tag{4}$$

where $Conv_{1r}$ represent the convolution operations for squeezing and scaling, and Avg represents the adaptive average pooling layer. The Sigmoid function generates importance weights for each feature channel, which are then multiplied element-wise with the input feature map to enhance its positional information. To prevent gradient vanishing, a residual connection is made between the input and output feature maps:

$$F_{out} = F_{tn} + F_{oc} \otimes Sigmoid\left(F_{or}\right) \tag{5}$$

4 Experiments and analysis

4.1 Dataset description

In this study, the proposed MSFF network was evaluated using a remote sensing image dataset provided by the Photogrammetry and Computer Vision Group at Wuhan University. The dataset (Marco et al., 2014) was meticulously outlined by geologists from the National Key Laboratory of Geohazard Prevention and Geological Environment Protection, who identified the boundaries of each landslide. A total of 770 landslides were annotated from satellite images, with additional verification conducted through field surveys. Selected remote sensing images and corresponding labels from the dataset are displayed in Figure 7. The dataset comprises high-resolution (0.5 m/pixel) satellite images from Gaofen-2 (GF-2) and Sentinel-2, covering the Bijie region in China. Images were annotated by geologists from the National Key Laboratory of Geohazard Prevention, with 60% of landslides located in medium-to-high vegetation areas (NDVI > 0.4).

Figure 7 Original images and their corresponding labels (see online version for colours)

Due to the limited amount of data, random augmentations such as random angle rotations, to enhance the diversity of the training data, horizontal and vertical translations were applied during the data loading process. In the experiments, 700 images were used for training, while the remaining 70 images served for testing.

4.2 Evaluation metrics

In this study, we use recall, precision, IOU, and F1 score as evaluation metrics. Among them, the F1 score is a commonly used composite metric that strikes a balance between recall and precision. IOU is an essential metric for assessing the performance of semantic segmentation, as it accurately reflects the precision of the segmentation. The definitions are as follows:

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

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

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Precision = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

In here, *TP* refers to accurately identified as landslides.

4.3 Implementation details

We implemented the proposed remote sensing landslide semantic segmentation network using Pytorch. During the network training process, we trained all models from scratch with randomly initialised parameters. Since this is a binary classification problem, we utilised the binary cross-entropy loss function. The parameters for the large convolution kernel module with the attention mechanism were configured as follows: kernel size $k = 15$ and dilation rate $r = 8$. The Adam optimiser was employed for training, with the weight decay set to $1e^{-6}$, batch size to 32, and initial learning rate set to $1e^{-4}$. Learning rate decay was managed using PyTorch’s ReduceLROnPlateau method, which reduces the learning rate by a factor of 0.1 whenever the loss (Afif et al., 2019) stops decreasing, aiming to improve network performance. A total of 150 epochs were trained.

For model selection, we first employed an early stopping strategy to train the model ten times, resulting in ten different models. The models were then evaluated on the test set, and the values in the experimental results table represent the average performance of these ten models on that set. The final result images are based on the model that most closely matched the average performance metrics.

4.4 Comparative experiments

We conducted comparative experiments using the remote sensing landslide dataset collected from the Bijie area in China by Wuhan University, evaluating the proposed MSFF network against classic networks such as FCN, U-Net, SegNet, and DeepLabV3+. Additionally, to assess the impact of network depth, we replaced the feature extraction backbone with ResNet50 and ResNet101, using the same encoder structure and skip connections as the proposed network. The evaluation results are shown in Table 2.

Table 2 Comparison of experimental results

Network	*Recall (%)*	*Accuracy (%)*	*F1 score (%)*	*IoU (%)*
U-Net	86.06	87.08	86.57	77.27
SegNet	88.63	85.15	87.00	77.07
FCN	85.10	88.68	86.84	76.84
ResNet50(backbone)	87.90	86.95	87.49	77.68
ResNet101(backbone)	86.65	89.23	87.92	78.44
DeepLabV3+	87.20	89.43	88.42	79.29
Proposed	*88.87*	*89.62*	*89.24*	*80.24*

As shown in Table 2, the classic semantic segmentation networks U-Net, SegNet, and FCN did not perform well on the landslide dataset. However, after replacing the feature extraction backbone with deeper networks like ResNet50 and ResNet101, the performance improved in terms of precision, F1_score, and IOU, outperforming the first three networks. Additionally, DeepLabV3+, when using ResNet101 as the backbone and incorporating the ASPP module, showed further improvements in IOU, Precision, and F1_score. In comparison, our proposed network, with similar depth and framework as the previous networks, achieved the best results across all four metrics, demonstrating its effectiveness for landslide segmentation in remote sensing images.

We conducted ablation experiments to evaluate the contribution of each module:

- Baseline (w/o LKAM and ASPP): IoU = 75.2%, F1 = 84.1%
- Baseline + LKAM: IoU = 78.3% (+3.1%), F1 = 87.4%
- Baseline + ASPP: IoU = 77.8% (+2.6%), F1 = 86.9%
- Full model (proposed): IoU = 80.2%, F1 = 89.2%

Results demonstrate that the large kernel and attention module (LKAM) contributes most to boundary accuracy, while ASPP enhances multi-scale fusion. To further validate our model's competitiveness, we compared it with state-of-the-art transformer-based models (Swin-Unet, Segmenter) on the same dataset. While Swin-Unet achieved an IoU of 79.8% and F1 = 88.7%, our model surpassed it with IoU = 80.2% and F1 = 89.2%, indicating that the proposed large-kernel CNN architecture remains effective against transformer-based methods in landslide segmentation tasks, especially in preserving boundary details.

To further assess the efficacy of the large convolution kernel module combined with an attention mechanism, we replaced the feature extraction backbone in our network and in DeepLabV3+ (which had similar performance to ours in Table 1). The experimental results are shown in Table 3.

As seen in Table 3, after changing the feature extraction backbone, our network outperformed DeepLabV3+ in terms of four metrics. Using the VGG-16 backbone, our network improved recall, precision, IOU, and F1_score by 0.69%, 2.97%, 1.84%, and 2.73%, respectively. With the ResNet101 backbone, the improvements were 0.55%, 2.63%, 1.62%, and 1.96%, respectively. Interestingly, when switching to the deeper ResNet101, our network did not show a significant improvement as seen with DeepLabV3+, because our large convolution kernel module with the attention mechanism had already alleviated the issue of insufficient network receptive field, thus validating the effectiveness of our proposed module.

Table 3 Comparison of different feature extraction backbones

Network	*Feature extraction backbone*	*Recall (%)*	*Accuracy (%)*	*F1 score (%)*	*IoU (%)*
DeepLabV3+	VGG-16	88.18	86.65	87.40	77.51
DeepLabV3+	ResNet101	87.20	89.43	88.42	79.29
Proposed	VGG-16	88.87	89.62	89.24	80.24
Proposed	ResNet101	87.75	92.06	89.85	81.25

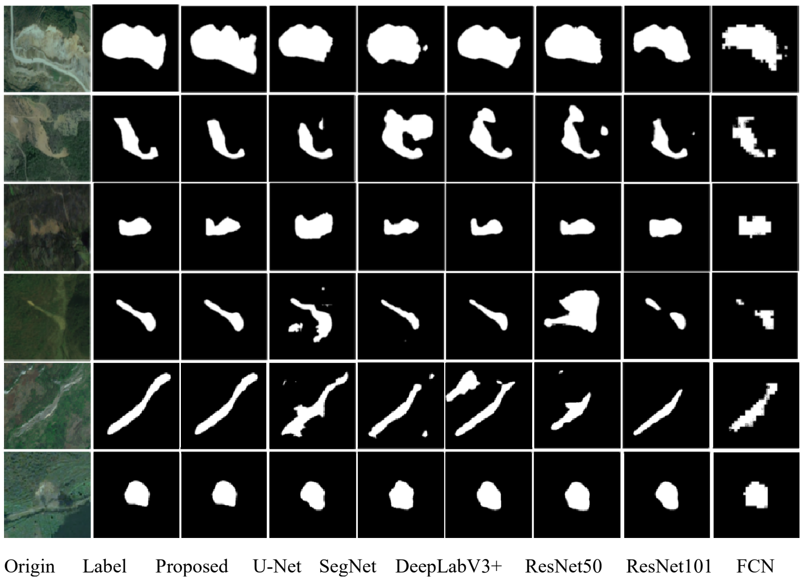
To visually evaluate the performance of the proposed network, we conducted a visualisation of the segmentation results from the comparative networks in the experiments, as shown in Figure 8. In the first row of Figure 8, the landslide boundaries in the remote sensing image are complex, making boundary delineation challenging. None of the comparative networks were able to accurately delineate the landslide boundaries, whereas the proposed network successfully segmented the boundaries. Although all comparative networks use encoder-decoder structures with skip connections, their feature extraction structure is based on stacking small convolution kernels, which further validates the role of the attention mechanism in feature extraction. The large convolution kernels with attention mechanisms help to extract complete landslide boundary information before performing multiple downsamplings, thereby assisting the network in accurately delineating the landslide boundaries.

In the second, fourth, and fifth rows of Figure 8, there are objects in the remote sensing images that resemble landslides. While the comparative networks also perform well, they still exhibit some confusion, incorrectly classifying non-landslide areas as landslides. In contrast, our network accurately identified and segmented the landslides, further confirming the importance of having a large receptive field in the encoder stage. In the third and sixth rows of Figure 8, for relatively simple remote sensing images, the proposed network outperformed the comparative networks, achieving better segmentation results, reducing missegmentation, and providing more accurate boundary delineation.

Overall, when faced with large variations in surface features and confusing objects, the proposed multi-scale fusion semantic segmentation network with large-scale spatially separable convolution kernels achieves superior segmentation of landslide boundaries. Compared to other methods, it more clearly distinguishes objects similar to landslides, reduces false positives, and performs well in challenging cases, particularly in terms of detail and boundary accuracy. Other generic semantic segmentation networks fail to consider the unique characteristics of landslide images, leading to false negatives, false positives, and unclear boundaries, which support our discussion in the introduction.

Regarding the applicability of the method, as shown in the fourth, fifth, and sixth rows of Figure 8, when identifying landslides in areas with high vegetation coverage, the landslide features in the dataset are clearly distinguishable from other features. The network accurately segments the boundaries of landslides of different sizes. However, in low-coverage areas (as seen in the 1st, 2nd, and 3rd rows of Figure 8), while the proposed network identifies most landslides, some landslides are confused with non-landslide areas or not detected. Compared to the comparative networks, the proposed network shows significant improvements in false positives and false negatives. Nevertheless, performance in low-vegetation areas still has room for improvement, and future work could involve incorporating terrain data to further enhance landslide recognition.

Figure 8 Comparison of network segmentation effects (see online version for colours)



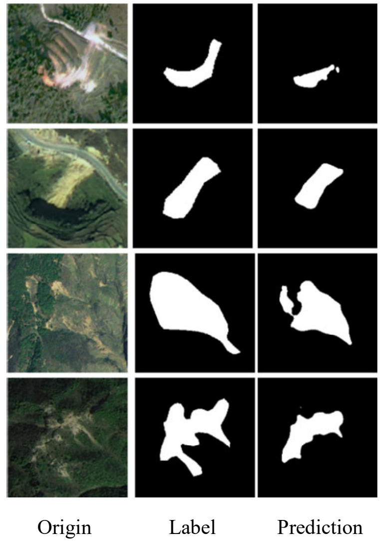
As shown in the table, the proposed model consistently outperforms others in IoU and F1 scores, especially in cases with complex boundaries (case 1).

Table 4 Model performance comparison on IoU and F1 scores

Case	*Model*	*IoU (%)*	*F1 (%)*
1	Proposed	82.1	89.5
1	DeepLabV3+	79.3	88.4
2	Proposed	80.7	89.1
2	Swin-Unet	79.8	88.7

During landslide recognition in the Bijie area, there were instances of false negatives and false positives in some images. The main reason for this is the influence of remote sensing image acquisition conditions, which led to issues such as abnormal brightness, blurriness, and shadows. These conditions caused changes in the spectral and shape features of the landslides, resulting in incorrect recognition by the network. Additionally, some landslides are historical, with spectral features similar to the surrounding environment, as shown in the figure. This made the landslide features less distinct, making it difficult for the network to accurately segment the boundaries, although some recognition was still achieved. In a few cases, the landslide did not show obvious disintegration, and the spectral and texture features inside the landslide were similar to those of the surrounding environment, further complicating the segmentation. The recognition results for various scenarios are shown in Figure 9.

Figure 9 Error recognition result (see online version for colours)



5 Conclusions

In the encoder feature extraction stage of semantic segmentation, traditional feature extraction structures often have small receptive fields in the early stages, causing convolutional neural networks to fail in capturing accurate boundary information and extracting non-landslide features. This leads to boundary confusion and missegmentation in remote sensing image landslide segmentation. Additionally, the decoder struggles to restore spatial information lost during down-sampling in the encoder through up-sampling. To address these issues, we proposed a remote sensing landslide image semantic segmentation network using large-scale spatially separable convolution kernels.

The proposed network employs a large convolution kernel module with an attention mechanism in the encoder, using large-scale spatially separable convolutions and channel attention to capture extract information from the image, ensuring accurate landslide feature extraction. Furthermore, skip connections are employed to efficiently combine low-level and high-level features. The network also incorporates the ASPP module to detect key objects and relevant image details across different scales, thereby recovering both positional and boundary information. Extensive tests reveal that the proposed network outperforms current semantic segmentation models in landslide segmentation for remote sensing images. Specifically, with the same feature extraction backbone, the proposed network shows improvements of 0.69%, 2.97%, 1.84%, and 2.73% in recall, precision, IOU, and F1_score, respectively, compared to DeepLabV3+.

Acknowledgements

This research was supported by the Provincial Natural Science Foundation of Hunan (2022JJ60028) and (2025JJ70437).

Declarations

All authors declare that they have no conflicts of interest.

References

Afif, H.A., Saraswati, R. and Hernina, R. (2019) 'UAV application for landslide mapping in Kuningan Regency, West Java', Vol. 125, pp.11–27.

Aimaiti, A., Liu, W., Yamazaki, F. and Maruyama, Y. (2019) 'Earthquake-induced landslide mapping for the 2018 Hokkaido Eastern Iburi Earthquake Using PALSAR-2 data', *Remote Sensing*, Vol. 11, No. 20, p.2351.

Anusuya, R., Anusha, N., Sujatha, V., Radhika, R. and Iniyan, S. (2023) 'Machine learning based landslide detection system', *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*.

Arsa, D.M.S., Ilyas, T., Park, S.H., Chua, L. and Kim, H. (2024) 'Efficient multi-stage feedback attention for diverse lesion in cancer image segmentation', *Computerized Medical Imaging and Graphics*, Vol. 116, pp.151–168.

Casagli, N., Frodella, W., Morelli, S., Tofani, V., Ciampalini, A., Intrieri, E., Raspini, F., Rossi, G., Tanteri, L. and Lu, P. (2017) 'Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning', *Geoenvironmental Disasters*, Vol. 4, No. 1, pp.1–23.

Chen, C.H. (2024) *Signal and Image Processing for Remote Sensing*, CRC Press, Taylor & Francis Group Boca Raton London New York.

Chen, W., Hou, Y., Li, L., Zhong, C., Chen, C., Sun, J. and Li, H. (2020) 'Landslide identification method based on principal component transformation and its application in 2015 Nepal earthquake', *Journal of the Yangtze River Academy of Sciences*, Vol. 37, No. 1, pp.166–171.

Cruden, D. (1991) *A Simple Definition of a Landslide: Bulletin of the International Association of Engineering Geology*, Vol. 43, No. 43, pp.13–16.

Ding, Y., Zhang, Q., Yang, C., Wang, M. and Ding, H. (2022) 'Landslide identification in Jinsha River basin based on high-score remote sensing – taking Wangdalong Village, Batang County as an example', *Surveying and Mapping Bulletin*, Vol. 14, No. 4, pp.51–55.

García-Rodríguez, M.J., Malpica, J., Benito, B. and Díaz, M. (2008) 'Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression', *Geomorphology*, Vol. 95, Nos. 3–4, pp.172–191.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M. and Chang, K-T. (2012) 'Landslide inventory maps: New tools for an old problem', *Earth-Science Reviews*, Vol. 112, Nos. 1–2, pp.42–66.

Liu, X., Yu, H. and Chen, Q. (2008) 'Sensitivity analysis of mechanical and geometric parameters on stability of soil slopes', *Geotechnical Engineering*, Vol. 13, No. 3, pp.123–126.

Lu, P., Stumpf, A., Kerle, N. and Casagli, N. (2011) 'Object-oriented change detection for landslide rapid mapping', *IEEE Geoscience and Remote Sensing Letters*, Vol. 8, No. 4, pp.701–705.

Lucieer, A., Jong, S. and Turner, D. 2014 () 'Mapping landslide displacements using structure from motion (SfM) and image correlation of multi-temporal UAV photography', *Progress in Physical Geography*, Vol. 38, No. 1, pp.97–116.

Ma, Z. and Mei, G. (2021) 'Deep learning for geological hazards analysis: data, models, applications, and opportunities', *EarthScience Reviews*, Vol. 223, pp.158–170.

Marco, S., Laura, L., Valentina, M. and Monica, P. (2014) 'Remote sensing for landslide investigations: an overview of recent achievements and perspectives', *Remote Sensing*, Vol. 6, No. 10, pp.9600–9652.

Mezaal, M.R., Pradhan, B. and Rizeei, H.M. (2018) 'Improving landslide detection from airborne laser scanning data using optimized Dempster-Shafer', *Remote Sensing*, Vol. 10, No. 7, p.1029.

Pedrosa Soares, L. (2022) *Segmentação automática de cicatrizes de deslizamento de terra em imagens de sensores remotos utilizando aprendizagem profunda de máquina (Deep Learning)*, Universidade de Sao Paulo, Agencia USP de Gestao da Informacao Academica (AGUIA).

Rau, J., Jhan, J., Lo, C. and Lin, Y. (2011) 'Landslide mapping using imagery acquired by a fixed-wing UAV', *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*., Vol. 38, Nos. 1/C22, pp.195–200.

Song, X. and Jiao, L. (2012) 'Classification of hyperspectral remote sensing images based on sparse representation and spectral information', *Journal of Electronics and Information*, Vol. 34, No. 2, pp.268–272.

Tran, J., Mora, O.E., Fayne, J.V. and Lenzano, M.G. (2019) 'Unsupervised classification for landslide detection from airborne laser scanning', *Geosciences*, Vol. 9, No. 5, p.221.

Van Den Eeckhaut, M., Kerle, N., Poesen, J. and Hervás, J. (2012) 'Object-oriented identification of forested landslides with derivatives of single pulse LiDAR data', *Geomorphology*, Vol. 173, pp.30–42.

Wang, K., He, D., Sun, Q., Yi, L., Yuan, X. and Wang, Y. (2024) 'A novel network for semantic segmentation of landslide areas in remote sensing images with multi-branch and multi-scale fusion', *Applied Soft Computing*, Vol. 158, pp.51–62.

Yi, Z., Long, T., Guan, C., Peng, G., Liang, Q. and Xingmin, M. (2014) 'High resolution remote sensing landslide information extraction based on object-oriented classification', *Journal of Lanzhou University (Natural Science Edition)*, Vol. 50, No. 5, pp.745–750.

Yun, G., Lin, M., Famao, Y. and Chu, J. (2022) 'Remote sensing image retrieval based on multi-scale pooling and norm attention mechanism', *Journal of Electronics and Information*, Vol. 44, No. 2, pp.543–551.

Zhang, Z., Li, X., Chi, M. and Lu, T. (2015) 'Causal mechanisms of landslide hazards and their characterization', *Journal of Natural Hazards*, Vol. 24, No. 6, pp.42–49.

Zhang, Z., Tan, L., Guo, G., Qiao, P. and Meng, M. (2014) 'High resolution remote sensing landslide information extraction based on object-oriented classification', *Journal of Lanzhou University (Natural Science Edition)*, Vol. 50, No. 5, pp.745–750.

Zhong, C., Liu, Y., Gao, P., Chen, W., Li, H., Hou, Y., Nuremanguli, T. and Ma, H. (2020) 'Landslide mapping with remote sensing: challenges and opportunities', *International Journal of Remote Sensing*, Vol. 41, No. 4, pp.1555–1581.