Change detection of built-up areas based on ensemble learning

Lei Chen1,2,3,*, Yingcheng Li2,3,4, Zhongyuan Geng2,3, Xilin Li3,5, Yanli Xue2,3,4, Guangliang Wang3,4 and Yanhui Wang1

1College of Resource Environment and Tourism, Capital Normal University, Beijing, 100048, China
2Key Laboratory for Aerial Remote Sensing Technology of Ministry of Natural Resources, Beijing, 100039, China
3China TopRS Technology Co. Ltd., Beijing, 100039, China
4Beijing Low Altitude Remote Sensing Data Processing Engineering Technology Research Center, Beijing, 100039, China
5Wuhan Sunmap Remote Sensing Technology Co. Ltd., Wuhan, 430223, China

* Corresponding author: cl_toprs@hotmail.com

Abstract. The paper presents a new method for the change detection of built-up areas based on ensemble learning. The method selects the feature generation part of VGG network to replace the encoder of UNet network, and makes full use of the advantages of VGG network in feature interpretation to construct a new network VGG-UNet, so as to improve the information extraction accuracy of built-up areas. First, the slice sample sets containing built-up areas and non-built-up areas were established. Secondly, the model was trained and tested to obtain several optimal models with high precision. Thirdly, these optimal models were used to extract the change information of built-up areas in the experimental area. Finally, the results were ensemble optimized and the accuracy was verified by comparing with the field measurements. The experimental results in Xinjiang demonstrate the advantages, feasibility and reliability of the proposed method. Moreover, the ensemble optimization of multiple model results can further improve the change detection accuracy of built-up areas.

1. Introduction

Built-up areas refer to the non-agricultural production and construction areas formed by human activities, including houses, structures, road networks, etc., which are characterized by frequent changes and environmental impacts [1]. In order to make rational use of the limited land resources, the administrative departments need to extract the change information of built-up areas and provide the scientific guidance for urban and rural planning periodically.

Change detection methods using remote sensing images are as follows: 1) Manual interpretation method, based on the data processing experience of technical personnel, adopts human-computer interaction to extract the change information. 2) Machine learning algorithm, which can automatically interpret the shallow features of remote sensing images and complete the change detection [2].

With the launch of more and more satellites, a large number of remote sensing images can be obtained by the domestic and foreign scholars, but the above methods can’t meet the increasing
demand of remote sensing images processing. Deep learning algorithm has a more complex network model, which can automatically process large amounts of image features and obtain the optimal model parameters. Nowadays, it is widely used in land use and land cover classification, object detection, scene classification and change detection [3, 4].

In terms of the change detection of built-up areas, the existing researches focus more on the change detection of single type of built-up area, and less on the comprehensive extraction of multiple types of built-up areas information. In literature [5], the deep convolutional neural network was proposed to integrate panchromatic and multispectral images to extract the change information of built-up areas, but the accuracy was poor in the boundary, shadow and vegetation covered areas. In literature [6], the experimental results showed that the deep learning method had a high accuracy in the extraction of urban buildings and roads, but poor accuracy for rural areas, scattered buildings, moving blocks. In literature [7], the scholars pointed out that the representative samples were the foundation of deep learning, and the difference of built-up areas had a significant impact on the prediction accuracy of the model.

In this paper, we proposed a new method for the information extraction of built-up areas based on ensemble learning, where the encoder of UNet [8] was modified combining the advantages of feature generation network VGG [9] in image classification to improve the change detection accuracy of built-up areas.

The rest of this paper is organized as follows. Section 2 will describe the proposed method in detail. Section 3 will present the experimental results and discussions in Xinjiang using multi-source remote sensing images. In section 4, the conclusions are drawn and the future research directions are proposed.

2. Method

The method we proposed is shown as follows in Fig. 1.

![Flow chart of the method proposed](image)

**Fig. 1. Flow chart of the method we proposed.**

2.1. Sample sets production

Multi-source remote sensing images covering the experimental area were collected, and the label files containing the built-up areas and the non-built-up areas were made by combining the vector polygon. Then the corresponding relations between the label files, the range files and the images were
constructed. \( N \) points were selected for sampling and the slice of \( 256 \times 256 \) pixels were made with each point as the center. Finally, data enhancement (translation, flipping, scaling, etc.), normalization were carried out to produce the train, valid and test sample sets containing the built-up and non-built-up regions.

2.2 Model construction

UNet network is a encoder-decoder structure [8], which is mainly composed of convolution layers, pooling layers, activation functions and deconvolution layers. This paper takes UNet as the basic framework to construct the model. The feature generation network VGG is brought into the encoder to improve the efficiency and precision of image features interpretation, so as to improve the semantic segmentation precision of UNet:

1) The first 6 convolution layers of VGG11 network are taken as the encoder of UNet network. The input terminal was set to 3 channels (the input image was in RGB format), the two \( 3 \times 3 \times 1024 \) convolution of the bottleneck part was convered to one \( 3 \times 3 \times 512 \) convolution operation, and the output channel of the first upampling of the decoder was changed from 512 to 256. The model obtained is called as VGG11-UNet. It can effectively reduce the number of convolution layers and the model parameters that need to be trained.

2) The first 9 convolution layers of VGG16 network are taken as the encoder of UNet network and the input terminal was also set to 3 channels. The model obtained is called as VGG16-UNet. The advantages of this network in image interpretation can be fully utilized to reduce the influence of other factors on the accuracy of model interpretation.

2.3 Image mosaicking

For the test of a certain region, the test sample sets are made by sequentially cropping the images, and there is an overlapping area between two adjacent slices. The test results are grayscale images, which need to be mosaicked to get the ground feature information in the whole region. Finally the slices are geocoded and are mosaicked according to the geographic coordinates, in which the pixel gray value of the overlapping area is the average value of the pixel gray value of the adjacent slices.

2.4 Ensemble optimization

In order to improve the generalization of deep learning model, the ensemble optimization strategy is adopted to vote on the ground features obtained by each model of each pixel, so as to judge whether the ground features in the pixel are built-up areas. In this paper, the two groups of models established above were used for testing, and the grayscale images of the two optimal results of each model were selected for ensemble optimization, so as to obtain the scopes of built-up areas.

2.5 Image post-processing

There are many independent built-up areas information in the semantic segmentation results based on pixels. In this paper, morphological corrosion and expansion are adopted to carry out post-processing of the built-up areas classification map, and vectorization of raster data is used to extract the boundary information. Finally, the change information of built-up areas can be obtained by comparing these classification maps.

2.6 Precision evaluation

In order to evaluate the reliability of the proposed method, this paper uses confusion matrix to estimate the detection accuracy of built-up areas. The confusion matrix of the built-up areas in the experimental area is shown in Table 1 below:
Table 1. Confusion matrix of the built-up areas.

|                     | Predicted built-up area | Predicted non-built-up area |
|---------------------|-------------------------|-----------------------------|
| Actual built-up area| \( N_{T,T} \)           | \( N_{T,F} \)               |
| Actual non-built-up area| \( N_{F,T} \)           | \( N_{F,F} \)               |

where \( N_{T,T} \) means the number of pixels where the actual surface is the built-up area and the model prediction is also the built-up area, \( N_{T,F} \) means the number of pixels where the actual built-up area and predicted non-built-up area, \( N_{F,T} \) means the number of pixels where the actual non-built-up area and predicted built-up area, \( N_{F,F} \) means the number of pixels where the actual non-built-up area and predicted non-built-up area.

The follows precision indexed were calculated to quantitatively evaluate the test precision of the proposed model:

\[
\text{True positive rate (TP)} = \frac{N_{T,T}}{N_{T,T} + N_{F,T}} \\
\text{False positive rate (FP)} = \frac{N_{F,T}}{N_{F,T} + N_{F,F}} \\
\text{True negative rate (TN)} = \frac{N_{T,F}}{N_{T,F} + N_{T,F}} \\
\text{Accuracy} = \frac{N_{T,T} + N_{F,F}}{N_{T,T} + N_{F,T} + N_{T,F} + N_{F,F}}
\]

3. Results & Discussions

Xinjiang has a vast territory and rich land resources. In recent years, the process of industrialization and urbanization in Xinjiang promotes the continuous expansion of built-up areas, which aggravates the contradiction between built-up areas and non-built-up areas [10].

In order to extract the change information of built-up areas effectively, multi-source remote sensing images (GF-1, GF-2, GJ-1, BJ-2, etc) covering 8 counties were collected as the experimental data in this paper. 100 thousands 3-channel RGB image slices with a resolution of 1m and the size of 256×256 pixels were made by combining with the polygon vector files of each county. Enhancement was performed on the slices, that is, the slices were randomly translated, flipped, rotated, scaled and other operations to improve the diversity of the samples. The slices were normalized, that is, the mean value and variance of the pixel gray value of the sample slices were set as a unified value, so as to reduce the influence of hue on training. Finally, 80 thousands slices were selected as train sample sets, 10 thousands slices as valid sample sets and 10 thousands slices as test sample sets to train and test the deep learning model and obtain the optimal parameters.

In this paper, two regions in Urumqi County of Xinjiang were taken as experimental areas, and the remote sensing images acquired in 2017 and 2019 covering the experimental areas were used to verify the efficiency and accuracy of the proposed method. Several optimal models above obtained were used to extract the built-up areas and non-built-up areas in the experimental area, so as to obtain the change information of built-up areas. Then, the above experimental results were ensemble optimized and the accuracy was verified by comparing with the field measurements.

3.1 Experiment 1

Two remote sensing images of experiment 1 are shown in Fig. 2. Figure (a) is the image acquired in 2017 and Figure (b) is the image acquired in 2019.
UNet, VGG11-UNet, VGG16-UNet network were used to extract the change information of built-up areas in experiment 1, respectively. The results of two VGG-UNet were ensemble optimized to obtain the modified change information. Experimental results are shown in Fig. 3.

| Method        | TP (%) | FP (%) | TN (%) | Accuracy (%) |
|---------------|--------|--------|--------|--------------|
| UNet          | 93.43  | 6.57   | 1.36   | 97.74        |
| VGG11-UNet    | 93.80  | 6.60   | 1.37   | 97.72        |
| VGG16-UNet    | 93.89  | 6.10   | 1.26   | 97.90        |
| Proposed method | 94.67  | 5.33   | 1.11   | 98.17        |

In order to analyze the accuracy of the experimental results, four precision indexes in Section 2.6 were calculated as shown in the Table 2. In this experiment, the difference between built-up areas and non-built-up areas is obvious, and the detection accuracy of VGG-UNet is only slightly improved comparing with the results of UNet. The ensemble learning of VGG-UNet can improve the detection accuracy of change information to a certain extent.
3.2 Experiment 2
Two remote sensing images of experiment 2 are shown in Fig. 4. Figure (a) is the image acquired in 2017 and Figure (b) is the image acquired in 2019.

![Fig. 4. Two remote sensing images of experiment 2.](image)

The change information of built-up areas was extracted using the same method as Experiment 1. Experimental results are shown in Fig. 5.

![Fig. 5. Change detection results obtained by four methods.](image)

Four precision indexes were calculated as shown in the Table 3. In this experiment, the image features of some built-up areas are not obvious. The change detection accuracy of VGG-UNet is obviously better than that of UNet, and the accuracy of ensemble learning is also slightly improved compared with that of VGG-UNet alone.

| Method     | TP (%) | FP (%) | TN (%) | Accuracy (%) |
|------------|--------|--------|--------|--------------|
| UNet       | 94.34  | 5.99   | 5.31   | 94.36        |
| VGG11-UNet | 97.05  | 3.04   | 2.69   | 97.15        |
| VGG16-UNet | 98.87  | 1.14   | 1.01   | 98.93        |
| Proposed method | 99.92 | 0.07   | 0.06   | 99.93        |
4. Conclusions
In this paper, we have proposed a new ensemble learning technology to extract the change information of built-up areas. Experimental results show that the precision of this method in distinguishing built-up areas from non-built-up areas is better than that of the single modified VGG-UNet model, which can also effectively improve the change detection accuracy of built-up areas.

Although we have made progress in the change detection of built-up areas, there are still some problems to be studied further. In the future, we will research and construct the deep learning model to simultaneously analyze two remote sensing images acquired at different times and directly extract the change information of built-up areas. In addition, the change information will be studied to determine the type of built-up areas.

Acknowledgments
This work was supported by the National Key Research and Development Program of China (2018YFF0215301), Yunnan Provincial Key Laboratory of Forensic Science (2020SKF01).

References
[1] Du, P. J., Liang, H., Wang, X., Li, Y. F. (2019) A new urban built-up land extraction method based on ensemble learning. Environmental Monitoring and Forewarning, 11(05): 39-45.
[2] Zhang, Y. H., Zhang, J. X., Zhang, X. F., Wu, H. A., Guo, M. (2015) Land cover classification from polarimetric SAR data based on image segmentation and decision trees. Can. J. Remote Sens., 41(1): 40-50.
[3] Ma, L., Liu, Y., Zhang, X. L., Ye, Y. X., Yin, G. F., Johnson, B. A. (2019) Deep learning in remote sensing applications: A meta-analysis and review. ISPRS J. Photogramm., 152, 166-177.
[4] Vetrivel, A., Gerke, M., Kerle, N., Nex, F., Vosselman, G. (2018) Disaster damage detection through synergistic use of deep learning and 3D point cloud features derived from very high resolution oblique aerial images, and multiple-kernel-learning. ISPRS J. Photogramm., 140, 45-59.
[5] Tan, Y. H., Ren, F. F., Xiong, S. Z. (2017) Automatic extraction of built-up area based on deep convolution neural network. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth. pp. 3333-3336.
[6] Li, W., Liu, H., Wang, Y., Li, Z., Jia, Y., Gui, G. (2019) Deep learning-based classification methods for remote sensing images in urban built-up areas. IEEE Access, 7: 36274-36284.
[7] Wu, H. P., Huang, S. C. (2019) Research on new construction land information extraction based on deep learning: Innovation exploration of the national project of land use monitoring via remote sensing. Remote Sensing for Land and Resources, 31(4): 159-166.
[8] Ronneberger, O., Fischer, P., Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 [cs.CV].
[9] Simonyan, K., Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. Comput. Sci..
[10] Han, D. J., Du, H. R. (2020) Suitability Analysis and Spatial Allocation Optimization Strategy of Xinjiang Construction Land. Areal Research and Development, 39(4): 122-126.