

Landslide detection of optical remote sensing image based on attention and u-net

Riya Su, and Yanming Yang*

Earthquake Agency of Inner Mongolia Autonomous Region, Hohhot, China

Abstract. With the improvement of remote sensing image technology, researchers pay more and more attention to detecting landslides in optical remote sensing images. In this paper, the landslide is detected by semantic segmentation model based on deep learning, U-shaped network is used to enhance the extraction ability of landslide features, and the model pays more attention to landslide area through attention mechanism, so as to make the model detect landslide more accurately. Through experiments on the Bijie Landslide Dataset, the values of OA and mIoU in this model are increased by 1% and 16% respectively. The boundary of landslide is more straightforward and more accurate.

1 Introduction

Landslides are one kind of geological hazard, which will cause severe consequences and cause considerable losses to the society, environment, life, and property [1]. It is urgent to detect landslides quickly and accurately and reduce landslide damage [2]. Optical remote sensing image, as an essential earth observation tool, can intuitively identify the information of ground objects that have been successfully applied in landslide detection [3]. However, due to the variety of size, shape, and color of landslides, it is still a challenging but significant task to detect landslides from optical remote sensing images accurately.

In recent years, researchers have proposed various methods of landslide detection from optical remote sensing images, which can be roughly divided into two strategies: traditional methods and deep learning methods. The traditional methods usually rely on low-level visual features such as the shape and color of optical remote sensing images to detect landslides. Hu et al. [4] based on the feature analysis of landslide scars from these remote sensing imageries, the object-oriented landslide detection methodology is presented concerning remote sensing imageries. Yu et al. [5] further extracted landslides by object-based contours using morphological operations based on the saliency calculations. Due to the weak representation ability of low-level visual features, traditional methods are time-consuming, and their interpretation accuracy may be poor. Over the past few years, with the rapid development of deep learning, the outstanding feature extraction ability makes the deep learning method can detect landslides quickly and automatically. Ye et al. [6] proposed a deep belief network with constraints to extract the features of the landslide from remote sensing imageries. Cheng et al. [7] proposed a YOLO-SA landslide detection model

* Corresponding author: yym_2005@163.com

by joint CNN and attention. These deep learning methods can accurately detect landslides, but there is still much room for improvement, especially to determine the boundary of landslide accurately. As shown in Fig. 1, some examples of landslides in optical remote sensing images. The size, shape, and color of the landslide are very different, and they are too similar to neighboring features, making it difficult to distinguish the landslide boundary.

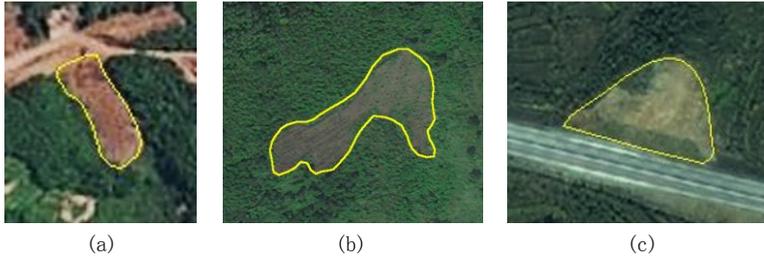


Fig. 1. Some examples of landslides in optical remote sensing images. (The area marked with line is a landslide.)

To address the need for more accurate boundary of landslide, the present study aimed to propose a novel landslide detection module based on attention and u-net for segmentation of optical remote sensing image. The structure of the model in this work is shown in Fig. 2. This model adopts a u-shape encoder-decoder structure, which can achieve pixel-level segmentation and capture semantic features more accurately, so that the relationship between the pixels is more apparent, to get more accurate landslide boundaries. The contributions of this work primarily include the following two points:

- (1) A more efficient structure is used as the backbone of the model in the encoder part, to enhance the ability of the model to capture information. It should be noted that there are not many parameters of this efficient structure, which will not make the model difficult to train.
- (2) Enhance the attention of the model to landslide through an attention mechanism.

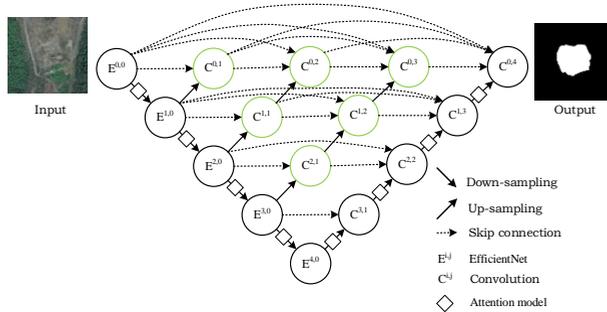


Fig. 2. The structure of the model in this work.

2 Methodology

2.1 Backbone for model

Unet++[8] is a high-performance semantic segmentation model, which based on nested and dense skip connections. The encoder and decoder of Unet++ are connected through a series of nested thick convolution blocks, and use skip connection to associate feature maps with different depths. It is conducive to improving the accuracy of semantic segmentation that fully integrates multi-scale feature maps. However, the computational resources required

increase with the number of parameters to increase. To solve the problem of parameters, EfficientNet[9], which is a more efficient structure, is used as the backbone of the model in this work. EfficientNet improves performance by optimizing the width, depth, and resolution of the model, which is described as:

$$\begin{aligned}
 & \max_{d,w,r} \text{Accuracy}[N(d,w,r)] \\
 & \text{s.t. } N(d,w,r) = \bigodot_{i=1-s}^s \hat{F}_i^{d \cdot \hat{L}_i} \left[X_{(r \cdot \hat{H}_i, r \cdot \hat{W}_i, w \cdot \hat{C}_i)} \right] \\
 & \text{Memory}(N) \leq \text{target_memory} \\
 & \text{FLOPS}(N) \leq \text{target_flops}
 \end{aligned} \tag{1}$$

where d , w , and r represent the coefficients of the width, depth, and resolution of the network model respectively, N represents the network model, \hat{F}_i , \hat{L}_i , \hat{H}_i , \hat{W}_i , and \hat{C}_i are predefined parameters. target_memory and target_flops are threshold value of parameter and floating-point operation quantity.

By optimizing and adjusting the coefficients of the width, depth, and resolution of the network model for many times, EfficientNet can effectively use the resources of the model, and has the powerful ability to capture the feature of the landslide with fewer parameters.

2.2 Attention mechanism

The attention mechanism is similar to the human observation environment. It enhances the expression ability of the feature and reduces the noise interference by paying more attention to the important local information through the weighted processing of the feature map. SGE attention module are adopted in this paper.

Based on making full use of the global information, the SGE attention module gives each group the spatial position weight according to the contribution of semantic information. It enhances the semantic feature of the region of interest and is conducive to establishing context connection in the local area.

3 Experimental results

3.1 Datasets

The model in this work was validated on Bijie Landslide Dataset[11]. This dataset was captured over Bijie city in China, which has a resolution of 0.8 m/pixel with tiles of approximately 61~1239 pixels \times 61~1197 pixels. The dataset contains 770 remote sensing images of landslides, all of the image are annotated with white (landslides) and black (non-landslide). The annotation dataset is divided into the training set, test set, and validating set according to the ratio of 7:2:1. The dataset has only 770 images, which is challenging to train. Therefore, it is expanded to 2000 images through vertical flip, horizontal flip, and distortion.

3.2 Experimental setting

This work is based on the public PyTorch deep learning framework of version 1.7. All experiments were tested on a workstation with Windows10. The workstation contains an Intel Xeon Gold CPU and an NVIDIA Quadro RTX 5000 GPU card with 16GB. During the training of the model, Adam is set as the optimizer, the initial learning rate of the model is

0.0001. The focalloss function evaluates the similarity between the results and the label. All models need train 200 epochs, the learning rate strategy uses the way of step decline, which reduces by 5% every 20 epochs. The overall accuracy (OA) and the mean intersection over union (mIoU) evaluate the performance of the model, to ensure the accuracy of the test, the average value of 5 repeated tests is used as the final result.

3.3 Results and discussion

The experimental results on the Bijie Landslide Dataset are shown in table 1. The performance of the proposed model gets the highest value, the improvements are more than 1% in OA and 16% in mIoU compared with Unet++. Visualization of the results of different methods are provided in Fig. 3, the results shows that the proposed model can achieve more accurate segmentation, especially the boundaries of landslides are clearer.

Table 1. The experimental results on Bijie Landslide Dataset.

Method	OA	mIoU	Parameter
Unet++	96.17%	64.76%	26.79MB
The proposed model	97.42%	81.69%	20.01MB

4 Conclusions

In this work, to improve the accuracy of the landslide of optical remote sensing images, we joint U-Net and attention mechanism to design a novel model. The model adopts the u-shape structure to enhance the feature extraction ability. In addition, the attention mechanism makes the model pay more attention to landslides. Experiments Bijie Landslide Dataset show that the model in this paper has a significant improvement in the segmentation of the landslide boundary.

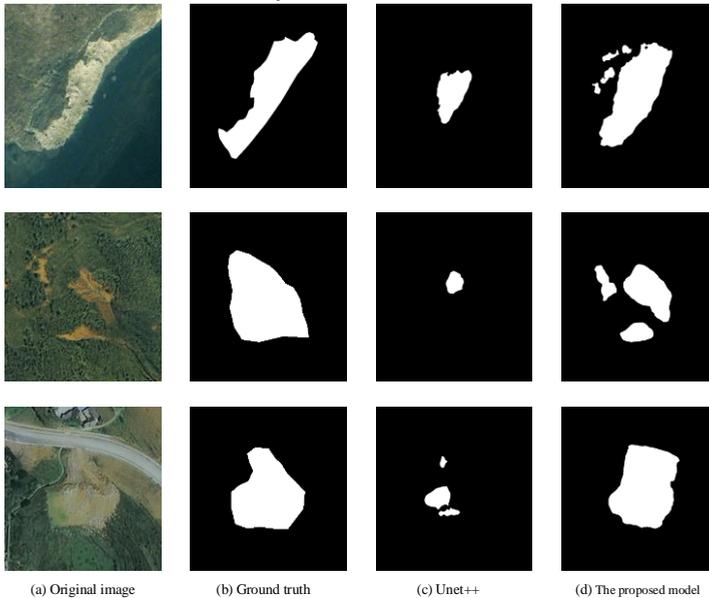


Fig. 3. Visualization of results on the Bijie Landslide Dataset.

This work was supported by the Key tasks of earthquake emergency youth of China Seismological Bureau (CEA_EDEM-202103).

References

1. Z. Ma, G. Mei, *Earth-Science Reviews*, **223** (2021)
2. R. Su, Q. An, Y. Yang, F. Xue, *Seismological and Geomagnetic Observation and Research*, **03**, (2019)
3. Y. Liu, W. Zhang, X. Chen, *IEEE International Geoscience and Remote Sensing Symposium*, 4091-4094 (2021)
4. D. Y. Hu, J. Li, W. J. Zhao, G. X. Peng, *Journal of Natural Disasters*, **06**, 46-50 (2008)
5. B. Yu, F. Chen, *J. Appl. Rem. Sens.* **10**, 045013 (2016)
6. C. M. Ye, Y. Li, P. Cui, L. Liang, *IEEE J-STARS*, **11**, 1-14 (2019)
7. L. Cheng, J. Li, P. Duan, M. Wang, *Landslides*, **18**, 2751-2765 (2021)
8. Z. Zhou, MMR. Siddiquee, N. Tajbakhsh, J. Liang, *4th Deep Learning in Medical Image Analysis*, **11**, 3-11, (2019)
9. M. Tan, Q. Le, *Proceedings of International Conference on Machine Learning*, 6105-6114 (2019)
10. X. Y. Wang, Z. Y. Cui, Z J Cao, *IEEE International Geoscience and Remote Sensing Symposium*, 1244-1247 (2020).
11. S. Ji, D. Yu, C. Shen, W. Li, Q. Xu, *Landslides*, **6** (2020)