Multispectral U-Net: A Semantic Segmentation Model Using Multispectral Bands Fusion Mechanism for Landslide Detection

Lin Bai[1,2,3], Weile Li[1,4,*], Qiang Xu[1,4], Weihang Peng[2,3], Kai Chen[2,3], Zhenzhen Duan[5] and Huiyan Lu[1,4]

[1]*State Key Laboratory of Geohazard Prevention and Geoenvironment Protection, Chengdu University of Technology, No.1 East Third Road, Erxianqiao, Chenghua District,Chengdu, 610059, China*

[2]*College of Mathematics and Physics, Chengdu University of Technology, No.1 East Third Road, Erxianqiao, Chenghua District,Chengdu, 610059, China*

[3]*Geomathematics Key Laboratory of Sichuan Province, Chengdu University of Technology, No.1 East Third Road, Erxianqiao, Chenghua District,Chengdu, 610059, China*

[4]*College of Environment and Civil Engineering, Chengdu University of Technology, No.1 East Third Road, Erxianqiao, Chenghua District,Chengdu, 610059, China*

[5]*College of Computer Science and Cyber Security, Chengdu University of Technology, No.1 East Third Road, Erxianqiao, Chenghua District,Chengdu, 610059, China*

Abstract

Each image patch of the Landslide4Sense dataset has 14 bands, which is different with the RGB satellite imagery. A new model called Multispectral U-Net was proposed. We conduct experiments and apply two modifications to traditional U-Net to improve the performance. Firstly, the skip-connection and spectral normalization regularization are used to enhance the model ability of feature extraction. Secondly, an extra Inverted Residuals and Linear Bottlenecks branch is introduced for 10 meters resolution bands. We split the official dataset into two parts, with 3539 images for training and 260 images for testing. Multispectral U-Net accomplishes the best performance, with an F1-score of 77.83%, followed by the baseline U-Net and Deeplabv3+. The model has better performance on multi-spectral data and small objection for landslide detection.

Keywords

Landslide detection, deep learning, multispectral imagery, semantic segmentation

1. Introduction

Landslides are the most frequent geological disasters in nature, which lead to heavy damage to the infrastructure every year. It is very important to detect the landslide position and take some measures to avoid more loss, and landslide interpretation from satellite images has already become a reliable technique for landslide investigation [1]. However, manual landslide interpretation is difficult, as it is time-consuming and over-reliance on professional experience.

In recent years, the deep learning methods have made a large number of achievements in most computer vision tasks, including classification, object detection and semantic segmentation [2]. Semantic segmentation model requires to obtain the exact shape of the object, which is widely used in high precision tasks, such as medical image segmentation and the road area segmentation. For example, Ronneberger at et. [3] proposed a symmetrical and malleable structure, called U-Net, for the biomedical image segmentation.

With the development of astronautical technology, it is easy to access to the remote sensing data for a wide range of area. Meanwhile, a growing number of landslide datasets are released, which make it possible to adopt the deep learning methods in landslide detection, such as [4, 5, 6] and etc. The most common datasets provide the RGB image to keep the same input format with the original deep learning methods. It can take some advantages, for example, researchers can adopt these methods without any changes, and the pre-trained weights can shorten the training time. However, most satellite imagery has more than three bands, for example, Landslide4Sense [7] provides the data has 14 bands. If the model only uses the RGB bands, it can lead to information loss. In this paper, we will try to use the multi-spectral data for landslides detection tasks.

CDCEO 2022: 2nd Workshop on Complex Data Challenges in Earth Observation, July 25, 2022, Vienna, Austria

bailin@cdut.edu.cn (L. Bai); liweile08@cdut.edu.cn (W. Li); xq@cdut.edu.cn (Q. Xu); pangdarren@outlook.com (W. Peng); 15196231187@163.com (K. Chen); 201813160831@stu.cdut.edu.cn (Z. Duan); luhui8535@gmail.com (H. Lu)

0000-0002-1005-0875 (L. Bai); 0000-0002-3741-8801 (W. Li)

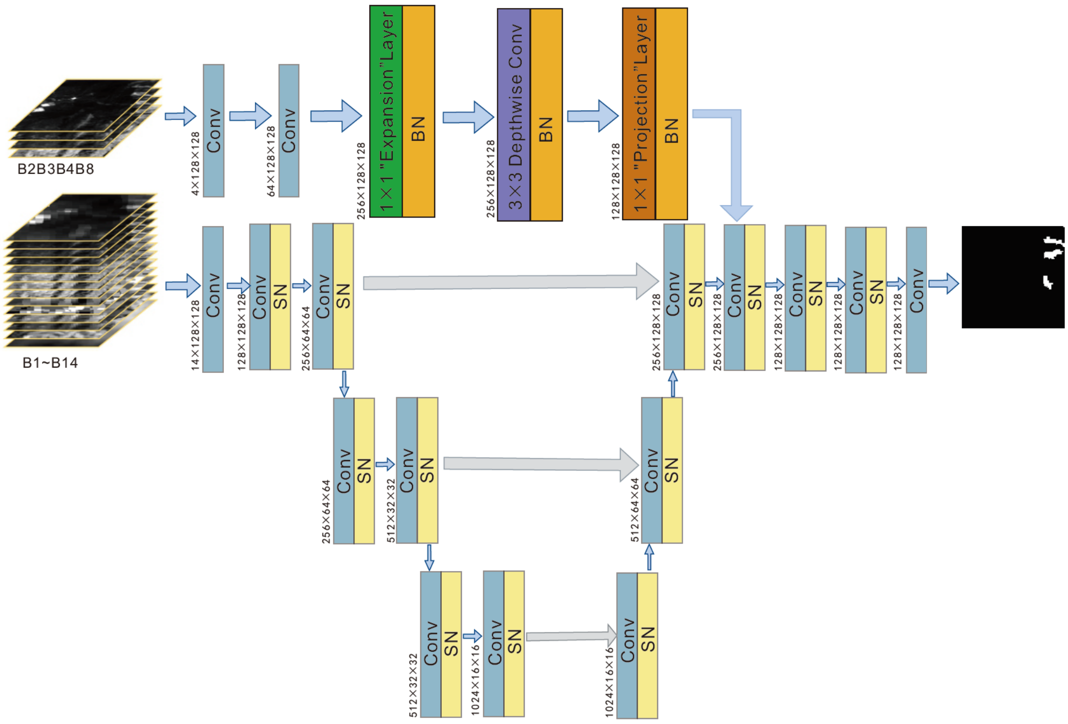

Figure 1: The architecture of Multispectral U-Net.

2. Methodology

The Landslide4Sense data has more spectral bands than RGB image. Considering the differences, we introduce some modification on U-Net, so that it can adopt the multi spectral satellite imagery.

2.1. Multispectral bands selection

Landslide4Sense [7] dataset provides remote sensing image patches with 14 bands. One of the general methods is that feeding all the bands together to the model, but could not achieve a high score in our experiment. In the validation dataset, U-Net [3], Deeplabv3 [8] and Deeplabv3+ [9] can achieve 0.65, 0.66 and 0.67 f1 scores respectively. The multispectral data from Sentinel-2 has different Spatial resolution [10], as table 1.

2.2. Multispectral U-Net

We introduce a novel network called Multispectral U-Net, which aim to enrich the semantic features extracted from the 14 bands. Multispectral U-Net's architecture is illustrated in figure 1. The model structure includes two parts, Inverted Residuals and Linear Bottlenecks (upper part) and U-Net (lower part).

Inverted Residuals and Linear Bottlenecks was introduced in the MobileNetv2 [11]. The high-dimensional feature maps contains manifolds, which can be compressed until they span the whole space. However, it the dimension is reduce a lot, Relu will cause information loss. The features propagate from low-dimension to high-dimension by the expansion layer, and the projection layer does the opposite. Projection convolution uses the linear activation instead of Relu to alleviate the conflict between dimension reduction and non-linear optimization. In the Multispectral U-Net the features in this branch will be fused with U-Net features, which can get more refined feature maps to improve the model's performance.

Inspired by the Real-ESRGAN [12], we use U-Net with skip-connection as the backbone, and use the spectral normalization regularization for a stable training process. Due to the data size is only 128 pixels, we only use three downsampling layers to avoid information loss. Additionally, we change the activation function to SMU, which can improve the model's performance in this dataset.

In the Multispectral U-Net model, we input all 14 bands to U-Net branch and 10 meters resolution bands (B2, B3, B4, B8) to Inverted Residuals and Linear Bottlenecks branch, and this strategy can get the best results.

Table 1
The spatial resolution of SENTINEL-2

Spatial resolution	10m	20m	60m
Bands	B2, B3, B4, B8	B5, B6, B7, B11, B12	B1, B9, B19

Table 2
Quantitative Results of Landslide Detection (%)

	Recall	Precision	F1
Deeplabv3+	72.54	74.22	73.37
U-Net	71.34	**81.82**	76.22
Multispectral U-Net	**75.41**	80.41	**77.83**

3. Experiment

We train the model on NVIDIA GeForce RTX 3090 GPU, and the training parameters are 8 batch size, adam optimizer, cosine learning rate (0.00001) with warmup and restarts and the cross-entropy loss. As we explore the common image data augmentation strategies, such as flip and rotate, and do not get an improvement on the model performance, we do not add the augmentation in the following experiments.

Then, we split the official dataset into two parts, with 3539 images for training and 260 images for testing. After training 200 epochs, we evaluate the model with recall, precision and f1 score metrics, which is illustrated in the table-2. The recall of the Multispectral U-Net is significant higher than other two models, with 75.41, and by this reason, although the precision of Multispectral U-Net is lower than U-Net, its f1 score is also better than that of U-Net. Deeplabv3+ do not have a very high precision, and we think the reason is that the more downsampling layers cause the edge information loss.

Figure 2 shows the prediction of different models, and each row represents a single data. In the first data, the landslide segmentation results of the three models are similar, and we can see that the area predicted by Deeplabv3+ is bigger but coarse. This is a representative example that Deeplabv3+ tend to have a high recall but not precision. In the second data, although the landslide morphology cannot be seen from the RGB image, all three models can detect landslide. This means besides from the RGB bands (B4, B3, B2), which can be observed by humans, other bands also make contribution to the landslide detection. The third data is a very complex landslide scenario. It is showed that the three models can detect the landslide successfully, but Multispectral U-Net has better performance in some details.

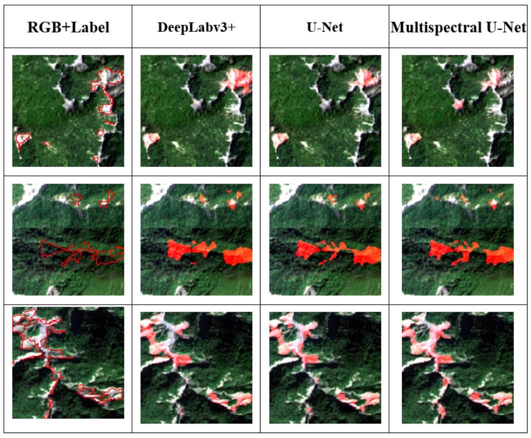

Figure 2: The landslide detection maps obtained by different deep learning methods.

4. Conclusion

In this study, we first adopt three different models, including U-Net, Deeplabv3, Deeplabv3+, to train the multi-spectral remote sensing data without any changes and pre-trained weights. It shows that there is only slightly different of the models' performance. Then, we first proposed a new U-Net-based approach denoted as Multi-spectral U-Net for the better use of the multi-spectral data. It combines two branches, called Inverted Residuals and Linear Bottlenecks and improved U-Net, and the features from two branches will be fused by concatenation in the final predicting stages. The former pay attention to extract the features with high resolution, with 10 meters/pixel in our experiments while the latter is used to capture the main features with a large number of convolution layers. Besides, to make the model more adaptable for the dataset, we reduce the number of downsampling layers and add more skip-connection and use the SMU activation instead of the Relu. To evaluate the performance quantitatively, our model is compared with U-Net and Deeplabv3+ in experiments. The results show that Multispectral U-Net have the significant improvement in the recall metrics, with 75.41 and 2.87 absolute percent greater than the second place, and the f1 score of our model is also the best.

Overall, this study demonstrated the potential usage of multi-spectral satellite imagery in landslide detection.

A new model namely Multispectral U-Net is proposed, and achieve a better performance compared with the traditional deep learning models.

References

[1] X. Qiang, Understanding and consideration of related issues in early identification of potential geohazards, 2020. URL: http://ch.whu.edu.cn/en/article/doi/10.13203/j.whugis20200043. doi:10.13203/j.whugis20200043.

[2] S. Ghosh, N. Das, I. Das, U. Maulik, Understanding deep learning techniques for image segmentation, ACM Comput. Surv. 52 (2019). URL: https://doi.org/10.1145/3329784. doi:10.1145/3329784.

[3] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.

[4] D. V. Dao, A. Jaafari, M. Bayat, D. Mafi-Gholami, C. Qi, H. Moayedi, T. V. Phong, H.-B. Ly, T.-T. Le, P. T. Trinh, C. Luu, N. K. Quoc, B. N. Thanh, B. T. Pham, A spatially explicit deep learning neural network model for the prediction of landslide susceptibility, CATENA 188 (2020) 104451. URL: https://www.sciencedirect.com/science/article/pii/S0341816219305934. doi:https://doi.org/10.1016/j.catena.2019.104451.

[5] O. Ghorbanzadeh, A. Crivellari, P. Ghamisi, H. Shahabi, T. Blaschke, A comprehensive transferability evaluation of u-net and resu-net for landslide detection from sentinel-2 data (case study areas from taiwan, china, and japan), Scientific Reports 11 (2021) 14629. URL: https://doi.org/10.1038/s41598-021-94190-9. doi:10.1038/s41598-021-94190-9.

[6] J. M. Habumugisha, N. Chen, M. Rahman, M. M. Islam, H. Ahmad, A. Elbeltagi, G. Sharma, S. N. Liza, A. Dewan, Landslide susceptibility mapping with deep learning algorithms, Sustainability 14 (2022). URL: https://www.mdpi.com/2071-1050/14/3/1734. doi:10.3390/su14031734.

[7] O. Ghorbanzadeh, Y. Xu, P. Ghamis, M. Kopp, D. P. Kreil, Landslide4sense: Reference benchmark data and deep learning models for landslide detection, ArXiv abs/2206.00515 (2022).

[8] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, ArXiv abs/1706.05587 (2017).

[9] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: V. Ferrari, M. Hebert, C. Sminchisescu, Y. Weiss (Eds.), Computer Vision – ECCV 2018, Springer International Publishing, Cham, 2018, pp. 833–851.

[10] Y. Du, Y. Zhang, F. Ling, Q. Wang, W. Li, X. Li, Water bodies' mapping from sentinel-2 imagery with modified normalized difference water index at 10-m spatial resolution produced by sharpening the swir band, Remote Sensing 8 (2016). URL: https://www.mdpi.com/2072-4292/8/4/354. doi:10.3390/rs8040354.

[11] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520. doi:10.1109/CVPR.2018.00474.

[12] X. Wang, L. Xie, C. Dong, Y. Shan, Real-esrgan: Training real-world blind super-resolution with pure synthetic data, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021, pp. 1905–1914. doi:10.1109/ICCVW54120.2021.00217.