



Technical Note

A Novel Weakly Supervised Remote Sensing Landslide Semantic Segmentation Method: Combining CAM and cycleGAN Algorithms

Yongxiu Zhou ¹, Honghui Wang ^{1,2,3,*} , Ronghao Yang ⁴ , Guangle Yao ^{2,3}, Qiang Xu ³ and Xiaojuan Zhang ⁴

¹ Key Laboratory of Earth Exploration and Information Technology of Ministry of Education, Chengdu University of Technology, Chengdu 610059, China; zhouyongxiu@stu.cdut.edu.cn

² College of Computer Science and Cyber Security, Chengdu University of Technology, Chengdu 610059, China; yaoguangle19@cdut.edu.cn

³ State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, Chengdu 610059, China; xq@cdut.edu.cn

⁴ College of Earth Sciences, Chengdu University of Technology, Chengdu 610059, China; yangronghao@cdut.edu.cn (R.Y.); zxj6120@126.com (X.Z.)

* Correspondence: wanghh@cdut.edu.cn

Abstract: With the development of deep learning algorithms, more and more deep learning algorithms are being applied to remote sensing image classification, detection, and semantic segmentation. The landslide semantic segmentation of a remote sensing image based on deep learning mainly uses supervised learning, the accuracy of which depends on a large number of training data and high-quality data annotation. At this stage, high-quality data annotation often requires the investment of significant human effort. Therefore, the high cost of remote sensing landslide image data annotation greatly restricts the development of a landslide semantic segmentation algorithm. Aiming to resolve the problem of the high labeling cost of landslide semantic segmentation with a supervised learning method, we proposed a remote sensing landslide semantic segmentation with weakly supervised learning method combining class activation maps (CAMs) and cycle generative adversarial network (cycleGAN). In this method, we used the image level annotation data to replace pixel level annotation data as the training data. Firstly, the CAM method was used to determine the approximate position of the landslide area. Then, the cycleGAN method was used to generate the fake image without a landslide, and to make the difference with the real image to obtain the accurate segmentation of the landslide area. Finally, the pixel-level segmentation of the landslide area on remote sensing image was realized. We used mean intersection-over-union (mIOU) to evaluate the proposed method, and compared it with the method based on CAM, whose mIOU was 0.157, and we obtain better result with mIOU 0.237 on the same test dataset. Furthermore, we made a comparative experiment using the supervised learning method of a u-net network, and the mIOU result was 0.408. The experimental results show that it is feasible to realize landslide semantic segmentation in a remote sensing image by using weakly supervised learning. This method can greatly reduce the workload of data annotation.

Keywords: landslide semantic segmentation; remote sensing; weakly supervised learning; CAM; cycleGAN



Citation: Zhou, Y.; Wang, H.; Yang, R.; Yao, G.; Xu, Q.; Zhang, X. A Novel Weakly Supervised Remote Sensing Landslide Semantic Segmentation Method: Combining CAM and cycleGAN Algorithms. *Remote Sens.* **2022**, *14*, 3650. <https://doi.org/10.3390/rs14153650>

Academic Editors: Sansar Raj Meena, Filippo Catani, Mario Floris and Yunus P. Ali

Received: 27 June 2022

Accepted: 26 July 2022

Published: 29 July 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 one of the most widespread natural hazards [1]. The early spotting of landslide disasters in remote areas through remote sensing images can effectively reduce the occurrence of secondary disasters. Due to the problems of slow speed and the expertise required for the manual interpretation of remote sensing images, researchers have started to apply deep learning techniques to the automatic interpretation of remote sensing images [2].

Since the release of the ImageNet dataset [3] in 2009, deep learning has been rapidly developed in the field of computer vision. With the development of deep learning techniques in the fields of image classification, detection, and segmentation, the detection or segmentation of landslide areas in remote sensing images has also started to explore the use of deep learning methods to solve this problem. However, unlike other remote sensing image recognition tasks, there have been relatively few studies using deep learning to detect or segment landslide areas at this stage [4]. Sameen et al. [5] used deep residual detection based on a feature fusion network to detect the remote sensing images of landslides in the Kinmallan plateau, Malaysia, and improved the F1 score by 0.13 and mIOU by 0.1296 compared with the method of convolutional layer stacking. Chen et al. [6] proposed a change detection method based on deep convolutional network and obtained a false recognition rate of 0.176. Cheng et al. [7] proposed a YOLO-SA model using Qiaojia and Ludian counties in Yunnan Province, China, as the study area, and improved the recognition accuracy to 0.9408.

The remote sensing landslide detection algorithm can output the coordinate position of the landslide, while the remote sensing landslide segmentation algorithm can accurately segment the boundary of the landslide. Therefore, the remote sensing landslide segmentation algorithm can be used to study the landslide area change and landslide area calculation. Additionally, there have been even fewer studies related to remote sensing image segmentation tasks due to the lack of datasets. Soares et al. [8] used DEM information as training data and used the U-Net model to automatically segment landslides in the city of Novo Fribourg, located in the mountains of Rio de Janeiro, southeastern Brazil, and obtained F1 scores of 0.55 and 0.58 on two different test sets. Du et al. [9] compared six commonly used deep learning semantic segmentation models based on a self-built Yangtze River coastal landslide dataset, and finally obtained 0.542 and 0.740 mIOU accuracy on the GCN and deeplabv3 models, respectively. Prakash et al. [10] used an improved u-net network to experiment on the state-wide landslide dataset in Oregon. This method finally achieved a detection rate of 0.72, which is better than the traditional method. Bo et al. [11] used the deep learning semantic segmentation model to accurately detect the landslide area of remote sensing images in Nepal in 2015, and achieved a recall rate of 0.65 and an accuracy rate of 0.55.

The remote sensing landslide segmentation algorithm is identical to the ordinary semantic segmentation task. At present, the mainstream method is supervised learning, which requires a large number of pixel-level labeled training data. Pixel-level annotated datasets often require significant labor costs. For example, in the Bijie landslide dataset [12], the annotators used irregular polygons to annotate the landslide area in each remote sensing image. Annotators need to mark an average of 29 coordinate points on each image. If more accurate annotation results are required, the annotator needs to annotate more coordinate points and more accurate coordinate point locations. The datasets of landslide remote sensing images are too expensive to produce, which is a great challenge for performing research with remote sensing landslide images. The weakly supervised semantic segmentation algorithm based on image-level annotation can realize the study of landslide region segmentation on image-level annotated datasets, which provides a new idea for the study of a landslide region segmentation algorithm in remote sensing images.

As a commonly used weakly supervised learning algorithm, CAM can be used for extracting the geographic objects supervised by image-level labels [13]. Feng et al. [14] used self-matching CAM to provide a novel and accurate explanation of CNN for SAR image interpretation. While cycleGAN is usually used for virtual data generation, Park et al. [15] used cycleGAN to generate wildfire images as the training data for Densenet, and demonstrated a high wildfire detection accuracy. However, there has been no research to date that has used CAM and cycleGAN in the field of remote sensing landslides via image segmentation.

To this end, we propose a weakly supervised approach to landslide image segmentation training using only image-level labeled data as training data. First, we use the

CAM method to locate the approximate location of the landslide area. Then, we use the cycleGAN method to generate the image in which there is no landslide and make the difference with the current image to realize the fine segmentation of the landslide area. Finally, the pixel-level segmentation of the landslide area on the remote sensing image is performed on the dataset based on image-level annotation.

Our main contributions are the following:

- (1) Proposing a cycleGAN-based method for the weakly supervised training of image-level labeled remote sensing landslide images, and achieving the fine segmentation of landslide regions;
- (2) Combining the CAM and cycleGAN methods to improve the segmentation accuracy of weakly supervised learning algorithms on remote sensing landslide images.

2. Dataset

2.1. Data Sources

The remote sensing landslide image dataset we used was from the Bijie landslide dataset [12] (Figure 1). This dataset uses remote sensing images of Bijie City of Guizhou Province, in China. The dataset was acquired by the TripleSat remote sensing satellite between May and August 2018, and the spatial resolution of the images is 1 meter. This dataset contains 770 remote sensing landslide images and 2003 remote sensing non-landslide images.

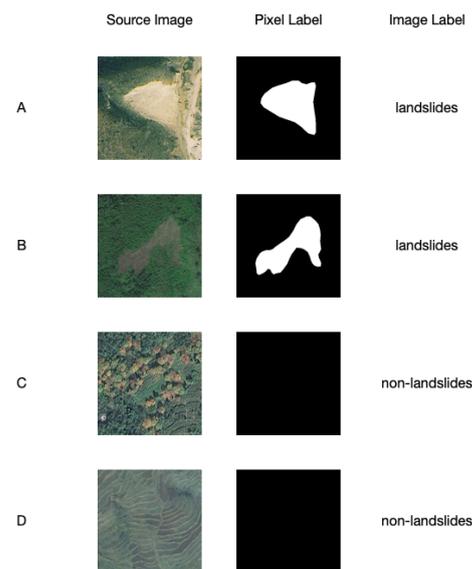


Figure 1. Some images and labels of the Bijie dataset. (A,B) are samples with landslides, (C,D) are samples without landslides.

The Bijie dataset consists of slices of landslide images and non-landslide images, but the town buildings are only included in the non-landslide images. Since town building images in non-landslide images would affect the training of cycleGAN, we manually removed the non-landslide images which contain town buildings, and finally, 958 non-landslide remote sensing images were left as negative samples and 770 remote sensing landslide images were left as positive samples of the dataset. We randomly chose 78 images (approximately 10% of the total) among the positive samples and 97 images among the negative samples as the test set.

2.2. Pixel-Level Annotation and Image-Level Annotation

Pixel-level annotation is to determine whether each pixel in an image belongs a certain category, with the aim of annotating specific category regions in the image. Since the class to be labeled usually has irregular shapes, the labeling is labor-intensive, such as

the COCO dataset [16]. Pixel-level labeled datasets are usually used for the training of semantic segmentation networks. The pixel-level labels that we used are shown in the pixel label column in Figure 1.

Image-level annotation only needs to annotate whether the image contains the category, and usually one image corresponds to one or more labels, so the annotation is relatively simple. Image-level annotated datasets are often used to train classification models, such as the ImageNet dataset [3]. The weakly supervised learning approach we used trains a semantic segmentation task with image-level labeled datasets. The image-level label we used is shown in the image label column in Figure 1.

3. Method

3.1. CAM-Based Weakly Supervised Algorithm

The class activation mapping (CAM) approach was first proposed by Zhou et al. [17] for the study of weakly supervised target localization and neural network visualization.

As shown in Figure 2, for a trained landslide classification network, the final classification result depends on the weight corresponding to the category class of the landslide multiplied by the final feature value, which in turn, is obtained by averaging the feature layers. Therefore, the weights of the feature layers directly affect the final classification results.

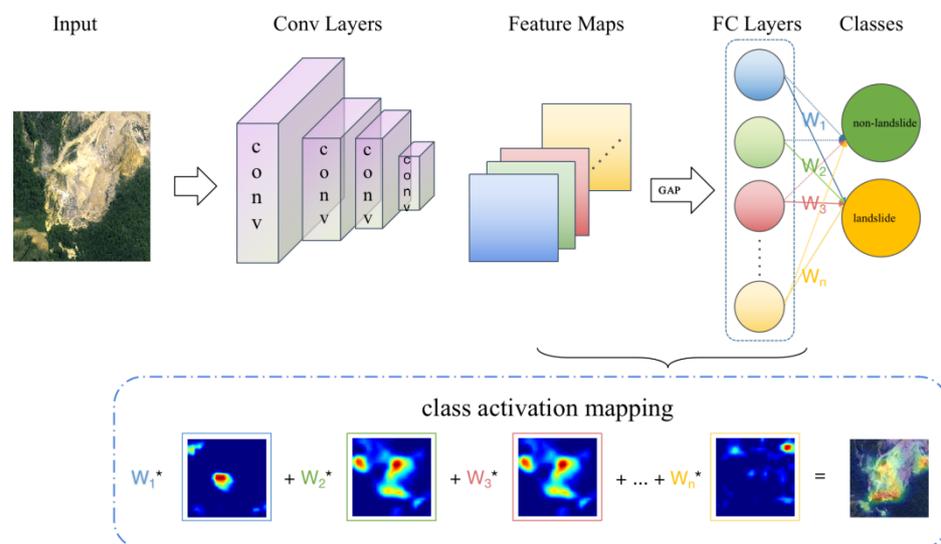


Figure 2. CAM schematic.

We used the weights of the landslide category to linearly weight the feature layer to obtain the class activation map for landslide classification, upsampled the class activation map to the size of the original image, and then superimposed it with the original image so that we could observe which regions of the original image are used by the network to make judgments about the category to which the image belongs.

Because the CAM method can draw the heat map of the target class of interest on the dataset based on image-level annotation, we can realize the rough segmentation of the remote sensing landslide image based on the heat map obtained by CAM.

CAM can be expressed by Equation (1).

$$M_c(x, y) = \sum_k w_k^c f_k(x, y) \quad (1)$$

where $M_c(x, y)$ denotes the final obtained class activation map of category c , k denotes the number of channels of the feature map after the last convolution layer, w_k^c denotes the weight of k channels of the fully connected layer corresponding to category c , and $f_k(x, y)$ denotes the feature map after the last convolution layer.

Since the average image size of the Bijie dataset is 282×272 , we resized the images to 256×256 as model inputs. When using the CAM approach, we first trained the training dataset for classification using the ResNet 50 model [18], and then calculated the class activation mapping maps for the test images based on the trained classification network. Since we trained the classification model with only two classes, namely landslide and non-landslide, we substituted the class corresponding to the landslide class directly into Equation (1) to obtain the class activation map corresponding to the landslide image.

3.2. cycleGAN-Based Weakly Supervised Algorithm

3.2.1. Generate Images before Landslides with cycleGAN

GAN networks [19] were first used to randomly generate images of a specific style, and cycleGAN networks [20] can achieve the conversion of styles between different domains while keeping the subject content of the original image unchanged.

Inspired by the function of style migration implemented by the cycleGAN network, we consider the images with landslide as one style of domain and the images without landslide as another style of domain. Additionally, we can train a generation network that can convert the landslide images to non-landslide images, and finally realize the generation of virtual non-landslide images from the landslide images.

Compared to the pix 2 pix adversarial generative network algorithm [21], the training data used by cycleGAN does not require the paired images of two domains, which provides a great convenience in obtaining training data. As we used the landslide images and non-landslide images, we do not need to obtain pairs of remote sensing images before and after landslides, but only one part of the dataset is landslide images and the other part is non-landslide images.

The goal of cycleGAN is to obtain, by training the data of two domains X and Y , the generator G , which can realize the style migration from domain X to domain Y ; the generator F , which can realize the style migration from domain Y to domain X ; the discriminator D_x , which can identify whether the image is a real image in X or a fake image generated by the image in Y through F ; and the discriminator D_y , which can identify whether the image is a real image in Y or a fake image generated by the image in X through G . The flow chart of cycleGAN training is shown in Figure 3.

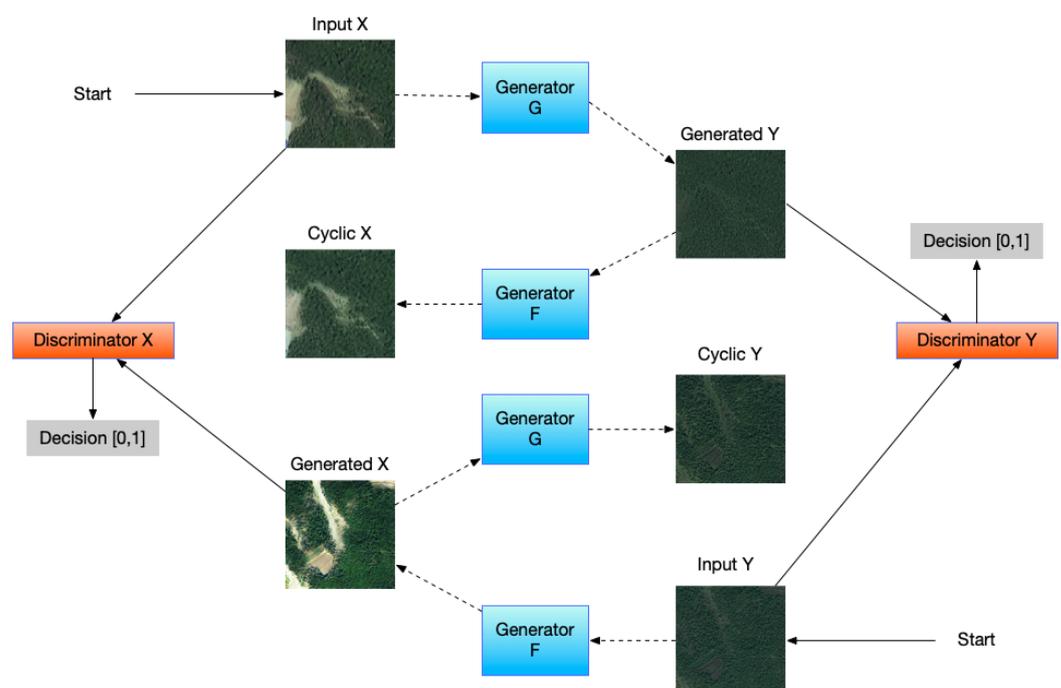


Figure 3. Flow chart of cycleGAN training.

The loss function of the cycleGAN network mainly consists of two parts: one part is the adversarial loss, which is similar to the loss function of the classical GAN network and is designed to make the style of the fake samples generated by generators G and F close to the style of the real samples; and the other part is the cyclic consistency loss, which is used to constrain that the fake sample $G(x)$ (obtained by real sample x through generator G) can be recovered into x again through generator F , i.e., $F(G(x)) = x$. Similarly, the real sample y can be recovered into y by $G(F(y))$. Therefore, the total loss function of the cycleGAN network can be expressed as Equations (2)–(4).

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{cyc}(G, F) \quad (2)$$

Additionally:

$$L_{GAN}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} |\log D_Y(y)| + \mathbb{E}_{x \sim p_{data}(x)} |\log (1 - D_Y(G(x)))| \quad (3)$$

$$L_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \quad (4)$$

When using cycleGAN to train remote sensing images, we consider the landslide image as domain X and the non-landslide image as domain Y . Through the loss function, we obtain the generator G that can convert the landslide images into the non-landslide image and the generator F that can convert the non-landslide image into the landslide image.

3.2.2. Difference Method to Obtain Landslide Area

The difference method can analyze the degree of difference between two images by finding the difference value between the two images pixel by pixel. By setting a proper threshold value, the part of the two images with large differences can be further analyzed.

From Figure 4, we can see that the difference between the real landslide image and the generated non-landslide image mainly exists in the landslide area. Therefore, we can obtain the specific area where the landslide occurs in the landslide image by a simple difference method.

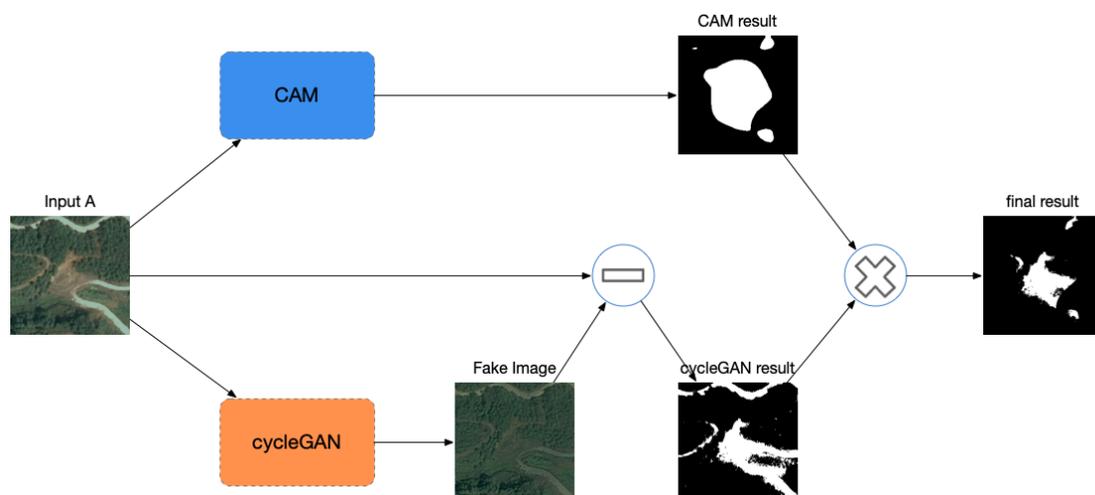


Figure 4. Flowchart combining CAM and cycleGAN methods.

Before segmenting the landslide image using the difference method, the color image needs to be converted to grayscale, as shown in Equation (5):

$$Gray = R \times 0.3 + G \times 0.59 + B \times 0.11 \quad (5)$$

Then, the pixel-by-pixel difference between the landslide image and the non-landslide image is obtained to obtain the difference image of the two images, as shown in Equation (6).

$$D(x, y) = |f_{landslide}(x, y) - f_{non-landslide}(x, y)| \quad (6)$$

Finally, a proper threshold T is chosen to perform a pixel-by-pixel analysis of the differential image to find the landslide area in the remote sensing image, as shown in Equation (7).

$$R(x, y) = \begin{cases} 1, & D(x, y) > T \\ 0, & \text{others} \end{cases} \quad (7)$$

3.3. Method of Combining cycleGAN and CAM

By comparing the segmentation results of the CAM-based method and the cycleGAN-based method, we can find that the CAM-based method can determine the approximate location of the landslide region in the image, but it is not very accurate in segmenting the edges of the landslide; however, the cycleGAN-based method is better at handling the details of the image, and thus more accurately segments the boundary of the landslide region in the image, although it can easily mis-segment the non-landslide region (such as the road in the images).

Additionally, we sought to verify whether the results of both methods contained the correct segmentation results. Therefore, we combined the CAM and cycleGAN methods and made an intersection of the results of both so as to improve the accuracy of the segmentation results.

The flow of the remote sensing landslide image segmentation using the combined CAM-based and cycleGAN-based methods is shown in Figure 4. First, we used the original image to obtain the result using the CAM-based method in Section 3.1; then, we used the original image to obtain the result using the cycleGAN-based method in Section 3.2; finally, we intersected the CAM result and the cycleGAN result pixel by pixel to obtain the final result.

4. Results and Discussion

4.1. Model Evaluation Method

In landslide remote sensing image semantic segmentation tasks, test results are usually evaluated using metrics such as precision, recall, and mIOU [8]. Precision, recall, and mIOU are mainly obtained by calculating TP, FP, TN, and FN [22]. The true positives (TPs) denote positive samples correctly predicted by the model; the false positives (FPs) denote positive samples which were incorrectly predicted by the model; true negatives (TNs) denote negative samples correctly predicted by the model; and false negatives (FNs) denote negative samples which were incorrectly predicted by the model.

Precision: indicates how many of the positive samples predicted by the model are correct, i.e., the percentage of the predicted landslide area that was correctly predicted. Precision is calculated using Equation (8).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

Recall: indicates how many of the actual positive samples were correctly predicted, i.e., the percentage of the area where the landslide actually occurred that the model was able to correctly predict. Recall is calculated using Equation (9).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

mIOU: it is usually used to comprehensively evaluate the performance of the segmentation model. IOU is calculated by dividing the intersection set by the union set of ground truth and prediction results, and mIOU requires averaging the calculated results for each category [23]. mIOU can be expressed by Equation (10).

$$\text{MIoU} = \frac{A \cap B}{A \cup B} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (10)$$

False positive rate: in the process of the landslide segmentation of remote sensing images using the landslide region segmentation algorithm, the number of negative samples (non-landslide images) is usually much higher than the number of positive samples (landslide images), and if the false positive rate of the algorithm is excessively high, a large number of false positive events will be generated in the practical application, resulting in a decrease in the practicality of the algorithm. Therefore, we add the non-landslide image dataset as a test data to calculate the false positive rate of the algorithm. The calculation of the false positive rate can be expressed as Equation (11).

$$\text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}} \quad (11)$$

4.2. Evaluation Results

We evaluated the precision, recall, and mIOU with the CAM-based method, cycleGAN-based method, and the combined CAM and cycleGAN methods. The test results are shown in Table 1.

Table 1. Comparison of test results between weakly supervised learning and supervised learning.

| Method | | Precision | Recall | mIOU | FPR |
|----------------------------|----------------|-----------|--------|-------|-------|
| weakly supervised learning | CAM | 0.692 | 0.593 | 0.159 | 0.054 |
| | cycleGAN | 0.845 | 0.404 | 0.184 | 0.042 |
| | CAM + cycleGAN | 0.924 | 0.383 | 0.237 | 0.004 |
| supervised learning | U-Net | 0.955 | 0.555 | 0.408 | 0.011 |

4.2.1. CAM Results Analysis

For the CAM-based method, from the overall test results, the precision was 0.692, the recall was 0.593, and the false positive rate was 0.054. Analyzing the test results of a single sample, the weakly supervised learning of the CAM-based method can roughly locate the location of the landslide area (as shown in Figure 5).

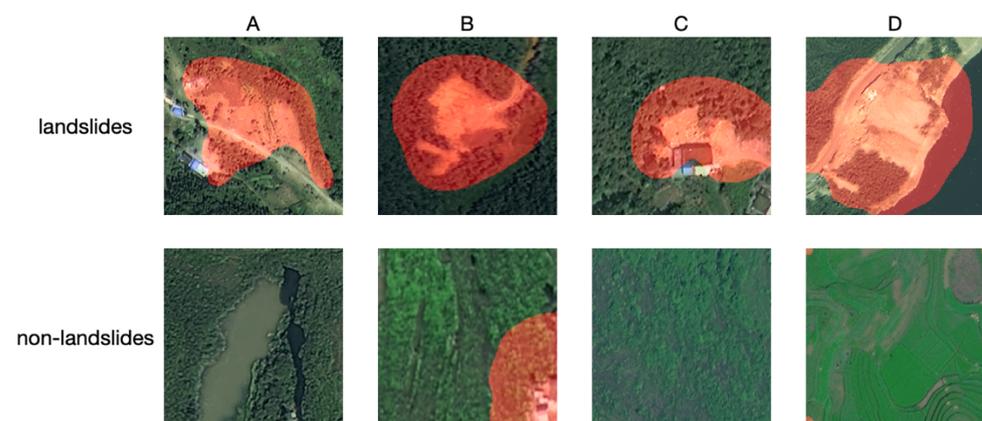


Figure 5. Segmentation results of CAM method. We have used 4 pairs of samples for test. (A,C,D) show that the results are all right, and (B) shows that the algorithm incorrectly recognized the house as landslide.

We can also see from the prediction results that the CAM-based method will predict as many positive samples as possible, and the actual landslide area is almost completely included in the predicted area, so a higher recall (0.593) can be obtained using the CAM-based method. However, the method can hardly determine the boundary of the landslide area, so it cannot be used to directly segment the contour of the landslide area.

On the other hand, because the CAM-based method is too sensitive to abnormal regions in the image, there are more false detections, for example, column B of the non-

landslides in Figure 5, the algorithm incorrectly recognized the house as landslide, and the correct rate is also the lowest among all methods (0.692). Comparing several other methods, the CAM-based method has the highest false positive rate of 0.054, indicating that it is more likely to produce false detections when predicting images without landslides compared to several other methods.

4.2.2. cycleGAN Results Analysis

We choose the optimal threshold $T = 32$ with Equation (7). The test results using the cycleGAN-based method are shown in the second column of Table 1, with a correct rate of 0.845, a recall rate of 0.404, and a false positive rate of 0.042, all of which are intermediate values for the three weakly supervised algorithms.

The segmentation results of the cycleGAN-based method are shown in Figure 6. As can be seen from the results generated by the model in the second column, comparing the original landslide remote sensing images (e.g., A and B in Figure 6), the fake images generated by cycleGAN not only change the color, but can also generate the texture of the landslide area similar to the non-landslide area.

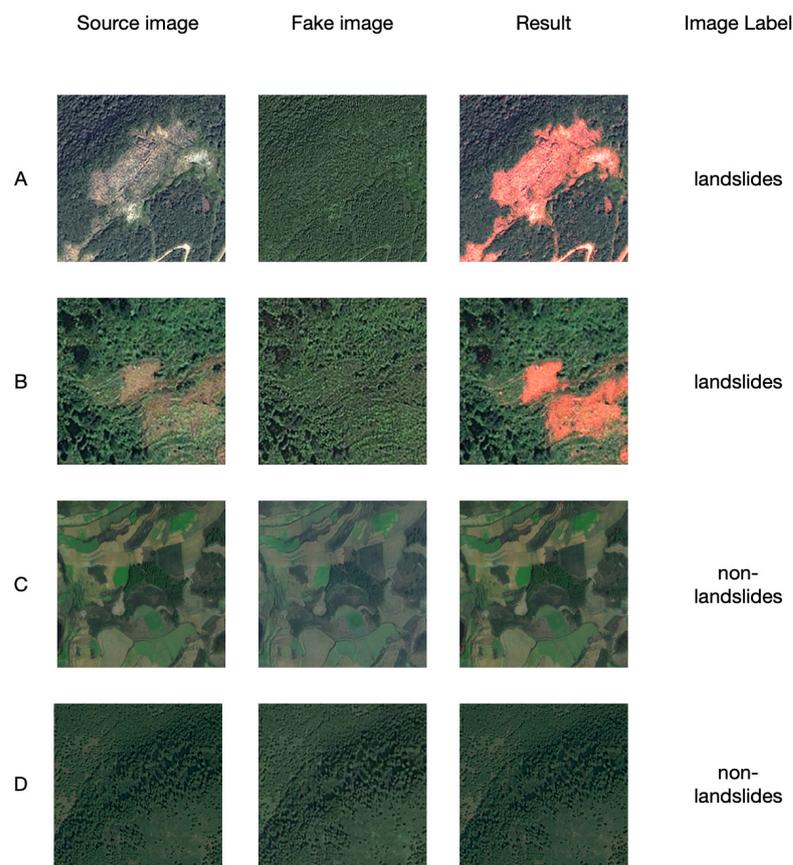


Figure 6. Segmentation results of cycleGAN method. We have used 4 samples for test. (A,B) are the landslides samples, (C,D) are the non-landslides samples. The results show that the 4 samples are all segmented correctly.

For the non-landslide remote sensing images (such as C and D in Figure 6), the fake images generated by cycleGAN can almost be consistent with the original images, especially when comparing the original and generated images of sample C in Figure 6, we can find that, although there is a yellow non-landslide area similar to the landslide in the original image, the image generated by cycleGAN do not process this area, which indicates that the model does not only judge whether an area is a landslide area from the color, but also has the ability to distinguish the real landslide area.

After obtaining the fake image generated by cycleGAN, we will use the difference method to judge each pixel of the image, so compared with the CAM-based method, this method can obtain a more accurate segmentation boundary. On the other hand, due to reasons such as the lesser number of learning samples, the model will cause the false detection of regions such as the road in the samples (sample A in Figure 6), which is the most important reason for the existence of false detection in the cycleGAN-based method.

4.2.3. Analysis of Combined Method Results

As can be seen in the third column of Table 1, the precision of the combined CAM-based and cycleGAN-based methods is 0.924, the recall is 0.383, and the false positive rate is 0.004.

By comparing the shortcomings of the CAM-based and cycleGAN-based methods, we found that the segmentation results of both methods contain the correct detection results, and the reasons for the existence of misdetection are not the same. The misdetection of the CAM-based method is due to the inability of the model to correctly partition the boundary of the landslide area. However, the misdetection of the cycleGAN-based method was due to the fact that the model misidentifies areas such as roads as landslide areas. Therefore, we combined the above two methods and took the intersection of the two results to make the prediction results for positive samples more reliable, and finally, obtained the highest accuracy (0.924) among the three weakly supervised algorithms.

Because the combined method result is obtained by taking the intersection of the results from CAM and cycleGAN, it is the most stringent compared to the single algorithm. This method sacrifices recall while improving precision, resulting in the lowest recall (0.383) among the three methods.

In addition, the combined method has the lowest false positive rate of 0.004 among all methods, which is even lower than the fully supervised method, indicating that this method has the lowest false alarm rate in the prediction of non-landslide images.

In order to evaluate the performance of the three methods together, we tested the test set using mIOU. The test results show that the combined method could obtain an mIOU of 0.237, which is better than the mIOU of 0.159 for the CAM-based method and 0.184 for cycleGAN-based method.

4.2.4. Comparison of Weakly Supervised Method and Supervised Method

Today, there are a few open source landslide datasets. As the first open remote sensing landslide dataset [12], the Bijie dataset is the only landslide dataset that we have access to. In order to compare the weakly supervised learning approach with the supervised learning approach, we trained the Bijie data with supervised learning using the U-Net network [24], and the segmentation results using the U-Net network are shown in Figure 7.

From the comparison test results in Table 1, we can see that the weakly supervised algorithm combining the CAM-based and cycleGAN-based methods is already very close to the supervised algorithm in terms of model accuracy, and the main gap is reflected in the recall rate of the model. Compared with the mIOU of 0.408 obtained by the supervised method, the mIOU of our weakly supervised algorithm is 0.237, which still has a certain gap.

We also need to note that the cost of labeling the dataset is low due to the weakly supervised algorithm compared to the supervised learning algorithm. For this dataset, the average workload of labeling a supervised training datum is 29 coordinate points, while labeling a weakly supervised training datum requires only 1 label selection. Since the workload of labeling coordinate points is greater than the workload of labeling a selection, the workload of labeling a supervised training datum exceeds the workload of labeling a weakly supervised training data by a factor of 29. With continuously increasing data, we can easily improve the algorithm accuracy using the weakly supervised learning algorithm. While improving the supervised algorithm accuracy, we need to invest a greater labeling cost.

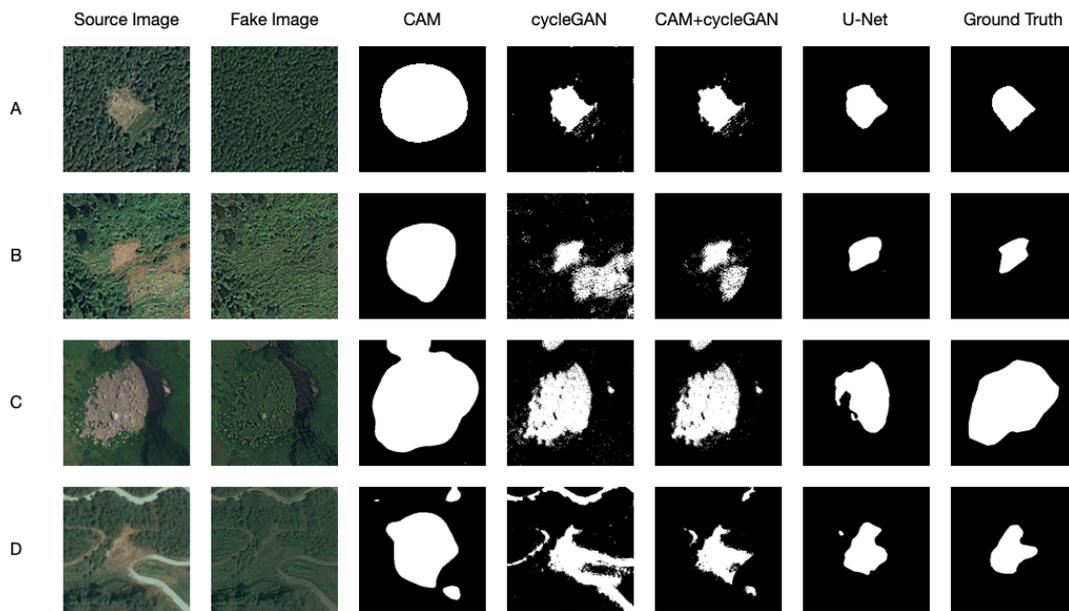


Figure 7. Results of remote sensing landslide image segmentation by different methods. We have used 4 samples with landslides (A–D) for test. The data are shown in Table 1, the mIOU of our method is 0.237, and the aggregate performance is lower than the supervised learning method but higher than other weakly supervised learning methods.

5. Conclusions

We studied a weakly supervised learning approach combining the CAM-based method and cycleGAN-based method for the semantic segmentation task of remote sensing landslide images, and compared the test results with the supervised learning approach.

The experimental results show that the mIOU of the combined method is higher compared with the single weakly supervised method, and the method can be used for remote sensing landslide area segmentation. Specifically, the accuracy of our model is 0.924, which is higher than 0.692 and 0.845 of CAM and cycleGAN, respectively, indicating that our model is more accurate in recognizing landslides. The FPR of our model is 0.004, which is lower than the 0.054 and 0.042 of CAM and cycleGAN, respectively, indicating that our model has a low probability of incorrect recognition. With the weakly supervised remote sensing landslide image segmentation algorithm, we can obtain a landslide semantic segmentation model based on image-level annotated training data. Compared with other models that must require pixel-level annotated training data, our model can greatly reduce the workload of annotators.

On the other hand, by comparing the weakly supervised learning method with the supervised learning method, we found that the experimental results of the method are still some distance away compared with the supervised learning method. However, since the weakly supervised learning method has the advantage of a low labeling cost, we can increase the training data to improve the accuracy of the algorithm. Another existing problem is that since the training data do not contain town buildings or rivers, this means that our model cannot be applied to remote sensing images containing town buildings or rivers. In the future, we can consider adding town buildings and rivers as training data in both the landslide dataset and non-landslide dataset to improve the accuracy of the model in remote sensing images of towns and rivers. For the model structure, using an end-to-end model for training is a worthwhile direction of investigation.

Author Contributions: Conceptualization, Y.Z. and H.W.; Formal analysis, R.Y.; Funding acquisition, H.W. and Q.X.; Methodology, Y.Z.; Resources, G.Y.; Software, Y.Z.; Supervision, H.W.; Writing original draft, Y.Z.; Writing—review & editing, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China, grant number 2019YFC1509602, the National Natural Science Foundation of China, grant number 41521002, the Sichuan Science and Technology Program, grant number 2021YFS0324 and 2021YFG0377, and the State Key Laboratory of Geohazard Prevention and Geoenvironment Protection Independent Research Project, grant number SKLGP2019Z012.

Data Availability Statement: The experimental datasets are accessed from open source data: Bijie Landslide Dataset (http://gpcv.whu.edu.cn/data/Bijie_pages.html, accessed on 26 June 2022).

Acknowledgments: The authors thank Dalan Xie for providing test datasets. They would also like to thank all the reviewers and editors for their great helps and useful suggestions.

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

References

1. Lukić, T.; Bjelajac, D.; Fitzsimmons, K.E.; Marković, S.B.; Basarin, B.; Mlađan, D.; Micić, T.; Schaetzel, R.J.; Gavrilov, M.B.; Milanović, M.; et al. Factors Triggering Landslide Occurrence on the Zemun Loess Plateau, Belgrade Area, Serbia. *Environ. Earth Sci.* **2018**, *77*, 519. [[CrossRef](#)]
2. Mohan, A.; Singh, A.K.; Kumar, B.; Dwivedi, R. Review on Remote Sensing Methods for Landslide Detection Using Machine and Deep Learning. *Trans. Emerg. Telecommun. Technol.* **2021**, *32*, e3998. [[CrossRef](#)]
3. Deng, J.; Dong, W.; Socher, R.; Li, L.-J.; Li, K.; Fei-Fei, L. Imagenet: A Large-Scale Hierarchical Image Database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 248–255.
4. Zhong, C.; Liu, Y.; Gao, P.; Chen, W.; Li, H.; Hou, Y.; Nuremanguli, T.; Ma, H. Landslide Mapping with Remote Sensing: Challenges and Opportunities. *Int. J. Remote Sens.* **2020**, *41*, 1555–1581. [[CrossRef](#)]
5. 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](#)]
6. Chen, Z.; Zhang, Y.; Ouyang, C.; Zhang, F.; Ma, J. Automated Landslides Detection for Mountain Cities Using Multi-Temporal Remote Sensing Imagery. *Sensors* **2018**, *18*, 821. [[CrossRef](#)] [[PubMed](#)]
7. Cheng, L.; Li, J.; Duan, P.; Wang, M. A Small Attentional YOLO Model for Landslide Detection from Satellite Remote Sensing Images. *Landslides* **2021**, *18*, 2751–2765. [[CrossRef](#)]
8. 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.
9. Du, B.; Zhao, Z.; Hu, X.; Wu, G.; Han, L.; Sun, L.; Gao, Q. Landslide Susceptibility Prediction Based on Image Semantic Segmentation. *Comput. Geosci.* **2021**, *155*, 104860. [[CrossRef](#)]
10. Prakash, N.; Manconi, A.; Loew, S. Mapping Landslides on EO Data: Performance of Deep Learning Models vs. Traditional Machine Learning Models. *Remote Sens.* **2020**, *12*, 346. [[CrossRef](#)]
11. Bo, Y.A.; Fang, C.; Chong, X.D. Landslide Detection Based on Contour-Based Deep Learning Framework in Case of National Scale of Nepal in 2015. *Comput. Geosci.* **2020**, *135*, 104388.
12. 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](#)]
13. Su, Q.; Zhang, X.; Xiao, P.; Li, Z.; Wang, W. Which CAM Is Better for Extracting Geographic Objects? A Perspective from Principles and Experiments. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 5623–5635. [[CrossRef](#)]
14. Feng, Z.; Zhu, M.; Stanković, L.; Ji, H. Self-Matching CAM: A Novel Accurate Visual Explanation of CNNs for SAR Image Interpretation. *Remote Sens.* **2021**, *13*, 1772. [[CrossRef](#)]
15. Park, M.; Tran, D.Q.; Jung, D.; Park, S. Wildfire-Detection Method Using Densenet and CycleGAN Data Augmentation-Based Remote Camera Imagery. *Remote Sens.* **2020**, *12*, 3715. [[CrossRef](#)]
16. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft Coco: Common Objects in Context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
17. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning Deep Features for Discriminative Localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.
18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
19. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. *Adv. Neural Inf. Process Syst.* **2014**, *27*, 2672–2680.

20. Zhu, J.-Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
21. Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A.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.
22. 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](#)]
23. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation. *arXiv* **2017**, arXiv:1704.06857.
24. 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: Berlin/Heidelberg, Germany, 2015; pp. 234–241.