

Article

Automatic Identification of Landslides Based on Deep Learning

Shuang Yang ¹, Yuzhu Wang ^{1,*} , Panzhe Wang ², Jingqin Mu ^{1,3}, Shoutao Jiao ⁴, Xupeng Zhao ¹, Zhenhua Wang ¹ , Kaijian Wang ⁵ and Yueqin Zhu ⁶

¹ School of Information Engineering, China University of Geosciences, Beijing 100083, China

² School of Geophysics and Information Technology, China University of Geosciences, Beijing 100083, China

³ Department of Computer Science, Tangshan Normal University, Tangshan 063000, China

⁴ Development and Research Center, China Geological Survey, Beijing 100037, China

⁵ China Aero Geophysical Survey and Remote Sensing Center for Natural Resources, China Geological Survey, Beijing 100083, China

⁶ National Institute of Natural Hazards, Ministry of Emergency Management, Beijing 100085, China

* Correspondence: wangyz@cugb.edu.cn

Abstract: A landslide is a kind of geological disaster with high frequency, great destructiveness, and wide distribution today. The occurrence of landslide disasters bring huge losses of life and property. In disaster relief operations, timely and reliable intervention measures are very important to prevent the recurrence of landslides or secondary disasters. However, traditional landslide identification methods are mainly based on visual interpretation and on-site investigation, which are time-consuming and inefficient. They cannot meet the time requirements in disaster relief operations. Therefore, to solve this problem, developing an automatic identification method for landslides is very important. This paper proposes such a method. We combined deep learning with landslide extraction from remote sensing images, used a semantic segmentation model to complete the automatic identification process of landslides and used the evaluation indicators in the semantic segmentation task (mean IoU [mIoU], recall, and precision) to measure the performance of the model. We selected three classic semantic segmentation models (U-Net, DeepLabv3+, PSPNet), tried to use different backbone networks for them and finally arrived at the most suitable model for landslide recognition. According to the experimental results, the best recognition accuracy of PSPNet is with the classification network ResNet50 as the backbone network. The mIoU is 91.18%, which represents high accuracy; Through this experiment, we demonstrated the feasibility and effectiveness of deep learning methods in landslide identification.

Keywords: deep learning; semantic segmentation; PSPNet; landslide



Citation: Yang, S.; Wang, Y.; Wang, P.; Mu, J.; Jiao, S.; Zhao, X.; Wang, Z.; Wang, K.; Zhu, Y. Automatic Identification of Landslides Based on Deep Learning. *Appl. Sci.* **2022**, *12*, 8153. <https://doi.org/10.3390/app12168153>

Academic Editor: Rubén Usamentiaga

Received: 13 July 2022

Accepted: 11 August 2022

Published: 15 August 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 common and frequent geological hazards around the world. The occurrence of landslides affects the terrain and causes different degrees of damage [1]. Furthermore, when residential areas or public buildings are close to a landslide site, the event is often accompanied by emergencies. Landslides in specific areas need to be identified within a short period to intervene and resolve the crisis [2,3]. With the rapid development of remote sensing technology, these methods have been widely used [4,5]. Currently, the landslide identification methods based on remote sensing images are mainly divided into visual interpretation, pixel-based and object-oriented landslide identification methods.

Visual interpretation is the earliest landslide identification method applied to remote sensing images. Visual interpretation is when the interpreter extracts the landslide shape from the remote sensing image according to his or her professional knowledge and related research materials. Data extracted by this method have high accuracy. However, the visual interpretation also has the disadvantages of being time-consuming, having an overreliance on manual discrimination and inefficient [6].

Pixel-based landslide identification methods focus on pixel values and pixel value changes in remote sensing images and classify them according to pixel changes or change characteristics [7]. Although this method overcomes the shortcomings of visual interpretation, it is only judged by a single pixel value; moreover, it does not consider the correlation between pixels, resulting in an indistinct recognition of landslide edge regions and poor performance [8].

Object-oriented landslide identification methods utilize the attribute features of raw data (such as texture, spectrum, etc.) to classify remote sensing images by one or more attributes [9]. However, this method uses one or more attribute features to set thresholds, and the process is complicated and cannot process large-scale remote sensing images in time.

With the continuous increase in high-resolution remote sensing images, how to quickly and efficiently identify targets from massive remote sensing images has become a key problem for scholars and experts to study. In recent years, with the rapid development of deep learning, computer vision and image processing technology have been introduced into the field of remote sensing as a new method for remote sensing image classification and target detection [10]. Research has shown that deep learning methods do not require many of the data provided in traditional landslide identification methods to help identify landslides; they only need enough landslide images as samples and training, which greatly simplifies the complex calculation process, makes up for the shortcomings of the above-mentioned traditional landslide identification methods, and realizes automatic identification of landslides. At the same time, deep learning methods are higher accuracy than the traditional landslide identification methods [11]. The main contributions of this article are as follows:

- (1) We process the landslide data in the Bijie landslide dataset, create a landslide dataset, and preprocess the dataset (data cleaning, data enhancement);
- (2) On the landslide dataset, we use three models (U-Net, DeepLab v3+ and PSPNet) to conduct experiments and test the performance changes in the models when different classification networks are used as the backbone network;
- (3) We use the above pretrained model to test the landslide test set and use mIoU, precision, and recall to evaluate the model performance to obtain the optimal model for landslide identification performance.

The remainder of this paper is structured as follows. Section 2 introduces related work. Section 3 shows the data and methodology we used. Section 4 presents and analyzes the experiment results. Section 5 presents our conclusions.

2. Related Work

Convolutional neural networks (CNN) [12] have achieved great success in the field of image processing because of their nonlinear learning ability [13], driving the rapid development of computer vision [14,15]. Based on CNN studies, various models have been developed for image classification [16], object detection [17,18], semantic segmentation [19], etc., where semantic segmentation performs pixel-level segmentation of the images. These models have achieved satisfactory results in traditional vision tasks. Therefore, people have begun to apply these deep learning models to landslide identification in remote sensing images in the past few years.

Ye et al. (2019) proposed a constrained deep learning model, applied it to identifying landslides in hyperspectral images, and compared the results with the support vector machine–spectral information divergence–spectral angle matching method. They that the extraction of high-level features by deep learning has great potential for improving the accuracy of landslide identification [20]. Ghorbanzadeh et al. (2019) used CNNs for Himalayan landslide identification and compared them with state-of-the-art machine learning methods (artificial neural networks, support vector machines, and random forests); the results show that deep learning is superior to machine learning in landslide identification experiments [21]. Prakash et al. (2020) proposed an improved U-Net model that uses ResNet34 blocks for feature extraction and enables landslide identification in Douglas County, south of Portland, Oregon, USA [22]. Zhu et al. (2020) proposed a method based

on U-Net architecture to fuse local and nonlocal features, upsampling by dilated convolution, and the corresponding spatial pyramid expanded receptive field and scale attention mechanism to identify the landslide caused by the earthquake in Jiuzhaigou, China [23]. Ji et al. (2020) developed a CNN-based spatial channel attention mechanism to classify and identify landslides in Bijie City, China, from available satellite imagery and DEM datasets; this experiment concludes that the attention mechanism and DEM data can effectively improve the accuracy of landslide identification [24]. Liu (2020) proposed to use ResU-Net to identify earthquake landslides in Jiuzhaigou, Sichuan Province, China, and obtained an F1 value of 93.3%, and an mIoU value of 87.5% [25]. Ju et al. (2020) identified old loess landslides on Google Earth images using the two-stage algorithm Mask R-CNN; although the accuracy rate did not reach a high level, it confirmed the feasibility of Mask R-CNN to identify old landslides [26]. Dai et al. (2021) proposed an improved U-Net neural network and completed the automatic identification of the deformation features of the landslide time series [27]. Ullo et al. (2021) used the Mask R-CNN model with ResNet101 as the backbone network and the transfer learning algorithm to complete landslide recognition in digital images of hilly areas obtained by drones, and the results show that the method is superior to the existing research in both algorithm performance and robustness [28]. Liu (2021) proposed to use an improved Mask R-CNN to identify earthquake landslides in the Jiuzhaigou area of Sichuan Province, China, and obtained an F1 value of 94.5% and an mIoU value of 89.6% [29]. Ghorbanzadeh (2022) combined the object-based image analysis (OBIA) approach with the fully convolutional network (FCN) model to complete the landslide detection in Sentinel-2 images and verified the method's feasibility [30].

Therefore, deep learning methods have been applied to landslide identification. However, due to the diversity and complexity of landslides, these methods still have many problems to be solved. For example, in order to improve the ability to identify landslides, the model needs to learn a large number of data [31]; this is a key issue because there is very little landslide data currently available. In addition, for this work, more data information can help the model to better improve the landslide recognition accuracy [32]; however, the acquisition, retrieval, and annotation of datasets is often a difficult point in landslide identification tasks. Therefore, to address the issues mentioned above, we conducted this study. Since there are few publicly available landslide datasets and their quality is uneven, to have a good experimental basis, we selected the Bijie dataset published by Ji et al. (2020) [24]. The Bijie dataset is the first large-scale, public remote sensing landslide dataset and has a double check; more detailed dataset information will be introduced in Section 3.1. At the same time, since the interpretation of landslide images has very high professional requirements, in the re-labeling of samples, we strictly follow the samples provided by the Bijie dataset to ensure the reliability of the data. In the final sample set, we expand the sample set through the data augmentation method and finally obtain a data set containing 2500 landslide images.

3. Materials and Methods

3.1. Data Source

Remote sensing datasets of landslide are difficult to obtain, we used the open source Bijie landslide dataset [24]. The Bijie landslide dataset is the first open remote sensing landslide dataset with careful triple inspection; the data set was proposed by scholars such as Ji et al. (2020), and the classification research of landslides was carried out on it. Its study area is located in Bijie City, Guizhou Province, China, with about 26,853 square km and an altitude ranging from 457 m to 2900 m. The soil on the slopes caused by the perennial rainfall is soft and prone to landslides, and it is one of the most prone areas in China.

The remote sensing images in the Bijie landslide dataset were captured by the TripleSat satellite, and the RGB images have a resolution of 0.8 m. Seven hundred and seventy landslide images and two thousand and three other types of images were intercepted from the captured remote sensing images. The dataset consists of satellite optical images and label files. In the process of making the dataset, two methods were adopted to interpret the

landslide images to ensure the reliability of the database: One is the visual interpretation by geologists through optical remote sensing images; the other is based on residents' reports and field surveys. Throughout the work, the shapes of landslide samples were drawn with the help of ArcGIS.

3.2. U-Net

U-Net is a semantic segmentation network proposed by Olaf Ronneberger in the ISBI Cell Segmentation Competition in 2015. It utilizes a U-shaped network structure to capture contextual information and location information. It was initially used to solve medical image segmentation problems, especially cell-level segmentation tasks, and was gradually used to solve problems in other fields [33].

The network structure of U-Net, which is an encoder-decoder structure, is shown in Figure 1. The encoder utilizes the idea of stacking convolutional layers, downsampling the feature map through convolution and pooling, and performing four total pooling operations. After each stacking convolution layer operation, the size of the feature map is halved and, at the same time, the pooling result of each step is passed to the decoder; in the decoder, the feature map is first upsampled or deconvolved and then concatenated on the channel with the previous feature map of the same size. Convolution and upsampling is then performed, and after upsampling four times, an output result of the same size as the original image is obtained.

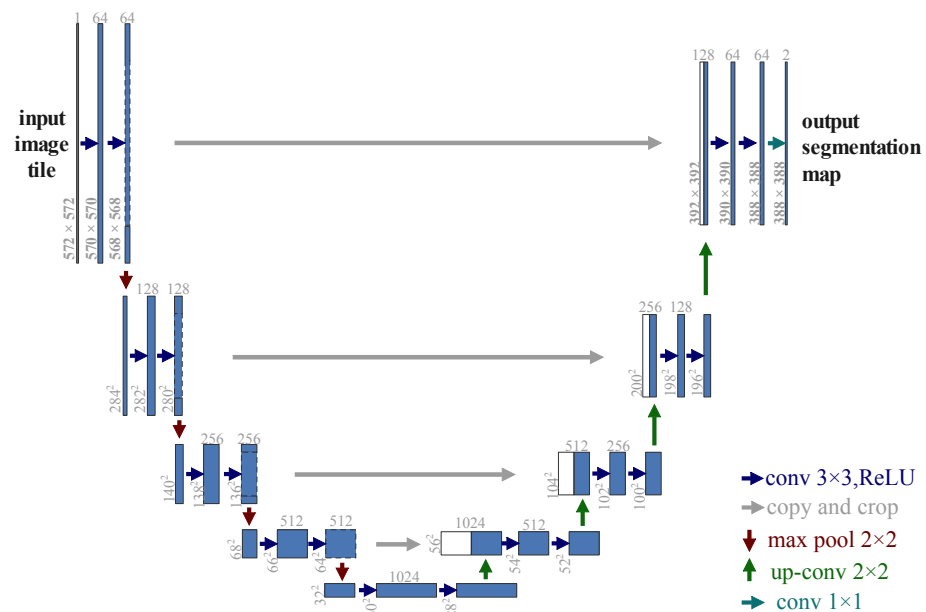


Figure 1. U-Net architecture. The blue box represents the multichannel feature layer. The channel number is shown at the top of the box. The white boxes represent the replicated feature maps. The arrows represent operations on the feature layers.

3.3. DeepLab v3+

DeepLabv3+ was proposed by the Google team in 2018 and is the DeepLab series model [34]. DeepLabv3+ is based on DeepLabv3 and improves it. It uses Deeplabv3 as the encoder, introduces atrous convolution in the encoder for downsampling and uses the spatial pyramid pooling module to extract multiscale information, which improves the accuracy by fusing low-level and high-level features.

Its specific structure is shown in Figure 2. The encoder extracts image features through a deep convolutional neural network (DCNN), and the extracted feature layers are input to the decoder for 1×1 convolution. Meanwhile, the feature layers extracted by DCNN use 1×1 , 3×3 , 3×3 , and 3×3 atrous convolutions for downsampling and pooling, where the expansion rates of atrous convolution are 1, 6, 12 and 18, respectively. Then, the obtained

new feature layer is concatenated, its channel is changed to $1/5$ of the original through 1×1 convolution, and the decoder is entered for upsampling. The unsampled feature layer is concatenated with the feature layer in the decoder, and then the 3×3 convolution and upsampling are gone through to obtain the final prediction map.

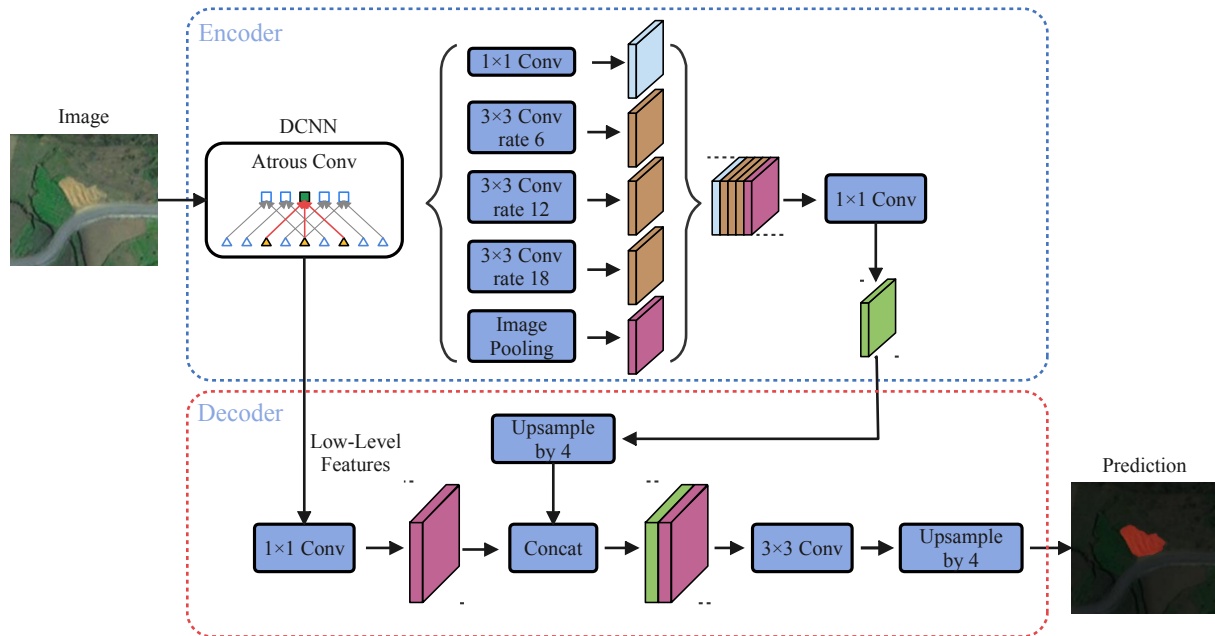


Figure 2. DeepLabv3+ architecture. DCNN stands for deep convolutional neural network, and Atrous Conv stands for atrous convolution. Cyan, orange, pink and green represent the extracted feature layers. The dark blue boxes represent the operations taken on the feature layers.

3.4. PSPNet

PSPNet is a semantic segmentation model jointly proposed by the Chinese University of Hong Kong and Shangtang Technology and it won the championship in the 2016 ImageNet Challenge [35]. The original intention of PSPNet was to improve the FCN. The most prominent feature of PSPNet is that it adds a PSP module between the encoder and the decoder, which is also the main difference between it and the FCN.

The structure of PSPNet is shown in Figure 3; the input layer obtains the feature layer of the input image through CNN, and the feature layer size is changed to $1/5$ of the original through. Then, the obtained feature map is input to the pyramid pooling module. First, this module divides the input feature layer into 6×6 , 3×3 , 2×2 and 1×1 sized areas; the ave-pooling operation is performed in the divided area to obtain four feature layers of different sizes (corresponding to the green, blue, orange and red outputs in Figure 3, respectively); then, 1×1 convolution operations are performed on these feature layers; next, the number of channels of the feature layer is changed to one-fourth of the original; and finally, the feature layer is up-sampled by bilinear interpolation. The upsampled feature map and the feature layer obtained by CNN are concatenated, and finally, the final output is obtained through the convolution operation.

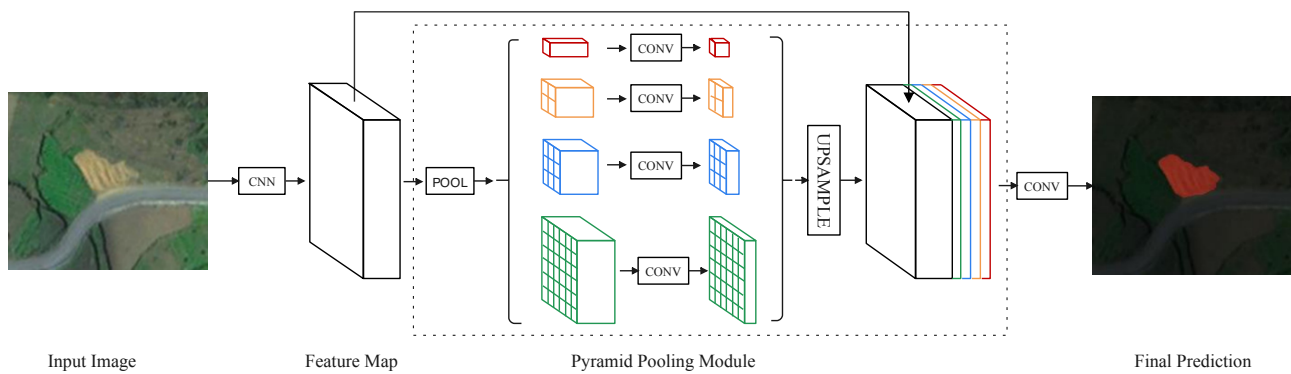


Figure 3. PSPNet architecture. First, a CNN is used to obtain the last convolutional feature map given an input image. A pyramid parsing module is then applied to collect different subregion representations, followed by upsampling and concatenation layers to form the final feature representation. Finally, the representation is fed into convolutional layers to obtain the final per-pixel predictions.

3.5. Evaluation Metrics

In deep learning methods, recall rate and accuracy rate are indicators that can evaluate the recognition effect of the model, and they are associated with the confusion matrix; As Table 1 shows, T and F represent the prediction of true or false; P (positive) and N (negative) represent the type of prediction; TP (true positive), TN (true negative), FP (false positive), and FN (false negative) are used to classify pixels. Taking this article as an example, TP means that the pixel is identified as a landslide pixel; the identification is correct. TN means that the pixel is identified as a nonlandslide pixel; the identification is correct. FP indicates that the pixel was identified as a landslide pixel and identified incorrectly. FN indicates that the pixel was identified as a nonlandslide pixel and identified incorrectly.

Table 1. Confusion matrix of classification results.

Actual Values	Predicted Values	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Precision represents the proportion of the actual landslide pixels in the pixels predicted by the model as landslides, as shown in Equation (1):

$$Precision = \frac{TP}{FP + TP}, \quad (1)$$

Recall represents the proportion of landslide pixels predicted by the model in all actual landslide pixels, as shown in Equation (2):

$$Recall = \frac{TP}{FN + TP}, \quad (2)$$

In addition, mIoU is a widely used metric in semantic segmentation tasks and is used as a standard measure to measure semantic segmentation models. Intersection over union (IoU) represents the ratio between the intersection and union of the predicted results of landslide pixels and the actual landslide pixels, and mIoU represents the average of all categories of IoU. The IoU is shown by Equation (3):

$$mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij} + \sum_{j=0}^k p_{ji} - p_{ii}}, \quad (3)$$

In this formula, i represents the real pixel, j represents the predicted pixel, p_{ij} represents that i will be predicted as j , p_{ii} represents that i will be predicted as i and p_{ji} represents that j will be predicted as i .

4. Results and Discussion

4.1. Data Preprocessing

Data preprocessing includes relabeling samples and implementing data augmentation strategies on samples. Due to different target tasks, we only selected 770 landslide samples in the Bijie landslide dataset as the original data. In the labeling work, in order to obtain reliable data, we relabeled according to the original labels provided in the Bijie dataset. At the same time, we processed the data according to the research goals and finally obtained 510 landslide samples. Among these obtained data, 90% of the data was used for training and validation of the model, and 10% of the data was used for testing the model.

The deep learning model learns the features of the image through the provided image samples during the training process. Therefore, the more samples that are provided to the model, the more the model can learn the features of such images and the better its predictions. For the model, if an image is rotated, cropped and then passed into the model, the model will consider it a new image, so we can enhance the sample by expanding it. We implemented a data enhancement strategy for the existing training set according to the characteristics of sample landslides with different directions, different structures and different boundary shapes. We enhance the dataset's quality through augmentation techniques, we thereby improved the model's training effect. However, excessive rotation and flipping of images will cause overfitting of the model, thereby reducing its generalization and causing the model to achieve high accuracy on the training set but not achieve very high accuracy on the test set. Considering these problems in the process of data expansion, we rotate, scale and flip the sample set according to a certain probability based on experience. These expanded data were all used in our model training; after expansion, the training samples were changed from the original 510 to 2500. For more details on data preprocessing, see Table 2.

Table 2. Data augmentation.

Method	Probability of Execution	Specific Operations
random rotation	50%	rotate 20°, +90°, −90°
left-right flipping	100%	flip the image left and right
image cropping	100%	original image 0.7 × dimension

4.2. Training

The training effect of the deep learning model is closely related to the accuracy of the data set, suitable parameters, and training methods. Therefore, in this study, we selected three models, U-Net, DeepLab v3+, and PSPNet, and used two different classification networks as the backbone network of each model; the backbone network selection of the model is shown in Table 3. We record these models as U-Net (VGG), U-Net (ResNet50), DeepLab v3+ (MobileNet), DeepLab v3+ (Xception), PSPNet (MobileNet), and PSPNet (ResNet50).

In the experiment, we set and adjusted the training parameters uniformly for all the models, as shown in Table 4. The input image size is fixed at 473 × 473. The classification of pixel types is landslide and background. The model training adopts the method of freezing training, which divides the training into two stages: freezing and unfreezing. In the freezing stage, the backbone of the model is frozen, and the feature extraction network does not change. At this time, 50 rounds of fine-tuning are performed on the network. The video memory occupied in the freezing phase is small, so the batch_size and learning rate are set larger. In the unfreezing stage, the backbone network of the model is unfrozen, and the feature extraction network will change; at this time, the network is trained for

100 rounds. Due to a large amount of video memory occupied, the batch_size is set to 0.5 times that of the frozen phase, and the learning rate is reduced.

Table 3. The backbone network selection of the models.

Model	Backbone
U-Net	VGG
U-Net	ResNet50
DeepLab v3+	MobileNet
DeepLab v3+	Xception
PSPNet	MobileNet
PSPNet	ResNet50

Table 4. Parameters for our models.

Hyper-Parameter	Parameter Values
input_shape	[473, 473]
classes	landslide, background
freeze_Train	True
pretrained weights	True
datasets used for pre-training	VOC data set
Init_Epoch	0
downsample_factor	8
freeze_epoch	50
unfreeze_epoch	100
freeze_learning_rate	10^{-4}
freeze_batch_size	8
unfreeze_batch_size	4
unfreeze_learning_rate	10^{-5}
focal_loss	True
dice_loss	True
eager pattern	False
aux_branch	False
early_stopping	True
num_workers	1
cls_weights	np.array([1, 2], np.float32)

In addition, during model training, we use the pre-trained weights obtained in the VOC dataset as the initial parameters of the model and set dice_loss to balance the number of training categories. We set focal_loss to balance positive and negative samples and use the NumPy form to give different loss weights to the background and landslides so that the model focuses on landslide pixels. In order to improve the recognition effect, we set the downsample_factor to 8, but it will also occupy much memory; therefore, in order to reduce the occupation of video memory, we do not use aux_branch by default, do not use multi-threading to read data and use the early stop strategy to save computing resources.

The hardware environment of this experiment: GPU: 4*NVIDIA Tesla K80, CPU: 32*Intel (R) Xeon (R) CPU E5-2620 v4 @ 2.10GHz, OS: CentOS 8.3. The software environment of this experiment: CUDA 11.2, Python 3.6, PyTorch 1.10.1, Tensorflow 2.2.0.

4.3. Experimental Results and Analysis

Pretrained models obtained by the experiment were used to predict the test set. The prediction results show that these models can effectively identify landslides. At the same time, according to the experimental results, we conclude that the PSPNet model using ResNet50 as the backbone network has the best recognition effect. We will discuss the results first from the perspective of image recognition and then from the perspective of index evaluation.

A total of 51 remote sensing images of landslides were included in the test set; all of the images were from the eastern part of the Qinghai–Tibet Plateau in Bijie City, Guizhou Province, China. As shown in Figure 4, we show some landslide images in the test set, and none of the images in the test set participated in the training, through which we evaluated the performance of each model. We have selected some of these samples for analysis, as shown in Figure 5. The figure includes four samples of landslide images that occurred in different places, marked as Landslide I, Landslide II, Landslide III, and Landslide IV. Among them, I, II, and III are new landslides with various characteristics and shapes of landslides, and IV is an old landslide with inconspicuous characteristics of landslides. In addition, the figure also includes the label file of the landslide, which was used for comparison with the predicted map. We use the pretrained model to obtain the predicted map; we will analyze the predicted maps of the four landslide samples separately.



Figure 4. Examples of landslide samples in the test set, all from Bijie, Guizhou, China.

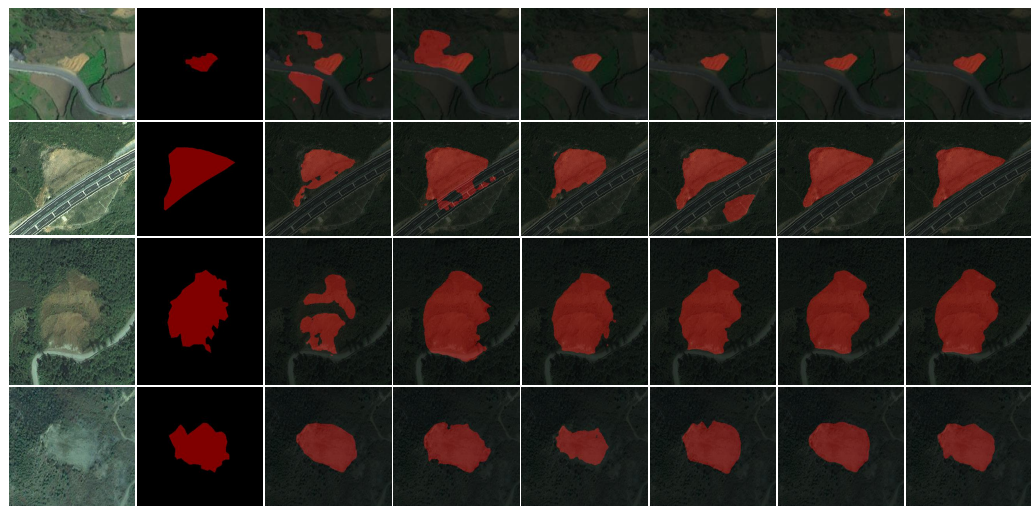


Figure 5. Model prediction of landslide results. Horizontally, they are Landslide I, Landslide II, Landslide III and Landslide IV. Vertically, they are the landslide image, label map (for comparison), U-Net (VGG) recognition map, DeepLabv3+ (Xception) recognition map, DeepLab v3+ (MobileNet) recognition map, U-Net (ResNet50), the recognition map of PSPNet (MobileNet) and the recognition map of PSPNet (ResNet50).

Landslide I: U-Net (VGG) has a high error rate in identifying landslide images; it identifies the background associated with the landslide color as a landslide. However, the majority of these environments are land and cultivated land. There are 15 additional prediction maps for the same situation. DeepLab v3+ (Xception) is similar to U-Net (VGG). DeepLab v3+ (MobileNet), U-Net (ResNet50), PSPNet (MobileNet), and PSPNet (ResNet50) can more accurately identify landslides; in contrast, PSPNet (ResNet50) has better performance in recognizing landslides.

Landslide II: U-Net (VGG) and DeepLab v3+ (MobileNet) cannot fully identify landslides, so the effect is poor. DeepLabv3+ (Xception) recognizes roads with similar colors as landslides. U-Net (ResNet50) recognizes other objects with similar shapes to landslides as landslides and cannot distinguish the landslide from other content in the background. PSPNet (MobileNet) and PSPNet (ResNet50) perform better.

Landslide III: There is a chasm in identifying the landslide by U-Net (VGG); the color of the chasm part is darker, so the background color is closer, which is caused by the new vegetation growing on the landslide, the model cannot recognize this and thus identifies errors. The situation of DeepLab v3+ (Xception) is just the opposite. Although it can distinguish landslides from vegetation, it cannot distinguish the features between landslides and roads. DeepLabv3+ (MobileNet), U-Net (ResNet50), PSPNet (MobileNet) and PSPNet (ResNet50) can identify landslides more accurately; however, their sensitivities to landslide boundaries vary.

Landslide IV: Landslide IV has been formed for a long time, the landslide has been covered with vegetation, and the entire landslide is green. Because its landslide characteristics are not prominent, they are not easy to separate. This type of landslide is too complex for the model to recognize. Fortunately, according to the recognition effect of these models on Landslide IV, the model can distinguish the landslide from the features of the surrounding vegetation. However, the segmentation effect of the boundary is not accurate enough.

We evaluated the model using the metrics in Section 3.5, and the evaluation results are shown in Table 5. Observing the recall index and precision index values, we found that the recall rate of all models is higher than the precision; this means that these models can identify real landslide pixels but also identify many non-landslide pixels as landslide pixels. It can also be seen from the analysis of Figure 6 that when identifying Landslide I and Landslide II, U-Net (VGG), DeepLabv3+ (Xception) and U-Net (ResNet50) easily confuse other objects with similar colors and shapes to landslides.

In Table 5, mIoU values are used to evaluate model performance comprehensively. Among these pretrained models, PSPNet (ResNet50) produced the best landslide recognition effect, with an mIoU value of 91.18%, and obtained the highest precision index (93.76%); this means that the model has a good effect on the recognition of landslide pixels. Followed by PSPNet (MobileNet) and U-Net (ResNet50), the mIoU values of which are 89.11% and 88.75%, respectively; PSPNet (MobileNet) obtained the highest recall index (97.39%), which means that the model can identify most of the landslide pixels. U-Net (VGG) has the worst landslide recognition effect, with a mIoU of 81.64% and a recall and precision of 89.34% and 89.22%, respectively.

Table 5. The results of the suggested model. Numbers in bold represent the best model for identifying landslides (with mIoU metric as the final criterion).

Model	Backbone	mIoU	Recall	Precision
U-Net	VGG	81.64%	89.34%	89.22%
DeepLab v3+	Xception	86.15%	92.26%	92.20%
DeepLab v3+	MobileNet	87.06%	94.06%	91.64%
U-Net	ResNet50	88.75%	96.15%	91.82%
PSPNet	MobileNet	89.11%	97.39%	92.61%
PSPNet	ResNet50	91.18%	96.9%	93.76%

Below, we combine the chart to discuss the recognition effect of PSPNet on landslides when MobileNet and ResNet50 are used as the backbone network, respectively. As shown in Table 6, we used precision, recall and IoU to evaluate the model's ability to recognize landslide and background pixels, respectively, when using two different backbone networks. P is the abbreviation for precision, R is the abbreviation for recall and IoU represents intersection over union. In identifying background pixels, when ResNet50 is used as the backbone network, the IoU value is 97.76%, which is 14.79% higher than when MobileNet is used as the backbone network; this can be seen intuitively from the landslide prediction map. In Figure 6, the blue boxes mark the parts of PSPNet (ResNet50) and PSPNet (MobileNet) that misidentify the background pixels as landslide pixels. Compared with PSPNet (MobileNet), PSPNet (ResNet50) is less likely to mistakenly identify content in the background (such as roads, green vegetation, bare land) as landslides. At the same time, the precision and recall of ResNet50 are higher than MobileNet. In identifying landslide pixels, when ResNet50 is used as the backbone network, the IoU value is 84.6%, which is 3.45% higher than when MobileNet is used as the backbone network. At the same time, the precision is 5.35% higher than that of PSPNet (MobileNet), and the recall is lower than 1.9%. Therefore, its landslide recognition effect is better.



Figure 6. The comparison results of landslide identification of PSPNet using MobileNet and ResNet50 as the backbone network, respectively. Behaviors 1, 3 and 5 use MobileNet as the identification result of the backbone network, and behaviors 2, 4 and 6 use ResNet50 as the identification result of the backbone network. The blue box marks the background pixels that the model misidentified as part of the landslide.

Table 6. The performance of PSPNet under different backbone networks. The numbers in bold are the highest IoU values obtained for landslide and background recognition, respectively.

Backbone	Evaluate	Landslide	Background
MobileNet	P	82.79%	97.08%
	R	97.37%	97.41%
	IoU	81.15%	82.97%
ResNet50	P	88.14%	99.41%
	R	95.47%	98.34%
	IoU	84.6%	97.76%

4.4. Discussion

In previous research on landslide recognition based on deep learning, Ghorbanzadeh used CNN to train and test landslides in the Himalayas and obtained an F1 value of 87.8% and an mIoU value of 78.26% [21]. Liu (2020) proposed to use ResU-Net to identify earthquake landslides in Jiuzhaigou, Sichuan Province, China, and obtained an F1 value of 93.3% and an mIoU value of 87.5% [25]. Ullo (2021) applied Mask R-CNN to landslide recognition in digital images of target hilly areas acquired by drones; and when ResNet101 was used as the backbone network, the obtained F1 value was 97% [28]. Liu (2021) proposed to use an improved Mask R-CNN to identify earthquake landslides in the Jiuzhaigou area of Sichuan Province, China, and obtained an F1 value of 94.5% and an mIoU value of 89.6% [29]. Ghorbanzadeh (2022) obtained an F1 of 84.03% and an mIoU value of 72.49% when using ResU-Net and OBIA for landslide detection in multitemporal Sentinel-2 images [21]. Our proposed PSPNet, using the classification network ResNet50 as the backbone network, achieves a 91.18% mIoU value on the Bijie landslide dataset; however, due to the differences in datasets and evaluation metrics, we cannot compare it with other models; however, according to the current experimental results, the method proposed in this paper is effective for landslide recognition.

Although PSPNet (ResNet50) achieves good results in landslide identification, it still has some shortcomings. For example, its segmentation of landslide boundaries still needs further improvement. Landslide images are different from traditional remote sensing images. In addition to the information of the images themselves, remote sensing images also contain rich geological information. For example, a digital elevation model (DEM) can reflect local terrain features at a specific resolution. Therefore, we should further combine deep learning with remote sensing to maximize the role of remote sensing data. If the DEM data and remote sensing images are fused, we can obtain the local terrain information of the landslide from the DEM, which will help the model to improve the segmentation accuracy of the landslide boundary.

At present, the automatic identification of landslides based on deep learning still presents research challenges and problems to be solved. For example, the scarcity of open source code for landslide identification research and the lack of high-resolution public landslide remote sensing image datasets and validation areas have brought great difficulties to such research. At the same time, the further improvement in landslide identification accuracy remains to be explored. Given these problems, we need to continue research on automatic landslide identification. In terms of datasets, we will try to integrate the DEM data into remote sensing images so that the datasets contain more information about landslides, thereby improving the accuracy of landslide identification. In terms of models, we will further discuss the influence of model structure on landslide identification and then improve the model to improve the effect of landslide identification.

5. Conclusions

In this study, we proposed a deep learning-based research method for the automatic identification of landslides and obtained good results on the Bijie landslide dataset. First, we reconstructed the dataset for semantic segmentation and preprocessed the landslide data

based on the Bijie landslide dataset. We then used three models: U-Net, DeepLabv3+ and PSPNet, each using two classification networks as the backbone network, and completed training and validation on the Bijie landslide dataset on these models. Finally, we used the experimentally obtained pretrained model for landslide recognition and used mIoU, precision and recall to evaluate the model's recognition effect on landslides. According to the experimental results, we obtained the best recognition effect of the PSPNet model with ResNet50 as the backbone network, with an mIoU of 91.18%. This experimental result shows that it is feasible to detect the automatic identification of landslides by using the deep learning method; simultaneously, our proposed PSPNet method with ResNet50 as the backbone network can effectively identify landslides.

Based on the above experimental results, we believe that this research will be helpful to landslide relief operations in real life. The automatic identification method can effectively make up for the shortcomings methods, which are time-consuming, labor-intensive and highly dependent on labor; save much time and human resources for emergency rescue work; and reduce the loss of life and property. At the same time, it can help geologists significantly improve their work efficiency and allow them to spend more time on work that requires more geologists. Therefore, this research has significant practical potential.

In future work, we will contribute to the lack of high-resolution landslide remote sensing image datasets and open source code for landslide identification research. We will try to enrich the landslide data set by adding more factors so that it contains more landslide information for the model to improve the recognition accuracy of landslides further. In terms of models, we will try to explore, for example, the self-attention mechanism is introduced into the model so that the model can learn the characteristics of landslides in a more targeted way, thereby improving its accuracy. In addition, compared with new landslides, the characteristics of old landslides are less obvious, and it is not easy to realize automatic identification. Thus, we will also try to apply the training model of new landslides to the extraction of old landslides through transfer learning, thereby improving the accuracy of deep learning methods in the automatic identification of old landslides.

Author Contributions: Conceptualization, Y.W. and Y.Z.; methodology, S.Y. and Y.W.; software, S.Y., P.W. and X.Z.; validation, S.Y. and P.W.; writing—original draft preparation, S.Y.; writing—review and editing, Y.W., J.M., S.J., Z.W. and K.W.; supervision, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 41872253 and in part by the GHFUND B of China under Grant ghfund202107021958.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Hacıfendioğlu, K.; Demir, G.; Başağa, H.B. Landslide detection using visualization techniques for deep convolutional neural network models. *Nat. Hazards* **2021**, *109*, 329–350. [\[CrossRef\]](#)
2. Voigt, S.; Kemper, T.; Riedlinger, T.; Kiefl, R.; Scholte, K.; Mehl, H. Satellite image analysis for disaster and crisis-management support. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1520–1528. [\[CrossRef\]](#)
3. Plank, S.; Tuele, A.; Martinis, S. Landslide mapping in vegetated areas using change detection based on optical and polarimetric SAR data. *Remote Sens.* **2016**, *8*, 307. [\[CrossRef\]](#)
4. Czikhhardt, R.; Papco, J.; Bakon, M.; Liscak, P.; Ondrejka, P.; Zlocha, M. Ground stability monitoring of undermined and landslide prone areas by means of sentinel-1 multi-temporal InSAR, case study from Slovakia. *Geosciences* **2017**, *7*, 87. [\[CrossRef\]](#)
5. Rosi, A.; Tofani, V.; Tanteri, L.; Tacconi Stefanelli, C.; Agostini, A.; Catani, F.; Casagli, N. The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution. *Landslides* **2018**, *15*, 5–19. [\[CrossRef\]](#)

6. 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. [\[CrossRef\]](#)
7. Li, Z.; Shi, W.; Myint, S.W.; Lu, P.; Wang, Q. Semi-automated landslide inventory mapping from bitemporal aerial photographs using change detection and level set method. *Remote Sens. Environ.* **2016**, *175*, 215–230. [\[CrossRef\]](#)
8. 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. [\[CrossRef\]](#)
9. Chen, T.; Trinder, J.C.; Niu, R. Object-oriented landslide mapping using ZY-3 satellite imagery, random forest and mathematical morphology, for the Three-Gorges Reservoir, China. *Remote Sens.* **2017**, *9*, 333. [\[CrossRef\]](#)
10. Vaduva, C.; Gavat, I.; Datcu, M. Deep learning in very high resolution remote sensing image information mining communication concept. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012; pp. 2506–2510.
11. Peña, J.M.; Gutiérrez, P.A.; Hervás-Martínez, C.; Six, J.; Plant, R.E.; López-Granados, F. Object-based image classification of summer crops with machine learning methods. *Remote Sens.* **2014**, *6*, 5019–5041. [\[CrossRef\]](#)
12. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
13. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
14. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
15. Ghorbanzadeh, O.; Tiede, D.; Dabiri, Z.; Sudmanns, M.; Lang, S. Dwelling extraction in refugee camps using cnn—First experiences and lessons learnt. In Proceedings of the ISPRS TC I Mid-term Symposium “Innovative Sensing—From Sensors to Methods and Applications” Conference, Karlsruhe, Germany, 10–12 October 2018.
16. 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.
17. Natarajan, A.; Bharat, K.; Kaustubh, G.R.; Moharir, M.; Srinath, N.; Subramanya, K. An Approach to Real Time Parking Management using Computer Vision. In Proceedings of the 2nd International Conference on Control and Computer Vision, Jeju Island, Korea, 15–18 June 2019; pp. 18–22. [\[CrossRef\]](#)
18. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [\[CrossRef\]](#)
19. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [\[CrossRef\]](#)
20. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Landslide Detection of Hyperspectral Remote Sensing Data Based on Deep Learning with Constrains. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 5047–5060. [\[CrossRef\]](#)
21. 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. [\[CrossRef\]](#)
22. 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. [\[CrossRef\]](#)
23. Zhu, Q.; Chen, L.; Hu, H.; Xu, B.; Zhang, Y.; Li, H. Deep Fusion of Local and Non-Local Features for Precision Landslide Recognition. *arXiv* **2020**, arXiv:2002.08547.
24. 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. [\[CrossRef\]](#)
25. 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. [\[CrossRef\]](#)
26. 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. [\[CrossRef\]](#)
27. Dai, B.; Wang, Y.; Ye, C.; Li, Q.; Yuan, C.; Lu, S.; Li, Y. A Novel Method for Extracting Time Series Information of Deformation Area of A single Landslide Based on Improved U-Net Neural Network. *Front. Earth Sci.* **2021**, *9*, 1139. [\[CrossRef\]](#)
28. Ullo, S.L.; Mohan, A.; Sebastianelli, A.; Ahamed, S.E.; Kumar, B.; Dwivedi, R.; Sinha, G.R. A new mask R-CNN-based method for improved landslide detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 3799–3810. [\[CrossRef\]](#)
29. Liu, P.; Wei, Y.; Wang, Q.; Xie, J.; Chen, Y.; Li, Z.; Zhou, H. A research on landslides automatic extraction model based on the improved mask R-CNN. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 168. [\[CrossRef\]](#)
30. Ghorbanzadeh, O.; Gholamnia, K.; Ghamisi, P. The application of ResU-net and OBIA for landslide detection from multi-temporal sentinel-2 images. In *Big Earth Data*; Taylor & Francis: Abingdon, UK, 2022; pp. 1–26. [\[CrossRef\]](#)
31. Dahmane, M.; Foucher, S.; Beaulieu, M.; Riendeau, F.; Bouroubi, Y.; Benoit, M. Object detection in pleiades images using deep features. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 1552–1555. [\[CrossRef\]](#)
32. Längkvist, M.; Alirezaie, M.; Kiselev, A.; Loutfi, A. Interactive learning with convolutional neural networks for image labeling. In Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), New York, NY, USA, 9–15 July 2016.

33. 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; pp. 234–241. [[CrossRef](#)]
34. 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. [[CrossRef](#)]
35. 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, 21–16 July 2017; pp. 2881–2890.