Remote Sensing Image Building Extraction Method Based on Deep Learning

Minshui Wang, Mingchang Wang, Guodong Yang and Ziwei Liu

Department of Surveying and Mapping Engineering, Jilin University, College of Geo-Exploration Science and Technology, Room 137, Geological Palace, 938 West Democratic Street, Chaoyang District, Changchun City, Jilin Province, China
Email: 1543383519@qq.com

Abstract. Using random patches and deeplabV3+ network can effectively improve the building extraction accuracy and ensure the integrity of building. First, acquiring the image of a 5000x5000 pixel one, and using the random Patch Extraction datastore function to create a number of random patches with the size of 224x224 pixels as network input images. Second, creating a convolutional neural network based on resnet50 by using the deeplabv3plusLayers function, and then projecting the learned discrimination features with lower resolution to the pixel space with higher resolution, to realise the automatic extraction of the building. Third, two images were input to verify the extraction accuracy of the trained network. The results showed that the Pixel accuracy of image 1 and image 2 reached 97.98% and 92.59%. Compared with other building extraction algorithms, this method has higher extraction accuracy. This method has strong expansibility and It can be used for automatic extraction of other feature types.

Keywords. High-resolution remote sensing image; deep learning; extraction of buildings; random patch.

1. Introduction

Buildings are an important part of national geographic information. It has important significance in digital city construction, land use survey, urban change monitoring and urban area planning. How to quickly, precisely and automatically extract buildings in complex urban scenes has become a research focus. The traditional building extraction method does not make full use of the texture, shape and other features of buildings, and it is not adaptable to the extraction of buildings in complex environment, and It is easy to miss and Error extraction. with the development of UAV and photogrammetric technology, high-resolution images can be easily and quickly acquired. High-resolution images contain rich details and provide a lot of effective information. Meanwhile, the development of artificial intelligence and improvement of computer performance, the automatic extraction of ground objects by semantic segmentation has achieved remarkable results. Fan et al. [1] used principal component transform unsupervised pre-training network and adaptive pooling model to obtain texture features to participate in building extraction. Xu et al. [2] proposed a new multiscale semantic segmentation network (MSSNet), it shows that MSSNet can accurately segment similar objects and fine-structured objects. And it has the advantages of simple training process and easy to use. Geng et al. [3] proposed a semantic segmentation method to improved the classification accuracy of semantic segmentation network through the introduction of unlabeled data and Reduce the network learning performance when there is too little labelled data. Compared with the traditional ground
object classification algorithm, deep convolutional neural network (DNN) requires no manual intervention and has a strong feature extraction capability, which is widely used in the field of image processing. In general, before the semantic segmentation network training, it is necessary to cut the large image into several small size images, and select the samples from the images for training, the selection of training samples directly affects the accuracy of image segmentation [4-6]. Deep learning for large images is the research focus of this article.

2. Principles and Methods
Semantic segmentation is to assign each pixel to the object category to achieve pixel level classification. It is a new attempt to solve the problem of building extraction by deep learning image segmentation. Aiming at the problems existing in remote sensing image processing, in this section, we will introduce the semantic segmentation convolutional neural network [7-8], and detailed description of the semantic segmentation method based on random patches proposed in this paper. At last, describe the extraction process of image building based on semantic segmentation model.

2.1. Semantic Segmentation
Semantic segmentation is to assign each pixel in the image to a semantic category, which requires that the neural network not only has the ability to judge the pixel category, but also needs to project features into the pixel space, which is generally understood as an encoder - decoder network. The encoder trains a classification network, and the decoder projects the lower resolution discrimination features learned by the encoder into the higher resolution pixel space to achieve dense classification. Common semantic segmentation networks [9] include FCN, Deeplab series, SegNet, u-net, mask-rcnn, etc. This paper is based on Deeplabv3plus network, Deeplabv3plus is a deep network structure that combines encoder-decoder and spatial pyramid pooling module (SPP) for semantic segmentation. Encoder-decoder will obtain more boundary information. The spatial pyramid pooling module (SPP) uses the feature processing of different resolutions of multiple proportions and multiple effective sensing fields to mine multi-scale context information. Feature extraction can use transfer learning to obtain initialization weights from the pre-training network resnet50, or other pre-training basic network models can be selected according to experience. The schematic diagram of network structure is shown in figure 1.

![Figure 1. Schematic diagram of deeplabv3plus convolution network.](image_url)

2.2. Implementation of Building Extraction
In order to avoid image clipping, this study directly input a large range of 5000x5000 pixel images, and adopted the method of creating random patches for network training, so as to improve the efficiency and accuracy of image processing and maintain image integrity. Patch is a method to prevent large image from running out of memory and effectively increase the amount of available
training data. Remote sensing image extraction of buildings [10-11] is essentially a dichotomy problem, creating labels into two categories: background and buildings. The implementation steps: using randomPatchExtractionDatastore function creates data set, pick a certain number of random patches from an image data store containing real images and a pixel label store containing pixel labels. The feature extraction is carried out by using the pre-trained convolutional neural network, and then the feature reconstruction is carried out by deconvolution to realize the extraction of buildings. The processing process is shown in figure 2.

Figure 2. Technical route.

2.3. Semantic Segmentation Accuracy Evaluation
The semantic segmentation precision evaluation mainly includes the following methods:

- Pixel accuracy (PA), pixel accuracy is the ratio of correct classification pixel number to total pixel number. The equation is as follows:

\[
PA = \frac{\sum_{i=0}^{n} P_{ii}}{\sum_{i=0}^{n} \sum_{j=0}^{n} P_{ij}}
\]  

(1)

- Mean pixel accuracy (MPA) [12-13]. For remote sensing images containing multiple categories, calculate the ratio of the number of correctly classified pixels in each category to the total number of pixels in that category, and then take the average value after summing.

\[
MPA = \frac{1}{k+1} \sum_{i=0}^{k} \frac{P_{ii}}{\sum_{j=0}^{n} P_{ij}}
\]  

(2)

- Mean Intersection over Union (MIoU) [14] is the sum of the mean ratios of the intersection and union of the predicted results and the true values of each type. Equation is as follows:

\[
MIoU = \frac{1}{n+1} \sum_{i=0}^{n} \frac{P_{ii}}{\sum_{j=0}^{n} P_{ij} + \sum_{j=0}^{n} P_{ji} - P_{ii}}
\]  

(3)

- Weighted Intersection over Union (WIoU)is the weighted sum of the IoU for each class according to the frequency of its occurrence

\[
WIoU = \frac{1}{\sum_{i=0}^{k} \sum_{j=0}^{k} P_{ij}} \sum_{i=0}^{k} \sum_{j=0}^{k} \frac{P_{ij}}{P_{ij} + \sum_{j=0}^{n} P_{ji} - P_{ii}}
\]  

(4)
• **BF Score [15]:** The boundary contour matching score refers to the degree of alignment between the predicted boundary of each category and the true boundary. The average BF score is the average BF score of all categories in the remote sensing image.

In the equation, \( k + 1 \), \( k \) represents the target class and 1 represents the background class, \( P_{ii} \) represents true positives, \( P_{ij} \) represents false positives, \( P_{ji} \) represents false negatives.

3. **Test and Result Analysis**

3.1. **Data Set**

In this paper, INRIA aerial image data set [16] was used to evaluate the semantic segmentation network. Data set introduction:

The data set refers to different residential area, including densely-populated area and high mountain town, the total area is 810 km\(^2\) and the data contain training data and test data. The area of training data is 405 km\(^2\), the area of verify data is 405 km\(^2\). The images include two kinds, building and non-building. The image spatial resolution is 0.3 m after aerial color calibration.

3.2. **Test Design**

A remote sensing image of 5000x5000 pixels is selected and named image1 and label1. The building is marked with white and the background is black. Define the input patch size of 224x224, and the number of patches is 10000. Use the deeplabv3plusLayers function to create a semantic segmentation network based on resnet-50. Random gradient descent method (SGDM) was used for the training option, momentum was 0.9, initial learning rate was 0.01, and cross entropy was used for the loss function. The number of iterations is 18,750, totaling 30 rounds, with each round being 625 times. The hardware configuration is NVIDIA RTX2060 graphics card, 16 GB of running memory, Intel i7-9700F processor.

In this study, 2 images of 5000x5000 were selected for network accuracy test. Image 1 is a remote sensing image used for network training. The selected image 2 has different textures and higher vegetation coverage than image 1. From the experimental results, we can see that the extraction method of buildings in this study can obtain clear and complete boundary of buildings, and buildings with small area can be well recognized. It can be found from the accuracy evaluation table 1 that the Pixel accuracy of image 1 reaches 97.98\%, with an Mean Pixel accuracy of 97.97\% and the Mean IoU of 96\%. The Pixel accuracy of image 2 is 92.59\%, the Mean Pixel accuracy is 88.97\%, and the Mean IoU is 82.68\%. Compared with other methods, the extraction accuracy and processing efficiency of buildings are improved. The detailed results are shown in figure 3 and table 1.

|        | PA   | MPA  | MIoU  | WIoU  | MBFS  |
|--------|------|------|-------|-------|-------|
| Image1 | 0.9798 | 0.9797 | 0.95998 | 0.96042 | 0.99924 |
| Image2 | 0.92593 | 0.88972 | 0.8268 | 0.8624 | 0.98301 |
4. Conclusions and Prospects
The deep learn-based high resolution remote sensing image building extraction method can effectively solve the problem of building extraction in complex urban scenes. In the training process, the random patch method is used to process the whole image and obtain the result of building segmentation. Compared with other researches, the method of building extraction in this paper has a great improvement in processing speed and accuracy, and also guarantees the image integrity. At the same time, the convolutional neural network used in this study has strong expansibility in the field of intelligent recognition of high-resolution remote sensing image feature information, and can be applied to other feature extraction. Of course, due to the limitations of the author's knowledge reserve, the network and processing method used in this paper inevitably have some shortcomings, which will be further corrected in the follow-up research.

Acknowledgments
This research was funded by National natural science foundation of China (41472243); Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, MNR (KF-2018-03-020; KF-2019-04-080); Shanghai institute of geological survey (key laboratory of land subsidence detection and prevention, ministry of land and resources) open fund (KLLSMP201901); Scientific research project of the 13th five-year plan of Jilin province education department (JJKH20200999KJ).

References
[1] Fan R S, Chen Y and Xu Q 2019 A high-resolution remote sensing image building extraction method based on deep learning *Journal of Surveying and Mapping* 48 (01) 34-41.
[2] Xu Z J, Yang X B and He L M 2020 Multiscale remote sensing semantic segmentation network Computer Engineering and Applications 1-9.

[3] Geng Y L, Tao C and Shen J 2020 High-resolution remote sensing image semantic segmentation based on semi-supervised full convolution network method Acta Geodaetica et Cartographica Sinica 49 (04) 499-508.

[4] Chai D, Newsam S and Huang J 2020 Aerial image semantic segmentation using DCNN predicted distance maps ISPRS Journal of Photogrammetry and Remote Sensing 161 309-322.

[5] Lv D S 2019 An improved U-net remote sensing image building extraction Surveying and Mapping 42 (05) 231-234.

[6] Wang Z W 2019 Research and Application of Semantic Segmentation of High-Resolution Remote Sensing Image Based on Deep Learning (Beijing University of Posts and Telecommunications).

[7] Li Z Q 2019 Research on Urban Building Extraction from Deep Learning Approach (Beijing University of Civil Engineering and Architecture).

[8] Zhu G Y 2019 Research on Building Extraction from Remote Sensing Image Based on Deep Learning (Zhejiang University).

[9] Zhao W F 2018 Automatic Building Extraction from High-Resolution Optical Remote Sensing Imagery Using Deep Learning (Nanjing Normal University).

[10] Fu F, Wei J Y and Zhang L N 2018 A study of building extraction from remote sensing imagery based on convolution network Software Engineering 21 (06) 4-7.

[11] Gao Y 2018 Building Extraction from High Resolution Remote Sensing Images Based on Convolution Neural Network (Nanjing University).

[12] Zhang T 2018 Building Extraction from High Resolution Remote Sensing Images Based on Deep Learning and Line Detection (Wuhan University).

[13] Wu L Q 2018 Automatic Building Extraction of Remote Sensing Images Based on Fused Multi-Scale and Multi-Feature (Wuhan University).

[14] Long J, et al. 2014 Fully convolutional networks for semantic segmentation IEEE Transactions on Pattern Analysis & Machine Intelligence 39 (4) 640-651.

[15] Yang L, Wang H and Yan K 2019 Building extraction of multi-source data based on deep learning IEEE Vision and Computing pp 296-300.

[16] Maggiori E, Tarabalka Y, Charpiat G, et al, 2017 Can semantic labeling methods generalize to any city? The Inria aerial image labeling benchmark IEEE International Geoscience and Remote Sensing Symposium pp 3226-3229.