Remote Sensing Image Building Extraction Based on Deep Convolutional Neural network

Yang qun yi1,a, Wang wei an2, Ma xiao dong3
1Tongji University No. 1239, Siping Road, Shanghai, China 18321792670, 86
2Tongji University No. 1239, Siping Road, Shanghai, China
3Tongji University No. 1239, Siping Road, Shanghai, China
691530291@qq.com

ABSTRACT Segmentation Building extraction in high resolution remote sensing image is difficult due to different object has the same spectral feature. In this paper, we build a convolutional neural network RSB D4 base on theory of deep learning. We also proposes the overlap split method to solve the problem when we split image to speed up compute process and it causes the loss of edge information. We conduct experiments on large area Quick Bird image. Result shows that the proposed method extracts building well and has great practical value.

1. INTRODUCTION
The extraction of buildings from remote sensing images is of great significance. The fields of military investigation, urban planning, disaster emergency assessment, etc. need to quickly and accurately extract building targets from remote sensing images [1]. With the continuous improvement of the resolution of remote sensing images, remote sensing images show a lot of new features, such as rich geometry, structure, texture features, spectral refinement, multi-scale of ground objects, etc. [2]. High-resolution remote sensing images bring more details, but also bring a serious of problems, such as increasing of the noise in image, and limit the accuracy of the extraction [3]. Related researchers have been looking for algorithms with higher precision and automation for building extraction.

[4] uses the characteristics of the spectrum and shape of the building, plus the building template as an auxiliary means to extract the buildings in the Quick Bird image. Literature [5] uses the combination of various features of building spectrum, shape and texture features to extract buildings in remote sensing images with an accuracy rate of 72.7%. [6] cites the morphological building index and the shadow index, combined with the object-oriented extraction method, extracts buildings and predicts the height of the building on the 5.8m and 2.1m resolution fused images.

In [7], the deformable component model is used to treat the building as a combination of deformable components. The corresponding parameter template is obtained through training, and the effectiveness of the algorithm is proved in the image of 0.5 m resolution. [8] uses deep convolutional neural network, and the patch-based idea was used. They compare the detection effects of different network structures and extract buildings in high-resolution images.

Based on the principle of deep learning algorithm and the excellent structure of convolutional neural network in computer vision, this paper builds a CNN network for high-resolution remote sensing image building extraction. The strategy of “overlap split” is proposed to optimize the loss of edge information.
caused by splitting image to parallel computing. The algorithm is tested in a certain area of Shanghai suburbs, which has achieved relatively satisfactory results in accuracy and reliability.

2. Algorithm Flow

2.1 Convolutional neural network structure

The network structure of RSBD4 is shown in Figure 1.

![Figure 1. RSBD4 network structure.](image)

In the figure 1, C represents a convolution operation, and C 3*3*128 indicates that the convolution kernel has a size of 3*3 and outputs 128 channels. P is the pooling layer operation. In P 3*3 S2, 3*3 represents the pooling window in size of 3*3, and S2 represents the pooling stride is 2. D is atrous convolution layer, D0 3*3*1024 indicates the hole ratio is 0, convolution kernel size is 3*3, and has output of 1024 channels. Eltwise represents the addition operation. Interp represents an interpolation operation. Num represents the number of extracted categories.

The convolutional layer is the most basic part of a deep convolutional neural network. In convolutional layer, a convolution kernel of a specified size will simultaneously slid over each channel on the input image in a certain step.

In order to bring nonlinear classification to the network, each convolution layer usually has an activation function. The data given in [9] shows that relu can greatly shorten the learning period of the network compared with other activation functions. The network of this paper also adopts relu as the activation function. The formula is as follows:

\[ y = \max(0, x) \]

The first two sets of convolutional layers of RSBD4 are followed by a pooling layer. There are two main types of pooling methods, average pooling and max pooling. The average pooling can retain more background information of the image, and the max pooling can retain more texture information. In this
paper, the strong texture features of the building are considered, and the max pooling is selected as the pooling method. Assuming that the window size of a max pool is n×n, the output of the window conforms to the following formula:

\[ y = \max(x_1, x_2, x_3, \ldots, x_n^2) \]

The convolutional neural network can extract the features of an image, regardless of which position of the input image the feature point is located, and therefore has spatial position invariance, but the traditional neural network has poor adaptability to the scale change of the input image [10]. The size of buildings on remote sensing images is not fixed, which requires the network to have the ability to identify targets of different scales. In deeplabV2, the author used Atrous convolution layer [11] to solve this problem. Atrous convolution layer setting different hole rates, which makes network have the ability to identify targets of different scales. Considering that the size of the building in the remote sensing image is not fixed, RSBD4 use Atrous convolution layers with and the hole ratios are set to 0, 3, and 6, respectively.

The end of the network uses an Eltwise operation to combine the outputs of the three sets of convolutional layers. The Interp operation was used to restore the convolutional layer results to the original image size, and the bilinear interpolation algorithm was used in the experiment.

2.2 Overlap split method

The amount of data faced by remote sensing image processing tasks is usually huge. In order to solve the problem of slow processing speed caused by massive data operations, many remote sensing image processing methods adopt distributed computation [12]. This paper also uses this processing method and try to make full use of the hardware resources of the computer. The detection process is as figure 3:

In order to execute distributed computation, we first split the input image into some blocks, then use multiple compute units to process them, each compute unit will be assigned same number of blocks.

In the experiment, we found that if the block is located in the edge area of the input image, the network is not ideal for the extraction of such block, because some building is cut into two pieces and is difficult to extract such information-incomplete buildings, so this paper proposes overlap split method in the process, which avoids the loss of precision caused by splitting image.
Figure 2. Process procedure.

A block of overlap split method are shown in figure 3. Each block has two kind of region, valid region and overlap region. Although all of the buildings in the block will be extracted, but only extraction result in valid region will be adopted. Overlap region is repeated in two interfacing blocks,
which can retain the whole information of buildings in the edge of valid region. Although overlap region is repeated, valid region is unique in different blocks, they can be merged as the final result.

3. Experimental analysis and discussion
In order to verify the effectiveness of the algorithm, we do experiment in Quick Bird high-resolution remote sensing image of Shanghai suburb, and measure results in terms of correctness and completeness.

3.1 Dataset and experiment environment
The experiment is based on the ubuntu16.04 operating system and the open source deep learning framework caffe. We use NVIDIA GTX1080 graphics card for deep learning training, and the 8-core Dell T3600 workstation for various test experiments.

The remote sensing image data used in the experiment is high-resolution remote sensing image acquired by Quick Bird satellite. It has multi-spectral band of 0.6 m resolution and a full-color band of 2.4 m resolution. The training data in the suburbs of Shanghai is about 24 square kilometers. The test data is about 12 square kilometers. The ground truth of the building is manually annotated using ARCGIS 10.1. The test image is shown in figure 4:

![Figure 4. Test image.](image)

3.2 Evaluation method
For remote sensing image extraction buildings, there is currently no uniform accuracy evaluation standard. The literature [17] uses the correctness and completeness to measure the results of the segmentation:

\[
\text{correctness} = \frac{TP}{TP + FP} \\
\text{completeness} = \frac{TP}{TP + FN}
\]

Where TP represents a positive-class pixel point that is correctly classified, FP is a positive-class pixel point that is misclassified, FN is a negative-class pixel point that is misclassified. The literature [5] has adopted accuracy evaluation standards similar to the correctness rate and completeness rate. Some people also consider the particularity of remote sensing images, and adopt the relaxation accuracy rate and the relaxation recall rate for accuracy evaluation [9] [8], which relaxes the constraint conditions on the conventional accuracy rate and recall rate.

We consider that the extraction of buildings requires the estimation of the area and the extraction of the precise contours in many applications, so we adopt correctness rate and the completeness rate to evaluate experiment.

3.3 result analyze
The test is performed on the above data set by using the network RSBD4 and the overlapping sliding window detection method proposed in this paper. At the same time, in the same data set and environment, the deep learning network segmentation network deeplabV2[10] and the SVM-based object-oriented extraction algorithm and rule-based object-oriented extraction algorithm in ENVI software were selected for comparison experiments. Among them, deeplabV2 selects the best experimental results under the condition of network training convergence. Based on SVM and rule-based object-oriented extraction algorithm, after repeated debugging, the optimal parameters are selected to obtain experimental results. The specific data is shown in Table 1.

Table 1. Comparison between RSBD4 and deeplabv2

| Network   | correctness | completeness |
|-----------|-------------|--------------|
| deeplabV2 | 78.93%      | 79.06%       |
| RSBD-4    | 80.54%      | 83.18%       |

The experimental results show that the correctness rate of RSBD4 reaches 80.54%, and the completeness rate reaches 83.18%. Compared with the deeplabV2, the two have increased by nearly 2 percentage points and 4 percentage points respectively.

By observing the extraction result, we found that compared with deeplabV2, the extracted building edge is more refined. For small buildings, as shown in Figure 6, there is a better extraction result.

4. Conclusion

We build a deep convolutional neural network RSBD4 for the extraction of high-resolution remote sensing image buildings, use the excellent structure of other networks, such as the atrous convolution layer of deeplabV2. We split image into blocks and use distributed computing to accelerate computation. We propose overlap split method to solve the problem of edge information loss caused by splitting the image. The experiment shows that our correctness rate is 80.54% and completeness rate is 83.18%. In next step we can increase the amount of data for training, adjust the network structure, speed up the training time, and improve the efficiency of detection.
REFERENCES

[1] Wang Jun, Qin Qi ming, Ye Xin, et al. A Survey of Building Extraction Methods from Optical High Resolution Remote Sensing Imagery [J]. Remote Sensing Technology and Application, 2016, 31

[2] Li D R, Tong Q X, Li R X, et al. Current issues in high-resolution Earth observation technology. Sci China Earth Sci, 2012, 55: 1043–1051, doi: 10.1007/s11430-012-4445-9

[3] ZHANG Qing yun, ZHAO Dong. Research on Methods of Building Extraction from High Resolution Remote Sensing Images [J] GEOMATICS & SPATIAL INFORMATION TECHNOLOGY, 2015, 38(04): 74-78

[4] WU Wei, LUO Jiancheng, SHEN Zhanfeng, ZHU Zhiwen. Building Extraction from High Resolution Remote Sensing Imagery Based on Spatial Spectral Method [J] Geomatics and Information Science of Wuhan University, 2012, 37(07): 800-805.

[5] Xiaoying Jin, Curt H. Davis. Automated Building Extraction from High-Resolution Satellite Imagery in Urban Areas Using Structural, Contextual, and Spectral Information [J] EURASIP Journal on Advances in Signal Processing, 2005, Vol.2005 (14), pp.1-11

[6] Fu Qiankun, Wu Bo, Wang Xiaqin, et al. Building Extraction and Its Height Estimation over Urban Areas based on Morphological Building Index [J]. Remote Sensing Technology and Application, 2015, 30(1): 148 – 154

[7] SHEN Jiajie, PAN Li, HU Xiangyun, Building Detection from High Resolution Remote Sensing Imagery Based on a Deformable Part Model [J] Geomatics and Information Science of Wuhan University 2017, 42(09): 1285-1291.

[8] Mnih, Volodymyr. "Machine Learning for Aerial Image Labeling." Doctoral (2013).

[9] Xu L, Choy C S, Li Y W. Deep sparse rectifier neural networks for speech denoising [C] IEEE International Workshop on Acoustic Signal Enhancement. IEEE, 2016: 1-5.

[10] LIAN Zifeng, JING Xiaojun, SUN Songlin, HUANG Hai. Multi-Scale Convolutional Neural Network Model with Multilayer Maxout Networks [J] Journal of Beijing University of Posts and Telecommunications. 2016, 39(05): 1-5+32

[11] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. 2016, PP(99): 834-848.

[12] YANG Haiping, SEHN Zhanfeng, LUO Jiancheng and WU Wei, Recent Developments in High Performance GeoComputation for Massive Remote Sensing Data [J] JOURNAL OF GEO-INFORMATION SCIENCE. 2013, 15(01): 128-136