



Article

Study of the Automatic Recognition of Landslides by Using InSAR Images and the Improved Mask R-CNN Model in the Eastern Tibet Plateau

Yang Liu ^{1,2,3,4}, Xin Yao ^{2,3,4,*}, Zhenkui Gu ^{2,3,4}, Zhenkai Zhou ^{2,3,4,5}, Xinghong Liu ^{2,3,4,6}, Xingming Chen ^{2,3,4,7} and Shangfei Wei ⁸



Citation: Liu, Y.; Yao, X.; Gu, Z.; Zhou, Z.; Liu, X.; Chen, X.; Wei, S. Study of the Automatic Recognition of Landslides by Using InSAR Images and the Improved Mask R-CNN Model in the Eastern Tibet Plateau. *Remote Sens.* **2022**, *14*, 3362. <https://doi.org/10.3390/rs14143362>

Academic Editor: Francesca Ardizzone

Received: 21 May 2022

Accepted: 4 July 2022

Published: 12 July 2022

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

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.



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 Faculty of Engineering, China University of Geosciences, Wuhan 430074, China; liuyang2020@cug.edu.cn

² Institute of Geomechanics, Chinese Academy of Geological Sciences, Beijing 100081, China; guzhenkui15@mails.ucas.ac.cn (Z.G.); zzk114043@cugb.edu.cn (Z.Z.); 2019010059@mail.hfut.edu.cn (X.L.); chenxm@cug.edu.cn (X.C.)

³ Key Laboratory of Active Tectonics and Geological Safety, Ministry of Natural Resources, Beijing 100081, China

⁴ Research Center of Neotectonism and Crustal Stability, China Geological Survey, Beijing 100081, China

⁵ School of Engineering and Technology, China University of Geosciences (Beijing), Beijing 100083, China

⁶ School of Resources and Environmental Engineering, Hefei University of Technology, Hefei 230009, China

⁷ School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China

⁸ Department of Artificial Intelligence, School of Informatics, Xiamen University, Xiamen 361005, China; weishangfei@stu.xmu.edu.cn

* Correspondence: yaixin@mail.cgs.gov.cn

Abstract: The development of landslide hazards is spatially scattered, temporally random, and poorly characterized. Given the advantages of the large spatial scale and high sensitivity of InSAR observations, InSAR is becoming one of the main techniques for active landslide identification. The difficult problem is how to quickly extract landslide information from extensive InSAR image data. Since the instance segmentation model (Mask R-CNN) in deep learning can provide highly robust target recognition, we select the landslide-prone eastern edge of the Tibetan Plateau as a specific test area. Introducing and optimizing this model achieves high-speed and accurate recognition of InSAR observations. First, the InSAR patch landslide instance segmentation dataset (SLD) is established by developing a common object in context (COCO) annotation format conversion code based on InSAR observations. The Mask R-CNN+++ is found by adding three functions of the ResNext module to increase the fineness of the network segmentation results and enhance the noise resistance of the model, the DCB (deformable convolutional block) to improve the feature extraction ability of the network for geometric morphological changes of landslide patches, and an attention mechanism to selectively enhance usefully and suppress features less valuable to the native Mask R-CNN network. The model achieves 92.94% accuracy on the test set, and the active landslide recognition speed based on this model under ordinary computer hardware conditions is 72.3 km²/s. The overall characteristics of the results of this study show that the optimized model effectively enhances the perceptibility of image morphological changes, thereby resulting in smoother recognition boundaries and further improvement of the generalization ability of segmentation detection. This result is expected to serve to identify and monitor active landslides in complex surface conditions on a large spatial scale. Moreover, active landslides of different geometric features, motion patterns, and intensities are expected to be further segmented.

Keywords: landslide; Mask R-CNN; InSAR; deep learning; instance segmentation

1. Introduction

Due to the rapid expansion of modern human activities and the increase in extreme meteorological events because of global climate change, the occurrence of landslides has also

increased significantly [1]. Recognizing and monitoring active landslides are increasing in importance, and traditional field surveys [2,3], optical remote sensing [4,5], and techniques that rely on global navigation satellite systems (GNSSs) [6] have difficulty meeting practical needs. Interferometric synthetic aperture radar (InSAR), with its all-weather, all-day, short-period, and high-precision observation performance [7,8], is rapidly developing in the application of surface observation and geohazard potential hazard recognition. Fruneau [9] first used this technique to accurately identify landslides in Italy in 1996. In recent years, with the increase in the number of radar data (including Sentinel-1A data in the C-band of ESA [10], COSMO-SkyMed data in the X-band of Italy [11], ALOS-2 in the L-band of Japan [12], and HJ-1-C data in the S-band of China [13]), InSAR observation technology in geological hazard applications has been promoted [14–20].

However, the InSAR interpretation process depends highly on experience and has a long decoding period. However, the increasing number of DL (deep learning) methods has shown excellent accuracy in medical image processing, text analysis, and other fields. Due to its data-driven approach, strong learning ability, and good adaptability, InSAR is used in automatic recognition tasks in several fields. DL has also been applied in geology, such as in the study of surface deformation caused by volcanic activity. For example, Anantrasirichai et al. [21,22] used a convolutional neural network (CNN) trained with dichotomous (i.e., “background” and “volcanic”) markers from InSAR observations to quickly identify ground deformations caused by volcanic activity; Valade et al. [23] used a CNN to learn interferogram features to detect intense deformations in real interferograms and demonstrated the CNN’s accuracy through a series of recent volcanic eruptions; Matthew et al. [24] developed a “two-headed model” that can locate and classify volcanic deformations in a single interferogram; and Clayton et al. [25] used CNN networks to classify surface deformations in synthetic interferograms of volcanoes and used CAMs (class activation maps) to show the location of surface deformations. Regarding landslide identification based on InSAR technology, KJ et al. [26] used the AlexNet model to classify the interferometric stripes of landslide motion and then predict the landslide boundaries. Chen et al. [27] proposed the DRs-UNet model to achieve semantic segmentation of potentially active landslides in InSAR images. These models provide a new idea of how to monitor landslides. Namely, since the deformation rate of active landslides is shown as a colour change in the InSAR monitoring result map [28], identifying landslides can be transformed into an image recognition problem. However, there are some challenging problems: first, compared with volcanic deformations, which are large and concentrated in location, common landslides are much smaller, much less intense, and highly dispersed in space, and both the spatial heterogeneity and complexity of geomorphic evolution increase the difficulty of identifying landslide deformations; second, when deep learning is used for image segmentation, not only the primary location of the landslide but also the spatial extent of the landslide’s abnormal deformation must be identified to serve as the necessary reference information for practical management.

In traditional digital image processing, thresholding segmentation [29], edge detection [30], region segmentation [31,32], and clustering [32] rely mainly on manually designed extractors to solve target segmentation tasks with poor generalization ability and robustness [33]. However, various deep learning models (including the Faster R-CNN [34] algorithm) with high robustness and good generalization ability propose using an RPN (region proposal network) instead of selective search methods for region selection, thereby enabling end-to-end target detection. However, the ROI (region of interest) pooling layer uses two quantization rounding operations, thereby leading to errors in the detection frame localization, thus bringing the problem of spatial location misalignment, and cannot achieve the target segmentation task. To achieve a high-quality pixel segmentation task, the Mask R-CNN [35] model solves the spatial pixel misalignment problem produced by the use of ROI pooling layer quantization in the Faster R-CNN algorithm by using ROIAlign pooling to achieve an accurate one-to-one correspondence between pixel points by preserving fractional information. In deep learning, the image instance segmentation model Mask

R-CNN [35] in deep learning [36] combines target detection and segmentation tasks in one network model by segmenting the target pixels within the detection frame while locating the target location. Therefore, this model is used as the basis for further improving the characteristics of InSAR patches to automatically recognize landslide patches in InSAR observation results.

This study selects part of the eastern edge of the Tibetan Plateau, which is located in landslide-prone areas, as a specific test area. This study takes InSAR observations as the analysis object based on the instance segmentation model in deep learning. (1) This study realizes the conversion code to automatically generate a standard COCO annotation format from InSAR images and corresponding vector files and establishes the SLD (InSAR Landslide Dataset); and (2) using the Mask R-CNN network as the base model, an instance segmentation model (Mask R-CNN+++) for InSAR landslide patches is established by replacing the convolutional blocks of the feature extraction network and adding an attention mechanism to the feature pyramid network to automatically recognize active landslides based on InSAR result maps. In conclusion, an automatic active landslide recognition method based on deep learning and InSAR observation results is established. A format conversion tool is developed using ArcPy [37] to form a complete image recognition and processing process that reduces unnecessary human intervention and thus improve work efficiency and facilitate large-scale and short-period recognition tasks.

2. Research Area

The study area (Figure 1) is located on the eastern edge of the Tibetan Plateau in the middle section of the Jinsha, Lancang, and Nujiang river basins, one of China's most developed landslide hazard areas [38]. The area was influenced by the collision between the Indian and Eurasian plates and experienced tectonic deformation stages, such as right-slip compressional torsion, large-scale slip extrusion, and left-slip tensional torsion during the Cenozoic, with complex tectonic activity. Since the Quaternary, the tectonic activity in the region has remained robust [39,40], and seismic activity has been frequent [41]. The strong and continuous tectonic activity and runoff erosion have led to the overall topography of the study area, which is characterized by high northwest, low southeast, and dense canyons [42]. The average elevation ranges from 4200 m elevation in the north to approximately 1800 m elevations in the south [43], with a maximum elevation difference of more than 3000 m per square kilometre. This topographic feature also importantly influences the regional climate distribution pattern, in which the plateau area has a subcold semihumid plateau climate with active freeze–thaw phenomena, permafrost, and seasonal permafrost development; the high mountain valley area is influenced by the near north–south trending mountains and thus atmospheric circulation, with significant temperature differences between day and night and reduced precipitation, thus manifesting as hot in river valley areas and cold in high mountain areas. This highly undulating terrain, variable climate, and tectonic movements lead to vigorous internal and external dynamic effects, a fragile geological environment, and frequent disasters in the area, thus posing severe threats to engineering construction and human safety.

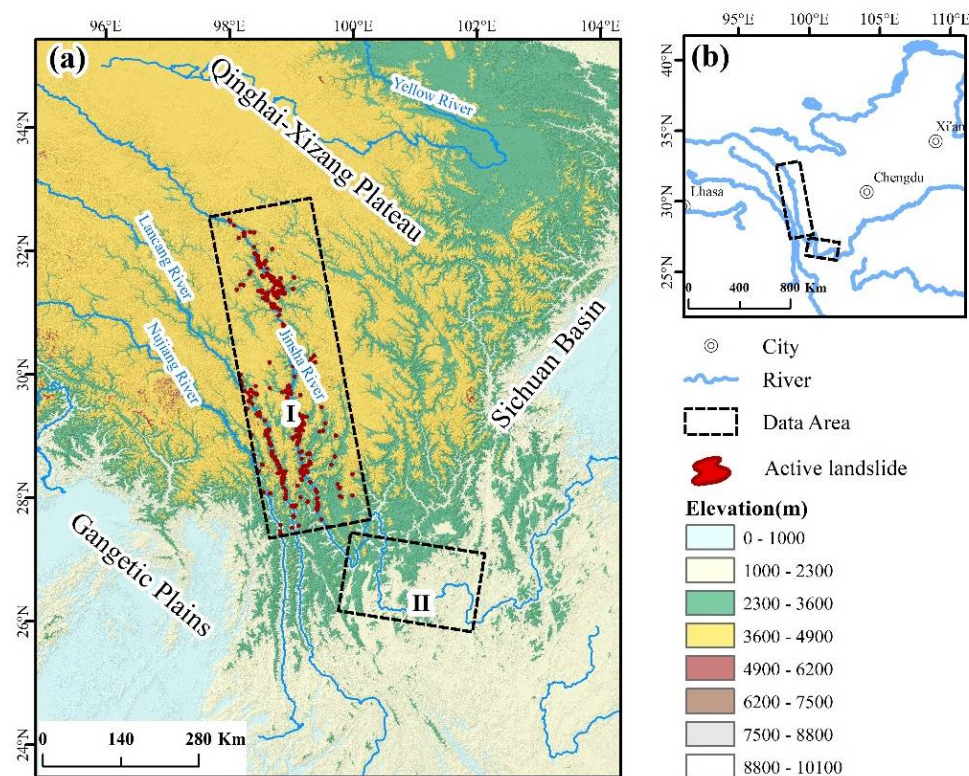


Figure 1. Location map of the study area. (a) Digital elevation model (I: the base data of the dataset, II: the base area to be identified). (b) Provincial administrative boundaries.

3. Data and Methods

3.1. InSAR Data Processing

The deep learning approach carried out in this study is based on InSAR observations. The InSAR observations are based on data from Sentinel-1A (<https://scihub.copernicus.eu>, accessed on 10 September 2021), which has a C-band wavelength of 5.6 cm. The observations cover IW (wide interferometric) mode radar images with 30 scenes of ascending and descending orbit images. The period is from January 2019 to March 2020, the polarisation mode is VV, the average incidence angle of the images is 42.5 degrees, the spatial resolution is $25\text{ m} \times 25\text{ m}$, and the azimuthal resolution is 13.8–13.9 m. The distance direction resolution is 2.3–2.4 m.

Sentinel-1A images must be processed before the start of the study process. The general processing flow of D-InSAR was used for data processing. The SAR data with short temporal and spatial baselines were first selected for D-InSAR calculation [44] to obtain high-quality relative interferometric data, followed by laminar atmosphere removal with a positive correlation to elevation [45]. Then, filtering [46], superposition enhancement, random phase error removal, and deformation phase resonance enhancement were performed sequentially. The surface deformation map of the study area is obtained and used as the base data for deep learning and the validation data with a resolution of 25 m. Finally, all InSAR observations of the region (II in Figure 1a) with different topographic conditions from the SLD are selected as the images identified from January 2019–March 2020 to further validate the Mask R-CNN+++ generalization capability.

3.2. Automatic Recognition Solutions

This study needs to interpret a certain number of active landslides from InSAR observations as a dataset, which is divided into three parts: a training set, validation set, and test set. This process uses the standard COCO [47] (common object in context) annotation format. During the implementation, different optimization methods are overlaid on the

original Mask R-CNN model, and the network is trained by migration learning and data augmentation. Then, the different optimization methods are evaluated from multiple perspectives based on evaluation metrics and expert assessments in computer vision, and iterative improvements are made to determine the most suitable model structure and values of hyperparameters for the optimization methods. Finally, the optimized model establishes the recognition process and completes the format conversion for the whole InSAR observation result map. An active landslide is identified in the study area. The specific research flow chart is shown below (Figure 2).

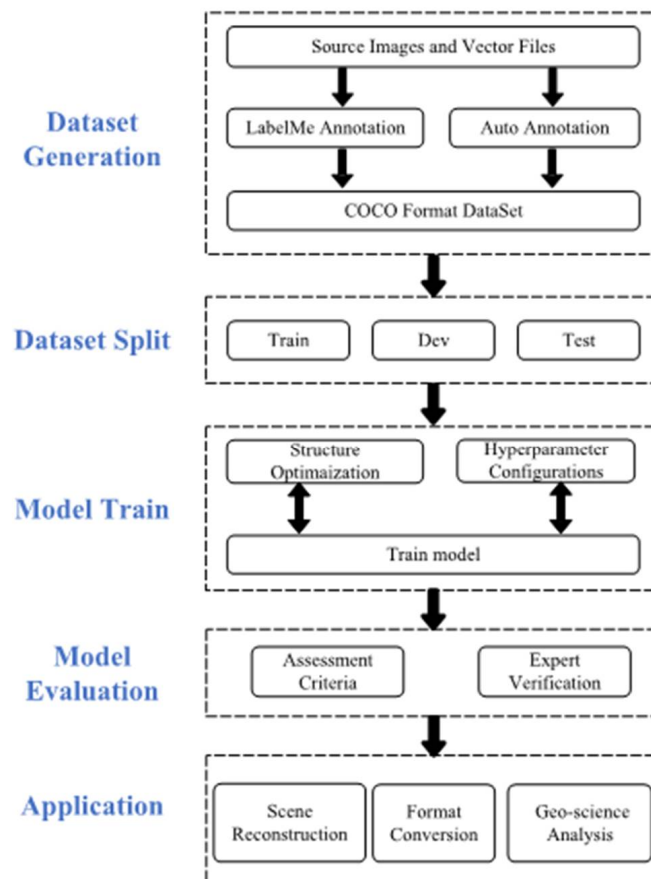


Figure 2. Methodological flowchart of deep learning instance segmentation.

3.2.1. Construction and Partitioning of the Dataset

The construction of a dataset is the basis of deep learning; the dataset construction is divided into two steps: data collection and data annotation. In ground deformation detection by using InSAR observations, a certain amount of landslide location and boundary information needs to be obtained in advance by manual decoding as the target learning object, and this process requires data annotation. The COCO annotation format followed by most instance segmentation algorithms is followed in the data annotation process. Therefore, this paper develops a method to generate a COCO annotation format from vector files and raw images automatically transformed for sample acquisition (constructing datasets); this method includes three steps: data preparation, format transformation, and data generation. That is, the image files (such as Image and Mask) are generated by aligning and cropping the images based on reading the original images and vector files; then, the storage paths and files (JSON) are generated by reading the image sample geometry information, label information, and category information; and finally, the files are converted into a dataset in COCO annotation format. The detailed steps are shown in Figure 3.

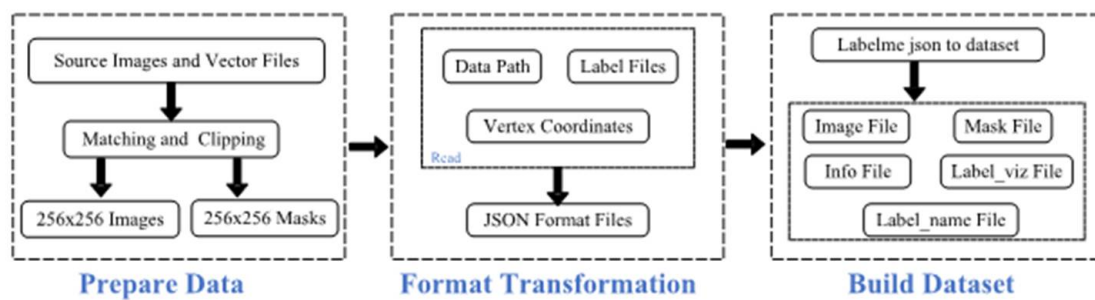


Figure 3. Flowchart to obtain samples for training and creating a Common Object in Context (COCO) annotation format.

The InSAR observations of the study area and the 421 active landslide boundary vectors obtained through decomposition are used to finally form a landslide dataset (SLD) with a sample size of 636 and a resolution of 30 m. A total of 380 images are randomly selected for training (60%), 128 images for testing (20%), and 128 images for validation (20%) according to the typical method of dividing the training, validation, and test segmentation datasets. In this case, the image size is 256×256 pixels, the input shape is $256 \times 256 \times 3$, and the segmentation process is carried out with the minimization of edge effects and computational power methods. In addition, the online augmentation method was applied to further increase the number of training datasets by performing operations such as flipping and brightening the datasets when training the model.

3.2.2. Model Description

The Mask R-CNN model is based on the original target detection model (Faster R-CNN) classification and local frame regression branch, with a mask branch, which can detect objects and segment instances [48–50]. The model first passes an image through the backbone network and then extracts feature maps C2, C3, C4, and C5 at different resolutions in different stages to form a “feature pyramid network” (FPN). C2, C3, C4, and C5 contain the feature information in bottom-up order, from high to low level. P2, P3, P4, P5, and P6 are obtained by the FPN structure, which can perform multiscale feature fusion and improve the scale robustness of the model. The model performs binary classification (foreground and background) and regression based on the anchor points generated by the RPN to filter out some suggestions. Then, the ROI is changed to a fixed size of 7×7 or 14×14 pixels by ROIAlign. Finally, the ROI is fed into the fully connected layer and FCN for classification, regression, and segmentation tasks. The above model uses ROIAlign instead of ROI pooling in Faster-RCNN and combines the residual network with a feature pyramid network (FPN) for feature extraction of images, enabling the network to segment the targets with high quality while detecting them. Although the model has achieved excellent results, it needs to be optimized for specific data in specific scenarios to meet the needs of different tasks.

In this paper, to use the model to better identify active landslides, the bottom ResNet block in the original feature extraction network is replaced with a ResNext block, which can increase the fineness of the network segmentation results and enhance the noise resistance of the model without a significant increase in computational effort. Second, this paper replaces a modulated deformable convolution in the higher levels of the feature extraction network to capture the implicit higher levels in the InSAR observations and improve the feature extraction capability of the network for geometrically changing objects. Finally, an attention mechanism is introduced in the feature pyramid construction to scale the different channels to enhance usefully selectively and suppress less valuable features. The attention mechanism models the pixel-level dense context-aware relationships by recalibrating the channel dependencies according to the global context to achieve advanced feature enhancement. The specific network structure diagram is shown below (Figure 4).

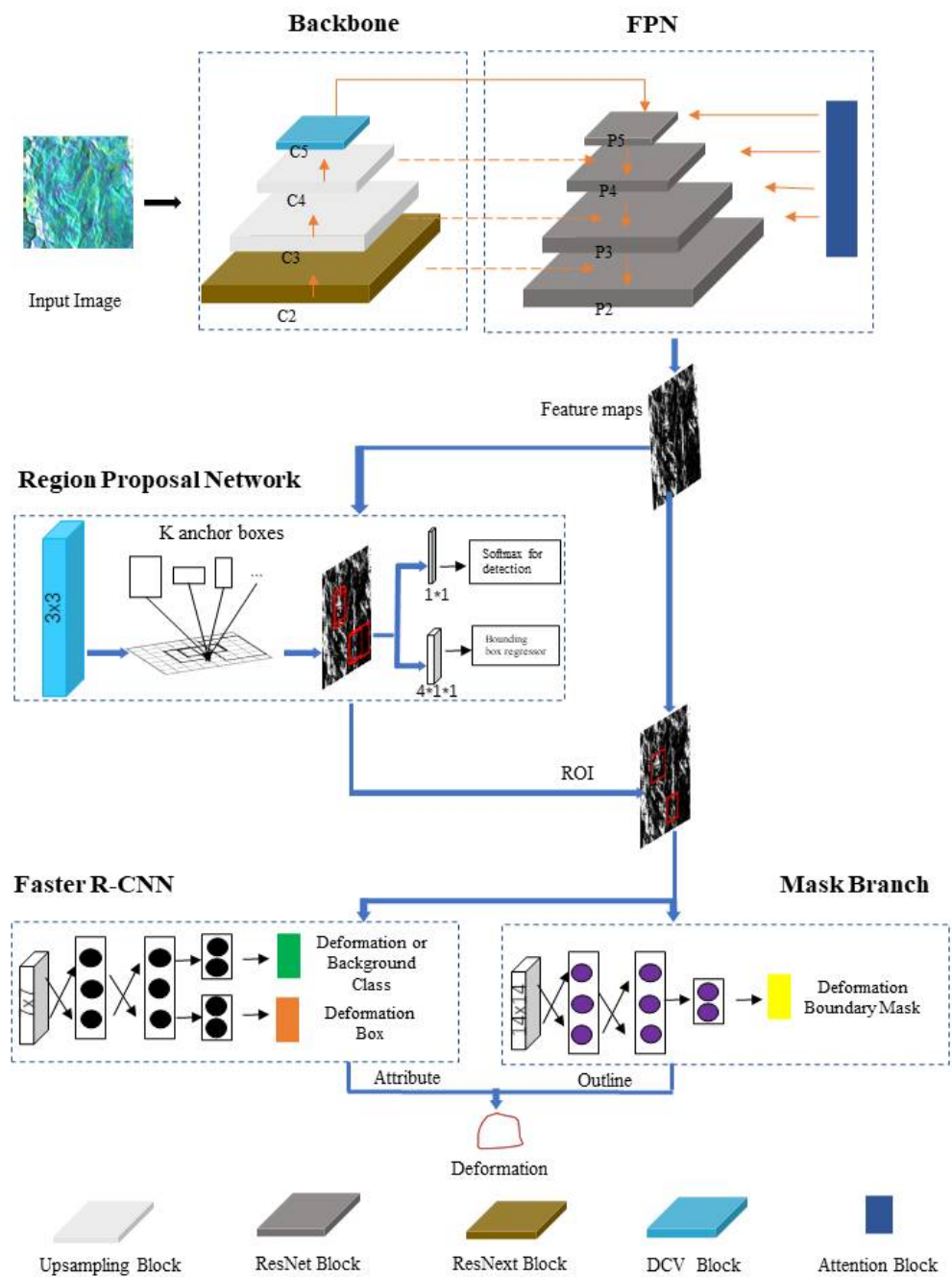


Figure 4. The overall architecture of Mask R-CNN+++. The backbone shows the modified architecture with the attention feature pyramid network and the predictions (three tasks).

1. ResNext Convolution Block

Deep neural networks have a robust feature extraction capability, but the depth of the network is not as deep as possible. The increase in network layers in the model may be accompanied by degradation, so more feature information cannot be obtained by directly increasing the number of network layers. The ResNext convolutional block [51] is an improvement on Inception and ResNet. The internal structure of the ResNext convolutional block is shown in Figure 5.

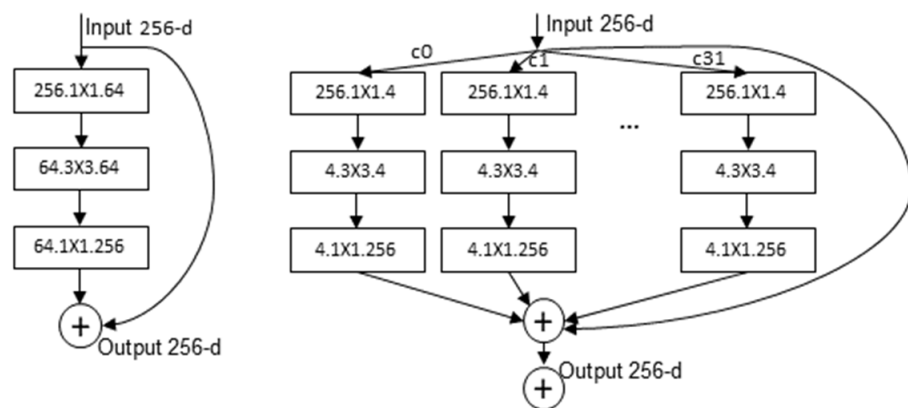


Figure 5. Left: A block of ResNet. Right: A block of ResNext. Layer is shown (as in channels, filter size, out channels). 256-d indicates that the dimension is 256.1×1 indicates that the convolution kernel size is 1×1 . The plus sign (+) indicates the corresponding addition of numbers.

The module is based on the “split-transform-merge” model of inception, where the 256 input channels are first split into 32 input channels. Then, each group is subjected to the same convolutional operation, and finally, the results of all the groups are fused with the original input. Second, this study inherits the repetitive layer strategy of ResNet. However, the difference is that the number of paths increases, and the same topology is used to form the ResNext module group convolution on each path. This unique structure allows the residual network (ResNext) to improve accuracy without increasing parameter complexity, while the same topology reduces the number of hyperparameters. Since Voulodimos et al. [52] verified ResNext on ImageNet [53], the top-5 error was reduced by 0.62% from the 50-layer residual network to the 101-layer residual network. However, the 152-layer residual network to the 101-layer residual network was reduced by only 0.11%, and the overall time and computational effort increased significantly. The ResNext module is selected to replace the underlying ResNet module in this paper. The purpose is to increase the fineness of the network segmentation results and enhance the noise resistance of the model.

2. Deformable Convolution Block

Due to the high variability of surface undulations, images captured by satellites often exhibit significant variations in geometric features. However, geometric variations do not affect the DCB (deformable convolutional block) [54]. In the deformable convolutional layer, an additional two-dimensional offset is added to the regular grid sampling locations in the standard convolution. See the schematic DCB structure below (Figure 6). For example, given a three $\times$ three kernel with an extension of 1, the receiver domain size and extension of the standard convolutional grid $\mathcal{R}$ can then be expressed as:

$$\mathcal{R} = \left\{ \begin{array}{ccc} (-1, -1) & (-1, 0) & (-1, 1) \\ (0, -1) & (0, 0) & (0, 1) \\ (1, -1) & (1, 0) & (1, 1) \end{array} \right\} \quad (1)$$

Thus, for every outcome y , we have:

$$y(p_0) = \sum_{p_n \in \mathcal{R}} w(p_n) \cdot x(p_0 + p_n) \quad (2)$$

where x represents the input feature map, w represents the weights of the sampled value, and p_n enumerates the locations. Although in deformable convolution, the regular grid

$\Re$ is augmented with offsets $N = |\Re|$. Therefore, the deformable convolution can be expressed as:

$$y(p_0) = \sum_{p_n \in \Re} w(p_n) \cdot x(p_0 + p_n + \Delta p_n) \quad (3)$$

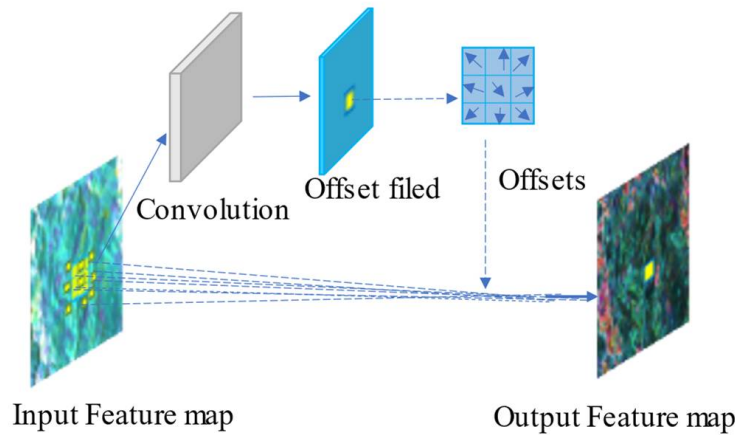


Figure 6. Illustration of a 3×3 deformable convolution. The offset field comes from the input feature map and has the same spatial resolution as the input.

Now, the free deformation is described by the irregular offset positions $p_n + \Delta p_n$. The offsets are learned from the previous feature maps by using additional convolutional layers in parallel. The free deformation has a $2N$ channel dimension that corresponds to N 2-D offsets. As the offset is usually a decimal, bilinear interpolation is introduced to revise the value of the sampled points after migration. Therefore, to improve the feature extraction capability of the backbone network for deformation change objects, this paper adopts the deformable convolution that can effectively simulate the geometric change of the target.

3. Attentional Mechanisms

Landslides in InSAR observations often exhibit different characteristics at different pyramid levels. The attention mechanism is called to shift attention to the most critical regions of an image and ignore irrelevant parts [55], thus allowing capturing critical information from complex graphical features by further weakening the requirement for training sets to construct semantic associations of individual pixel points in an image. The introduction of attention mechanisms in constructing feature pyramids can solve the problem of feature layer imbalance for different sizes of landslides. The attention mechanisms introduced in this study include the convolutional block attention module and self-attention module.

The convolutional block attention module (CBAM) [56] concatenates channel attention and spatial attention. See the schematic CBAM structure below (Figure 7). The algorithm decouples the channel attention map from the spatial attention map to improve computational efficiency and exploits the global spatial information by introducing global pooling. The CBAM has two sequential submodules, namely, channel and spatial. Given an input feature mapping $X \in \mathbb{R}^{C \times H \times W}$, the CBAM sequentially derives a one-dimensional channel attention vector $sc \in \mathbb{R}^C$ and a two-dimensional spatial attention mapping $ss \in \mathbb{R}^{H \times W}$. The CBAM combines channel attention and spatial attention sequences and uses the spatial and cross-channel relationships of features to tell the network what and where to pay attention. More specifically, the CBAM emphasizes proper channels and reinforces local areas of information.

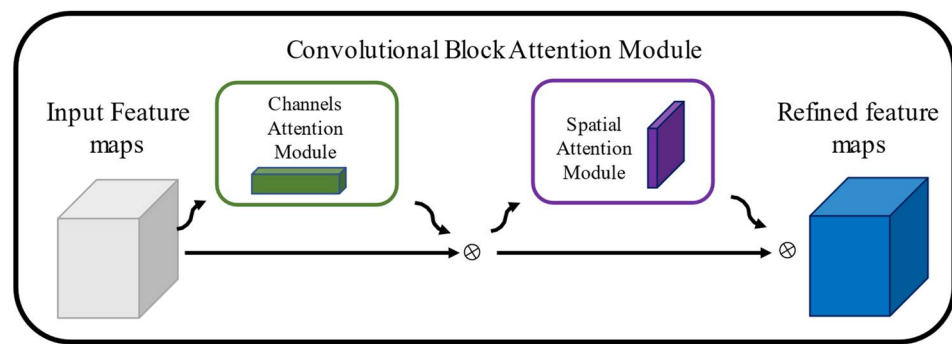


Figure 7. Overview of the CBAM. The module has two sequential submodules: channel and spatial. Different colors and shapes represent different convolution blocks.

The self-attention (SA) module [57] is a particular case of the attention mechanism that differs from the standard one. See the schematic SA structure below (Figure 8). The SA module aims to select from the global information the information that is more critical to the current task goal. Thus, the full features of the image can be better utilized. The module first performs a linear transformation and channel compression on the convolutional feature mapping by using two 1×1 convolutions. Then, the module converts the two tensors into matrix form, transposes and multiplies them, and then obtains the attention mapping by using softmax.

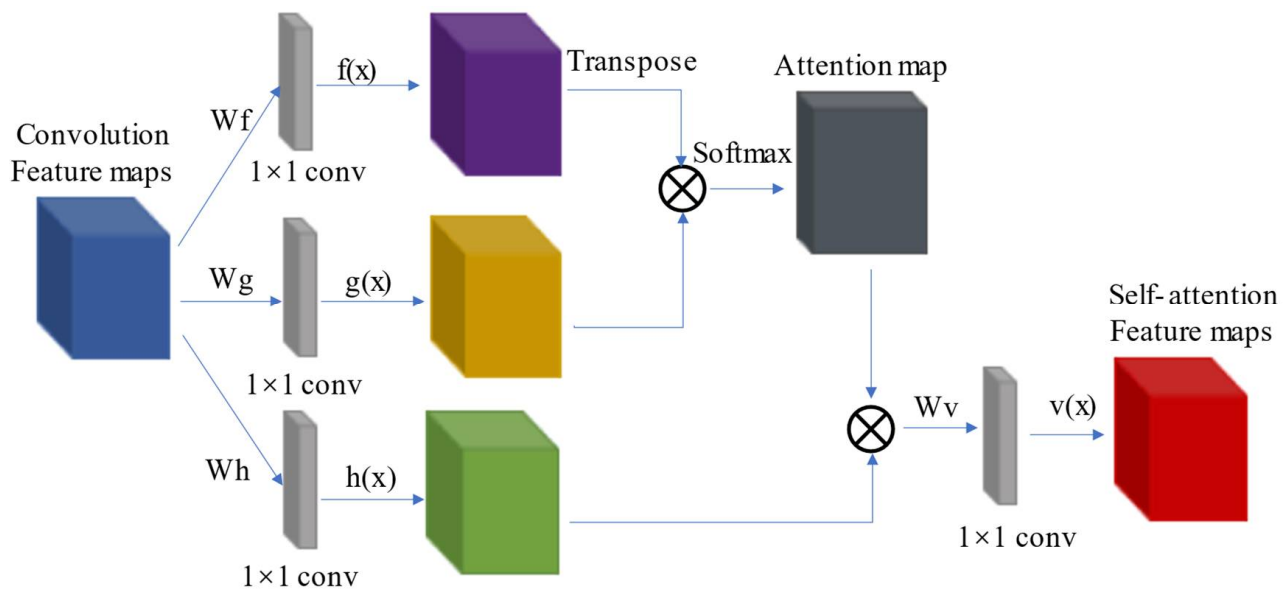


Figure 8. Self-attention module structure diagram. Different colors and shapes represent different convolution blocks.

Additionally, the original feature mapping is linearly transformed by using 1×1 convolution and then multiplied by the previously obtained attention mapping matrix, which is summed to obtain the self-attentive feature mapping. Finally, the self-attentive feature mapping and the original convolutional feature mapping are weighted and summed as the final output. Self-attentive feature mapping can be regarded as the product of feature mapping and its transposition. This operation can enhance the association between distant features in the image. The dependency between any two-pixel points can be learned, and then global features can be obtained.

3.3. Experiments

3.3.1. Model Training

The model built in this paper is implemented based on the deep learning frameworks TensorFlow and Keras. The experiments run on an NVIDIA GeForce GTX 2080Ti GPU with 16 GB of RAM and a dual-core Intel (R) Xeon (R) CPU E5-2637 computer on Windows OS. To enhance the generalization ability and robustness of the model, the following data enhancement and migration learning techniques are used in this paper: (1) Gaussian noise addition and sharpening in the images, (2) horizontal and vertical image flipping, and (3) generic initial parameters trained on the COCO dataset. Four hundred iterations were trained for the SLD dataset, with an initial learning rate of 0.01 for the first 200 iterations of the training network and a learning rate of 0.001 for 200 iterations of all networks. Finally, the completed training model was applied to the corresponding InSAR observations in the study area.

3.3.2. Evaluation Indicators

To quantitatively evaluate the recognition performance of network structures with different optimization approaches, this paper uses accuracy, the F1-score, and the mean cross-merge (mIoU) as quantitative metrics to evaluate the recognition results. That is:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (4)$$

$$F1 - score = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (5)$$

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

In Equations (4)–(6), TP denotes the number of correctly extracted landslides, FP denotes the number of incorrectly extracted landslides, and FN denotes the number of incorrectly extracted landslides. TN denotes the number of incorrectly extracted nonlandslides. The Index parameter's relation is shown in Table 1.

Table 1. Index parameter.

Predicted Results \ Real Results	Landslides	Others
	True Positive (TP)	False Positive (FP)
Landslides	True Positive (TP)	False Positive (FP)
Others	False Negative (FN)	True Negative (TN)

4. Results

4.1. Indicator Results

Since the model improvement is optimized for the feature extraction method, we use the feature extraction network ResNet101 + FPN as the basis, and the feature extraction methods of the ResNext module, DCB module, and attention mechanism (CBAM and SA) are used for experimental comparison. The results of their experimental parameters are shown in Table 2. By comparison, replacing only the bottom layer of the network with ResNext, that is, the original single-way convolution is transformed into a multi-way convolution for prediction, all three indices are improved; compared to adding only ResNext, replacing DCB with the extracted deformed object features at the top layer of the network improves the mIoU value by 2.17%; introducing the SA module in the FPN improves the F1-score value. However, the CBAM module combines spatial attention with channel attention, thereby improving some parameters, while mIoU decreases by 0.77% when the attention module is added. Therefore, the SA module is undoubtedly the best choice among the two modules by learning the dependencies between any two-pixel points and thus obtaining the global features. The results show that introducing the DCB and

SA modules in the base network for feature extraction is the optimal solution. In terms of model performance, the addition of the ResNext module can increase the perceptual field of the network, thereby significantly improving the generalization ability of the model segmentation detection.

Table 2. Metrics precision.

Method	Accuracy	F1-Score	mIoU
Mask R-CNN (Baseline)	89.19	81.24	84.38
Mask R-CNN-ResNext	90.44	81.63	85.02
Mask R-CNN-ResNext-DCB	91.37	82.58	87.19
Mask R-CNN-ResNext-DCB +CBAM	92.86	84.06	86.42
Mask R-CNN-ResNext-DCB +SA	92.94	84.12	90.26

4.2. Recognition Results

4.2.1. Test Set Recognition Results

The native model focuses only on the accuracy of the area and not much on the variation in landslide geometry; consequently, the detection results may portray the active landslide morphology inaccurately. Therefore, after adding ResNext, DCB, and attention mechanisms (CBAM and SA) to the model, the model becomes more suitable for capturing the specific morphological features of landslides in InSAR observation maps. The typical landslides identified with the help of the optimized model in this paper are shown in Figure 9, which reflects the recognition effect of Mask R-CNN+++ on different types of landslides under different combination conditions. The detection results of Mask R-CNN+++ proposed in this paper are accurate in outlining landslide boundaries and can better extract InSAR landslide patches, except for some more minor landslides that will be missed.

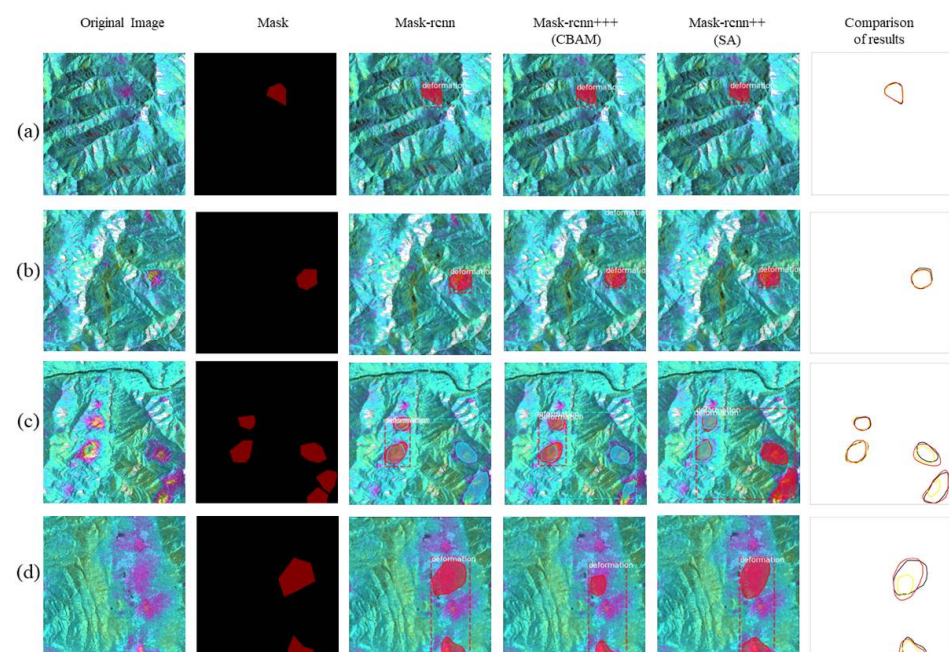


Figure 9. Comparison results of different combination models: (a–d) are cropped images; (a,b) are single targets in different terrains; and (c,d) are multiple targets in different terrains. In the column “Comparison of results”, black, yellow, and red represent the recognition results of Mask R-CNN, Mask R-CNN+++ (CBAM), and Mask R-CNN+++ (SA), respectively.

4.2.2. Application Recognition Results

This paper uses the established Mask R-CNN++ to identify the entire study area's InSAR images. The overall number of landslide deformations is finally predicted to be 891. The results are shown in Figure 10 below. The recognition results show that most of the active landslides around the rivers have been successfully detected, especially in the landslide-prone high mountain canyon areas, thus showing that the method proposed in this paper is very effective. The picture below shows the recognition results of the whole image, and the identified landslides are indicated in yellow, red, and black.

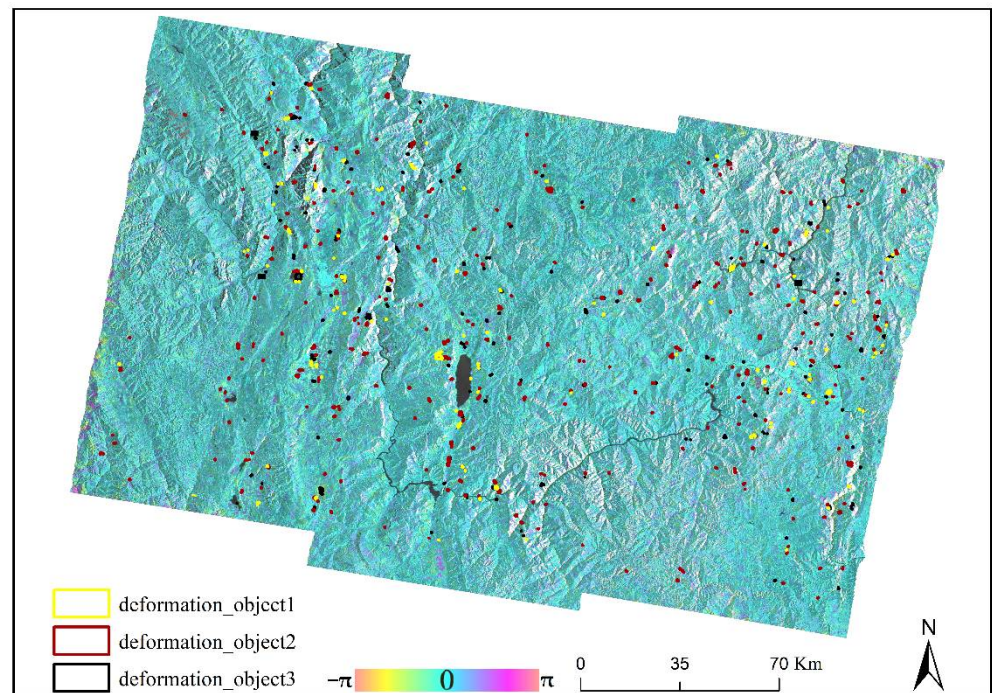


Figure 10. This figure shows the study area image obtained by using the proposed method. Red, yellow, and black represent the different objects of the same deformation class.

4.2.3. Identify the Preliminary Results of the Classification

The model used in this paper can perform the instance segmentation task, i.e., classifying pixels by using semantic segmentation and distinguishing different object instances. Mask R-CNN adds the mask branch for pixel segmentation to Faster-CNN. A fully convolutional network structure for predicting segmentation masks is introduced behind the ROIAlign layer and applied to a single ROI to predict the segmentation masks in pixel-to-pixel behaviour and decide the size and class of masks to use in combination with the predicted classes in the Faster branch. In this way, the model can segment different targets in the same class. With its advantage of surface coverage, InSAR deformation observation can observe various types of surface deformation. Different deformation features will show different InSAR patches, distinguished by colour, shape, intensity, texture, spatial structure combination, and geomorphic location. Therefore, in this paper, the deformation areas automatically identified by the above model are transformed into shapefiles and statistically analysed in terms of number, perimeter, area, and patch brightness and further divided into deformation characteristics of different target types in the same category by combining existing experience.

In terms of quantity, the number decreases sequentially from object 1 and object 2 to object 3; accordingly, the sum of the perimeter and area of each target type decreases, but each target type corresponds to a similar average value. Using the geometric morphological perspective, the perimeter divided by the area of the lowest target type is object 2, thus reflecting that mainly elongated patches are dominant. Additionally, the highest target

type is object 3, which is dominated by small circular patches, and object 1 is in the middle. The results are shown in Figure 11 below.

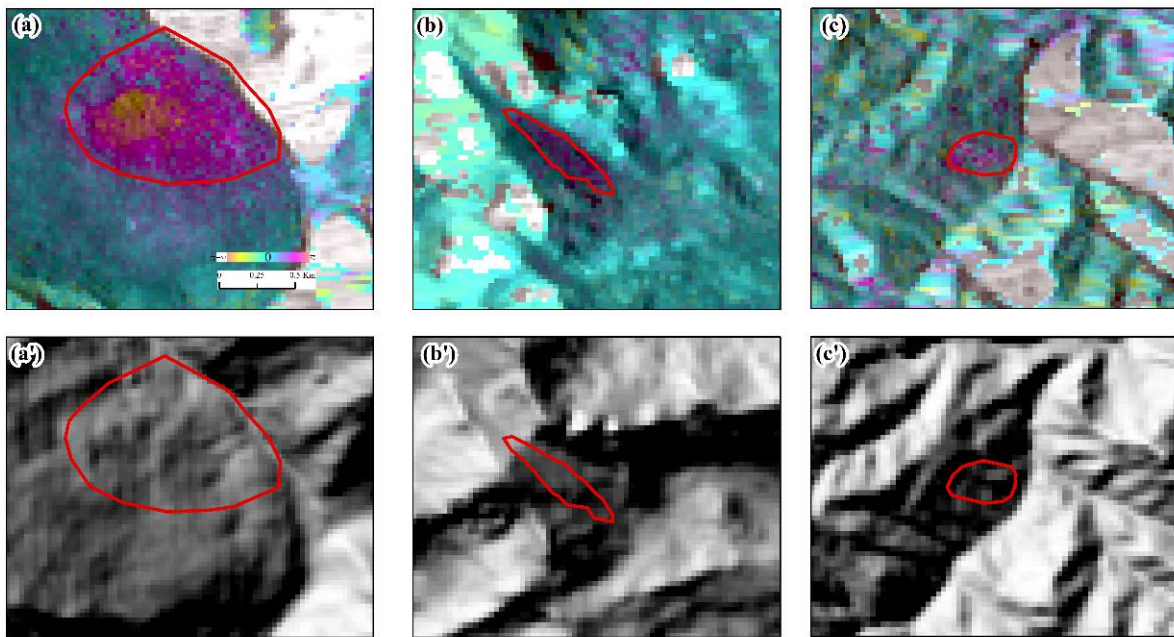


Figure 11. The above images (a–c) and (a'–c') show different patterns of plaques and landforms.

The information on surface change is directly expressed as the bands of InSAR patches into periodicity, and the patch brightness can be used to indicate the intensity of change. In terms of patch brightness, object 1 shows a higher brightness and a near-circular chair-like shape, while object 2 and object 3 have lower brightness and similar intensity. The results are shown in Figure 12 below.

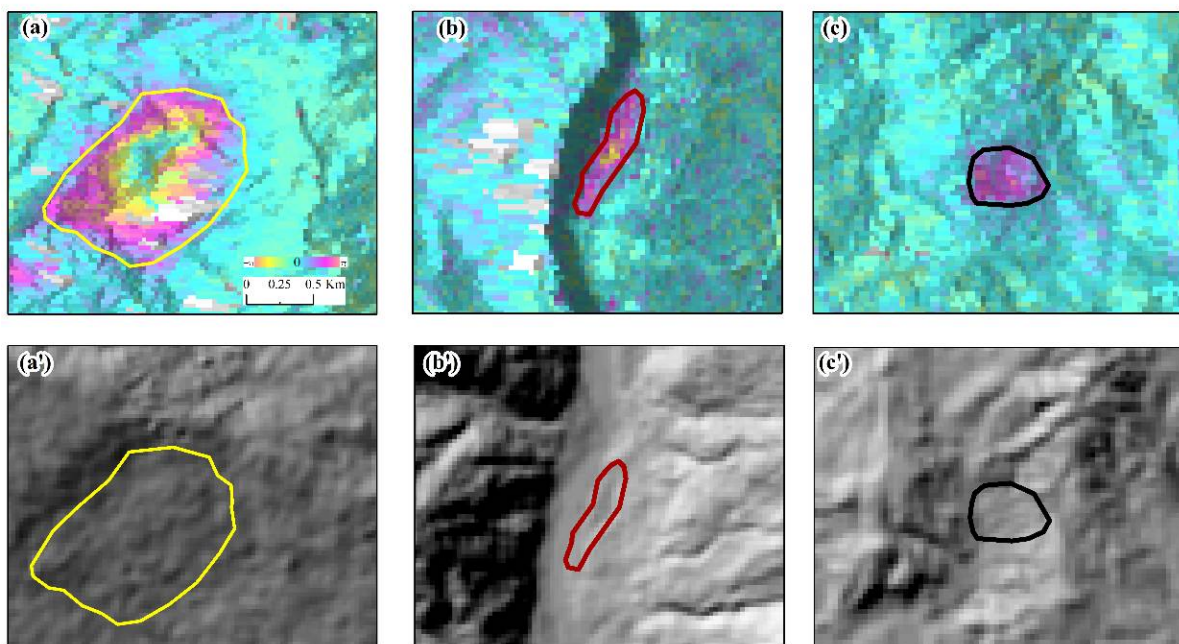


Figure 12. The above images (a–c) and (a'–c') show patches of different brightnesses and landforms.

In summary, from the perspective of landslide type classification, object 1 is a strongly deformed landslide, object 2 is a collapse-debris flow type, and object 3 is a small landslide with a relatively slow deformation rate. Object 1 is brighter in colour, thus reflecting multiple winding cycles, indicating that this object is the most robust deformation among the three. Objects 2 and 3 are darker and single in colour, thus reflecting that the deformation is relatively small in a single 2π cycle. Object 1 is mainly in topographic conditions with a considerable height and steep slope. The soft surface on its slope is in the prolapse state. The upper geotechnical body of the slope is in an unstable state, thus providing the driving force for the deformation destabilization of the landslide. Object 1 shows clear deformation patches of landslide form, thus reflecting obvious overall deformation and clear deformation boundaries. Therefore, this type is dominated by strongly deformed landslide accumulation deformation, with few newly developed bedrock deformations. Object 2 appears mostly in the form of debris flows and riverbank collapses. Debris flow is a strong fragmentation of the sloping rock in the moving process after the destabilization of the landslide source area and gradually disintegrates into debris particles whose particle size range usually spans several orders of magnitude and spreads in the form of fluid transport over long distances and large areas. Under hydrodynamic action, riverbank collapse impacts the soil particles on the bank slope. The cohesive material in the soil body is detached from the soil particles, followed by fissures, with the expansion of fissures leading to the dispersion of aggregates and disintegration. Concerning Object 3, a small landslide with low-velocity deformation is the main feature of this type. Most deformations are distributed as independent individuals, irregularly rounded or oval, generally from several tens to one hundred metres in diameter. The edges of these deformations are visible in the image. The morphology in the deformation patch is subcircular, and the deformation boundary is clear.

5. Discussion

5.1. Advantages of the Present Model in the Recognition Process

This paper uses a deep learning instance segmentation model that uses InSAR observations to identify the location of active landslides and their boundaries over a large area with high dispersion. In 2017, He et al. [35] proposed Mask R-CNN, an instance segmentation model in the Faster-RCNN framework based on target detection; Mask R-CNN can perform target classification, target detection, semantic segmentation, instance segmentation, and many other tasks. Compared with Mask R-CNN, Mask R-CNN+++ (ResNext-DCB + CBAM\SA), which is based on the Mask R-CNN proposed in this paper, improves accuracy, the F1-score, and mIoU by up to 3%. In addition, the landslide edges extracted by the optimized model are smoother and more accurate in scope. The overall time taken for the recognition process of the whole image of the test area ($5.32 \times 10^4 \text{ km}^2$) for the whole study area is 736 s, which is also more significant than the manual time efficiency. The improvement of the Mask R-CNN model can serve for active landslide investigation in complex geological environments and areas with a wide distribution, large scale and high risk of significant landslide hazards. This study finds that the attention mechanism, the addition of variable convolution, and the optimal input scale can effectively improve deep learning accuracy, but there is also room for improvement. First, although adding the attention mechanism improves the overall model accuracy to some extent, the increase in computational burden cannot be ignored, especially in the case of the self-attention mechanism, which has significantly improved the recognition effect. Second, including deformable convolution at the higher level of the feature extraction network allows adaptive changes by changing the sampling position of the input, i.e., the function of making adaptive changes, such as object magnification or selection by enhancing the sensory domain.

5.2. Impact of Instance Segmentation Model Optimization on the Recognition Process

Landslides develop at different scales, and small landslides are often more numerous, resulting in many small targets in the images. Affected by the resolution of remote sensing images, the deformation characteristics of such landslides are frequently blurred. Therefore, the labelling process in the premodelling stage will cause some information loss and lead to some offset in the boundary recognition in the detection process. Sun et al. [58] added deformable convolution at the top level of the instance segmentation model to enhance the recognition of the changing geometry of the target, but the backbone network used ResNet50 as the feature extraction network. Therefore, this paper replaces the top convolution of the feature extraction network with a deformable convolutional block that adaptively captures saliency according to the morphological features of the landslide. Additionally, this paper selects the backbone network as ResNet101 with a more robust extraction capability and replaces the bottom convolutional layer with ResNext to avoid information loss during sampling and increase the range of the perceptual field view. The comparison experiments show that replacing convolutional blocks in the network improves the F1-score value in the evaluation index by nearly 4%, thereby indicating that the difference between improving precision and recall is as small as possible. In the actual recognition process, the background information of remote sensing images can improve the model generalization ability, but this information also interferes with the determination ability in model detection. Therefore, in this case, the attention mechanism needs to be added to the feature pyramid network to select the information that strongly correlates with the detected target from considerable redundant information and strengthen the dependency relationship between features to achieve the ability to express the enhanced features. As the types of attention mechanisms are increasing in depth by researchers [59–61], two types of attention mechanisms, namely, the CBAM and the SA, are selected in this paper. Furthermore, the mIoU value of the evaluation parameter after the experiment is improved by nearly 7%, thus indicating that the gap between the predicted values of the output of the well-labelled accurate value model is reduced, where the SA is more suitable for the overall contour extraction of landslide types. Different types of convolutional kernels are used in the feature extraction network to extract features from landslide observation result images, which can be used to maintain the spatial relationship between image pixels. The need for data volume is also reduced by methods such as data augmentation and migration learning, while the recognition performance of the model is improved.

5.3. Uncertainty in Observation and Identification under Different Natural Conditions

Landslides are controlled by many factors, such as topography, lithology, tectonics, meteorology, hydrology, vegetation, and human activities. The combinations of these factors also vary, thus resulting in different landslide patterns reflected in the InSAR results as various features in the geometry and colour changes of landslide patches. For example, in the Jinsha River, the retrogressive landslide formed by the traction damage process of the landslide body triggered by the rising water level [62], the InSAR patches show large deformation at low levels and many colour change cycles, while the deformation in other parts is relatively small; the red-layered soft rock landslide between Nangqian town and Gama town on the south bank of the Lancang River [63] is characterized by minor colour differences in the strip patches, but clusters are developed. In this case, the recognition of all types of active landslides cannot be fully guaranteed, even if data enhancement techniques are used in training the model. An ideal landslide recognition model oriented to InSAR observation result maps should be able to solve the recognition and classification problems of different types of landslides. Although the current deep learning technology has a powerful ability to analyse large datasets, this technology cannot identify all landslides indiscriminately because the sample dataset needs to be continuously improved. In the actual sample collection process, only some representative landslide types are usually considered [64], and the classifier in the training model, in this case, can meet the practical needs, especially when it is difficult to obtain a sufficient number of less common landslide

samples. This choice is beneficial for conducting research. However, this can also lead to some undesirable results. The more frequent landslides are more easily learned by the model, while landslides that are uncommon or have problems with the original data are easily misidentified. The features learned by the model are more limited. In this case, it is necessary to combine geomorphological features and optical remote sensing features to supplement the judgement further.

5.4. Influence of the Input Scale on the Identification Results

The input scale plays a vital role in modern detection frameworks during experiments. It is logical that an optimal image training scale exists and that clarifying this scale plays a vital role in improving deep learning. The resolution of the input images can be very high; however, as a result, such images can eventually lead to a decision-maker whose judgements are highly dependent on image data analysis. However, due to memory and GPU capacity limitations, a whole remote sensing image cannot be directly used for detection or incorporated into segmentation frameworks. For example, Chen et al. [65] used the Mask R-CNN model for urban village recognition of whole optical images based only on cropping, without preserving the geographic coordinates in the image. This model can complicate the complete monitoring of large images but cannot extract the identified targets into a standard format for dissemination and more in-depth study. This paper addresses oversized image recognition by sliding shifting the high-resolution input into multiple subimages based on coordinate information, building up a deep learning channel, and fusing this subimage input to generate the final result. This approach requires high computer hardware performance and a long processing time, and it also needs to consider the strategy of cropping images. Therefore, the next step of this paper is to enable the fast acquisition of essential features from large images and the prediction of their results by using different sampling results for large images as input data for the network and constructing an optimization scheme for the detector corresponding to them.

5.5. Potential Uncertainty of Geological Body Characteristics and Landslide Type Recognition

Currently, instance segmentation model improvement focuses on improving recognition accuracy and real-time performance [66], and application in specific tasks is still being attempted. In landslide research, the size and motion pattern of InSAR deformation patches of landslides can be used to determine their main hazard formation patterns [67] and further analyse and evaluate landslide hazards. However, landslide classification needs to be based on geometric features, motion patterns, and intensity.

The geometry of anomalously deformed slopes in InSAR observations is controlled by complex geological conditions and topographic and geomorphological conditions. The spatial contours of these slopes are often closely associated with different types of active landslides. For example, semicircular landslides are semicircular at the upper boundary and slightly open at the lower part in the shape of a dustpan, which is a typical morphology of medium-sized landslides [68]. The rectangular landslide has a flat main scarp, and the overall length in the direction of sliding is smaller than the width. The morphology is transverse spreading, which is similar to a rectangle. A tongue-shaped landslide body is small in the upper part, significant in the lower part, and high in the middle, with a shape similar to a tongue, such as a translational landslide [69]. Irregular landslides are controlled by complex geological conditions, including topographic and geomorphological conditions. Such conditions form multistage landslides, multiple juxtaposed landslides, or irregular shapes.

Different forms of landslide movement will also have their characteristics in the InSAR deformation map; landslides can be divided into two basic types: retrogressive and thrust. For retrogressive landslides, the toe of the slope is more significant than that of the top, while the deformation of other parts is relatively small. The width of the landslide decreases from the back end to the front end. As two boundaries control the deformation area, the deformation in some parts of the boundary produces a significant abrupt change. In

contrast, the thrust-type landslide body has a large deformation in the high main scarp and a blocked weak deformation in the low area. The larger deformation areas are located near the landslide perimeter. This is because the rear section of the thrust-type landslides first sliding deformation due to the existence of a significant sliding thrust, and then with the deformation of the rear section of the geotechnical body continuously developed forward and on both sides.

The size of the InSAR deformation patch corresponds to the range of feature deformation. Small and medium landslides have a small surface deformation range, thus resulting in small deformation patches; therefore, identification is not dominant. The opposite is true for large landslides, which are formed after years of breeding and have a larger overall surface deformation range. Moreover, large landslides are more likely to cause human and economic losses, and their differentiation is of great significance in disaster recognition. The scientific, systematic, and practical classification of landslides can promote the understanding of landslides in the study area and provide essential reference values for future related landslide field surveys, monitoring and early warning, emergency prevention and control, and other related research work. For example, Xu et al. [70] achieved the mastery of spatial and temporal distribution and deformation patterns of significant geological hazards, such as landslides, collapses, and debris flows by using space-based InSAR and other monitoring means and established a comprehensive classification and early warning system based on deformation observation characteristics. This method can issue early warning information before the actual occurrence of geological hazards and evacuate the residents in the danger zone to protect their lives and properties. Xie et al. [71] revealed the spatial and temporal distribution and deformation patterns of significant geological hazards in the upper reaches of the Minjiang River by using a combination of InSAR monitoring and in situ monitoring in 2020. This method can further reduce the loss of life and property to landslides within the Minjiang River Gorge. Rosi et al. [72] used satellite SAR data processed with the continuous persistent scattering InSAR (PSI) technique to map density areas to highlight areas of different landslide densities and sizes. Therefore, the method constructed in this paper cannot identify all small slope bodies. However, because the method can accurately identify large and medium-sized landslides, it will effectively improve the identification efficiency of typical active landslides in mountainous areas to ensure greater confidence.

The instance segmentation model can further distinguish active landslides with different geometric features, motion patterns, and intensities of deformation by distinguishing different targets in the same class, thereby improving deep learning for landslide recognition. Due to the grey box characteristics of deep learning, although the model can distinguish between different targets, the meaning of the model's geological attributes cannot be given directly yet and needs to be further clarified by combining both geology and statistics. The preliminary statistics reveal the difference in the model's geometric features, motion patterns, and intensity. Moreover, according to the theoretical analysis, there is still room for excavation to distinguish more geological attributes of landslides; distinguishing more geological attributes of landslides is to be studied in the next step.

6. Conclusions

How to use deep learning to recognize active landslides is a challenging but exciting question. When using deep learning as an automatic method to solve InSAR landslide mapping, specific research on its three major components, namely, the dataset, the algorithm, and computing power, is still needed. In this study, a Mask R-CNN+++ model is proposed for the rapid recognition of active landslides in some areas of the eastern edge of the Tibetan Plateau, where landslides are highly prevalent. The dataset construction in this paper is based on an automatic dataset generation procedure for raw images and vector files and a landslide dataset (SLD) conforming to the COCO annotation format. The image instance segmentation Mask R-CNN model is optimized by introducing deformable convolutional layers, residual multibranch network ResNext blocks, and attention mecha-

nisms. Correspondingly, the model increases the fineness of the segmentation results. The noise resistance of the model improves the feature extraction ability of the network for geometrically changing objects. Moreover, it is ensured that the computational effort does not increase significantly, thus making the model identify smoother landslide edges, adapt to a broader range of geological environments, and identify more diverse landslide forms. The instance segmentation model can also distinguish different targets in the same category and differentiate active landslides with different scales and deformation degrees. Initially, we reveal the differences in the geometric features, motion patterns, and intensity of landslides combined with landslide theory. This paper solves the mismatch between the input image size of the model and the size of the whole remote sensing image and realizes the rapid recognition of active landslides in an extensive range of high mountain valley areas. Using this deep learning model, the identification results of general active deformation slopes in the InSAR monitoring results over an extensive spatial range are characterized by a clear delineation of deformation boundaries, high efficiency, and good accuracy. However, it is easy to ignore ancient landslide accumulations with blurred deformation patches and discontinuous and incomplete morphology, and small-scale deformation bodies, which need to be appropriately combined with geomorphological and optical remote sensing features for further supplementary judgement.

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

Funding: This research is supported by the National Key R&D Program of China (2018YFC1505002); China Three Gorges Corporation YMJ(XLD)/(19)110; National Science Foundation of China (41672359).

Data Availability Statement: The SAR data are available at <https://scihub.copernicus.eu> (accessed on 10 September 2021).

Acknowledgments: The authors thank the three anonymous reviewers for their helpful comments and suggestions. The authors also thank the Editor for the kind assistance and beneficial comments.

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

References

1. Dai, F.; Lee, C.; Ngai, Y.Y. Landslide risk assessment and management: An overview. *Eng. Geol.* **2002**, *64*, 65–87. [\[CrossRef\]](#)
2. Brardinoni, F.; Slaymaker, O.; Hassan, M.A. Landslide inventory in a rugged forested watershed: A comparison between air-photo and field survey data. *Geomorphology* **2003**, *54*, 179–196. [\[CrossRef\]](#)
3. Gu, Z.; Shi, C.; Yang, H.; Yao, H. Analysis of dynamic sedimentary environments in alluvial fans of some tributaries of the upper Yellow River of China based on ground penetrating radar (GPR) and sediment cores. *Quat. Int.* **2019**, *509*, 30–40. [\[CrossRef\]](#)
4. Cheng, G.; Han, J. A survey on object detection in optical remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2016**, *117*, 11–28. [\[CrossRef\]](#)
5. Gu, Z.; Zhang, Y.; Fan, H. Mapping inter-and intra-annual dynamics in water surface area of the Tonle Sap Lake with Landsat time-series and water level data. *J. Hydrol.* **2021**, *601*, 126644. [\[CrossRef\]](#)
6. Glabsch, J.; Heunecke, O.; Schuhbäck, S. Monitoring the Hornbergl landslide using a recently developed low cost GNSS sensor network. *J. Appl. Geod.* **2009**, *3*, 179–192. [\[CrossRef\]](#)
7. Zhenkai, Z.; Xin, Y.; Hongyan, L.; Kaiyu, R. Accurate Identification of Active Landslides in Region Composed with Glacier, Forest, Steep Valley: A Case Study in the Lantsang Meili Snow Mountain Section. *Adv. Eng. Sci.* **2020**, *52*, 61–74. [\[CrossRef\]](#)
8. Zhu, Y.; Yao, X.; Yao, L.; Yao, C. Detection and characterization of active landslides with multisource SAR data and remote sensing in western Guizhou, China. *Nat. Hazards* **2022**, *111*, 1–22. [\[CrossRef\]](#)
9. Achache, J.; Fruneau, B.; Delacourt, C. Applicability of SAR interferometry for monitoring of landslides. In Proceedings of the Second ERS Applications Workshop, London, UK, 6–8 December 1995; p. 15.
10. Torres, R.; Snoeijs, P.; Geudtner, D.; Bibby, D.; Davidson, M.; Attema, E.; Potin, P.; Rommen, B.; Floury, N.; Brown, M. GMES Sentinel-1 mission. *Remote Sens. Environ.* **2012**, *120*, 9–24. [\[CrossRef\]](#)
11. Covello, F.; Battazza, F.; Coletta, A.; Lopinto, E.; Fiorentino, C.; Pietranera, L.; Valentini, G.; Zoffoli, S. COSMO-SkyMed an existing opportunity for observing the Earth. *J. Geodyn.* **2010**, *49*, 171–180. [\[CrossRef\]](#)

12. Kankaku, Y.; Suzuki, S.; Osawa, Y. ALOS-2 mission and development status. In Proceedings of the 2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS, Melbourne, Australia, 21–26 July 2013; pp. 2396–2399.
13. Zhang, W.; Lin, Y. Application preliminary evaluation of HJ-1-C SAR satellite of S band. In Proceedings of the MIPPR 2015: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications, Enshi, China, 14 December 2015; pp. 295–299.
14. Ye, X.; Kaufmann, H.; Guo, X. Landslide monitoring in the Three Gorges area using D-InSAR and corner reflectors. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 1167–1172. [[CrossRef](#)]
15. Sun, Q.; Zhang, L.; Ding, X.; Hu, J.; Li, Z.; Zhu, J. Slope deformation prior to Zhouqu, China landslide from InSAR time series analysis. *Remote Sens. Environ.* **2015**, *156*, 45–57. [[CrossRef](#)]
16. Yao, X.; Li, L.; Zhang, Y.; Zhou, Z.; Liu, X. Types and characteristics of slow-moving slope geo-hazards recognized by TS-InSAR along Xianshuihe active fault in the eastern Tibet Plateau. *Nat. Hazards* **2017**, *88*, 1727–1740. [[CrossRef](#)]
17. Yao, C.; Yao, X.; Gu, Z.; Ren, K.; Zhou, Z. Analysis on the development law of active geological hazards in the Loess Plateau based on InSAR identification. *J. Geomech.* **2022**, *28*, 257–267. [[CrossRef](#)]
18. Li, L.; Yao, X.; Zhou, Z.; Wang, F. The applicability assessment of Sentinel-1 data in InSAR monitoring of the deformed slopes of reservoir in the mountains of southwest China: A case study in the Xiluodu Reservoir. *J. Geomech.* **2022**, *28*, 281–293. [[CrossRef](#)]
19. Zhu, Y.; Yao, X.; Yao, L.; Zhou, Z.; Yao, C.; Xiao, S. Identification and risk assessment of coal mining-induced landslides in Guizhou Province by InSAR and optical remote sensing. *J. Geomech.* **2022**, *28*, 268–280. [[CrossRef](#)]
20. Liu, X.; Yao, X.; Zhou, Z.; Li, L.; Y, J. Study of the technique for landslide rapid recognition by InSAR. *J. Geomech.* **2018**, *24*, 229–237. [[CrossRef](#)]
21. Anantrasirichai, N.; Biggs, J.; Albino, F.; Hill, P.; Bull, D. Application of Machine Learning to Classification of Volcanic Deformation in Routinely Generated InSAR Data. *J. Geophys. Res. Solid Earth* **2018**, *123*, 6592–6606. [[CrossRef](#)]
22. Anantrasirichai, N.; Biggs, J.; Albino, F.; Bull, D. A deep learning approach to detecting volcano deformation from satellite imagery using synthetic datasets. *Remote Sens. Environ.* **2019**, *230*, 111179. [[CrossRef](#)]
23. Valade, S.; Ley, A.; Massimetti, F.; D'Hondt, O.; Walter, T.R. Towards Global Volcano Monitoring Using Multisensor Sentinel Missions and Artificial Intelligence: The MOUNTS Monitoring System. *Remote Sens.* **2019**, *11*, 1528. [[CrossRef](#)]
24. Hooper, A.; Gaddes, M.; Bagnardi, M.; Albino, F. Towards Improved Forecasting of Volcanic Hazards Using Machine Learning Applied to InSAR Data. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Kuala Lumpur, Malaysia, 11–16 July 2021; pp. 8484–8486.
25. Brengman, C.M.; Barnhart, W.D. Identification of surface deformation in InSAR using machine learning. *Geochem. Geophys. Geosystems* **2021**, *22*, e2020GC009204. [[CrossRef](#)]
26. Kamiyama, J.; Noro, T.; Sakagami, M.; Suzuki, Y.; Yoshikawa, K.; Hikosaka, S.; Hirata, I. Detection of Landslide Candidate Interference Fringes in DInSAR Imagery Using Deep Learning. *Recall* **2018**, *90*, 95.
27. Chen, X.; Yao, X.; Zhou, Z.; Liu, Y.; Yao, C.; Ren, K. DRs-UNet: A Deep Semantic Segmentation Network for the Recognition of Active Landslides from InSAR Imagery in the Three Rivers Region of the Qinghai–Tibet Plateau. *Remote Sens.* **2022**, *14*, 1848. [[CrossRef](#)]
28. Riedel, B.; Walther, A. InSAR processing for the recognition of landslides. *Adv. Geosci.* **2008**, *14*, 189–194. [[CrossRef](#)]
29. Bhargavi, K.; Jyothi, S. A survey on threshold based segmentation technique in image processing. *Int. J. Innov. Res. Dev.* **2014**, *3*, 234–239.
30. Bhardwaj, S.; Mittal, A. A survey on various edge detector techniques. *Procedia Technol.* **2012**, *4*, 220–226. [[CrossRef](#)]
31. Masood, S.; Sharif, M.; Masood, A.; Yasmin, M.; Raza, M.J.C.M.I. A survey on medical image segmentation. *Curr. Med. Imaging* **2015**, *11*, 3–14. [[CrossRef](#)]
32. Dhanachandra, N.; Chanu, Y.J. A survey on image segmentation methods using clustering techniques. *Eur. J. Eng. Technol. Res.* **2017**, *2*, 15–20. [[CrossRef](#)]
33. Sreedhar, B.; BE, M.S.; Kumar, M.S. A comparative study of melanoma skin cancer detection in traditional and current image processing techniques. In Proceedings of the 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), Palladam, India, 7–9 October 2020; pp. 654–658.
34. 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](#)]
35. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 2017; pp. 2961–2969.
36. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
37. Toms, S. *ArcPy and ArcGIS—Geospatial Analysis with Python*; Kindle, Ed.; Packt: Birmingham, UK, 2015.
38. Peng, C.; Rong, C.; Lingzhi, X.; Fenghuan, S. Risk analysis of mountain hazards in Tibetan Plateau under global warming. *Adv. Clim. Chang. Res.* **2014**, *10*, 103. [[CrossRef](#)]
39. Wu, Z.; Long, C.; Fan, T.; Zhou, C.; Feng, H.; Yang, Z.; Tong, Y. The arc rotational-shear active tectonic system on the southeastern margin of Tibetan Plateau and its dynamic characteristics and mechanism. *Geol. Bull. China* **2015**, *34*, 1–31. [[CrossRef](#)]
40. Jun, D.; Chang-Meng, W.; Wen-Chang, L.; Li-Jiang, Y.; Qiang-Fei, W. The situation and enlightenment of the research of the tectonic evolution and metallogenesis in the Sanjiang Tethys. *Earth Sci. Front.* **2014**, *21*, 52. [[CrossRef](#)]

41. Deng, Q.-D.; Cheng, S.-P.; Ma, J.; Du, P. Seismic activities and earthquake potential in the Tibetan Plateau. *Chin. J. Geophys.* **2014**, *57*, 678–697. [\[CrossRef\]](#)
42. Liu, J.; Ding, L.; Zeng, L.-S. Large-scale terrain analysis of selected regions of the Tibetan Plateau: Discussion on the origin of plateau planation surface. *Earth Sci. Front.* **2006**, *13*, 285. [\[CrossRef\]](#)
43. Dai, F.; Deng, J. Development characteristics of landslide hazards in three-rivers basin of southeast Tibetan Plateau. *Adv. Eng. Sci.* **2020**, *52*, 3–15. [\[CrossRef\]](#)
44. Gabriel, A.K.; Goldstein, R.M.; Zebker, H.A. Mapping small elevation changes over large areas: Differential radar interferometry. *J. Geophys. Res. Solid Earth* **1989**, *94*, 9183–9191. [\[CrossRef\]](#)
45. Yao, J.; Yao, X.; Liu, X.; Chen, J.; Li, L.; Zhou, Z. Study on the atmospheric correction of d-insar removal by three-dimensional space multi-item model—a case study of qiaojia landslide deformation observation in jinshajiang. *J. Eng. Geol.* **2018**, *26*, 14–21. [\[CrossRef\]](#)
46. Goldstein, R.M.; Werner, C.L. Radar interferogram filtering for geophysical applications. *Geophys. Res. Lett.* **1998**, *25*, 4035–4038. [\[CrossRef\]](#)
47. 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, 5–12 September 2014; pp. 740–755.
48. Couteaux, V.; Si-Mohamed, S.; Nempont, O.; Lefevre, T.; Popoff, A.; Pizaine, G.; Villain, N.; Bloch, I.; Cotten, A.; Boussel, L. Automatic knee meniscus tear detection and orientation classification with Mask-RCNN. *Diagn. Interv. Imaging* **2019**, *100*, 235–242. [\[CrossRef\]](#)
49. Johnson, J.W. Automatic nucleus segmentation with mask-RCNN. In Proceedings of the Science and Information Conference, Las Vegas, NV, USA, 2–3 May 2019; pp. 399–407.
50. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [\[CrossRef\]](#)
51. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
52. Voulodimos, A.; Doulamis, N.; Doulamis, A.; Protopapadakis, E. Deep learning for computer vision: A brief review. *Comput. Intell. Neurosci.* **2018**, *2018*, 13. [\[CrossRef\]](#) [\[PubMed\]](#)
53. 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; pp. 248–255.
54. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
55. Guo, M.-H.; Xu, T.-X.; Liu, J.-J.; Liu, Z.-N.; Jiang, P.-T.; Mu, T.-J.; Zhang, S.-H.; Martin, R.R.; Cheng, M.-M.; Hu, S.-M. Attention Mechanisms in Computer Vision: A Survey. *arXiv* **2021**, arXiv:2111.07624. [\[CrossRef\]](#)
56. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
57. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Salt Lake City, UT, USA, 18–22 June 2018; pp. 7794–7803.
58. Sun, P.; Piao, J.-C.; Cui, X. Object Detection in Urban Aerial Image Based on Advanced YOLO v3 Algorithm. In Proceedings of the 2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE), Harbin, China, 25–27 December 2020; pp. 2191–2196.
59. Nie, X.; Duan, M.; Ding, H.; Hu, B.; Wong, E.K. Attention mask R-CNN for ship detection and segmentation from remote sensing images. *IEEE Access* **2020**, *8*, 9325–9334. [\[CrossRef\]](#)
60. Hu, X.; Zhang, Z.; Jiang, Z.; Chaudhuri, S.; Yang, Z.; Nevatia, R. SPAN: Spatial pyramid attention network for image manipulation localization. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 312–328.
61. Zhang, X.; An, G.; Liu, Y. Mask R-CNN with feature pyramid attention for instance segmentation. In Proceedings of the 2018 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018; pp. 1194–1197.
62. Sun, X.; Chen, J.; Bao, Y.; Han, X.; Zhan, J.; Peng, W. Landslide susceptibility mapping using logistic regression analysis along the Jinsha river and its tributaries close to Derong and Deqin County, southwestern China. *ISPRS Int. J. Geo Inf.* **2018**, *7*, 438. [\[CrossRef\]](#)
63. TIAN, Y.; CHEN, L.; HUANG, H.; GAO, B.; ZHANG, J. Origin and stability of landslides in Chaya County, Lancang River Basin, Tibet. *Geol. Bull. China* **2021**, *40*, 2034–2042.
64. Zhu, X.; Montazeri, S.; Ali, M.; Hua, Y.; Wang, Y.; Mou, L.; Shi, Y.; Xu, F.; Bamler, R. Deep learning meets SAR: Concepts, models, pitfalls, and perspectives. *IEEE Geosci. Remote Sens. Mag.* **2021**, *9*, 143–172. [\[CrossRef\]](#)
65. Chen, L.; Xie, T.; Wang, X.; Wang, C. Identifying urban villages from city-wide satellite imagery leveraging mask R-CNN. In Proceedings of the Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers; London, UK, 9–13 September 2019; pp. 29–32.
66. Gu, W.; Bai, S.; Kong, L. A review on 2D instance segmentation based on deep neural networks. *Image Vis. Comput.* **2022**, *120*, 104401. [\[CrossRef\]](#)

-
67. Li, L.; Yao, X.; Yao, J.; Zhou, Z.; Feng, X.; Liu, X. Analysis of deformation characteristics for a reservoir landslide before and after impoundment by multiple D-InSAR observations at Jinshajiang River, China. *Nat. Hazards* **2019**, *98*, 719–733. [[CrossRef](#)]
 68. Van Den Eeckhaut, M.; Poesen, J.; Gullentops, F.; Vandekerckhove, L.; Hervás, J. Regional mapping and characterisation of old landslides in hilly regions using LiDAR-based imagery in Southern Flanders. *Quat. Res.* **2011**, *75*, 721–733. [[CrossRef](#)]
 69. Hu, X.; Lu, Z.; Pierson, T.C.; Kramer, R.; George, D.L. Combining InSAR and GPS to determine transient movement and thickness of a seasonally active low-gradient translational landslide. *Geophys. Res. Lett.* **2018**, *45*, 1453–1462. [[CrossRef](#)]
 70. XU, Q.; DONG, X.; LI, W. Integrated space-air-ground early detection, monitoring and warning system for potential catastrophic geohazards. *Geomat. Inf. Sci. Wuhan Univ.* **2019**, *44*, 957–966. [[CrossRef](#)]
 71. Xie, M.; Zhao, W.; Ju, N.; He, C.; Huang, H.; Cui, Q. Landslide evolution assessment based on InSAR and real-time monitoring of a large reactivated landslide, Wenchuan, China. *Eng. Geol.* **2020**, *277*, 105781. [[CrossRef](#)]
 72. 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](#)]