remote sensing

MDPI

Technical Note

L-Unet: A Landslide Extraction Model Using Multi-Scale Feature Fusion and Attention Mechanism

Zhangyu Dong [1,2,3,*], Sen An [1,3], Jin Zhang [1,3], Jinqiu Yu [1,3], Jinhui Li [1,3] and Daoli Xu [1,3]

[1] School of Computer and Information, Hefei University of Technology, Hefei 230601, China; 2020111017@mail.hfut.edu.cn (S.A.); 2020171089@mail.hfut.edu.cn (J.Z.); 2020111049@mail.hfut.edu.cn (J.Y.); jinhui_li@mail.hfut.edu.cn (J.L.); 2020180040@mail.hfut.edu.cn (D.X.)
[2] Intelligent Interconnected Systems Laboratory of Anhui Province, Hefei 230009, China
[3] Anhui Province Key Laboratory of Industry Safety and Emergency Technology, Hefei 230601, China
* Correspondence: dzyhfut@hfut.edu.cn

Abstract: At present, it is challenging to extract landslides from high-resolution remote-sensing images using deep learning. Because landslides are very complex, the accuracy of traditional extraction methods is low. To improve the efficiency and accuracy of landslide extraction, a new model is proposed based on the U-Net model to automatically extract landslides from remote-sensing images: L-Unet. The main innovations are as follows: (1) A multi-scale feature-fusion (MFF) module is added at the end of the U-Net encoding network to improve the model's ability to extract multi-scale landslide information. (2) A residual attention network is added to the U-Net model to deepen the network and improve the model's ability to represent landslide features. (3) The bilinear interpolation algorithm in the decoding network of the U-Net model is replaced by data-dependent upsampling (DUpsampling) to improve the quality of the feature maps. Experimental results showed that the precision, recall, MIoU and F1 values of the L-Unet model are 4.15%, 2.65%, 4.82% and 3.37% higher than that of the baseline U-Net model, respectively. It was proven that the new model can extract landslides accurately and effectively.

Keywords: landslide; remote sensing; U-Net; attention mechanism; deep learning

Citation: Dong, Z.; An, S.; Zhang, J.; Yu, J.; Li, J.; Xu, D. L-Unet: A Landslide Extraction Model Using Multi-Scale Feature Fusion and Attention Mechanism. *Remote Sens.* **2022**, *14*, 2552. https://doi.org/10.3390/rs14112552

Academic Editors: Wanchang Zhang, Qiang Xu and Shunping Ji

Received: 8 April 2022
Accepted: 23 May 2022
Published: 26 May 2022

1. Introduction

Landslides are some of the most common and hazardous geological hazards. Although they often occur in mountainous areas, they seriously threaten human lives and property safety [1]. It is of great importance for disaster rescue, prevention and mitigation to obtain related information on landslides quickly and accurately after the occurrence of the hazards [2].

Satellite remote-sensing technology can accurately extract the data features of objects and capture information on earth changes in a timely manner. The technology has been widely used in geological disaster-related research. Satellite remote-sensing image data can cover an area of hundreds of square kilometers, which provides rich image information for landslide extractions [3]. Therefore, losses of human lives and properties can be effectively reduced by using remote-sensing technology to grasp landslide information in a timely manner and formulate reasonable rescue plans.

Visual interpretation is a commonly used landslide extraction method that mainly uses the spectral features and spatial features of images to extract landslides. This method has a high extraction accuracy and reliability, but it requires a significant investment of time and energy. Moreover, it requires a high level of professional knowledge and experienced interpreters, which make it difficult to meet the timeliness of disaster rescue.

The pixel-oriented landslide extraction method performs remote-sensing image analyses based on the spectral texture features of individual pixels to obtain information about landslides and achieve the automatic identification of such disasters. The commonly used

approaches used in this method include principal component analyses [4], support vector machines [5], maximum likelihoods [6] and so on. However, this method considers only the individual information of a single pixel point and ignores correlations between pixels, which greatly reduces its extraction accuracy.

With the improvement of remote-sensing image resolutions, objects on the ground in such images are becoming richer and more informative. Thus, the object-oriented image analysis (OBIA) approach has been applied to the information extraction of remote-sensing images and has been developed rapidly [7]. This method takes a target object as a basic unit after an image segmentation and determines the class in which the target object belongs by combining spectral, spatial, shape and contextual features. This method relies on the selection of a segmentation scale, but the existing segmentation algorithm cannot meet the demands of remote-sensing images [8].

Deep learning has been widely used in various fields, such as object detection [9], semantic segmentation [10], image classification [11] and so on. With a powerful feature extraction ability, it can quickly and efficiently extract landslide information, which provides a new method for landslide extractions from remote-sensing images. Sameen et al. [12] fused spectral and topographic information and trained it by using residual networks to achieve better results in landslide extraction. Wang et al. [13] fused the pre-disaster and post-disaster red, green, blue and near-infrared bands as well as NVDI data to obtain a total of nine bands for landslide extraction by using CNN. Ghorbanzadeh et al. [14] evaluated the ability of ANN, SVM, RF and CNN to detect landslides and concluded that the CNN approach is still in its infancy but has great potential. Zhang et al. [15] used the deep-learning module of ENVI to identify co-seismic landslides in the Hokkaido region of Japan. Lu et al. [16] combined transfer learning and OBIA methods to accurately extract landslides. Moreover, to emphasize landslide features from a complex background, Ji et al. [17] integrated spatial and channel attention to propose a novel attention module. This attention module significantly improves the ability of a model to extract landslides. Liu et al. [18] proposed an end-to-end landslide extraction method by improving the Mask R-CNN model.

At present, the use of deep learning for landslide extraction is still in its initial stage. How to finely extract landslide features and improve the accuracy of landslide extraction is the focus of research. U-Net [19] is a classical model in the field of semantic segmentation, and it has been widely used because of its simple structure and high recognition accuracy. Soares et al. [20] achieved the automatic extraction of landslides in the mountains of Rio de Janeiro, Brazil, by using the U-Net model. However, if U-Net is directly used for a landslide extraction from remote-sensing images, there are some problems. It is more difficult for the shallow U-Net model to learn landslide features with complex shapes and solve the problem of confusion between landslides and other background information. Therefore, Liu et al. [21] improved the U-Net model by adding residual learning units to the encoding and decoding network and expanding the input image to six channels by adding DSM, slope and aspect. They obtained fairly good results in the experiment. Ghorbanzadeh et al. [22] compared the results of the U-Net and ResU-Net models for extracting landslides and thought the performance of ResU-Net was better than that of U-Net. Following this, Ghorbanzadeh et al. [23] combined ResU-Net and OBIA approaches by using the OBIA approach to optimize the result maps generated by ResU-Net. The authors achieved accurate extractions of landslides. Although the above methods based on U-Net achieved better results, little attention has been paid to landslides showing different features at different scales in remote-sensing images. The simple skip connection of U-Net cannot meet the requirements of extracting multi-scale landslide information.

In order to solve the problems noted above, we improved the U-Net model and thus propose an automatic landslide extraction model named L-Unet. The main improvements are as follows: (1) A proposed MFF module based on dilated convolution is embedded at the end of the U-Net encoding network. (2) A residual attention network is added to the

U-Net network structure. (3) The upsampling process uses DUpsampling to replace the bilinear interpolation algorithm.

2. Model

2.1. U-Net

The U-Net model was proposed mainly for segmentation in the field of medical images and stitches features together in the channel dimension so that the obtained feature maps contain both high-level feature information and low-level feature information. The model can achieve a high accuracy even for small data sets. Our landslide dataset was small, so U-Net was chosen as a base model.

The U-Net model consists of two parts: encoding and decoding networks. The encoding network consists of four downsamplings, each consisting of two convolutional layers and a maximum pooling layer. The feature map size is reduced by half after each downsampling. The decoding network contains four upsamplings by which the feature-map size is recovered in relation to the feature-map size of the corresponding layer of the encoding network and the information is extracted by convolutional blocks after stitching. In the last layer, the output feature map is passed through a 1 × 1 convolutional layer to obtain a final result. The model structure is shown in Figure 1.

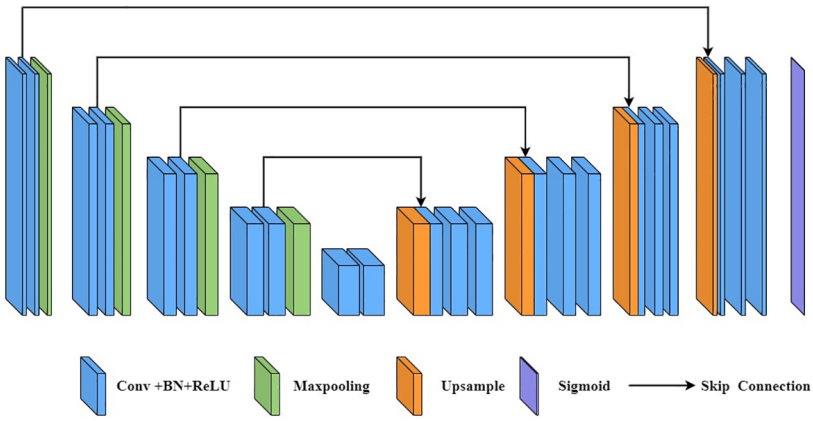

Figure 1. Structure diagram of U-Net model; consists of downsampling and upsampling.

2.2. L-Unet

The structure of the L-Unet model is shown in Figure 2.

In the encoding network, the residual network A is added after a 3 × 3 convolutional layer, batch normalization (BN) layer and ReLU activation function, and then a pooling layer is connected. Residual network A consists of residual blocks and the attention mechanism. The residual block can deepen the network so that the model can fully extract landslide features. Adding the attention mechanism can ensure more of a focus on landslide information and can suppress unimportant background information. The output of the last layer of the encoding network is passed through the MFF module to obtain fused features at different scales.

In the decoding network, the residual network B is added to each layer of U-Net after the upsampling and convolution. The feature-map size is recovered in relation to the output feature-map size of residual network A in the corresponding layer of the encoding network by using DUpsampling, and the stitching combines low-dimensional detail information and high-dimensional semantic feature information to allow for obtaining a more accurate feature map.

2.2.1. Residual Attention Network

Generally, the deeper the network is, the better the extraction of feature information and the better the model performance. However, experiments showed that when the

network is deepened to a certain level, its performance becomes worse. This is due to the problems of a vanishing gradient and an exploding gradient that occurs when the network deepens to a certain level, causing difficulties in model training. The proposed residual network [24] is a good solution to the problems caused by the deepening of the network.

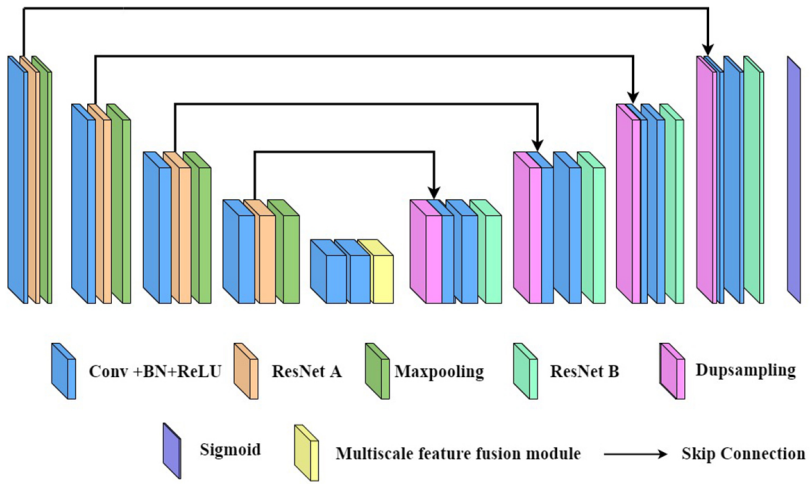

Figure 2. Structure diagram of L-Unet model; consists of downsampling, MFF module and upsampling.

The attention mechanism mimics human vision by focusing on a region of interest while ignoring other background information, thus achieving the effect of improving the performance of deep-learning models. Numerous experiments showed that the accuracy of the semantic segmentation model would be greatly improved by adding the attention mechanism. Co-ordinate attention (CoordAttention) [25] can focus on both position and channel information and bring little overhead to the network. It can aggregate features from horizontal and vertical directions, respectively. It not only retains critical position information when capturing channel information but also captures long-range dependent information.

We found that the low-resolution feature maps obtained during downsampling may lose position and channel information. In order to deepen the network while improving the model's ability to acquire position and channel information, the CoordAttention mechanism was combined with the residual block to form the residual attention module. The structures of the two residual networks are shown in Figure 3. In residual network A, the output of the input feature map X, after passing through two stacked residual blocks, is added with the output after the co-ordinate attention. In residual network B, the input feature map is directly output after two residual blocks. Since residual network A incorporates CoordAttention, it is added to the encoding network of the U-Net model to fully extract the landslide features and residual network B is added to the decoding network of the U-Net model.

2.2.2. MFF Module

Remote-sensing images contain many kinds of objects, among which landslides have complex and variable shapes, and such images' features are not the same at different scales. The size of the receptive field also affects the accuracy of landslide extraction to a certain extent. Although the U-Net model integrates both high-level feature information and bottom-level feature information, its extraction of landslide multi-scale information is limited. The dilated convolution inserts zero values on the normal convolution kernel to achieve the purpose of increasing the receptive field so that the output contains a larger range of information.

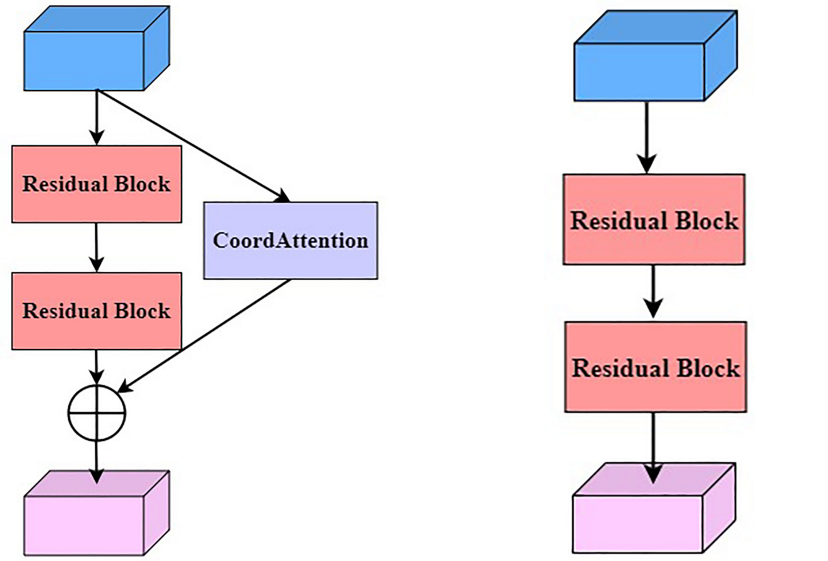

Figure 3. (**a**) Residual network A based on the residual block and the CoordAttention mechanism and (**b**) residual network B based on the residual block.

In the encoding networks, constant pooling operations can lead to the loss of information, which contains useful landslide information and can result in small landslides being missed and parts of large landslides being extracted incorrectly. The MFF module based on the dilated convolution is proposed to prevent the loss of useful information and to improve the ability of the model to learn landslide features at different scales, as shown in Figure 4.

The MFF module proposed in this paper contains five branches. The first branch uses a 1 × 1 normal convolution and the other four branches use different dilation rates of dilated convolutions. Due to the variable shapes of landslides, the set dilation rate is too large to effectively extract a small landslide. In order to enhance the correlation between features at different scales, the input of the third, fourth and fifth branches is the stitching of the output of the previous branch after the dilated convolution and the original input feature map. Then, the 1 × 1 convolution, BN layer and ReLU activation function are used to deepen the network while increasing nonlinear features. The output of each branch is defined as the following equation:

$$A_i = \begin{cases} D_{\epsilon_i}^{3\times3}(I), i = 2 \\ D_{\epsilon_i}^{3\times3}(C(A_{i-1}, I)), i = 3, 4, 5 \end{cases} \tag{1}$$

$$O_i = \begin{cases} f^{1\times1}(I), i = 1 \\ \eta(BN(f^{1\times1}(A_i))), i = 2, 3, 4, 5 \end{cases} \tag{2}$$

where I represents the input feature map, C represents stitching in the channel dimension, $D_{\epsilon_i}^{3\times3}$ represents the dilated convolution with the convolution kernel size of 3 × 3, $\epsilon = \{\epsilon_i | i = 2, 3, 4, 5\}$ is the set of the dilation rate of $D_{\epsilon_i}^{3\times3}$, η represents the ReLU activation function, A_i represents the output of the i-th branch-dilated convolution and O_i represents the output of the i-th branch.

The output feature maps of the five branches are fused to obtain a new feature map with rich multi-scale information. Finally, the new feature map is output by adjusting the number of channels through a 1 × 1 convolution layer. The MFF module's final output (Y) is written as follows:

$$Y = f^{1\times1}(C(O_1; O_2; O_3; O_4; O_5)) \tag{3}$$

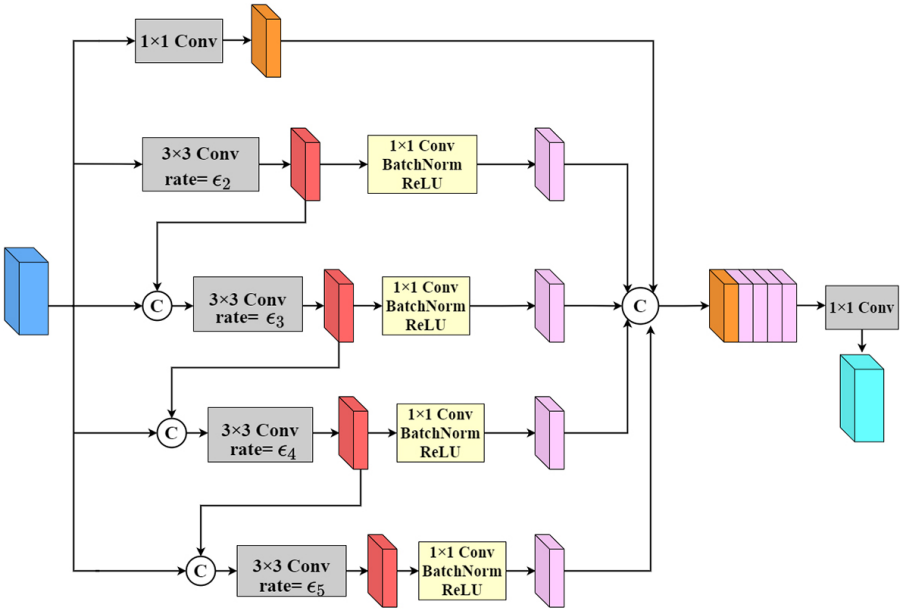

Figure 4. Multi-scale feature-fusion module based on dilated convolutions with different dilation rates.

2.2.3. DUpsampling

The bilinear interpolation method is data-independent, and the overly simple expansion does not consider the correlation between pixels, which affects the quality of the recovered feature maps. DUpsampling [26] can achieve accurate pixel-level predictions and improve segmentation accuracy when recovering image dimensions. Therefore, DUpsampling was used instead of the bilinear interpolation algorithm in the decoding network to avoid the loss of feature information in the bilinear interpolation process and better recover landslide information.

2.2.4. Loss Function

The binary cross-entropy was used as the loss function in the study. By letting N represent the number of samples, P represent the number of pixels in a single sample, $\widehat{y}_i^j$ represent the predicted value of the ith pixel of the jth sample and y_i^j represent the ground-truth value of the ith pixel of the jth sample, the binary cross-entropy loss function could be defined in the following formula:

$$L_{BCE} = \frac{1}{N}\frac{1}{P}\sum_{j=1}^{N}\sum_{i=1}^{P}(y_i^j \times log\widehat{y_i^j} + (1 - y_i^j) \times log(1 - \widehat{y_i^j})) \quad (4)$$

2.3. Evaluation Indicators

Precision, recall, F1 score (F1) and mean intersection over union (MIoU) were chosen to evaluate the performance of the model in the study. Precision refers to the proportion of pixels correctly detected as landslides to all pixels detected as landslides and recall refers to the proportion of all pixels correctly detected as landslides to all pixels labeled as landslides. MIoU is used to calculate the average of the ratio of the intersection of and the union of the two sets of true and predicted values. F1 is defined as the summed average of the precision and recall. The formulas for precision, recall, F1 and MIoU are as follows:

$$Precision = \frac{TP}{FP + TP} \quad (5)$$

$$Recall = \frac{TP}{FN + TP} \quad (6)$$

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

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

where true positive (TP) is the number of pixels correctly classified as landslide pixels, true negative (TN) is the number of background pixels correctly classified, false positive (FP) is the number of background pixels misclassified as landslide pixels and false negative (FN) is the number of landslide pixels misclassified as the background.

3. Experiment

3.1. Study Area

As shown in Figure 5, this study area was located in the northern mountainous area of Huizhou District, Huangshan City, Anhui Province, China. The study area is about 268.16 km^2. The whole study area is dominated by mountainous areas. Strong rainfall often occurs in the summer, and the stability of the mountain is poor; both factors create favorable conditions for the occurrence of landslides. Rapidly obtaining the locations of these landslides is of great importance for post-disaster rescues and reconstructions.

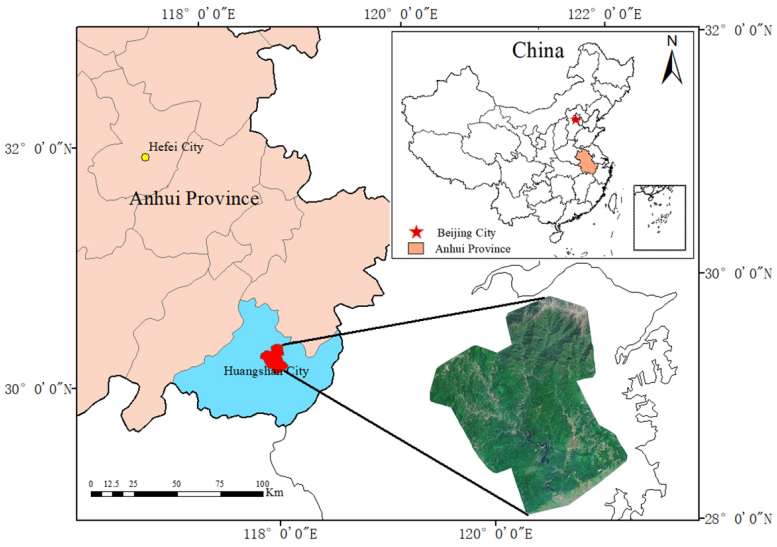

Figure 5. Geographic location and remote-sensing image of the study area.

3.2. Dataset

The data used in the experiments were obtained from Google Earth images. The imagery spatial resolution was 4 m. The main operations performed on the remote-sensing data of the study area included cropping, labeling and data enhancement. The large remote-sensing images were uniformly cropped to the size of 256 × 256 pixels. Labels were made for the small cropped images by using a manual interpretation method. The pixel value of the landslide area was set to 255, and the pixel values of other objects were set to 0. After that, the data were fed into the convolutional neural network for learning. This dataset has a total of 820 landslide images, a quarter of which were used for testing. Because the dataset is small, data enhancement operations were performed during the training stage, mainly including random rotation (90°, 180°, 270°), random vertical flip, random crop, random zoom in/out, random color transformation and random Gaussian noises.

3.3. Experimental Environment

All the code for this experiment was implemented in the deep-learning software framework Pytorch with the built-in Python version 3.6.0. The graphics card that was used

was a Telsa P100-PCIE-16GB, and the processor was an Intel(R) Xeon(R) Silver 4114 CPU @ 2.20 GHz.

In the training process, the batch size was set to 8. SGD was used as the optimizer. The initial learning rate was set to 0.001. If the accuracy of the validation set did not improve after ten training iterations, the learning rate decreased to 0.1 times the original rate. ϵ was $\{2, 3, 5, 7\}$. The total number of training iterations was 100.

3.4. Results

3.4.1. Comparison of L-Unet with the Baseline Model

In order to evaluate the landslide extraction ability of the L-Unet model proposed in this paper, some of the landslide extraction results of L-Unet for the test set are shown in Figure 6. It can be seen that the landslides in the images were extracted completely. Moreover, the comparison of the improved model with the base model is shown in Table 1 and Figure 7.

Table 1 and Figure 7 compare the model performance changes after adding different modules in turn. After adding the MFF module, the precision, recall, MIoU and F1 values of the model were 3.13% 0.53% 2.59% and 1.76% higher than those of the baseline U-Net, respectively. The addition of the residual attention network increased the precision, recall, MIoU and F1 values by 3.95% 1.76% 4.16% and 2.80% respectively, compared to the baseline U-Net values. Based on these findings, we used DUpsampling to replace the bilinear interpolation upsampling to refine the landslide information and obtain the L-Unet model in this paper. Its precision, recall, MIoU and F1 values were 4.15% 2.65% 4.82% and 3.37% higher than those of the baseline U-Net values, respectively. From the results, it can be seen that each improvement of the U-Net model improved the accuracy of landslide extraction in different degrees. Thus, the model proposed in this paper is feasible.

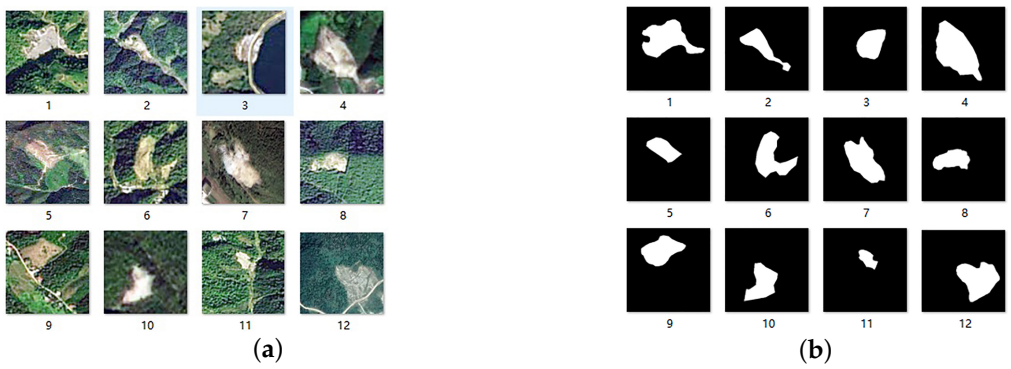

Figure 6. (**a**) Part of the landslide images for the test set and (**b**) extraction results of the L-Unet model.

Table 1. Comparison of the performance before and after improvement.

Model	Precision/%	Recall/%	MIoU/%	F1/%
U-Net	84.39	80.89	70.36	82.60
U-Net+MFF	87.52	81.42	72.95	84.36
U-Net+MFF+ResNet	88.34	82.65	74.52	85.40
L-Unet	88.54	83.54	75.18	85.97

3.4.2. Comparison of L-Unet with Other Models

To compare with other network models, we reproduced several state-of-the-art models: FCN-8s [27], SegNet [28], PspNet [29], HRNet [30], Deeplab v3+ [31], Liu et al. [21], DDCM-Net [32] and MACU-Net [33]. We trained these models on the same dataset, and the results of each model for the test set are shown in Table 2 and Figure 8.

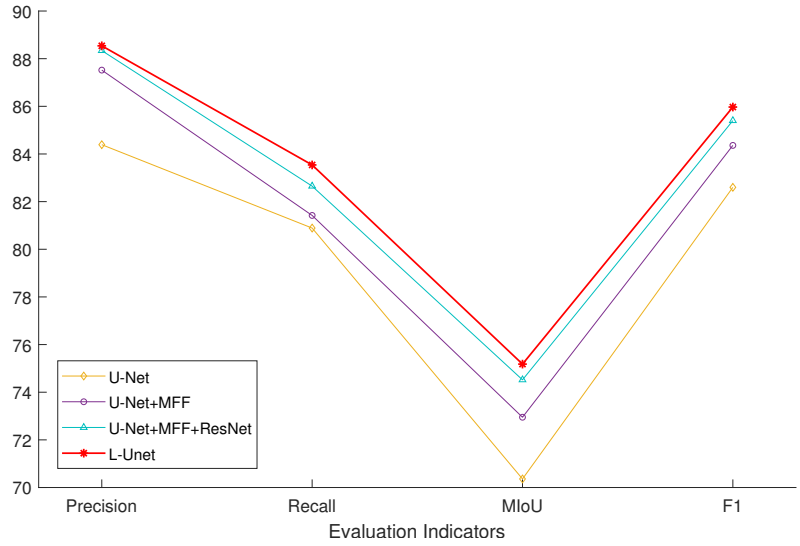

Figure 7. Comparison of the results of models U-Net, U-Net+MFF, U-Net+MFF+ResNet and L-Unet for the four metrics precision, recall, MIoU and F1.

Table 2. Comparison of the performance of different models.

Model	Precision/%	Recall/%	MIoU/%	F1/%
FCN-8s	82.93	81.01	69.43	81.96
SegNet	85.24	77.82	68.58	81.36
PspNet	80.58	83.27	69.35	81.90
HRNet	78.98	71.65	60.17	75.13
Deeplab v3+	83.36	85.84	73.20	84.58
Liu et al. [21]	83.51	82.62	71.04	83.07
DDCM-Net	86.21	83.28	74.06	84.72
MACU-Net	80.94	81.68	68.50	81.31
L-Unet	88.54	83.54	75.18	85.97

As can be seen from Table 2 and Figure 8, the L-Unet model obtained the highest values in the precision, MIoU and F1 metrics compared to the other models. The recall value of L-Unet was 2.3% lower than that of Deeplab v3+, but the precision value of L-Unet was 5.18% higher than that of Deeplab v3+. Because of the ambivalence between precision and recall, L-Unet uses a small decrease in recall in exchange for a large increase in precision. Finally, the F1 value of L-Unet improved by 1.39% compared to Deeplab v3+. It can be seen that although L-Unet has a reduction in recall metrics compared to Deeplab v3+, the former is still optimal.

3.4.3. Application Analysis

An image of Taoyuan Village in Qimen County, Huangshan City was selected for an application analysis, mainly for testing the extraction results of the model for multiple landslides. The image was from Google Earth with a spatial resolution of 4 m. This study area is mainly mountainous, and in addition, it contains houses, roads, bare land and so on. The baseline U-Net, Deeplab v3+, Liu et al. [21], DDCM-Net, MACU-Net and L-Unet models were selected for comparisons for this area, and the results are shown in Figure 9.

As can be seen from the results in Figure 9, except for Figure 9h, the extraction results of the other models had some obvious errors. The small negative samples in the images of the training set caused the models to mistakenly extract features with similar spectral information as landslides. As shown in Figure 9c,d,f,g, parts of the bare land were incorrectly extracted as landslides. In Figure 9e, some of the landslide pixels were missed and not extracted completely. L-Unet effectively avoided these misclassifications, so the extraction results of landslides were more accurate. It can be seen that the overall extraction

effect for landslides was satisfied. Moreover, it is worth mentioning that L-Unet extracted more detailed edge information than the other models.

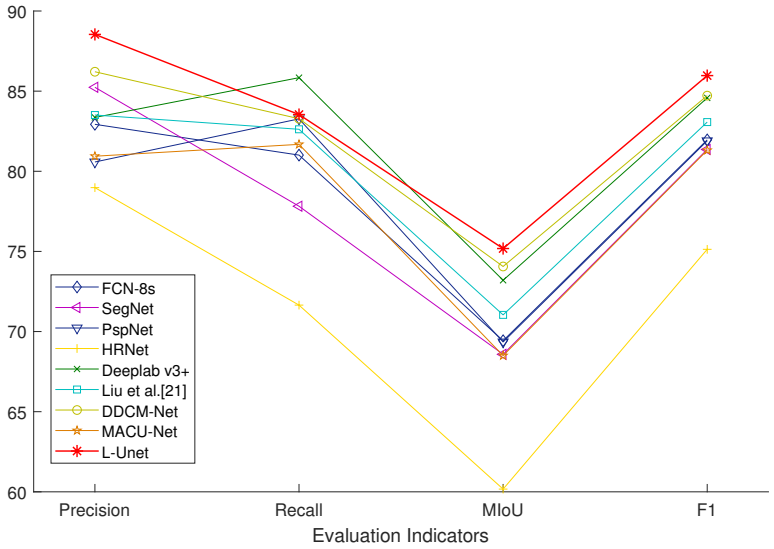

Figure 8. Comparison of the results of models FCN-8s, SegNet, PspNet, HRNet, Deeplab v3+, Liu et al. [21], DDCM-Net, MACU-Net and L-Unet for the four metrics precision, recall, MIoU and F1.

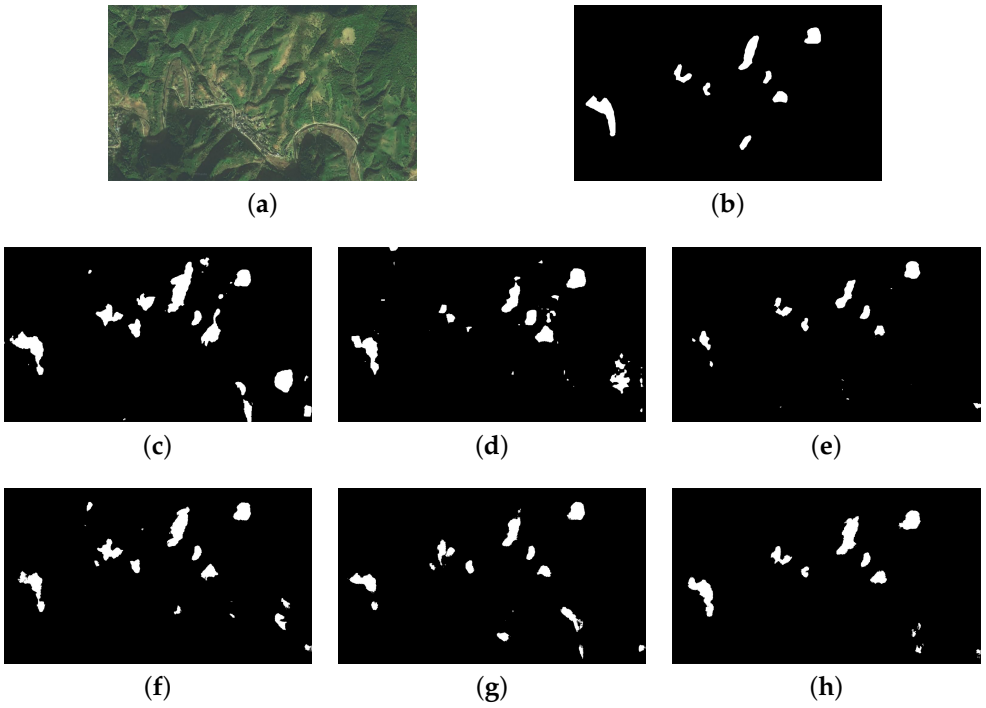

Figure 9. Extraction results: (**a**) image; (**b**) label; (**c**) extraction results of the U-Net model; (**d**) extraction results of the Deeplab v3+ model; (**e**) extraction results of Liu et al.'s [21] model; (**f**) extraction results of the DDCM-Net model; (**g**) extraction results of the MACU-Net model and (**h**) extraction results of the L-Unet model.

3.4.4. Comparison of L-Unet with Other Models on a New Dataset

To test the generalization ability of the model, the Google Earth images in Guichi District, Chizhou City, Anhui Province, were selected for validation. The imagery spatial resolution was 4 m. We created a small landslide dataset by cropping the large imagery into the images with a size of 512 × 512 pixels. This dataset has a total of 100 landslide images,

70 of which were used for training and 30 of which were used for testing. The ground truth was labeled by using a manual visual interpretation, the pixel value of the landslide area was set to 255 and the pixel values of other objects were set to 0. Data enhancement operations were performed, mainly including random 90°, 270° rotations, random crops, random color transformations and random Gaussian noises. The results obtained from the training on the original dataset were used as initialization weights to train the model on this small dataset, and the obtained results are shown in Table 3.

Table 3. Comparison of the performance of different models on a new dataset.

Model	Precision/%	Recall/%	MIoU/%	F1/%
U-Net	81.67	73.03	62.28	77.10
FCN-8s	78.34	72.76	60.58	75.45
Deeplab v3+	84.23	74.26	63.98	78.93
DDCM-Net	84.89	70.24	62.17	76.87
MACU-Net	80.37	74.03	62.69	77.07
L-Unet	86.24	76.82	66.03	81.26

From Table 3, it can be seen that our proposed model has a precision of 86.24%, a recall of 76.82%, an MIoU of 66.03% and an F1 of 81.26%, obtaining the highest values in precision, recall, MIoU and F1 metrics compared to other models. The L-Unet model is superior to other models. This is consistent with the results we obtained for the original dataset.

4. Discussion

The U-Net model is shallow, and the continuous simple downsampling is likely to lead to information loss and the inability to accurately locate the landslide areas. The L-Unet model solves such problems by embedding the residual attention module in the downsampling process. The residual unit can deepen the network to fully extract the sample features [21]. The combination with the CoordAttention mechanism enables the network to continuously capture the location and channel information during the downsampling process. Finally, the model uses the obtained information to precisely locate landslide areas.

In addition, the scales of landslides vary greatly in remote-sensing images. Accordingly, extracting the multi-scale features of images helps to reduce the loss of spatial information and improve the accuracy of landslide extraction. The MFF module is used in L-Unet to fuse the landslide features of different scales and improve the model's ability to extract multi-scale information. In Table 2, we can see that Deeplab v3+ can obtain such high recall values because the atrous spatial pyramid pooling (ASPP) [31] module can obtain the rich multi-scale features of a target. In our model, the MFF module not only obtains multi-scale information but also considers the correlation between convolutional layers with different dilation rates.

In Table 3, the L-Unet model obtained the best score on a new dataset compared to other state-of-the-art models. However, there are very few existing open landslide datasets at present. The spatial resolutions of images play dominant roles in the accuracy of landslide extraction. If we make a higher quality dataset to apply to our model, it will inevitably improve the accuracy and efficiency of the model's landslide extraction.

According to the above results, it can be seen that our model has advantages in the accuracy of landslide extraction and distinguishing background information that is similar to landslides. However, there are still a few errors in extraction. In fact, landslides usually occur in areas with large topographic fluctuations. We can consider extracting the topographic information from the digital elevation model (DEM) so as to further improve the accuracy of the model.

5. Conclusions

In this paper, we proposed a new model for landslide extractions from remote-sensing images based on the U-Net model that is named L-Unet. In the L-Unet model, the added

residual attention network not only deepens the network depth but also makes the model more perceptive of landslide features and reduces the interference of other features. The added MFF module fuses the landslide features of different scales, expands the receptive field and enhances the extraction ability of the model for multi-scale landslide information. The experimental results show that the L-Unet model obtains the best results for our dataset and effectively improves the accuracy of landslide extraction. For practical applications, we need to make a dataset, and then we can simply obtain extraction results. Disaster prevention and control departments can reduce losses of human lives and properties by responding quickly to obtained results. In our future work, we will make a better dataset with different resolutions and types and continue to evaluate the model using the dataset.

Author Contributions: All authors contributed in a substantial way to the manuscript. Conceptualization, Z.D.; methodology, Z.D.; software, S.A., J.Z., J.Y., J.L. and D.X.; data curation, Z.D. and S.A.; writing—original draft preparation, Z.D. and S.A.; writing—review and editing, S.A., J.Z. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the Anhui Province Key R&D Program of China (202004a07020030), the Fundamental Research Funds for the Central Universities (JZ2021HGTB0111) and the Anhui Province Natural Science Foundation (2108085MF233).

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, S.; Zhuang, J.; Zheng, J.; Fan, H.; Kong, J.; Zhan, J. Application of Bayesian Hyperparameter Optimized Random Forest and XGBoost Model for Landslide Susceptibility Mapping. *Front. Earth Sci.* **2021**, *9*, 617. [CrossRef]
2. Tien Bui, D.; Shahabi, H.; Shirzadi, A.; Chapi, K.; Alizadeh, M.; Chen, W.; Mohammadi, A.; Ahmad, B.B.; Panahi, M.; Hong, H.; et al. Landslide detection and susceptibility mapping by airsar data using support vector machine and index of entropy models in cameron highlands, malaysia. *Remote Sens.* **2018**, *10*, 1527. [CrossRef]
3. Hammad, M.; Leeuwen, B.V.; Mucsi, L. Integration of GIS and advanced remote sensing techniques for landslide hazard assessment: A case study of northwest Syria. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *6*, 27–34. [CrossRef]
4. Uddin, M.P.; Mamun, M.A.; Hossain, M.A. PCA-based feature reduction for hyperspectral remote sensing image classification. *IETE Tech. Rev.* **2021**, *38*, 377–396. [CrossRef]
5. Fu, W.; Hong, J. Discussion on application of support vector machine technique in extraction of information on landslide hazard from remote sensing images. *Res. Soil Water Conserv.* **2006**, *13*, 120–122.
6. Xu, C. Automatic extraction of earthquake-triggered landslides based on maximum likelihood method and its validation. *Chin. J. Geol. Hazard Control*. **2013**,*24*,19–25.
7. Li, X.; Cheng, X.; Chen, W.; Chen, G.; Liu, S. Identification of forested landslides using LiDar data, object-based image analysis, and machine learning algorithms. *Remote Sens.* **2015**, *7*, 9705–9726. [CrossRef]
8. Blaschke, T.; Feizizadeh, B.; Hölbling, D. Object-based image analysis and digital terrain analysis for locating landslides in the Urmia Lake Basin, Iran. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **2014**, *7*, 4806–4817. [CrossRef]
9. Yang, Y.; Xie, G.; Qu, Y. Real-time Detection of Aircraft Objects in Remote Sensing Images Based on Improved YOLOv4. In Proceedings of the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 12–14 March 2021; pp. 1156–1164.
10. Niu, R.; Sun, X.; Tian, Y.; Diao, W.; Chen, K.; Fu, K. Hybrid multiple attention network for semantic segmentation in aerial images. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–18. [CrossRef]
11. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 645–657. [CrossRef]
12. Sameen, M.I.; Pradhan, B. Landslide detection using residual networks and the fusion of spectral and topographic information. *IEEE Access* **2019**, *7*, 114363–114373. [CrossRef]
13. Wang, Y.; Wang, X.; Jian, J. Remote sensing landslide recognition based on convolutional neural network. *Math. Probl. Eng.* **2019**, *2019*, 8389368. [CrossRef]
14. Ghorbanzadeh, O.; Blaschke, T.; Gholamnia, K.; Meena, S.R.; Tiede, D.; Aryal, J. Evaluation of different machine learning methods and deep-learning convolutional neural networks for landslide detection. *Remote Sens.* **2019**, *11*, 196. [CrossRef]
15. Zhang, P.; Xu, C.; Ma, S.; Shao, X.; Tian, Y.; Wen, B. Automatic Extraction of Seismic Landslides in Large Areas with Complex Environments Based on Deep Learning: An Example of the 2018 Iburi Earthquake, Japan. *Remote Sens.* **2020**, *12*, 3992. [CrossRef]

16. Lu, H.; Ma, L.; Fu, X.; Liu, C.; Wang, Z.; Tang, M.; Li, N. Landslides information extraction using object-oriented image analysis paradigm based on deep learning and transfer learning. *Remote Sens.* **2020**, *12*, 752. [CrossRef]
17. Ji, S.; Yu, D.; Shen, C.; Li, W.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [CrossRef]
18. Liu, P.; Wei, Y.; Wang, Q.; Xie, J.; Chen, Y.; Li, Z.; Zhou, H. A research on landslides automatic extraction model based on the improved mask R-CNN. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 168. [CrossRef]
19. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
20. Soares, L.P.; Dias, H.C.; Grohmann, C.H. Landslide Segmentation with U-Net: Evaluating Different Sampling Methods and Patch Sizes. *arXiv* **2020**, arXiv:2007.06672.
21. Liu, P.; Wei, Y.; Wang, Q.; Chen, Y.; Xie, J. Research on post-earthquake landslide extraction algorithm based on improved U-Net model. *Remote Sens.* **2020**, *12*, 894. [CrossRef]
22. Ghorbanzadeh, O.; Crivellari, A.; Ghamisi, P.; Shahabi, H.; Blaschke, T. A comprehensive transferability evaluation of U-Net and ResU-Net for landslide detection from Sentinel-2 data (case study areas from Taiwan, China, and Japan). *Sci. Rep.* **2021**, *11*, 14629. [CrossRef]
23. Ghorbanzadeh, O.; Shahabi, H.; Crivellari, A.; Homayouni, S.; Blaschke, T.; Ghamisi, P. Landslide detection using deep learning and object-based image analysis. *Landslides* **2022**, *19*, 929–939. [CrossRef]
24. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
25. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13708–13717.
26. Tian, Z.; He, T.; Shen, C.; Yan, Y. Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 3121–3130.
27. Shelhamer, E.; Long, J.; Darrell, T. Fully convolutional networks for semantic segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 640–651. [CrossRef] [PubMed]
28. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef] [PubMed]
29. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
30. Sun, K.; Xiao, B.; Liu, D.; Wang, J. Deep high-resolution representation learning for human pose estimation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5686–5696.
31. Chen, L. C. Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the 2018 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
32. Liu, Q.; Kampffmeyer, M.; Jenssen, R.; Salberg, A. Dense Dilated Convolutions Merging Network for Semantic Mapping of Remote Sensing Images. In Proceedings of the 2019 Joint Urban Remote Sensing Event (JURSE), Vannes, France, 22–24 May 2019; pp. 1–4.
33. Li, R.; Duan, C.; Zheng, S.; Zhang, C.; Atkinson, P.M. MACU-Net for Semantic Segmentation of Fine-Resolution Remotely Sensed Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 1–5. [CrossRef]