



Article

Multisource Data Fusion and Adversarial Nets for Landslide Extraction from UAV-Photogrammetry-Derived Data

Haiqing He ^{1,*}, Changcheng Li ¹, Ronghao Yang ², Huaian Zeng ³, Lin Li ⁴ and Yufeng Zhu ¹

¹ School of Geomatics, East China University of Technology, Nanchang 330013, China; 201910816006@ecut.edu.cn (C.L.); yfzhu@ecut.edu.cn (Y.Z.)

² College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China; yangronghao@cdut.edu.cn

³ National Field Observation and Research Station of Landslides in the Three Gorges Reservoir Area of Yangtze River, China Three Gorges University, Yichang 443002, China; zhe@ctgu.edu.cn

⁴ Jiangxi Yufeng Intelligent Agricultural Technology Co., Ltd., Ganzhou 341309, China; 201910705033@ecut.edu.cn

* Correspondence: hehaiqing@ecut.edu.cn

Abstract: Most traditional methods have difficulty detecting landslide boundary accurately, and the existing methods based on deep learning often lead to insufficient training or overfitting due to insufficient samples. An end-to-end, semi-supervised adversarial network, which fully considers spectral and topographic features derived using unmanned aerial vehicle (UAV) photogrammetry, is proposed to extract landslides by semantic segmentation to address the abovementioned problem. In the generative network, a generator similar to pix2pix is introduced into the proposed adversarial nets to learn semantic features from UAV-photogrammetry-derived data by semi-supervised operation and a confrontational strategy to reduce the requirement of the number of labeled samples. In the discriminative network, DeepLabv3+ is improved by inserting multilevel skip connection architecture with upsampling operation to obtain the contextual information and retain the boundary information of landslides at all levels, and a topographic convolutional neural network is proposed to be inserted into the encoder to concatenate topographic features together with spectral features. Then, transfer learning with the pre-trained parameters and weights, shared with pix2pix and DeepLabv3+, is used to perform landslide extraction training and validation. In our experiments, the UAV-photogrammetry-derived data of a typical landslide located at Meilong gully in China are collected to test the proposed method. The experimental results show that our method can accurately detect the area of a landslide and achieve satisfactory results based on several indicators including the Precision, Recall, F1 score, and mIoU, which are 13.07%, 15.65%, 16.96%, and 18.23% higher than those of the DeepLabV3+. Compared with state-of-the-art methods such as U-Net, PSPNet, and pix2pix, the proposed adversarial nets considering multidimensional information such as topographic factors can perform better and significantly improve the accuracy of landslide extraction.

Keywords: landslide extraction; adversarial nets; unmanned aerial vehicle photogrammetry; semantic segmentation; topographic factor



Citation: He, H.; Li, C.; Yang, R.; Zeng, H.; Li, L.; Zhu, Y. Multisource Data Fusion and Adversarial Nets for Landslide Extraction from UAV-Photogrammetry-Derived Data. *Remote Sens.* **2022**, *14*, 3059. <https://doi.org/10.3390/rs14133059>

Academic Editor: Andrea Ciampalini

Received: 11 May 2022

Accepted: 22 June 2022

Published: 25 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslides are typical geological disasters which may cause serious damage to people's lives, property, and the natural environment [1,2]. Landslides also usually cause secondary disasters, in which a large amount of sand and gravel deposits may be produced and can further damage houses, destroy farmland, and block rivers. Landslides cause a large number of casualties and property losses every year globally. According to the statistics made by Froude et al. [3], nearly 60,000 people were killed in 4862 landslides all over the world from 2004 to 2016. Therefore, many studies have been conducted regarding

the mechanism and potential occurrence area of landslides to reduce the harm caused by landslides and their secondary disasters.

At present, various methods based on field investigation and remote sensing technologies are widely used in landslide extraction, potential landslide prediction, and disaster management. Field investigation is a traditional and simple means of finding potential landslides, collecting landslide information, and evaluating risk levels. However, it requires surveyors to work close to the adjacent areas of a landslide, which may lead to risks associated with exposing surveyors to danger during a field investigation. By contrast, remote sensing technologies, such as high-resolution optical remote sensing, synthetic aperture radar interferometry, and laser scanning, have already been broadly applied for landslide identification [4,5], deformation monitoring [6,7], and disaster early warning and prevention [8,9] because of their advantages in long-distance detection. In particular, very high resolution (VHR) satellite or aerial remote sensing imagery has been widely used to perform geological disaster interpretations such as landslide extraction.

Currently, landslide detection mainly depends on manual visual interpretation; it has high identification accuracy, but the interpreters must have professional knowledge and experience related to landslides. To accurately and automatically detect landslides from remote sensing images with complex backgrounds, many studies have been conducted and can be mainly classified into (1) object-based image analysis (OBIA) methods [10], (2) change detection methods [11], and (3) deep-learning-based methods [12].

Landslide detection via OBIA methods is achieved in terms of the characteristics of image color, texture, and spectrum, and some methods have been proposed to satisfy the requirement of landslide detection under different geological environments or from remote sensing data sources. Borghuis et al. [13] used SPOT-5 multispectral images to conduct supervised and unsupervised classification for the detection of landslides caused by typhoons, and the results show that unsupervised classification can identify 63% of landslide areas and even more small-area landslides than visual interpretation. Han et al. [14] used the segmentation scale of 25 pixels and 14 parameters of OBIA to extract landslides. However, the difficulty in OBIA lies in setting parameters and classification rules in the process of multiscale segmentation, which requires expert experience to adjust and optimize parameters iteratively. Specifically, it has difficulty obtaining optimal landslide detection results and is unsuitable for some special applications such as emergency monitoring. From the perspective of classification, the setting of artificial thresholds limits the improvement in the landslide extraction accuracy. Some methods have also been studied to detect the changed areas between multitemporal remote sensing images, and the changed areas can be considered landslide candidates through the change detection method. Lu et al. [15] proposed an object-oriented semi-automatic method for multitemporal landslide change detection, in which image texture, terrain, and spectral information for multiscale segmentation and optimization was used. Martha et al. [16] developed a multiresolution segmentation method, which combined the characteristics of spectrum, the vegetation index, and morphology and could detect landslides from multitemporal satellite remote sensing images. Markov-random-field-based change detection methods [17,18] have been also proposed to detect areas of landslide from multitemporal satellite optical images or aerial orthophotos (i.e., digital orthophoto maps (DOMs)). Unfortunately, multitemporal remote sensing images before and after a landslide in the same place are often difficult to obtain because the locations of landslides are not known in advance. Given that the occurrence of a landslide in a slope sliding process is closely related to the terrain, the abovementioned methods have difficulty distinguishing easily confused landslide areas (e.g., bare land) only based on two-dimensional images.

Various semantic segmentation methods, as inspired by the great success of deep learning in image feature extraction and classification, have been studied to significantly improve the performance of landslide detection by image segmentation. Multisource data such as topographic factors were introduced to perform landslide extraction together, and various methods considering specific data (e.g., the vegetation index, the digital elevation

model (DEM), slope, and aspect) have been also proposed [19,20]. Ji et al. [21] used an attention-enhanced convolutional neural network (CNN) to detect landslides from remote sensing images and the DEM. Shi et al. [22] combined a CNN and change detection to quickly identify landslides in Lantau, Hongkong. Ghorbanzadeh et al. [23] compared the performance of landslide detection using traditional methods (i.e., support vector machine and random forest) and deep-learning-based methods, and they concluded that CNN-based methods have great potential in landslide detection. Qi et al. [24] improved U-Net [25], namely ResU-Net, to extract landslides from remote sensing images with complex backgrounds (e.g., river valleys, beaches, and non-planted terraces). Asadang et al. [26] combined faster R-CNN and a classification decision tree to detect landslides from satellite images. Yu et al. [27] constructed an end-to-end semantic segmentation model and used multisource data (e.g., the vegetation index and the DEM) for landslide extraction. Liu et al. [28] proposed an improved U-Net semantic segmentation model, in which terrain factors such as the digital surface model (DSM), slope, and aspect were added to detect landslides after earthquakes. Xia et al. [29] proposed a fully convolutional spectral-topographic fusion network, which considered terrain factors (i.e., slope and aspect) and the normalized difference vegetation index, to detect landslides. In most cases, the scale of a landslide is small, and the distribution is fine and discrete. Although topographic factors used can effectively improve the robustness of landslide detection, the existing methods mainly depend on satellite images and low-precision topographic factors, which still do not satisfy the requirement of accurate and fine landslide characterization. In the aforementioned methods, the additional DEM is required if non-stereoscopic satellite-based remote sensing is used, and the DOM and DEM are inconsistent because of nonsynchronous acquirement.

In recent years, low-altitude photogrammetry (LAP) based on unmanned aerial vehicle (UAV) mapping has become a popular technical means to capture VHR aerial images for remote sensing detection due to its significant advantages such as higher resolution, lower cost, and faster response than satellite images. Synchronously, high-precision and dense three-dimensional point clouds can be intuitively derived from overlapped UAV images through dense matching (e.g., semi-global matching), and the geometric structure information of a landslide's surface is much richer than that derived by satellite-based mapping. LAP can also be an effective supplement to satellite remote sensing, especially in landslide extraction applications. However, the richer details brought by UAV images also lead to more complex image backgrounds, which greatly increase the difficulty of automatic landslide interpretation from UAV images. The commonly used CNNs or U-Net for landslide extraction are the typical discriminative networks. The training of these networks requires a large number of pixel-level labels, while reducing the number of labels may degrade their performance. Collecting a large number of landslide training samples is difficult because of the uncertainty of landslide occurrence. In contrast to the discriminative network, a generative adversarial network (GAN) is an unsupervised or semi-supervised network and requires a small number of samples to train. Therefore, GANs have also been widely used in the field of remote sensing object detection. For example, a conditional GAN called pix2pix [30] was used for remote sensing change detection [31,32].

In this study, an end-to-end, semi-supervised adversarial network based on multisource data fusion is proposed to perform landslide extraction to fully utilize high-resolution and high-precision three-dimensional information derived by UAV photogrammetry. A transfer learning generative network similar to the architecture in pix2pix, namely a generator, which considers visible spectra and topographic factors derived by UAV photogrammetry, is proposed to learn the semantic features by semi-supervised operation to reduce the training samples. These features are then used to separate landslides from complex backgrounds. In particular, topographic factors such as the DSM, slope, and aspect are selected to characterize the terrain closely related to the occurrence of landslides. To avoid CNN training from scratch, a state-of-the-art network with atrous spatial pyramid pooling (ASPP), namely DeepLabV3+ [33] (i.e., discriminator), is used to obtain pre-trained parameters of transfer learning for further fine-tuning. ASPP is helpful for multiscale

semantic segmentation from spectral information to characterize landslides, and high-precision topographic factors from UAV-photogrammetry-derived data are used to extract geometric structure information from landslides. DeepLabv3+ mainly focuses on the semantic segmentation of natural images, in which the object's overall structure and location information of a certain type can be extracted. However, DeepLabv3+ cannot characterize the relationship between two ground objects because of complex ground object distribution in remote sensing images. Therefore, DeepLabv3+ is improved by introducing multilevel skip connection architecture to characterize the spatial relationship among ground objects and retain the boundary information at all levels, and end-to-end mapping between input and output can be established through the generative and discriminative networks. Subsequently, multidimensional low- and high-level semantic multiscale features, which are learned from the proposed adversarial nets, are utilized to perform automatic landslide identification and boundary detection.

The main contribution of this work focuses on the design of an adversarial network framework for landslide extraction from UAV-photogrammetry-derived data. In this study, multisource data fusion enables semantic spectral and geometric features related to landslides to be extracted from complex VHR remote sensing image backgrounds. The generative network used is helpful in significantly reducing the training samples by semi-supervised operation. Improved DeepLabv3+ with multilevel skip connection architecture offers the possibility to capture low- and high-level features for landslide extraction via multiscale semantic segmentation.

The remainder of the paper is organized as follows: Section 2 describes the details of the proposed framework for landslide detection. Sections 3 and 4 present the comparative experimental results in combination with a detailed analysis and discussion. Section 5 concludes this study and discusses possible future work.

2. Methods

2.1. Overview of the Proposed Method

The workflow of the proposed method focuses on landslide extraction. For this purpose, the multisource data fusion model and adversarial nets were proposed in this study. The framework of landslide extraction mainly included three stages, namely adversarial net training, multisource data derived by UAV photogrammetry, and landslide extraction, as shown in Figure 1. In the adversarial net training stage, VHR remote sensing images of some typical landslides and their topographic factors were selected and labeled as examples to train the proposed adversarial nets including generative and discriminative networks. In the multisource data generation stage, UAV overlapping images were used to generate the DOM and DSM through several indispensable steps such as distortion correction, image matching, structure from motion (SfM), and bundle adjustment [34]; then, some topographic factors (e.g., slope and aspect) and the gamma-transform green leaf index (GGLI) were derived based on the previous studies [35]. In the landslide extraction, multisource data derived from the second stage were used to perform landslide extraction through the discriminative network trained in the first stage.

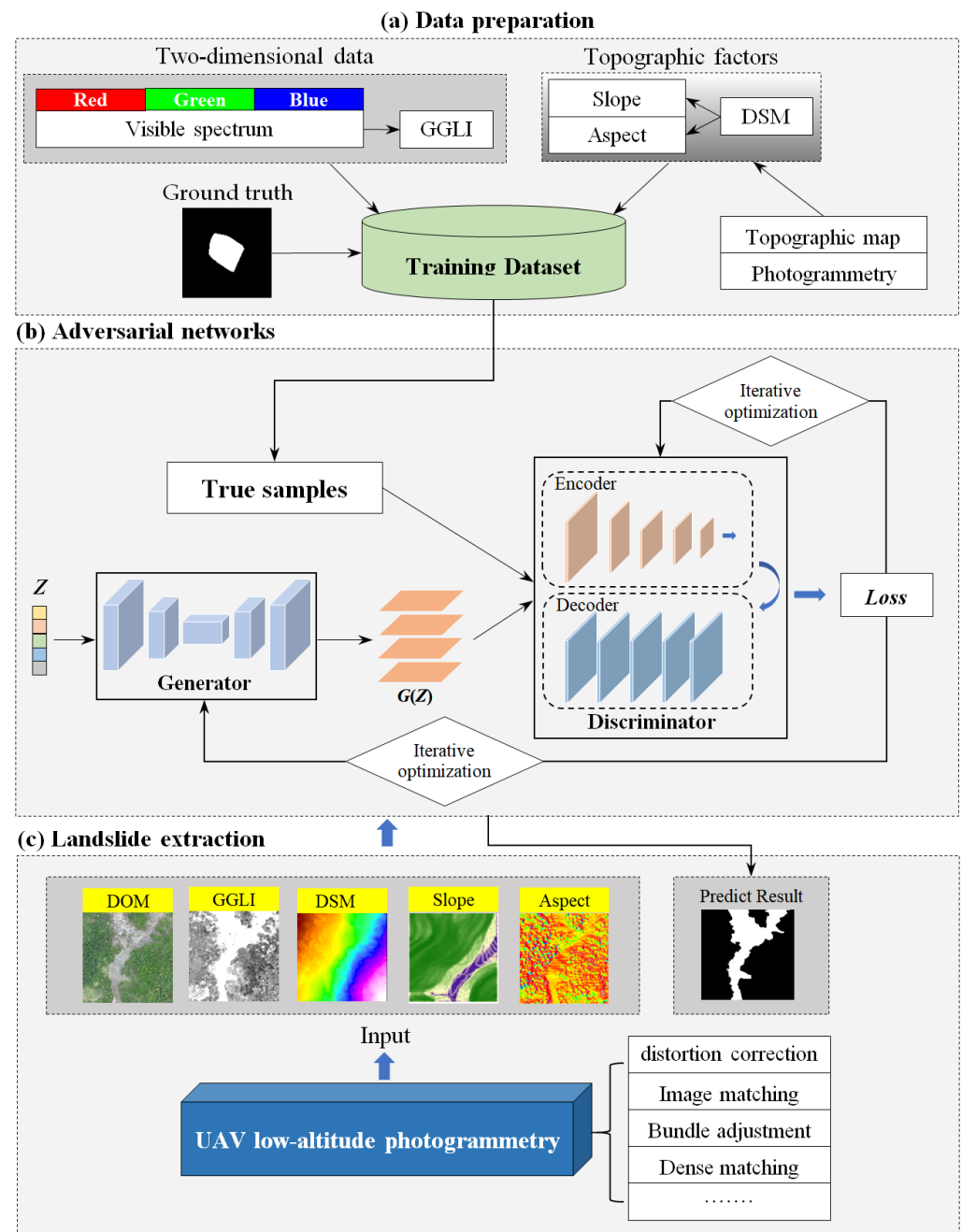


Figure 1. Workflow of the proposed method. (a) Data preparation; (b) Adversarial networks including generator and discriminator; (c) Landslide extraction from UAV-photogrammetry-derived data.

2.2. Proposed Adversarial Nets

In this study, we attempted to conduct CNN-based semantic segmentation for landslide extraction using adversarial networks. In most cases, training a very deep CNN is difficult because a large number of parameters need to be adjusted. In particular, collecting a large number of high-precision UAV data in a landslide area is a tedious task due to the unpredictability of landslide occurrence. Therefore, in our networks, we used the generator referred to as pix2pix to reduce the requirement of training samples. However, the limitation of pix2pix is that the region of interest is extracted by multiconvolutional operation, which usually leads to computational redundancy and difficulty obtaining a clear boundary contour of landslide. Thus, we performed semantic segmentation for landslide extraction by introducing ASPP and an encoder–decoder architecture similar to DeepLabV3+.

2.2.1. Multisource Data Fusion

Ideally, if a deep network is reasonable and fully trained, then vegetation information hidden in an RGB (i.e., red, green, and blue) image may be learned directly; similarly, slope and aspect hidden in the DSM or DEM may be derived. However, the following calculation formula of GGLI [35], slope, and aspect shows that it needs nonlinear implicit learning. To accelerate the convergence of the proposed network, multisource data, including spectral (RGB bands and GGLI) and topographic (DSM or DEM, slope, and aspect) information, were explicitly used in this study to feed into the proposed adversarial nets.

$$GGLI = 10^{\gamma} \cdot \left(\frac{2G - R - B}{2G + R + B} \right)^{\gamma}, \quad (1)$$

where γ denotes a gamma value, and R , G , and B are the three components of RGB color.

$$slope = \sqrt{slope_x^2 + slope_y^2}, \quad (2)$$

$$aspect = \frac{slope_y}{slope_x}, \quad (3)$$

where $slope_x$, $slope_y$ are the slopes in the direction x and y , respectively. The slope value can be calculated based on the elevation difference and horizontal distance.

Multisource data fusion for object detection based on deep learning can generally be classified into layer stacking [36] and feature fusion [37,38]. However, simple layer stacking cannot fuse spectral and topographic features, and two parallel branch networks, as shown in Figure 2, were designed for semantic feature representation. One branch network (i.e., DeepLabv3+ network) was used to extract complex nonlinear features from RGB images, and the other branch network with shallower architecture (including 15 convolutional layers) was explored to characterize the vegetation and terrain related to landslides.

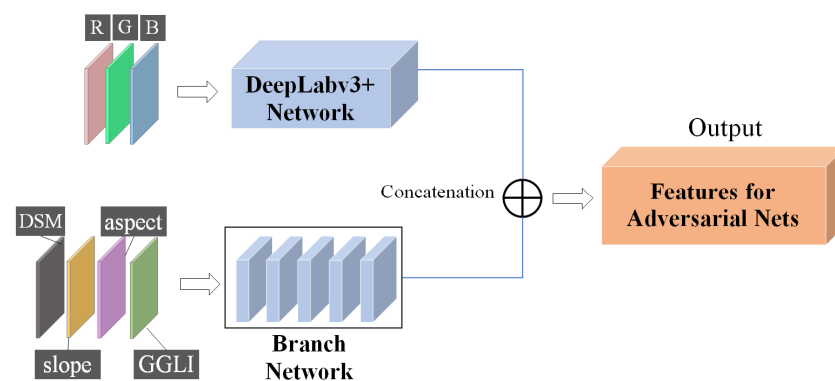


Figure 2. Architecture of multisource data fusion network.

2.2.2. Generative Network

As shown in Figure 1, the proposed adversarial net consisted of a generator and a discriminator; the former was used to generate simulated samples such that their distribution p_z was as similar as the true sample data distribution p_d as possible to deceive the discriminator D . The goal of the proposed adversarial nets was to distinguish the samples from the generator or true samples by the discriminator as much as possible and optimize the proposed adversarial nets through iterative machine learning with a confrontational strategy. A minimum–maximum competitive relationship existed between the generator and discriminator, in which loss function can be expressed as

$$\min_G \max_D V(G, D) = E_{x \sim p_d(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(x)))], \quad (4)$$

where x is the true samples and z is the random vector input into the generator. $D(x)$, $D(G(x))$ are the outputs of the discriminator when the input is true samples and the input is pseudo data generated by the generator, respectively. Here, the generator similar to pix2pix network was introduced into the proposed adversarial nets, and the architecture is given in Table 1. The skip connection was also used in the generator similar to U-Net, eight downsampling and upsampling steps were conducted, and a tanh activation function was added at the end. Then, the training samples could be extended by generating pseudo data, and the training process could be conducted via the semi-supervised operation to reduce the requirement of the number of labeled samples.

Table 1. Detailed configuration of the used generator.

Layer	Kernel	Stride	Padding	Feature	Activation
conv1	4×4	3	1	64	None
conv2	4×4	2	1	128	ReLU
conv3	4×4	2	1	256	ReLU
conv4-8	4×4	2	1	512	ReLU
dconv9-12	4×4	2	1	512	ReLU
dconv13	4×4	2	1	256	ReLU
dconv14	4×4	2	1	128	ReLU
dconv15	4×4	2	1	64	ReLU
dconv16	4×4	2	1	3	tanh

2.2.3. Improved DeepLabv3+ for Discriminator

Most VHR remote sensing semantic segmentation methods based on CNNs are achieved by (1) broadening the receptive field and combining multiscale operation for feature extraction [39–41], (2) fusing contextual information using skip connection [42], or (3) integrating multiple networks by transfer learning [43,44].

Meaningfully, a state-of-the-art network, namely DeepLabv3+, has an added ASPP module, which uses atrous convolutional operation to extract multiscale features in the larger receptive fields for capturing multiscale contextual information. In the encoder of DeepLabv3+, a deeper Xception network with separable convolutional operation is used to extract features and is helpful for retaining location information with unchanged semantic information. In the decoder, the low-level features are directly transmitted from the backbone network, and high-level features are fused with low-level features after a $4\times$ upsampling operation to obtain the output consistent with the input size. Thus, it can be inferred that DeepLabv3+ can perform well for VHR remote sensing semantic segmentation. However, the downsampling operation in DeepLabv3+ may weaken the extraction performance of location information and still make obtaining a clear boundary contour of landslide difficult.

In this study, multilevel skip connection architecture with an upsampling operation was designed to retain the boundary information at all levels as much as possible to address the aforementioned problem of DeepLabv3+ for landslide extraction. The architecture of improved DeepLabv3+ is given in Figure 3. Similar to DeepLabv3+, the improved model consisted of an encoder (Figure 3a) and decoder (Figure 3b), and the convolutional operation for landslide feature extraction for RGB images was consistent with DeepLabv3+. Unlike original DeepLabV3+, an architecture of topographic feature extraction, namely the topographic CNN shown in Figure 4, was inserted into the encoder to concatenate topographic features together with spectral features. Subsequently, low- and high-level features were concatenated through the upsampling module shown in Figure 3c to effectively obtain the contextual information, and it helped to transmit the detailed landslide features from lower to higher layers. Therefore, the spatial information of the landslide weakened by the pooling operation could be better restored, and the clear boundary contour of the landslide may have been delineated after the decoding operation in the final prediction.

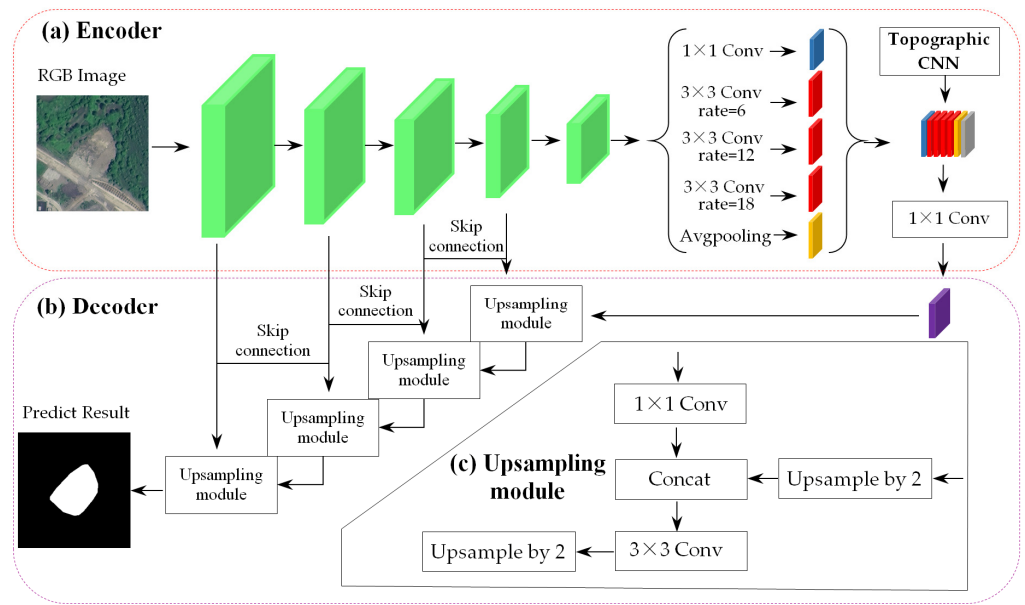


Figure 3. Architecture of the improved DeepLabv3+. (a) Encoder for landslide feature extraction; (b) Decoder for concatenating low- and high-level features; (c) Upsampling module.

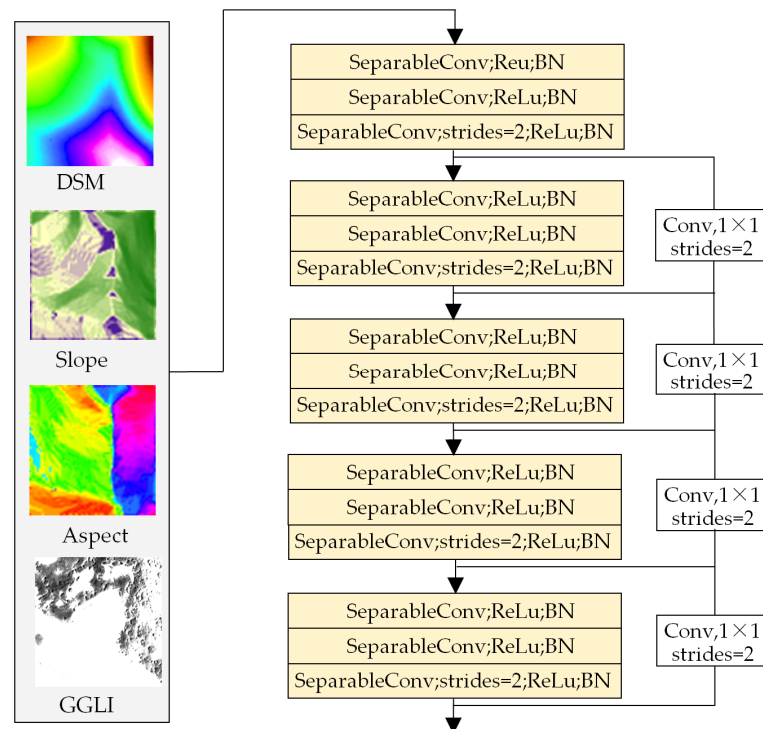


Figure 4. Architecture of topographic CNN. BN denotes the operation of batch normalization.

Here, the weight training of discriminator D needed to be realized by minimizing the cross-entropy loss function L_D as

$$L_D = - \sum ((1 - y_n) \log(1 - D(G(X_n))) + y_n \log(D(Y_n))), \quad (5)$$

where X_n is the input, and Y_n is the label corresponding to X_n . y_n denotes the input source of discriminator D , and $y_n = 0$ and $y_n = 1$ are the input from the output of generator G and the true sample, respectively. The traditional GAN is different in the weight training of

generator G with the proposed GAN for landslide extraction, and different loss functions need to be given to obtain different weights. That is, when generator G is trained, it should focus on the loss function calculated from the label data, and the adversarial and semi-supervised loss functions are relatively less important. Thus, the generator G was trained by minimizing a given multitask loss function L_G as

$$L_G = L_{seg} + \lambda L_{adv} + \beta L_{semi}, \quad (6)$$

$$L_{seg} = - \sum Y_n \log(G(X_n)), \quad (7)$$

$$L_{adv} = - \sum \log(D(G(X_n))), \quad (8)$$

$$L_{semi} = - \sum I(D(G(X_n)) > T_{semi}) \cdot Y_n \log(G(X_n)), \quad (9)$$

where L_{seg} and L_{semi} correspondingly denote the loss function of generator G training with labeled and unlabeled data, and L_{adv} denotes the adversarial loss function designed to confuse the discriminator D . λ and β are the weights corresponding to adversarial and semi-supervised loss function in multitask loss function, respectively. $I(\cdot)$ is an index function to obtain the image area involved in the back propagation operation. T_{semi} is an given threshold that is usually set to less than 0.3. The values of λ and β are set based on the following strategies: (1) the optimal λ is determined by training the generator network G (i.e., fine-tuning the pix2pix model) and the whole $G - D$ network using the labeled data. (2) The λ value is fixed on the basis of (1), T_{semi} is set to a certain value less than 0.3, and the optimal β is determined by training G and $G - D$ using half of the labeled and unlabeled data, respectively. (3) The values of λ and β are fixed first, which is similar to (2), and the optimal T_{semi} can then be determined.

2.2.4. UAV Photogrammetry

In this study, UAV LAP was used to generate the DOM and DSM synchronously, that is, RGB images and topographic factors could be derived simultaneously. On the basis of our previous studies [35], photogrammetric technologies were used to generate the DOM and DSM through a series of steps, such as distortion correction, image matching, SfM, bundle adjustment, and semi-global matching. Detailed technical details were also provided in our previously published literature [45].

3. Experiment Results and Analysis

3.1. Dataset

3.1.1. Training Dataset

In this study, the proposed adversarial nets were trained using the open-source dataset [21], which can be downloaded at http://gpcv.whu.edu.cn/data/Bijie_pages.html (accessed on 10 May 2022). This dataset consists of 770 samples, including optical remote sensing images, labels, patch boundaries, and the DSM related to landslides. Non-landslide samples were also provided and mainly composed of mountains, villages, roads, rivers, forests, and farmlands. These samples in the dataset were cropped from VHR satellite images with the spatial resolution of about 0.8 m, and the elevation accuracy of the DSM was about 2 m. The boundaries of landslides were delineated manually using ArcGIS software.

Although the number of these labeled samples appeared small, we increased the number of training samples through the data augmentation strategy similar to [46]. Meanwhile, the aforementioned transfer learning with the pre-trained parameters and weights, shared with pix2pix and DeepLabv3+, was used to perform landslide extraction training. On the basis of the dataset, the training samples derived from optical images and the DSM were used to generate serial patches with GGLI, slope, and aspect calculated using Equations (1)–(3), respectively. Several examples of landslide training samples are illustrated in Figure 5.

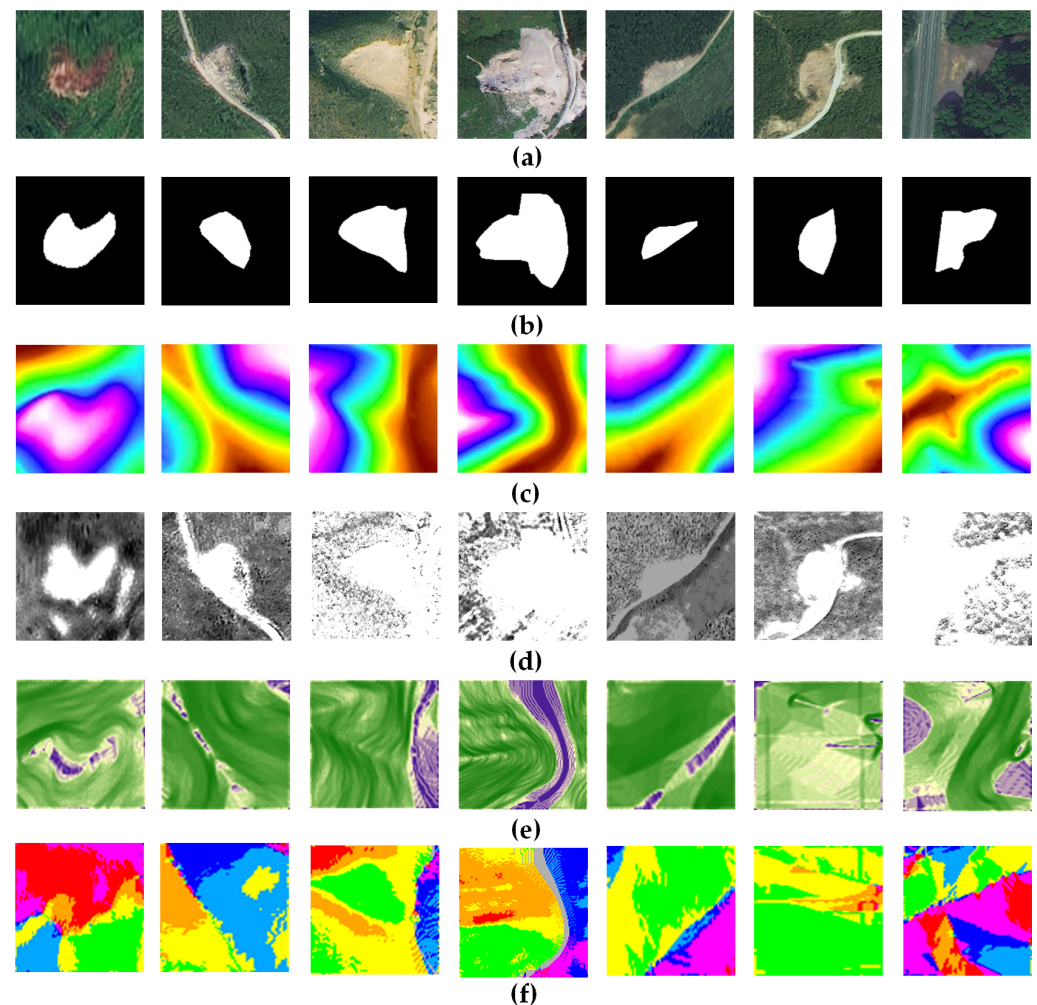


Figure 5. Examples of landslide training samples. (a) is the RGB images of seven landslides. (b–f) are the ground truth, DSM, GGLI, slope, and aspect corresponding to (a), respectively. Notably, the gray intensity or color rendering in (c–f) is produced by the value of DSM, GGLI, slope, and aspect, which are normalized to 0 to 1 for the training input of the proposed adversarial nets.

3.1.2. Test Dataset

A typical landslide in mountainous terrain (located in Danba County, Sichuan Province, China) was selected as the study area, as shown in Figure 6a,b. This landslide was located in the southeast of the Qinghai Tibet Plateau in China, with an altitude of 1700–5820 m. The study area belongs to a subtropical climate zone, with an annual average temperature of 14.2 °C, four distinct seasons, and concentrated rainfall. The topography of the study area includes high mountains and valleys, the geology is composed of a series of parallel linear folds, and the crustal activity is relatively strong and results in frequent geological disasters such as landslides.

In June 2020, heavy rainfall, mountain torrents, and debris flows in Meilong gully, Danba County led to large-area landslides shown in Figure 6c, which brought serious troubles to people's lives. The test dataset in this study was acquired in Meilong gully via UAVs in July 2020. Given the flexible operation of quadcopters, this study used a small quadcopter (DJI Phantom 4 RTK, DJI, Shenzhen, China), which has strong stability and can automatically plan routes and land at fixed waypoints. The maximum takeoff weight of this UAV is 1.4 kg, the longest flight time is about 30 min, the fastest flight speed is about 50 km/h, the equipped image sensor is a 1-inch complementary metal–oxide–semiconductor, the effective pixels are about 20 million, and the image storage format is JPEG.

The UAV image acquisitions were performed under good weather conditions, such as sunny and winds <10 m/s. The relative flight height was about 200 m, and the flight speed was 3 m/s. Figure 6c shows the UAV-photogrammetry-derived DOMs and DSMs of the study area.

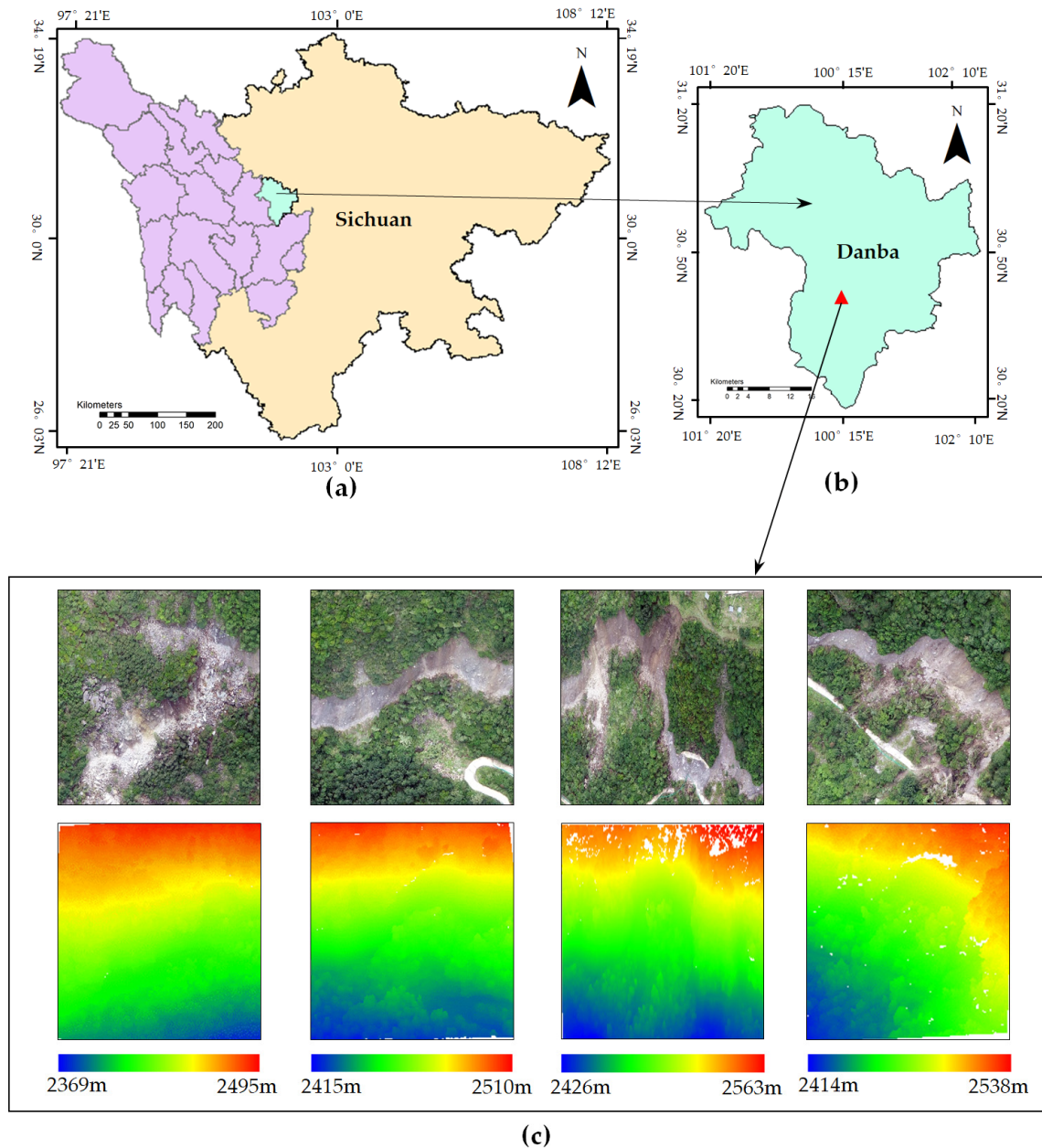


Figure 6. Study area located in Danba County, Sichuan Province, China. (a,b) show the location of the study area, and the DOMs and DSMs of the study area are illustrated in (c).

3.2. Evaluation Criteria of Landslide Extraction Performance

Four indicators, namely *Precision*, *Recall*, *F1_score*, and *mIoU*, were used to quantitatively evaluate the performance of the proposed network in the experiments. The four indicators can be calculated as follows:

$$Precision = \frac{TP}{TP + FP'} \quad (10)$$

$$Recall = \frac{TP}{TP + FN'} \quad (11)$$

$$F1_score = 2 \times \frac{precision \times recall}{precision + recall}, \quad (12)$$

$$mIoU = \frac{TP}{TP + FP + FN}, \quad (13)$$

where TP , FP , and FN denote true positive (i.e., the number of pixels correctly extracted as belonging to landslides), false positive (i.e., the number of non-landslide pixels incorrectly extracted as belonging to landslides), and false negative (i.e., the number of pixels incorrectly extracted as belonging to non-landslides). The method used in the experiments performed better when the four values of *Precision*, *Recall*, *F1_score*, and *mIoU* were larger.

3.3. Training and Validation

In the training stage, the training samples were extended to avoid overfitting by data augmentation [46], of which, 75% of the samples were used for training and 25% for validation. The initial learning rate of 0.001 and epoch of 200 were set to train using the Adam optimizer, and the learning rate was gradually reduced by a multiple of one tenth when the loss did not decrease after every 10 epochs. The resolution of landslide optical RGB images and derived additional data (i.e., GGLI, DSM, slope, aspect) in the input data was adjusted to 256×256 , and the input multisource data were normalized to a floating-point value between 0 and 1. The proposed network was trained in parallel on NVIDIA GPUs, and the Keras-TensorFlow library was introduced to perform convolutional operation.

Meanwhile, to evaluate the performance of single-source data (i.e., RGB images) and multisource data (i.e., RGB images, GGLI, DSM, slope, and aspect) for landslide extraction in the proposed network, two groups of comparative experiments were conducted in the same hardware and software conditions, and single data and multisource data were individually used to train the proposed network. The comparative training results of single data and multisource data for the proposed network are illustrated in Figure 7a,b, which indicate that the multisource data used could improve the training accuracy and accelerate the convergence of the proposed network. The use of multidata can provide additional information (e.g., three-dimensional surface), which enriches landslide characterization and offers opportunities for landslide extraction in complex terrains. In the proposed adversarial nets, improved DeepLabv3+ played a key role in landslide feature extraction. Training using the training dataset was performed to enable a comparison with the original DeepLabv3+, and its results are illustrated in Figure 7c,d. As observed, the proposed adversarial nets with the generator and discriminator could significantly improve the accuracy and accelerate the convergence of training. Therefore, the semi-supervised operation, which was guided by the generator network, could reduce the requirement of the number of labeled samples. Thus, it helped to avoid overfitting and accelerate training convergence. In contrast to the downsampling operation in DeepLabv3+, multilevel skip connection architecture with an upsampling operation in the discriminator network can retain the boundary information at all levels, which can perform better for landslide extraction than DeepLabv3+.

The prediction results obtained from the validation dataset were also compared with the ground-truth data to quantitatively evaluate the effectiveness of the proposed network. The aforementioned indicators were calculated through the confusion matrix, which is shown in Table 2. The accuracy indicators (i.e., *Precision*, *Recall*, *F1_score*, and *mIoU*) obtained from the proposed network with multisource data fusion were 6.22%, 4.08%, 3.83%, and 4.80% higher than those with single data, which indicated that multisource fusion could improve the performance of landslide extraction. Compared with DeepLabv3+, the accuracy indicators obtained from the proposed network with multisource data fusion were 13.07%, 15.65%, 16.96%, and 18.23% higher, which implied that the proposed network for landslide extraction using adversarial network architecture and an upsampling operation was much better than DeepLabv3+.

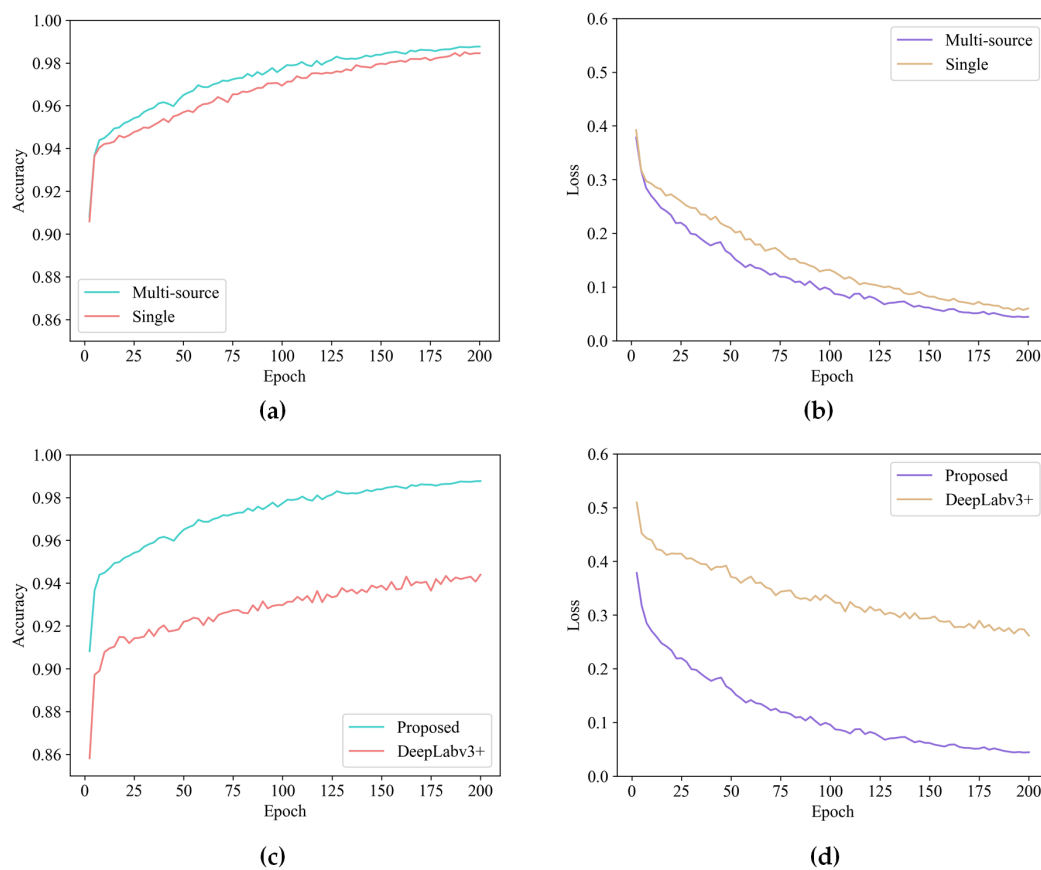


Figure 7. Comparisons of the training results. (a,b) are the comparative training results of single data and multisource data for the proposed network, and (c,d) are the comparative training results of the original DeepLabv3+ and the proposed network.

Table 2. Quantitative evaluation with the ground-truth data.

Network	Input	Precision	Recall	F1_Score	mIoU
DeepLabv3+	Multisource	0.7552	0.6689	0.6612	0.5482
Proposed	Single	0.8237	0.7846	0.7925	0.6825
	Multisource	0.8859	0.8254	0.8308	0.7305

Compared with the visualization of DeepLabv3+ and the proposed network, several examples shown in Figure 8a were selected to exhibit the experimental results shown in Figure 8c–e. The landslide extraction effect of the proposed network based on multisource fusion was better than that of the experiments performed based on single data. As shown in Figure 8c, compared with the ground-truth data shown in Figure 8b, the landslide regions obtained using DeepLabv3+ were broken, and it could not accurately characterize the boundaries of a landslide. In contrast, the proposed network, whether using single or multisource data, performed better in terms of region completeness and boundary accuracy, as shown in Figure 8d,e. Especially for complex terrain conditions, as shown in Figure 8d, the regions marked by red circles could not perfectly delineate the boundaries of landslides by the proposed network based on single data. On the contrary, as shown in Figure 8e, multisource data fusion combining spectral and topographical information could improve the performance of landslide detection and the characterization of landslide boundaries.

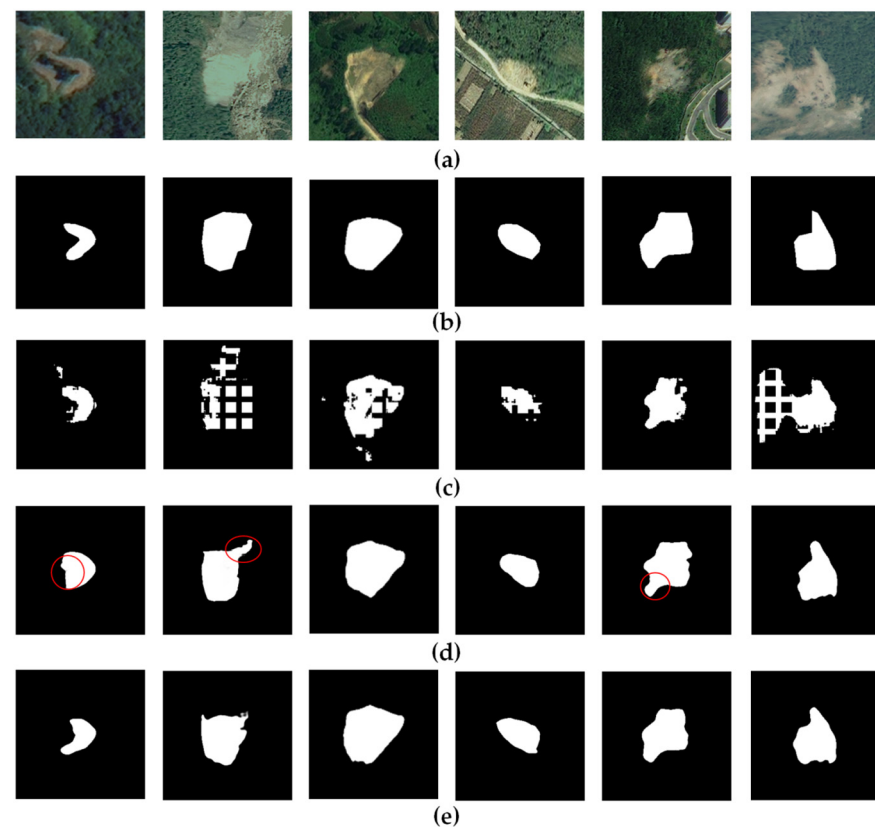


Figure 8. Examples of landslide extraction using DeepLabv3+ and the proposed network. White color denotes the region of landslide, and black color denotes the image background. (a) is the RGB images, and (b) is the ground-truth landslide boundaries corresponding to (a). (c–e) are the experimental results obtained using DeepLabv3+, the proposed network with single and multisource data, respectively. Red circle is used to highlight the comparative region.

In this study, the proposed network was designed with reference to two state-of-the-art deep learning networks (pix2pix and DeepLabv3+), and the architectures of the two networks were proven to be optimal in experiments [32,33]. Additionally, the architecture of the topographic CNN included five convolutional blocks, each of which included three convolutional layers defined through ablation studies, which were conducted by reducing (i.e., 4×3 layers) or adding to (i.e., 6×3 layers) the number of convolutional layers in the topographic CNN. The quantitative evaluation of the topographic CNN with the different number of layers is given in Table 3; it can be seen that the topographic CNN with 5×3 convolutional layers performed better than the other two.

Table 3. Quantitative evaluation of topographic CNN with the different number of convolutional layers.

Convolutional Layer	Precision	Recall	F1_Score	mIoU
4×3 layers	0.8488	0.7569	0.7534	0.6631
5×3 layers	0.8859	0.8254	0.8308	0.7305
6×3 layers	0.8900	0.8024	0.7756	0.6572

3.4. Comparisons with State-of-the-Art Methods

The test dataset acquired in the study area was also used to verify the effectiveness of landslide extraction using the adversarial nets to further evaluate the performance of the proposed network, except for the open-source dataset. In the experiments, the whole

remote sensing covering landslides could not be directly input into the proposed network for prediction due to the limited memory of the computer. Thus, several patches with sizes of 256×256 pixels were cropped to perform landslide extraction.

State-of-the-art methods based on deep learning, including U-Net [25], PSPNet [47], DeepLabv3+ [33], and pix2pix [30,48], were used for comparison with the proposed method. The comparisons of quantitative indicators of the five methods performed using the open-source dataset are shown in Figure 9, and the experimental statistical results using the test dataset are also given in Table 4. By contrast, U-Net, PSPNet, and DeepLabv3+ were relatively worse than the three other methods, and pix2pix and the proposed adversarial nets were significantly better than the three other networks in terms of *Precision*, *Recall*, *F1_score*, and *mIoU* values. Therefore, the improvement observed in the performance of the landslide extraction network by the generative network used in pix2pix or the proposed adversarial nets was better than that without generator. Overall, the proposed adversarial nets performed best in the open-source dataset and test data in terms of the indicators. Notably, the indicators (i.e., *Precision*, *Recall*, *F1_score*, and *mIoU* values) of landslide extraction from the Danba dataset were slightly lower than those obtained from the open-source dataset. In addition, through the visual comparison between Figures 8e and 10g obtained from the proposed network, it can be seen that Figure 8e is closer to the ground-truth landslides. The possible reasons for this are as follows: the proposed network was trained through satellite images rather than UAV images. Therefore, the generalization of the proposed network was affected by modeling training using the open-source dataset produced from satellite remote sensing images, which are different from UAV images in terms of imaging quality, resolution, and illumination. That is, the proposed network trained through satellite images may not be very applicable to UAV-photogrammetry-derived data.

The landslide boundaries obtained using the five methods are illustrated in Figure 10. In this study, the researchers conducted on-site investigations and delineated the ground-truth landslides through manual visual interpretation using ArcGIS software. The landslide areas extracted using U-Net and PSPNet were scattered, which was different from the ground-truth landslides. Although the landslide areas extracted using DeepLabv3+ and pix2pix could exhibit the basic shapes of landslides, the boundaries of landslides are not discriminative and cover many small non-landslide areas. The visual assessment shows that, although the experimental results obtained using the proposed adversarial nets were not perfect, the landslide regions extracted using the proposed network were highly consistent with the ground-truth regions compared with four other methods. Moreover, the proposed network, which is based on the integrity of spectral and topographic information, represents strong robustness and stability under complex backgrounds. This finding is attributed to two reasons. First, the generator with semi-supervised operation, similar to pix2pix, could effectively avoid overfitting and improve the generalization ability of the proposed adversarial nets. Second, multilevel skip connection architecture, which was used in the proposed network, is helpful in offering more possibilities to capture semantic features at all levels to clearly delineate the boundaries of landslides compared with DeepLabv3+. Therefore, in terms of the visual assessment and quantitative evaluation, the proposed network is useful and practicable in extracting landslides.

Table 4. *Precision*, *Recall*, *F1_score*, and *mIoU* values of U-Net, PSPNet, DeepLabv3+, pix2pix, and the proposed method performed using the test dataset.

Method	Precision	Recall	F1_Score	mIoU
U-Net [25]	0.6867	0.6515	0.6140	0.4767
PSPNet [47]	0.7362	0.6299	0.6271	0.4802
DeepLabv3+ [33]	0.7421	0.6830	0.6764	0.5800
pix2pix [30,48]	0.8052	0.8102	0.7740	0.6496
Proposed	0.8490	0.8204	0.8146	0.7002

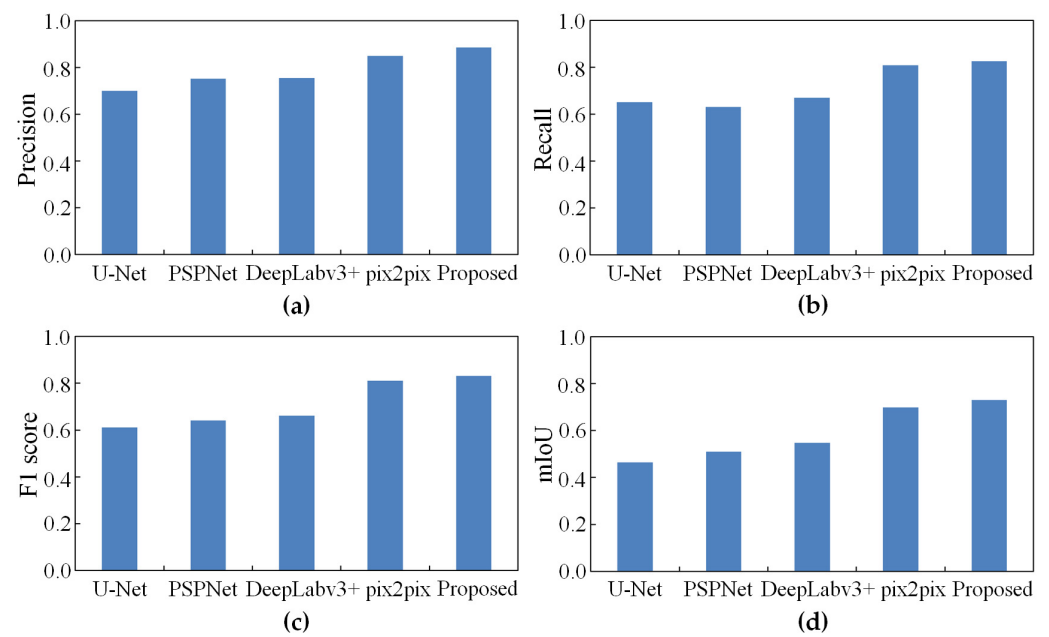


Figure 9. Comparisons of U-Net, PSPNet, DeepLabv3+, pix2pix, and the proposed method performed using the open-source dataset. (a–d) denote *Precision*, *Recall*, *F1_score*, and *mIoU* values of the five methods, respectively.

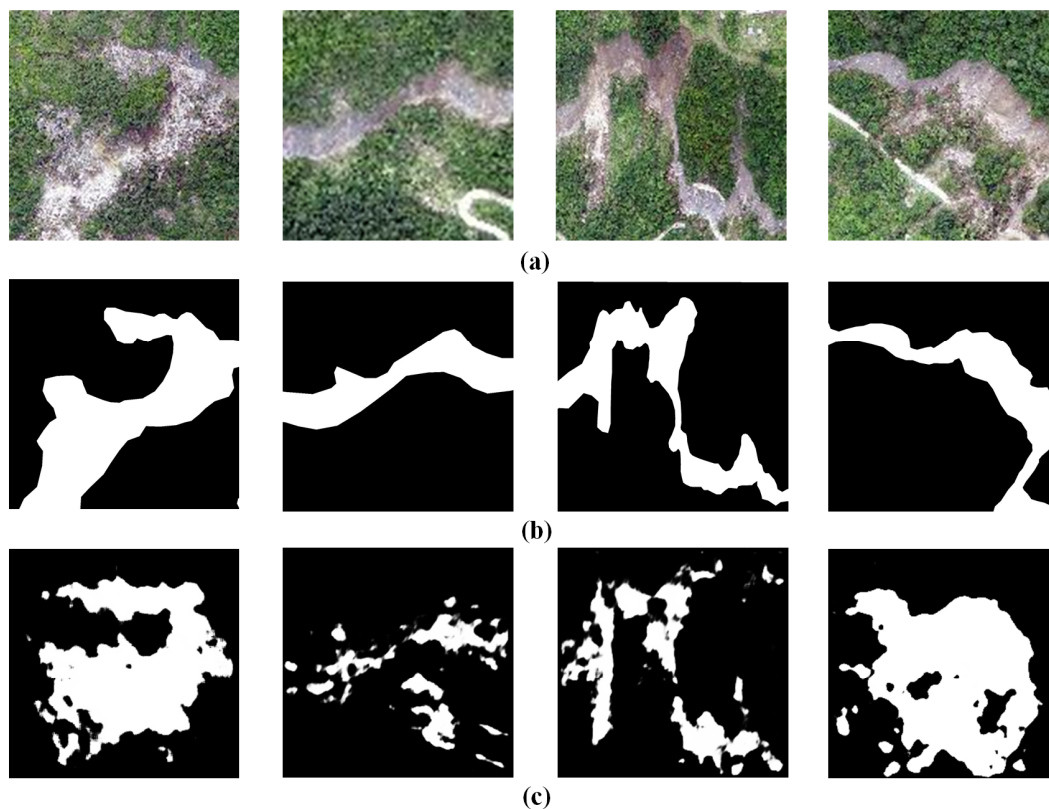


Figure 10. *Cont.*

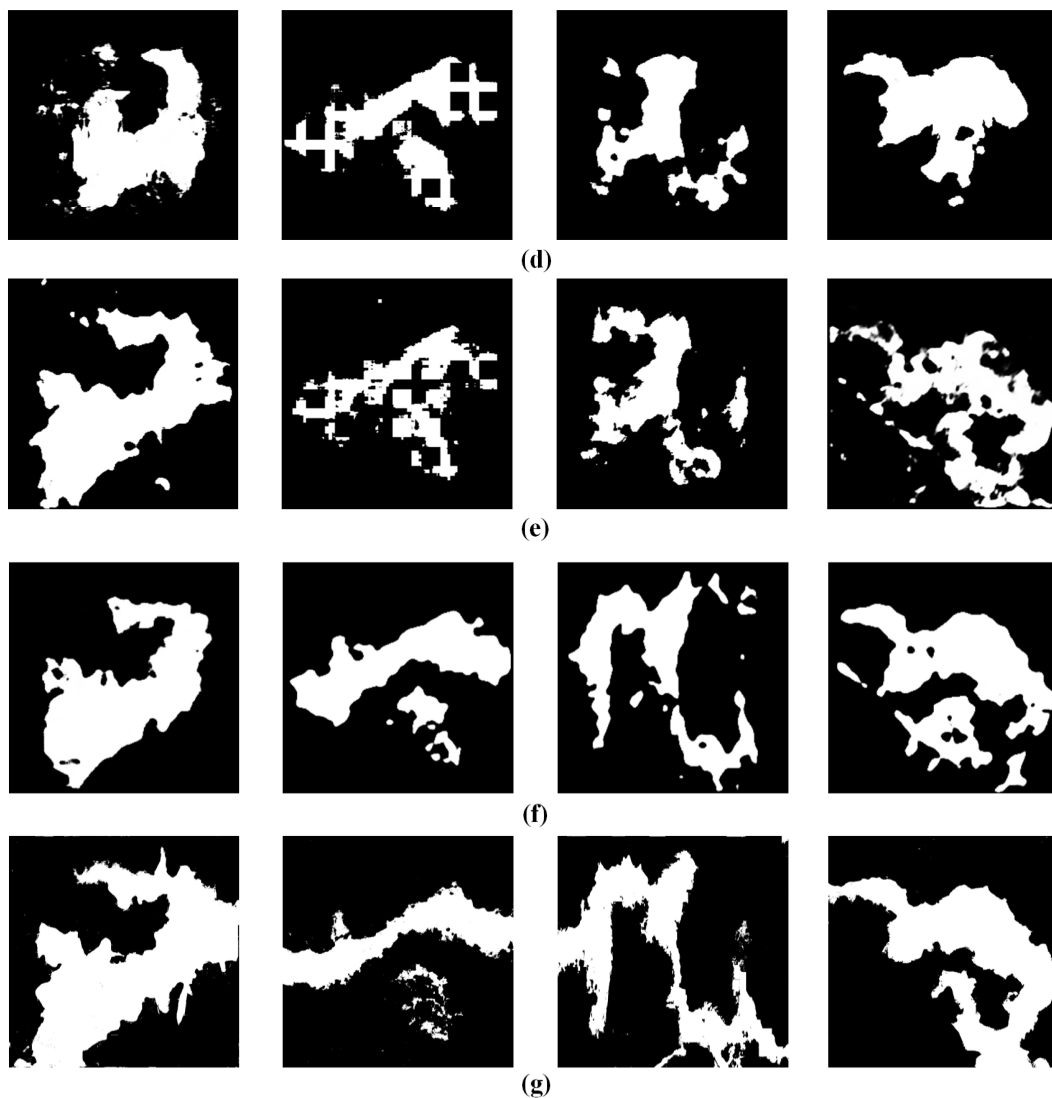


Figure 10. Examples of landslide extraction using test dataset. (a) is the RGB patches for visualization analysis, and (b) is the ground-truth landslides corresponding to (a). (c–g) are the experimental results obtained using U-Net, PSPNet, DeepLabv3+, pix2pix, and the proposed network, respectively.

4. Discussion

On the basis of the experimental results of landslide extraction, the proposed network can be considered a tentative and effective model for delineating landslide boundaries. In particular, the semi-supervised method means it is not imperative to avoid overfitting or convergence by a large number of labeled samples. The effectiveness of the proposed network can be explained by several reasons. First, VHR RGB DOM, which is illustrated in Figure 10a, derived from UAV LAP, can provide clear surface details, which offer the opportunity for the interpretation of a small landslide and its boundaries. Second, multisource fusion can enrich landslide presentation with additional information such as GGLI, DSM, slope, and aspect, which couples spectral and topographic information to improve the ability of landslide characterization, as verified in Figure 6 and Table 2. Third, the adversarial nets, especially the generative network, can contribute to reducing the requirement of the number of labeled samples by the semi-supervised operation. It can effectively alleviate the problem of network training difficulty caused by insufficient landslide samples. Four, multilevel skip connection architecture with an upsampling operation can retain the detailed boundaries at all levels and improve the performance of landslide extraction, which is at least 13% in terms of *Precision*, *Recall*, *F1_score*, and *mIoU* compared with DeepLabv3+.

Therefore, the proposed network combines the advantages of the following aspects: multi-source data fusion, the semi-supervised adversarial model, and multilevel skip connection architecture. These aspects make the network beneficial in capturing context information and performing semantic segmentation. Although the proposed network can obtain satisfactory landslide extraction results, it should be noted that landslides can only be extracted in areas where vegetation is destroyed. Some landslide bodies may not be completely broken, so there would still be good vegetation on the landslides, as shown in Figure 6, which makes them difficult to effectively extract because UAV photogrammetry cannot penetrate vegetation to obtain landslide information.

5. Conclusions

We presented an end-to-end, semi-supervised adversarial network based on multi-source data fusion for landslide extraction via semantic segmentation and landslide detection from UAV-photogrammetry-derived spectral and topographic data. First, multi-source data fusion, which was achieved by two branch networks, was proposed for semantic feature representation from spectral and topographic information. Second, a transfer learning network, that is, a pix2pix-based generator, was used to generate pseudo data for semi-supervised network training by reducing the requirement of the number of labeled samples. Third, a DeepLabV3+-based discriminator was improved to perform further fine-tuning by introducing ASPP and multilevel skip connection with upsampling operation to avoid CNN training from scratch. These two features contribute to characterizing the spatial relationship among ground objects and retaining the boundary information at all levels. End-to-end mapping between input and output could also be established through generative and discriminative networks. Finally, multidimensional low- and high-level semantic multiscale features, which were learned from the proposed adversarial nets, were integrated to perform automatic landslide extraction.

Although the generator used in the proposed adversarial nets could effectively reduce the requirement of the number of labeled samples, the quality and number of training samples still limited the improvement in the accuracy and reliability of the deep learning network for landslide extraction. In our experiments, the proposed network trained through satellite images was not very applicable to UAV-photogrammetry-derived data because nonlinear differences such as resolution, illumination, and ground object changes existed.

We currently lack sufficient UAV data related to landslides. Thus, UAV-derived data in this study were not used to train the proposed network. In future studies, we will attempt to collect more UAV data and improve the proposed adversarial nets by introducing UAV-photogrammetry-derived data for training and optimizing the proposed adversarial net parameters involved.

Author Contributions: Conceptualization, H.H. and C.L.; methodology, H.H. and R.Y.; software, R.Y. and H.Z.; validation, C.L., L.L. and Y.Z.; formal analysis, C.L.; investigation, H.H. and R.Y.; resources, H.Z.; data curation, L.L. and Y.Z.; writing—original draft preparation, H.H.; writing—review and editing, L.L.; visualization, H.H.; supervision, R.Y.; project administration, L.L.; funding acquisition, H.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the “National Natural Science Foundation of China, grant number 41861062 and 41401526”, “Fuzhou Youth Science and Technology Leading Talent Program, grant number 2020ED65”, “Science and Technology Project of Jiangxi Provincial Department of Water Resources, grant number 202123TGKT12”, and “Jiangxi 03 Special Project and 5G Project, grant number 20204ABC03A04”.

Data Availability Statement: Data are available upon request.

Acknowledgments: The authors thank Zhenjie Lai for providing the datasets.

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

References

1. Xu, D.; Peng, L.; Liu, S.; Wang, X. Influences of risk perception and sense of place on landslide disaster preparedness in southwestern China. *Int. J. Disaster Risk Sci.* **2018**, *9*, 167–180. [\[CrossRef\]](#)
2. Garnica-Peña, R.J.; Alcántara-Ayala, I. The use of UAVs for landslide disaster risk research and disaster risk management: A literature review. *J. Mt. Sci.* **2021**, *18*, 482–498. [\[CrossRef\]](#)
3. Froude, M.J.; Petley, D.N. Global fatal landslide occurrence from 2004 to 2016. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2161–2181. [\[CrossRef\]](#)
4. Bui, T.A.; Lee, P.J.; Lum, K.Y.; Loh, C.; Tan, K. Deep Learning for Landslide Recognition in Satellite Architecture. *IEEE Access* **2020**, *8*, 143665–143678. [\[CrossRef\]](#)
5. Ren, T.; Gong, W.; Gao, L.; Zhao, F.; Cheng, Z. An Interpretation Approach of Ascending–Descending SAR Data for Landslide Identification. *Remote Sens.* **2022**, *14*, 1299. [\[CrossRef\]](#)
6. Yang, H.; Omidalizari, M.; Xu, X.; Neumann, I. Terrestrial laser scanning technology for deformation monitoring and surface modeling of arch structures. *Compos. Struct.* **2017**, *169*, 173–179. [\[CrossRef\]](#)
7. Meng, Z.; Shu, C.; Yang, Y.; Wu, C.; Dong, X.; Wang, D.; Zhang, Y. Time Series Surface Deformation of Changbaishan Volcano Based on Sentinel-1B SAR Data and Its Geological Significance. *Remote Sens.* **2022**, *14*, 1213. [\[CrossRef\]](#)
8. Yan, Y.; Ashraf, M.A. The application of the intelligent algorithm in the prevention and early warning of mountain mass landslide disaster. *Arab. J. Geosci.* **2020**, *13*, 1–6. [\[CrossRef\]](#)
9. Casagli, N.; Frodella, W.; Morelli, S.; Tofani, V.; Ciampalini, A.; Intrieri, E.; Raspini, F.; Rossi, G.; Tanteri, L.; Lu, P. Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning. *Geoenviron. Disasters* **2017**, *4*, 1–23. [\[CrossRef\]](#)
10. Stumpf, A.; Kerle, N. Object-oriented mapping of landslides using Random Forests. *Remote Sens. Environ.* **2011**, *115*, 2564–2577. [\[CrossRef\]](#)
11. Ma, S.; Qiu, H.; Hu, S.; Yang, D.; Liu, Z. Characteristics and geomorphology change detection analysis of the Jiangdingya landslide on July 12, 2018, China. *Landslides* **2021**, *18*, 383–396. [\[CrossRef\]](#)
12. Ju, Y.; Xu, Q.; Jin, S.; Li, W.; Su, Y.; Dong, X.; Guo, Q. Loess Landslide Detection Using Object Detection Algorithms in Northwest China. *Remote Sens.* **2022**, *14*, 1182. [\[CrossRef\]](#)
13. Borghuis, A.M.; Chang, K.; Lee, H.Y. Comparison between automated and manual mapping of typhoon-triggered landslides from SPOT-5 imagery. *Int. J. Remote Sens.* **2007**, *28*, 1843–1856. [\[CrossRef\]](#)
14. Han, Y.; Wang, P.; Zheng, Y.; Yasir, M.; Xu, C.; Nazir, S.; Hossain, M.S.; Ullah, S.; Khan, S. Extraction of Landslide Information Based on Object-Oriented Approach and Cause Analysis in Shuicheng, China. *Remote Sens.* **2022**, *14*, 502. [\[CrossRef\]](#)
15. Lu, P.; Stumpf, A.; Kerle, N.; Casagli, N. Object-oriented change detection for landslide rapid mapping. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 701–705. [\[CrossRef\]](#)
16. Martha, T.R.; Kamala, P.; Jose, J.; Kumar, K.V.; Sankar, G.J. Identification of new landslides from high resolution satellite data covering a large area using object-based change detection methods. *J. Indian Soc. Remote Sens.* **2016**, *44*, 515–524. [\[CrossRef\]](#)
17. Li, Z.; Shi, W.; Lu, P.; Yan, L.; Wang, Q.; Miao, Z. Landslide mapping from aerial photographs using change detection-based Markov random field. *Remote Sens. Environ.* **2016**, *187*, 76–90. [\[CrossRef\]](#)
18. Lu, P.; Qin, Y.; Li, Z.; Mondini, A.; Casagli, N. Landslide mapping from multi-sensor data through improved change detection-based Markov random field. *Remote Sens. Environ.* **2019**, *231*, 111235. [\[CrossRef\]](#)
19. Mondini, A.C.; Guzzetti, F.; Reichenbach, P.; Rossi, M.; Cardinali, M.; Ardizzone, F. Semi-automatic recognition and mapping of rainfall induced shallow landslides using optical satellite images. *Remote Sens. Environ.* **2011**, *115*, 1743–1757. [\[CrossRef\]](#)
20. Fanos, A.M.; Pradhan, B.; Mansor, S.; Yusoff, Z.M.; Abdullah, A.F. A hybrid model using machine learning methods and GIS for potential rockfall source identification from airborne laser scanning data. *Landslides* **2018**, *15*, 1833–1850. [\[CrossRef\]](#)
21. Ji, S.; Yu, D.; Shen, C.; Li, W.; Xu, Q. Landslide detection from an open satellite imagery and digital elevation model dataset using attention boosted convolutional neural networks. *Landslides* **2020**, *17*, 1337–1352. [\[CrossRef\]](#)
22. Shi, W.; Zhang, M.; Ke, H.; Fang, X.; Zhan, Z.; Chen, S. Landslide recognition by deep convolutional neural network and change detection. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 4654–4672. [\[CrossRef\]](#)
23. 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. [\[CrossRef\]](#)
24. Qi, W.; Wei, M.; Yang, W.; Xu, C.; Ma, C. Automatic Mapping of Landslides by the ResU-Net. *Remote Sens.* **2020**, *12*, 2487. [\[CrossRef\]](#)
25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Springer: Cham, Switzerland, 2015; pp. 234–241.
26. Tanatipuknon, A.; Aimmanee, P.; Watanabe, Y.; Murata, K.T.; Wakai, A.; Sato, G.; Hung, H.V.; Tungpimolrut, K.; Keerativittayanun, S.; Karnjana, J. Study on Combining Two Faster R-CNN Models for Landslide Detection with a Classification Decision Tree to Improve the Detection Performance. *J. Disaster Res.* **2021**, *16*, 588–595. [\[CrossRef\]](#)
27. Yu, B.; Chen, F.; Xu, C. Landslide detection based on contour-based deep learning framework in case of national scale of Nepal in 2015. *Comput. Geosci.* **2020**, *135*, 104388. [\[CrossRef\]](#)

28. 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. [\[CrossRef\]](#)
29. Xia, W.; Chen, J.; Liu, J.; Ma, C.; Liu, W. Landslide Extraction from High-Resolution Remote Sensing Imagery Using Fully Convolutional Spectral–Topographic Fusion Network. *Remote Sens.* **2021**, *13*, 5116. [\[CrossRef\]](#)
30. Lebedev, M.A.; Vizilter, Y.V.; Vygodov, O.V.; Knyaz, V.A.; Rubis, A.Y. Change Detection in Remote Sensing Images Using Conditional Adversarial Networks. *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *42*, 565–571. [\[CrossRef\]](#)
31. Zhao, W.; Mou, L.; Chen, J.; Bo, Y.; Emery, W. Incorporating metric learning and adversarial network for seasonal invariant change detection. *IEEE Trans. Geosci. Remote Sens.* **2019**, *58*, 2720–2731. [\[CrossRef\]](#)
32. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
33. 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 European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
34. Agarwal, S.; Snavely, N.; Seitz, S.M.; Szeliski, R. Bundle adjustment in the large. In Proceedings of the European Conference on Computer Vision, Crete, Greece, 5–11 September 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 29–42.
35. He, H.; Zhou, J.; Chen, M.; Chen, T.; Li, D.; Cheng, P. Building Extraction from UAV Images Jointly Using 6D-SLIC and Multiscale Siamese Convolutional Networks. *Remote Sens.* **2019**, *11*, 1040. [\[CrossRef\]](#)
36. He, N.; Fang, L.; Li, S.; Plaza, A.; Plaza, J. Remote sensing scene classification using multilayer stacked covariance pooling. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6899–6910. [\[CrossRef\]](#)
37. Liu, D.; Han, G.; Liu, P.; Yang, H.; Sun, X.; Li, Q.; Wu, J. A Novel 2D-3D CNN with Spectral-Spatial Multi-Scale Feature Fusion for Hyperspectral Image Classification. *Remote Sens.* **2021**, *13*, 4621. [\[CrossRef\]](#)
38. Sameen, M.I.; Pradhan, B. Landslide detection using residual networks and the fusion of spectral and topographic information. *IEEE Access* **2019**, *7*, 114363–114373. [\[CrossRef\]](#)
39. Yu, B.; Yang, L.; Chen, F. Semantic segmentation for high spatial resolution remote sensing images based on convolution neural network and pyramid pooling module. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3252–3261. [\[CrossRef\]](#)
40. Bergado, J.R.; Persello, C.; Stein, A. Recurrent multiresolution convolutional networks for VHR image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6361–6374. [\[CrossRef\]](#)
41. Chen, G.; Zhang, X.; Wang, Q.; Dai, F.; Gong, Y.; Zhu, K. Symmetrical dense-shortcut deep fully convolutional networks for semantic segmentation of very-high-resolution remote sensing images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1633–1644. [\[CrossRef\]](#)
42. Pan, X.; Gao, L.; Marinoni, A.; Zhang, B.; Yang, F.; Gamba, P. Semantic Labeling of High Resolution Aerial Imagery and LiDAR Data with Fine Segmentation Network. *Remote Sens.* **2018**, *10*, 743. [\[CrossRef\]](#)
43. Marmanis, D.; Datcu, M.; Esch, T.; Stilla, U. Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geosci. Remote Sens. Lett.* **2015**, *13*, 105–109. [\[CrossRef\]](#)
44. Marmanis, D.; Schindler, K.; Wegner, J.D.; Galliani, S.; Datcu, M.; Stilla, U. Classification with an edge: Improving semantic image segmentation with boundary detection. *ISPRS J. Photogramm. Remote Sens.* **2018**, *135*, 158–172. [\[CrossRef\]](#)
45. He, H.; Yan, Y.; Chen, T.; Cheng, P. Tree Height Estimation of Forest Plantation in Mountainous Terrain from Bare-Earth Points Using a DoG-Coupled Radial Basis Function Neural Network. *Remote Sens.* **2019**, *11*, 1271. [\[CrossRef\]](#)
46. He, H.; Chen, M.; Chen, T.; Li, D. Matching of Remote Sensing Images with Complex Background Variations via Siamese Convolutional Neural Network. *Remote Sens.* **2018**, *10*, 355. [\[CrossRef\]](#)
47. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890.
48. Souly, N.; Spampinato, C.; Shah, M. Semi supervised semantic segmentation using generative adversarial network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5688–5696.