Accurate landslide identification by multisource data fusion analysis with improved feature extraction backbone network

Yuhui Jin¹, Xin Li¹, Sainan Zhu², Bin Tong², Fang Chen¹, Ru Cui¹ and Jian Huang¹

¹State Key Laboratory of Software Development Environment, Beihang University, Beijing, China; ²China Institute of Geo-Environmental Monitoring, Beijing, China

ABSTRACT

Traditional methods for landslide survey, whether field investigation or human remote sensing-based interpretation approaches, all require considerable labour costs and expert knowledge. Deep learning-based detection methods have significantly improved the speed of landslide recognition, but their accuracy still has much room for improvement. In our work, we propose SA-MFNet to achieve pixelwise landslide detection based on multisource data fusion analysis. On the one hand, we achieve improved feature extraction by utilizing an attention mechanism. On the other hand, based on raw sensing data and labeled results obtained from several regions, we propose a landslide detection model based on the fusion of multisource data, including digital elevation model (DEM), geological mapping data, river distribution data and other data related to earth observation information. We enhance the performance of the developed method via fusion analysis with features extracted from remote optical sensing images, thus achieving precise pixelwise landslide terrain classification and positioning. Experimental results demonstrate that the model proposed in this article is superior to the existing common baselines and can provide technical support for automatic landslide identification with practical value.

ARTICLE HISTORY

Received 2 June 2022
Accepted 18 August 2022

KEYWORDS

Pixelwise landslide detection; attention mechanism; multisource data fusion analysis; SA-MFNet

Introduction

Devastating natural hazards such as landslides are massive threats to lives and property around the globe, especially in mountainous regions (Haque et al. 2019). In recent years, due to the intensification of climate change, extreme weather has occurred more frequently, providing favorable inducing conditions for landslides (Gariano and Guzzetti 2016). To achieve accurate landslide prediction and reduce the losses caused by disasters, we need to conduct more in-depth research and understand the mechanism of landslides, in which practical landslide mapping from large
scale areas plays a major role (Yu et al. 2021). Landslides mapping is also crucial for the reconstruction and the validation of susceptibility, hazard and risk maps.

In recent years, with the full application of advanced observation methods represented by remote sensing, especially satellite-based spatial remote sensing, comprehensive and integrated earth observation data have been formed, laying a good data foundation for landslide identification and gradually eliminating the limitations of ‘field inspection’. Especially for landslide disasters occurring in remote mountainous areas, identification methods based on optical remote sensing images have more advantages in terms of human and material costs, safety over field inspection, wide coverage, etc. However, currently, the interpretation of remote sensing images mostly relies on manual expert interpretation, and the data processing speed of this approach is far inferior to the acquisition speed of earth observation data (Williams et al. 2018). To solve this problem, many researchers have proven that deep learning techniques can be used to extract landslide features from remote sensing images to reduce the reliance on manual work and accelerate the processing of observation data. For example, Liu et al. (2021) proposed an end-to-end landslide extraction model based on the improved mask region convolutional neural network (Mask RCNN).

Under the circumstance that the landslide mapping is executed by experts and the landslide inventory can be guaranteed with a high level of accuracy, deep learning solves the problem of inefficient manual remote sensing image interpretation, but such approaches are all still able to achieve improved recognition accuracy for the following two reasons. The first reason relates to the model structure. For deep learning models, the better the model structure design is, the higher the recognition accuracy that can be achieved under the premise of possessing a sufficient amount of qualified training data. The exploration of better model structure designs for different tasks is a constant topic in deep learning. Second, although optical remote sensing images contain the main features of landslides, such as colors, textures, and shapes (Gorum et al. 2011; Martha et al. 2012; Xu et al. 2014), these features are not exhaustive, and many other elements, such as topographic relief, water distribution, soil properties, movement on faults and weathering processes, can have impacts on the formation of landslides (Li et al. 2017).

Chen et al. (2017) proposed a new deep neural network-based feature fusion framework that employs deep CNNs to efficiently extract features from lidar data and their proposed fusion framework can effectively improve the classification performance of the resulting model. In addition, other works have demonstrated the advantages of fusion models that combine remote sensing data with other datasets, but these existing fusion-based works utilized locally available data sources with more homogeneous data types, which are difficult to apply in different regions. In addition, the existing models are basically based on one single level of fusion processes, data fusion, feature fusion or decision fusion, which lack mutual feedback mechanisms and coordination between different fusion levels, thereby reducing the reliability of the model analysis results.

In summary, on the one hand, our method improves the structures of landslide recognition models based on optical remote sensing images, imitates the cognitive process of experts with respect to landslides, optimizes the feature extraction network
of the developed model by introducing an attention mechanism (Vaswani et al. 2017), and improves the feature extraction capability of the model. On the other hand, the model realizes the recognition of landslides by fusing optical remote sensing data and factors that influence landslides such as topography, geomorphology, rock formations, and water system work, and a new model branching structure and fusion module are designed based on the idea of data fusion. Furthermore, to train and validate the model, remote sensing images from two regions, the Loess Plateau and Jinsha River Basin, are selected and annotated to enhance the generalization ability of the model. Some existing landslide identification methods are compared with each other and the proposed method to analyze the advantages and disadvantages of the various approaches.

The main contributions of our manuscript are as follows:

- We optimize the feature extraction network of the proposed model by introducing an attention mechanism.
- We explore the impact of multisource data on landslide identification methods based on remote sensing imagery.
- We propose a new semantic segmentation network structure based on the idea of multisource data fusion.
- We use the results of field work to create a small dataset for model validation.

Related works

Traditional landslide survey methods, including field investigation and human remote sensing-based interpretation approaches, all require professionals to provide judgments through expertise (Marcelino et al. 2009; Tsai et al. 2010). These methods are reasonably accurate but overly reliant on expertise, and the information processing procedure is cumbersome. Manual collection and judgment methods also require much work strength and bring high risks of danger. In general, traditional landslide detection methods are not suitable for landslide identification over a wide range of areas. Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance; this technique can help sense objects from long distances with a multidimensional views. Remote sensing is widely used for landslide identification due to its inherent advantage (Ding et al. 2016; Xiao et al. 2018; Yu et al. 2020). With the continuous development of remote sensing technology, the collection and acquisition speeds of earth observation data are increasing, and the method of artificial interpretation has not been able to process new data in time. Therefore, deep learning-based methods, which have stronger feature extraction and modeling capacities and higher calculation speeds than existing automatic landslide detection methods, have become a new research direction for landslide detection.

Deep learning methods have the ability to self-study. With the support of complete and large training data, deep learning-based methods can achieve better prediction performance than traditional machine learning methods. In addition, deep learning-based methods can be effectively expanded and easily migrated. Many studies have
been conducted on landslide analysis. Yu et al. (2020) constructed an end-to-end profile-based detection framework for landslide detection in Nepal and achieved a recall of 65% and an accuracy of 55.35%. Ding et al. (2016) used a convolutional neural network (CNN) based approach to detect landslides from China’s GF-1; their method could detect 72% of total landslides, and the detection accuracy was 67%. Xiao et al. (2018) built a sample dataset containing 3,800 remote sensing images and used a long short-term memory (LSTM) model to predict landslides in Xingdu Kush-Himalaya; the accuracy of the method reached 81.2%.

The methods discussed above are all based on remote sensing images, while multi-source data fusion analysis can extract fully stereo landslide features, reduce data blur, and achieve improved overall detection capabilities (Jeon and Landgrebe 1999; Abdulhafiz and Khamis 2013; Pradhan et al. 2016). In general, multi-source data fusion methods can be divided into data hierarchy, feature hierarchy, and decision fusion steps. Since insufficient data sources were available in the early stage, most relevant researches conducted fusion analysis at a particular hierarchy. In recent years, the contents of landslides not only include the volume of observational data but also include data type have become increasingly rich (including low-level sensing data, high-level semantic information and even prior knowledge), and a single level of fusion is no longer meet the demands of detection.

Eight different methods are compared for image fusion to show their ability to fuse multitemporal and multi-sensor image data in Zhang (2010). A series of eight multitemporal multispectral remote sensing images is fused with a panchromatic Ikonos image and a TerraSAR-X radar image as a panchromatic substitute. Mezaal et al. (2018) proposed an effective architecture that used Dempster-Shafer theory (DST) to produce final decisions based on the probability outputs of a support vector machine (SVM), a random forest and a k-nearest neighbors (KNN) classifier. Experiments showed that the DST method had high detection accuracy along with high computational efficiency and strong robustness. These fusion analysis approaches adopt probability-based, evidence-based, and knowledge-based fusion techniques. However, they are sensitive to noise and have difficulty handling high-dimensional, complex data. Therefore, deep learning-based fusion models are more suitable for the analysis of earth observation data, and an end-to-end fusion architecture is also easier to implement for automatic analysis. Sameen and Pradhan (2019) compared a single CNN layer with two deeper networks and compared the rest of the network with two fusion strategies (layer stacking and feature class fusion). They detected landslides in the Malaysian Camarn Highland and achieved high accuracy. Liu et al. (2018) combined quantitative analysis with trend analysis based on a back-propagation (BP) network to detect landslides. These research results have verified the application potential of deep learning-based fusion methods.

**Data preparation**

**Multisource data**

As described above, remote sensing images reflect the most intuitive characteristics of landslides, including brightness, intensity, hue, shape, size, position, spectral texture...
information, and geometric relationship information. At the same time, factors such as geological structures, terrain, landforms, rock groups and water systems all have impacts on landslides, and these factors are used in traditional methods (Nowicki Jessee et al. 2018). In our work, we make full use of the existing data resources by considering remote sensing optical images, digital elevation model (DEM), landform data, geological mapping data, hydrological distribution data, and active fault zones data, which are collected from the Loess Plateau and Jinsha River Basin. Since these data have different sources, standards, and storage formats, how to build data consistency relationships and extract features from different data that meet the same criteria must be solved first.

Remote sensing optical images
For optical remote sensing data, we obtain the complete 18-level images of the two aforementioned areas from Map World, which is the National Platform for Common Geospatial Information Services. The spatial resolution of the image is approximately 0.59 m per pixel and the data samples are shown in Figure 1.

DEM
Our DEM data are obtained from Shuttle Radar Topography Mission (SRTM) data, which is provided at a resolution of approximately 30 m (Farr et al. 2007). A visualization of the DEM data is shown in Figure 2 and darker colors represent lower
elevation levels. The DEM data are standardized and normalized, and the pixel values are divided into $[0, 255]$. 

**Landform distribution data**

Based on the specific regional morphology, formation causes and development trends, China is divided into 25 types of landforms; according to the elevation and terrain corresponding to a given area, a value in the range of 11–74 is specified for representation purposes (Cheng et al. 2011; Wang et al. 2020). The associated data attribute results are shown in Table 1. The attributes are expressed in two digits: the first digit represents the average altitude of the area, and the second digit represents the terrain. Each pixel is assigned a different value based on its attribute value, which is first converted to a single-channel grayscale.

**Geological mapping data**

Geological zoning is based on natural and geological conditions and is a zoning method that combines geological features, geothermal fields, and masked thermophysical properties. The SHP format raw data contain a total of 20 attributes, among which the character attribute is the most representative (Pang et al. 2018); the data are numerically processed according to the content of this attribute.

**Hydrological distribution data**

Since the flow of water also has a great effect on landslide hazards along the coast, information of river in study areas is collected, and the level attribute which comprehensively reflects area, length, navigation capabilities, importance, density, etc., is
used as the key feature (Zhang et al. 2013). During the subsequent conversion to numeric data, the fold line corresponding to the river is thickened, and different pixel values are given according to their levels.

**Active fault zones**
Active fault zones often have strong tectonic deformation, which can easily change the rock structure and cause the land to be looser, thus accelerating the development of landslide hazards. The raw active fault zone data are in SHP format and type attribute is selected as the key feature. By analyzing the type properties of the fault, a total of 160 different values are obtained; the numerical processing of the fracture data is completed by linear mapping, and the numerical range is expanded to 0-255.

**Data consistency construction**
A series of preprocessing steps is required before feeding the multisource data to the model to not only satisfy the semantic requirements of deep neural networks but also denoise.

**Data standardization**
For the use of raster data in descriptive text for zone division, the region division feature is converted to numeric data, and a plurality of metric hierarchies are set. Thus, raster data can be converted to the regional division factor values. The standardization and normalization of the attributes are performed with the raster data of the regional division factor, satisfying the requirement of the image format in which images must be composed of numerical values.

### Table 1. Geomorphology partition properties.

| OID | Value | Type                                      |
|-----|-------|-------------------------------------------|
| 1   | 11    | Plain terrain with low altitude          |
| 2   | 12    | Plain terrain with intermediate altitude  |
| 3   | 13    | Plain terrain with high altitude         |
| 4   | 14    | Plain terrain with extremely high altitude|
| 5   | 21    | Tableland with low altitude              |
| 6   | 22    | Tableland with intermediate altitude      |
| 7   | 23    | Tableland with high altitude             |
| 8   | 24    | Tableland with extremely high altitude   |
| 9   | 31    | Hilly terrain with low altitude          |
| 10  | 32    | Hilly terrain with intermediate altitude  |
| 11  | 33    | Hilly terrain with high altitude         |
| 12  | 34    | Hilly terrain with extremely high altitude|
| 13  | 41    | Low mountainous area with low relief      |
| 14  | 42    | Middle mountainous area with low relief   |
| 15  | 43    | High mountainous area with low relief     |
| 16  | 44    | Extremely high mountainous area with low relief|
| 17  | 51    | Low mountainous area with intermediate relief|
| 18  | 52    | Middle mountainous area with intermediate relief|
| 19  | 53    | High mountainous area with intermediate relief|
| 20  | 54    | Extremely high mountainous area with intermediate relief|
| 21  | 62    | Middle mountainous area with large relief |
| 22  | 63    | High mountainous area with large relief   |
| 23  | 64    | Extremely high mountainous area with large relief|
| 24  | 73    | Extremely high mountainous area with large relief|
| 25  | 74    | Extremely high mountainous area with extremely large relief|
Geometric correction processing for remote sensing images helps to eliminate all kinds of image distortions generated during image acquisition and provides accurate geographic locations for various types of objects on the image. In addition, geometric correction can guarantee a certain planar accuracy.

Projection transformation
Dataset building requires a transformation of the grid data contained in the dot matrix file of an optical remote sensing image with a labeled image format to the same geospatial coordinate system, thereby achieving geographic location.

Image cropping and bitmap generation
Since the grid data, which include national information, are stored as a dot matrix file and formed by geometric objects, they cannot meet the model vector input and correspondence requirements. The raster data must be converted into a bitmap format, and the corresponding optical remote sensing images should be matched. We first burn the attribute feature value of each grid-grid data point to the pixel of interest in the geometry, then select the content of interest from the image (that is, each optical remote sensing image), and then remove the irrelevant areas. Finally, bitmap-format data with attribute characteristics are formed and fed into the multisource fusion model for feature extraction.

The completed data are shown in Figure 3. The visualized figure is a grayscale map with a single channel (except for remote sensing images) and is directly fed to the model.

Our constructed dataset
In our dataset, there are totally 290 landslides located in two regions, the Loess Plateau and Jinsha River Basin. For the Jinsha River Basin, 90 landslides are selected within the area between 96°22'23"E and 99°27'10"E, 27°48'5"N and 33°46'3"N. For

Figure 3. Data processing results. (a)–(f) represents the topic data we use after data consistency construction and (g) represents the ground truth mask of the sliding area.
the Loess Plateau, 200 landslides are selected within the area between 103°12′6″E and 111°56′40″E, 33°25′37″N and 37°18′32″N. 80% of the landslides are divided into training set and the remaining 20% are used for test. Therefore, we use 232 landslides to train our model and use 28 landslides for evaluation. We use a random split strategy that all landslides are mixed together and then are divided randomly into training or test set.

For remote sensing images, we manually annotate the images under the guidance of experts. For the use of raster data of other topic data, the attribute can be converted to the regional division factor values. The standardization and normalization of the attributes are then performed with the numeric data, satisfying the requirement of the image format. Our remote sensing images are collected in a relatively unvegetated environment. Additionally, most remote sensing images in our dataset contains no cloud and only six images contain clouds.

**Methods**

Among all the collected data, the optical remote sensing images contain the most important features, including textures, shapes, colors, etc., and play a major role in the identification of landslides. However, at the same time, many other factors also inevitably affect changes in geological environment. As shown in Figure 4, to make full use of multisource data, our proposed end-to-end network adopts a multistage fusion procedure, which starts from raw data, extracts features and obtains the final semantic results through the decision module, and each step contains a fusion mechanism. We focus on the extraction of optical remote sensing image features and the extraction and fusion of these features with other topic data characteristics. The details are as follows.

**Single remote sensing image-based model for landslide detection**

The features extracted from remote sensing images can be used as a portion of the multisource data fusion and identification model. In cases where other sensing data are not applicable, an upsampling module can be added to the image features; thus, the identification model can be implemented separately based on the given remote

![Figure 4](image-url). Landslide detection framework based on data fusion analysis. The green mask denotes the identification result.
sensing images. As a practical application of semantic segmentation technology, the proposed recognition model, which provides a specific landslide area range from remote sensing images, requires a high recall to avoid false negatives, while large-scale land investigation requires high computational efficiency. The recognition accuracy and calculation efficiency are contradictory. An accuracy improvement is often accompanied by more layers, more complex connections, and more parameters, i.e., a more complex model, while bringing higher computational complexity. On the one hand, the proposed model is based on U-shaped structure (Ronneberger et al. 2015) and fully convolutional networks (FCNs) (Long et al. 2015), and combined with spatial pyramid pooling (He et al. 2015) and self-attention mechanism to improve the recognition accuracy. On the other hand, the model utilizes local connections, weight sharing, and group mechanisms to reduce its complexity, thus simultaneously meeting the requirements of a high recall and high calculation efficiency.

We utilize ResNet (He et al. 2016) as our backbone for the extraction of features from raw remote sensing images and utilize atrous spatial pyramid pooling (ASPP) (Chen et al. 2014; Chen et al. 2017; 2018) as the upsampling module to obtain outputs with the same resolution. Compared with traditional CNN networks, ResNet has a deeper structure, better learning efficiency and a faster convergence speed (He et al. 2016). The basic residual block adds additional skip connections on the basis of a normal convolution layer, which means that each layer feeds into the next layer and directly into the layers that are approximately 2-3 hops away. Residual blocks can help the model learn simpler feature mapping functions $F(x)$, alleviate the degradation problem and enhance the model performance. We also introduce ASPP to further extract multiscale information and control the size of the receptive field. The ASPP module has one convolution layer with a $1 \times 1$ kernel and three dilated convolutions with $3 \times 3$ kernels. Each convolution is followed by a batch normalization layer, and the number of output channels is 256. After a $1 \times 1$ convolution is completed, bilinear interpolation is used to expand the feature resolution. Dilated convolution is utilized as an improved method on the basis of discrete convolution, and the process is defined as follows:

$$F * k(p) = \sum_{s+t=p} F(s)k(t).$$  \hspace{1cm} (1)

where $k$ represents the convolution kernel and $F$ represents the original image. Given the image $F$, a convolution operator is used for signal processing. For a certain pixel on the image, the point of the kernel with coordinate $t$ and the corresponding point on the image with coordinate $s$ are multiplied and added. $p$ denotes the coordinate of the resulting point (Chen et al. 2018). Dilated convolution is a type of convolution that inflates the kernel by inserting holes between the kernel elements. The calculation process changes to

$$F * k_l(p) = \sum_{s+l=t} F(s)k(t).$$  \hspace{1cm} (2)

When identifying landslides, experts typically observe overall structures and ignore the areas that do not contain landslide; they then examine a suspicious area to
Inspired by this process, we optimize ResNet by modifying the bottleneck structure. First, a 1×1 convolution is performed, and then an attention block is added. The bottleneck ends with another 1×1 convolution. The structure is shown in Figure 5. Two different groups of independent subnetworks focus on learning different features and are complementary to each other. The feature extraction ability can be enhanced under the premise of controlling the computational consumption.

For the grouping mechanism, AlexNet (Krizhevsky et al. 2012) achieves improved accuracy without increasing the complexity of the resulting model. AlexNet divides the network into two groups. One group mainly learns black and white information, and the other group mainly learns color information, thereby effectively taking the depth and training speed of the model into account when the computing power is insufficient. Inspired by AlexNet, we introduce the cardinal architecture to divide the network into several branches, each of which serve as a subspace and increase the width of the network. The final ResNeXt is conducted for feature extraction.

The integrated attention mechanism and grouping mechanism constitute the full structure of the final ResNet50-SA network, as shown in Figure 6, where S represents the grouping mechanism (split) and A represents the attention mechanism.

**Multisource data fusion-based landslide identification**

After acquiring the features of the given remote sensing image, the extraction and feature fusion of other special topic data should be considered and performed. Due to the incompleteness and unavailability of other topic data, no existing method handles different types of input and makes use of all topic data. Therefore, it is necessary to concatenate all the input data except for remote sensing images to form a grayscale feature map with higher dimensionality to avoid frequently altering the network structure subject to the input variance. Moreover, a single model is able to deal with different inputs by using FCNs to handle their dimensions.
The information contained in earth observation data varies greatly. Therefore, overfitting or underfitting may occur when these data have the same structures and weights, causing bias in the results. To solve this, we propose a model with branches that accordingly applies a residual network with different depths and widths, with the aim of addressing different intrinsic data features. For optical images, we use ResNet50-SA as mentioned above. For other special topic data, we implement the ResNet34 structure to obtain features and fuse them together. We also apply feature fusion to all the topic data for comparability:

\[ T(X_t) = [H(X_t)D(X_t)Z(X_t)R(X_t)L(X_t)], \]

where \( X_t \) represents the fused feature tensor and \( H(X_t), D(X_t), Z(X_t), R(X_t), \) and \( L(X_t) \) represent an altitude feature, a landform feature, a geological feature, a river distribution feature and an active fracture, respectively.

The process of fusing the optical features and fused topic feature tensor is divided into three steps, and the details are shown in Figure 7. \( C \) and \( S \) represent an optical feature and a topic feature, respectively, \( i \) represents the initial fusion result, and \( U \) and \( S \) indicate the results of the channel extraction and relationship determination, respectively. The reconstructed feature vector \( Z \) is obtained via the fusion of \( I \) and \( S \) and is then added with the initial feature to obtain \( V \).

The first step is feature map fusion. We combine an optical feature \( C(X_i) \) and a topic feature \( S(X_i) \) through an element-wise product to obtain \( I(X_i) \):

\[ I(X_i) = C(X_i) \ast \text{sigmoid}(S(X_i)). \]
The second step involves channel-wise relationship construction. We apply channel extraction $F_{avg}$, relation calculation $F_{FC}$ and combination $F_{Mul}$ operations to recalibrate $I(X_i)$. For channel extraction $F_{avg}$, the shape of $I$ changes from $h/C_3 \times w/C_3 \times c$ to $1 \times 1 \times c$:

$$F_{Avg}(I_C) = u_c = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} I_C(i,j)$$

(5)

For channel relationship calculation, $W_i$ represents a convolution kernel with a fully connected layer, and $\delta$ represents a rectified linear unit (ReLU) activation function:

$$F_{FC}(u) = s = \text{sigmoid}(W_2 \ \delta(W_1 u)),$$

(6)

where $F_{Mul}$ is calculated through the Hadamard product of the previous two results:

$$F_{Mul}() = z = I_C \ast S_C.$$  

(7)

The third step is residual layer connection. We add the initial feature and the reconstruction feature to retain the importance of the feature tensor and reduce the loss of background information.

In summary, we can design a comprehensive data fusion framework SA-MFNet, based on a model fusion network (MFNet), as shown in Figure 8. A parallel feature extraction branch is added in addition to the improved ResNet50-SA. The features obtained from the two branches are fused before upsampling.
Experiments and comparisons

Implementation details

The training and testing of our model are conducted on the PyTorch 1.4.0 framework (Paszke et al. 2019) with Ubuntu 16.04, 64 GB of RAM and 4 Nvidia Titan (Pascal) GPUs. We utilize stochastic gradient descent (SGD) (Robbins and Monro 1951) as the optimizer, where the batch size, initial learning rate, and momentum are set to 64, 0.01 and 0.9, respectively. The effective learning rate is tuned by a polynomial decay policy with a decay rate of 0.0001. The initial number of channels, image batch size and crop size are set to 2048, 350 and 300, respectively. The model is trained for 250 epochs.

Metrics

Considering landslide recognition as a binary classification task, each pixel of an input remote sensing image is categorized into two classes: sliding and background pixels.

Figure 8. Structure of the feature extraction and detection model based on data fusion.
Table 2. The prediction results and the ground-truth values in the binary classification problem.

| Observed Positive | Detected Positive | Detected Negative |
|-------------------|-------------------|-------------------|
| True Positive (TP)| False Negative (FN)|                  |
| False Positive (FP)| True Negative (TN)|                  |

We choose the common binary classification metrics, that is, intersection over union (IOU), accuracy (ACC) and F1-score, as our evaluation metrics. In binary classification, four combinations can be formed by the ground-truth labels and results: true positive (TP), false positive (FP), false negative (FN), and true negative (TN) (Lin et al. 2014), as shown in Table 2. Positive represents a landslide, and negative represents the background area. Specifically, the chosen metrics are calculated according to the combinations above, the first two of which are determined by the following equations:

\[
IOU = \frac{TP}{TP + FN + FP} \tag{8}
\]

\[
ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{9}
\]

The F1-score can better reflect the comprehensive level of the model and is calculated based on recall and precision:

\[
Precision = \frac{TP}{TP + FP} \tag{10}
\]

\[
Recall = \frac{TP}{TP + FN} \tag{11}
\]

\[
F1 - score = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{12}
\]

**Quantitative and qualitative results**

Alongside the proposed optical remote sensing model and our data fusion model, we also test several common segmentation models, including U-Net (Ronneberger et al. 2015), the pyramid scene parsing network (PSPNet) (Zhao et al. 2017), DeepLab v3 (Chen et al. 2017), and DeepLab v3+ (Chen et al. 2018). The comparisons results are listed in Table 3.

As shown in Table 3, our proposed ResNet50-SA network achieves better performance than common semantic segmentation methods. The main progress in ResNet50-SA lies in the improvement of the feature extraction capability. The modified feature extraction module boosts performance by introducing a group-wise attention mechanism. Our model outperforms DeepLab v3 by 16.8% in terms of IOU. All of these factors provide supporting proof of the significance of feature extraction in
segmentation tasks. The visualizations of some recognition results are shown in Figure 9. DeepLab v3 is a representative semantic segmentation model, and it is compared to the improved identification model with an attention mechanism. For the first landslide located in Gang Su, due to the clear boundary between sliding area and non-sliding area, both models yield good segmentation results. DeepLab v3 has a certain degree of recognition error at the sharp angle. Due to the interference of clouds in the second landslide located in Ding Ba, DeepLab v3 induces a large error. The recognition accuracy of ResNet50-SA decreases somewhat but remains at a higher level overall. The results of DeepLab v3 on the third landslide located in Zeng Ti indicate that the model is more sensitive to color but ignores texture features during landslide detection, while ResNet50-SA can better capture the image texture information.

Furthermore, we not only collectively fuse every category of specific topic data, including elevation, geology, landform, water system, and active fracture data, but also test the omission case, leaving only remote sensing images and elevation data as inputs. The results are shown in Table 4. Analyzing the comparison results in Table 4, we can see that the fusion model is superior to the remote sensing image-only model, which confirms that incorporating expert knowledge in the geology domain can improve accuracy. Additionally, ablation studies further validate that the improvement of the data fusion model is mainly due to the contribution of terrain
data, and this finding adheres to our expectation. Comprehensive comparisons among common segmentation models and the proposed data fusion model show that our model achieves significant accuracy improvements. In particular, SA-MFNet outperforms DeepLab v3 by 42.2% in terms of F1-score. The visualizations of some recognition results are shown in Figure 10. We test the performance of the model in three cases: using remote sensing images only, combining several data (DEM) and integrating all the data. Geological expertise derived through multisource data fusion can effectively improve performance, especially for the second landslide located in Ningba and the third landslides located in Sekao and Dangdi. After multisource data are integrated, the model reduces its misjudgment of nonlandslide areas.

Our work still contains some limitations. Although we attempt to alleviate the serious data dependence of data-driven methods by obtaining annotations through field study, we face a lack of data relative to publicly available datasets, such as the Microsoft Common Objects in Context (COCO) dataset (Lin et al. 2014) and the Gaofen Image Dataset (GID) (Tong et al. 2020). Therefore, we are trying to build a landslide labeling platform to collect data, and we hope to realize the long-term and stable data production requirements of supervised learning.

**Conclusions**

Landslides have become massive threats to lives and property around the globe, especially in mountainous regions. This paper demonstrates the importance of landslide

---

**Table 4.** Quantitative results obtained based on multisource data fusion.

| Methods                              | IOU          | ACC          | F1-Score    |
|--------------------------------------|--------------|--------------|-------------|
| ResNet50-SA (Remote sensing image)   | 0.6582       | 0.7430       | 0.5844      |
| SA-MFNet (Remote sensing image + DEM)| 0.714        | 0.809        | 0.673       |
| SA-MFNet (all data)                  | **0.755**    | **0.850**    | **0.742**   |

---

**Figure 10.** Comparison results between the proposed multisource data fusion-based model and the single-view method.
prediction and analyzes the shortcomings of existing methods. Deep learning-based methods solve the problem of inefficient manual remote sensing image interpretation, but such approaches are still able to achieve improved recognition accuracy for the following two reasons. The first reason relates to the model structure. Second, multi-source data fusion can bring new landslide detection possibilities. On the one hand, we improve the single view-based feature extraction ability of our model by utilizing an attention mechanism and a group mechanism. Thereafter, detection is achieved via the fusion of multisource data, and a new semantic division-based model structure is designed and expanded. Experimental results demonstrate that the model proposed in this article is superior to the existing common baselines and can provide technical support for automatic landslide identification with practical value.

Disclosure statement

The authors declare no conflict of interest.

Funding

This research was funded by China Geological Survey with grant number DD20221939 and DD20190637.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. Remote sensing optical images are selected from National Platform for Common Geospatial Information Services and the link is https://www.tianditu.gov.cn/. The DEM data is obtained from Shuttle Radar Topography Mission (SRTM) data and the link is https://lpdaac.usgs.gov/products/srtmgl1v003/. Landform distribution data are collected from Resource and Environment Science and Data Center and the link is https://www.resdc.cn/data.aspx?DATAID=124. Geological mapping data are collected from National Geological Archives Data Center and the link is http://dc.ngac.org.cn/museumDetails?id=10.35080%2Ffn01.c.820021. Hydrological distribution data are collected from National Catalogue Service For Geographic Information and the link is https://www.webmap.cn/commres.do?method=result25W. Active fault zone data are collected from “Geocloud” system of China Geological Survey, and the link is https://geocloud.cgs.gov.cn/#/home.

References

Abdulhafiz WA, Khamis A. 2013. Bayesian approach to multisensor data fusion with Pre-and Post-Filtering. 2013 10th IEEE International Conference on Networking, Sensing and Control (ICNSC).
Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. arXiv:14127062.
Chen L-C, Papandreou G, Schroff F, Adam H. 2017. Rethinking atrous convolution for semantic image segmentation. arXiv:170605587.
Cheng W, Zhou C, Chai H, Zhao S, Liu H, Zhou Z. 2011. Research and compilation of the Geomorphologic Atlas of the People’s Republic of China (1: 1,000,000). J Geogr Sci. 21(1): 89–100.
Chen Y, Li C, Ghamisi P, Jia X, Gu Y. 2017. Deep fusion of remote sensing data for accurate classification. IEEE Geosci Remote Sens Lett. 14(8):1253–1257.
Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. 2018. Deeplab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected crfs. IEEE Trans Pattern Anal Mach Intell. 40(4):834–848.

Chen L-C, Zhu Y, Papandreou G, Schroff F, Adam H. 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. Proceedings of the European Conference on Computer Vision (ECCV); Munich, Germany.

Ding A, Zhang Q, Zhou X, Dai B. 2016. Automatic recognition of landslide based on CNN and texture change detection. 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC).

Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, Kobrick M, Paller M, Rodriguez E, Roth L, et al. 2007. The shuttle radar topography mission. Rev Geophys. 45(2).

Gariano SL, Guzzetti F. 2016. Landslides in a changing climate. Earth Sci Rev. 162:227–252.

Gorum T, Fan X, van Westen CJ, Huang RQ, Xu Q, Tang C, Wang G. 2011. Distribution pattern of earthquake-induced landslides triggered by the 12 May 2008 Wenchuan earthquake. Geomorphology. 133(3-4):152–167.

Haque U, da Silva PF, Devoli G, Pilz J, Zhao B, KHALOuA A, Wilopo W, Andersen P, Lu P, Lee J, et al. 2019. The human cost of global warming: Deadly landslides and their triggers (1995–2014). Sci Total Environ. 682:673–684.

He K, Zhang X, Ren S, Sun J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell. 37(9):1904–1916.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Las Vegas, NV.

Jeon B, Landgrebe DA. 1999. Decision fusion approach for multiphase classification. IEEE Trans. Geosci. Remote Sens. 37(3):1227–1233.

Krizhevsky A, Sutskever I, Hinton GE. 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems. 25.

Li Y, Huang J, Jiang S-H, Huang F, Chang Z. 2017. A web-based GPS system for displacement monitoring and failure mechanism analysis of reservoir landslide. Sci Rep. 7(1):1–13.

Lin T-Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P, Zitnick CL. 2014. Microsoft coco: common objects in context. European conference on computer vision; Zurich, Switzerland.

Liu J, Tang H, Li Q, Su A, Liu Q, Zhong C. 2018. Multi-sensor fusion of data for monitoring of Huagtupo landslide in the three Gorges Reservoir (China). Geomatics Nat Hazards Risk. 9(1):881–891.

Liu P, Wei Y, Wang Q, Xie J, Chen Y, Li Z, Zhou H. 2021. A research on landslides automatic extraction model based on the improved mask R-CNN. IJGI. 10(3):168.

Long J, Shelhamer E, Darrell T. 2015. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Boston, MA.

Marcelino EV, Formaggio AR, Maeda EE. 2009. Landslide inventory using image fusion techniques in Brazil. Int J Appl Earth Obs Geoinf. 11(3):181–191.

Martha TR, Kerle N, Van Westen CJ, Jetten V, Kumar KV. 2012. Object-oriented analysis of multi-temporal panchromatic images for creation of historical landslide inventories. ISPRS J Photogramm Remote Sens. 67:105–119.

Mezaal MR, Pradhan B, Rizeei HM. 2018. Improving landslide detection from airborne laser scanning data using optimized DempsterShafer. Remote Sens. 10(7):1029.

Nowicki Jesse M, Hamburger M, Allstadt K, Wald DJ, Robeson S, Tanyas H, Hearne M, Thompson E. 2018. A global empirical model for near-real-time assessment of seismically induced landslides. JGR Earth Surface. 123(8):1835–1859.

Pang J, Ding X, Han K, Zeng Y, Chen A, Zhang Y, Zhang Q, Yao D. 2018. The national 1:1000000 geological map spatial database. China Geology. 44(S1):8–18.

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L. 2019. Pytorch: An imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 32.
Pradhan B, Jebur MN, Shafri HZM, Tehrany MS. 2016. Data fusion technique using wavelet transform and Taguchi methods for automatic landslide detection from airborne laser scanning data and quickbird satellite imagery. IEEE Trans Geosci Remote Sens. 54(3): 1610–1622.

Robbins H, Monro S. 1951. A stochastic approximation method. Ann Math Stat. 22(3): 400–407.

Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention.

Sameen MI, Pradhan B. 2019. Landslide detection using residual networks and the fusion of spectral and topographic information. IEEE Access. 7:114363–114373.

Tong X-Y, Xia G-S, Lu Q, Shen H, Li S, You S, Zhang L. 2020. Land-cover classification with high-resolution remote sensing images using transferable deep models. Remote Sens Environ. 237:111322.

Tsai F, Hwang J-H, Chen L-C, Lin T-H. 2010. Post-disaster assessment of landslides in southern Taiwan after 2009 Typhoon Morakot using remote sensing and spatial analysis. Nat Hazards Earth Syst Sci. 10(10):2179–2190.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser u, Polosukhin I. 2017. Attention is all you need. Advances in Neural Information Processing Systems. 30.

Wang N, Cheng W, Wang B, Liu Q, Zhou C. 2020. Geomorphological regionalization theory system and division methodology of China. J Geogr Sci. 30(2):212–232.

Williams JG, Rosser NJ, Kincey ME, Benjamin J, Oven KJ, Densmore AL, Milledge DG, Robinson TR, Jordan CA, Dijkstra TA. 2018. Satellite-based emergency mapping using optical imagery: experience and reflections from the 2015 Nepal earthquakes. Nat Hazards Earth Syst Sci. 18(1):185–205.

Xiao L, Zhang Y, Peng G. 2018. Landslide susceptibility assessment using integrated deep learning algorithm along the China-Nepal highway. Sensors. 18(12):4436.

Xu C, Xu X, Yao X, Dai F. 2014. Three (nearly) complete inventories of landslides triggered by the May 12, 2008 Wenchuan Mw 7.9 earthquake of China and their spatial distribution statistical analysis. Landslides. 11(3):441–461.

Yu B, Chen F, Xu C, Wang L, Wang N. 2021. Matrix SegNet: A Practical Deep Learning Framework for Landslide Mapping from Images of Different Areas with Different Spatial Resolutions. Remote Sens. 13(16):3158.

Yu B, Chen F, Xu C. 2020. Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. Comput Geosci. 135:104388.

Zhang J. 2010. Multi-source remote sensing data fusion: status and trends. Int J Image Data Fus. 1(1):5–24.

Zhang N, Jianming MA, Wan J. 2013. Research on some issues of revision of Code for China River Name. Yangtze River.

Zhao H, Shi J, Qi X, Wang X, Jia J. 2017. Pyramid scene parsing network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Honolulu, HI.