Building Semantic Information Extraction Based on Full Convolution Neural Network and Parameter Transfer

Yue Zhang\textsuperscript{1,a*}, Congna Liu\textsuperscript{1,b} and Ruipeng Tang\textsuperscript{1,c}

\textsuperscript{1}College of Geomatics and Geoinformation, Guilin university of technology, Guilin, Guangxi, China
\textsuperscript{a}email: 2512919923@qq.com, \textsuperscript{b}email: 282842120@qq.com, \textsuperscript{c}email: 370514735@qq.com
\textsuperscript{*Corresponding author’s e-mail: chxy@glut.edu.cn}

Abstract. Aiming at the problem that the generalization ability of deep convolution neural network model is limited, this paper extracted the semantic information of buildings in different regions by combining the transfer learning method. Taking the Massachusetts aerial image building data set as an example, the layer depth structure of SegNet network model is improved, and a new model is trained by migrating the parameters of the improved model and combining with a small number of Guilin building image data samples. The extraction results were analyzed and compared. The results show that, by transferring the parameters of the trained model, the Kappa coefficient and F1 score of buildings in Guilin area reached about 0.8.

1. Introduction
The extraction of building information has a very important application value for urban and rural construction and management. Nowadays, deep convolution neural network is a hot spot in the field of image recognition. The training of deep neural network models often requires a large number of sample data, Simard et al. [1] improved the performance of the model through variant amplification of training samples in MNIST, which proved the effectiveness of the method, so it is a feasible method to expand and enhance the limited original data. Under the premise of given training samples, increasing the number of network layers can improve the performance of the model to some extent, but too deep model structure may lead to overfitting [2]. Therefore, how to ensure the accuracy of the model with less parameters and short training time is one of the important contents in this field [2,3].

When the difference between training set and test set is relatively large, the model is often unable to obtain a good classification effect. Therefore, the transfer learning method is introduced, which can build a more accurate model with less cost. Qiu Ningjia et al. [4] achieved good classification results by principal component analysis combined with transfer learning model. Teng Wenxiu et al. [5] take the network parameters trained on ImageNet as the initial parameters of large convolutional neural network model, and effectively extract the characteristics of tree species image. Zhang Wentian et al. [6] combined with small data samples and taking BP neural network as the framework, effectively classified the data in the target domain by transferring the useful parameter information from the source domain to the target domain.

According to the above rules, this paper takes the remote sensing images of buildings in Massachusetts as the experimental data to explore the SegNet layer depth structure which has the best
fitting effect on the data. Finally, by migrating the parameters of the trained model, the model can effectively extract the semantic information of buildings in Guilin by using only small samples.

2. Materials and Methods

2.1. Experimental process
As shown in the technical route in Figure 1, the experiment uses building image data from Massachusetts to explore the SegNet layer depth structure with the best fitting effect on the data by increasing or decreasing the number of layers of the model. Then the new model is trained by migrating the parameters and combining a small number of data samples in Guilin. Finally, this experiment designed a comparative experiment to explore the difference of model classification performance before and after parameter migration.

2.2. Experimental data

2.2.1. Data selection
The same Massachusetts aerial image building open data set as in reference [7] is selected as the experimental data of the improved layer depth structure. The data set consists of 1m high-resolution image and contains real label data on the ground, which is divided into "buildings" and "background". In order to ensure the validity of the experimental results, 10 of the 13 images with the size of 1500*1500 were selected as the original sample images, and 3 images were selected as the test set images.

High-resolution images of buildings in Guilin obtained by Google Earth were selected as the experimental data of parameter migration. The data resolution is 0.3m. The real label on the ground is drawn by using ENVI5.3 software through manual visual interpretation, which is divided into "building" and "background". 5 of the 8 images with the size of 1000*1000 were selected as the original sample images and 3 images were selected as the test set images.

2.2.2. Expansion and enhancement of sample set
By means of random cropping, the 10 original sample images of Massachusetts were expanded to 50000 samples and 5 original sample images in Guilin were expanded to 10000 samples. In order to ensure the richness of sample information, the cropping size is set to 256*256. The distribution was unified based on the ratio of 4/1 between the training set and the verification set.
The expansion and enhancement can enrich the information of the sample set. Thus reducing the occurrence of overfitting phenomenon to a certain extent and improving the classification performance of the network model. As shown in Figure 2, the image data of Massachusetts is taken as an example to enhance the sample image after cropping by means of rotation, mirroring, blurring and add noise.

![Figure 2. Sample image enhancement](image)

### 2.3. Experiment design

#### 2.3.1. Improvement of layer depth structure

Different classification tasks are suitable for different depth structure of network model layer. In order to explore the influence of model layer depth on building extraction effect, the SegNet convolution neural network models with 8, 10, 12, 14, 16, 18 layers were trained based on the experimental data of Massachusetts building images. According to the performance of computer hardware, the epochs is set as 15 and the batch size is set as 6 using the control variable method.

The building information is extracted from three images of test set by using different layer depth models trained. The extracted results was analyzed by using the confusion matrix. The average value of kappa coefficient of each test set image extraction result is calculated to determine the best network layer depth structure for Massachusetts buildings.

#### 2.3.2. Experiment of model parameter transfer

In order to explore the influence of parameter transfer on the classification performance of the model, three groups of comparative experiments were designed as follows:

1. The improved model obtained in 2.3.1 is used to extract the buildings in Guilin area directly, and evaluate the ability of this model to extract the buildings in different areas.

2. The retraining model is based on the image data of Guilin area. Since the training sample set of data in Guilin was relatively small, the epochs was set as 5 and the batch size was set as 6 to prevent overfitting. The number of model layers was consistent with the improved model of layer depth.

3. The parameters of the improved model trained by Massachusetts image data are extracted as the initial values of the parameters of the new model. And the network model was trained by combining the image data in Guilin. The settings of model layer depth, epochs and batch size are the same as the experiment (2).

The model obtained from the above experiment was used to extract buildings from three test images of Guilin image data. The extracted results was analyzed by the confusion matrix, and the
Kappa coefficient and F1 score were calculated. The influence of parameter transfer on the classification ability of the model was analyzed combined with error diagram.

### Table 1. Comparative experimental design

| Sample set     | Test set   | Number of samples | Epochs | Parameter transfer |
|----------------|------------|-------------------|--------|-------------------|
| Experiment 1   | Massachusetts | Guilin            | 50000  | 15                | No                 |
| Experiment 2   | Guilin     | Guilin            | 10000  | 5                 | No                 |
| Experiment 3   | Guilin     | Guilin            | 10000  | 5                 | Yes                |

3. Results & Discussion

#### 3.1. Result of improvement of layer depth structure

The Kappa coefficient changes of the classification results of the three test sets of images by different layer depth structure models are shown in Figure 3. The model layer depth with the best extraction effect on the test image A, B and C is 12 layers, 14 layers and 10 layers respectively, and the Kappa coefficient reaches 0.7176, 0.6468 and 0.7537 respectively. The results show that different test images may have different optimal layer depth models, so the comprehensive classification performance of each layer depth model is evaluated by calculating the average value. It is concluded that the average extraction effect of three test images is the best when the layer depth of the model is 14. The average Kappa coefficient reached 0.7036 and the average F1 score reached 0.7440. Therefore, the 14 layer network model is the improved layer depth structure model obtained in this experiment. It takes 530 minutes to train this model. The error diagram of the building extraction results of the three images of the test set in Massachusetts is shown in Figure 4.

![Figure 3. Extraction results of different layer depth models](image)
3.2. Comparison of model parameter transfer

The Kappa coefficient and F1 score of the extracted results from Experiment 1, 2 and 3 are shown in Table 2, and the error diagram is shown in Figure 5. The experimental results show that the improved layer depth model trained by Massachusetts image data has a good ability to extract buildings in this area. But due to the differences in the characteristics and distribution of buildings in Guilin and Massachusetts. Therefore, this model cannot extract buildings in Guilin effectively. The results of experiment 2 and experiment 3 were compared, it shows that the new model can improve the extraction ability of Guilin data by transferring the parameters trained from Massachusetts data, which also proves that the buildings in two different regions have similar characteristic information while there are differences. The training time of Experiment 2 and Experiment 3 was 35 minutes. Under the same training time, the Kappa coefficient and F1 score of the classification results were increased by about 0.1 through the transfer of model parameters.

| Test image of Guilin | Experiment 1   | Experiment 2   | Experiment 3   |
|---------------------|----------------|----------------|----------------|
|                     | Kappa  | F1 Score | Kappa  | F1 Score | Kappa  | F1 Score |
| A                   | 0.2569 | 0.3809   | 0.6765 | 0.7198   | 0.7909 | 0.8200   |
| B                   | 0.3007 | 0.4083   | 0.7745 | 0.8191   | 0.8154 | 0.8538   |
| C                   | 0.1577 | 0.4092   | 0.7649 | 0.8175   | 0.8118 | 0.8556   |

Figure 4. Error diagram of extraction result

Figure 5. Comparative experiments results
4. Conclusions
This paper focuses on the influence of SegNet model parameters migration on building extraction effect in different areas, and explores the influence of layer depth structure change on model performance. The results show that the 14 layer SegNet network model has the best extraction effect for buildings in Massachusetts. The Kappa coefficient and F1 score both reach about 0.7. It shows that the improvement of the layer depth structure can improve the ability of the model to extract specific data sets. Through the migration of model parameters, the extraction accuracy of Guilin area buildings is further improved. The Kappa coefficient and F1 score both reached about 0.8, which was about 0.1 higher than that without the migration of parameters. It shows that the migration of model parameters can enhance the extraction ability of the model for new data under the same cost.

In order to make the model fit the data more effectively, only the number of layers of the model was changed in this study. Therefore, the following work will further study the influence of convolution kernel size, sample size and epochs on the classification performance of the model. The effect of model parameter migration is also related to the sample size of new data and the epochs. If the sample size of new data is large and the epochs is large, the effect of the migration model parameters may not be to retrain a new model. Therefore, the next work will study the upper limit of the sample size and the epochs.

This study has a certain reference value for the effective extraction of information between different data. Through the improvement of layer depth structure and the transfer of model parameters, the classification performance of convolution neural network model can be enhanced to a certain extent, and the accuracy of building information extraction can be improved. Therefore, it provides important reference for urban and rural construction and planning.

Acknowledgments
This work was supported by the Guilin Scientific Research and Technology Development Plan(20190601), Guilin Guilin University of Technology Ph.D Scientific research initial funding(002401003316); Chongqing basic science and advanced technology research (cstc2015jcyjBX0023). Special thanks to the College of Geomatics and Geoinformation, Guilin University of Technology for the support of our work.
References
[1] Simard, P.Y., Steinkraus, D., Platt, J.C. (2003) Best practices for convolutional neural networks applied to visual document analysis. Proceedings of the International Conference on Document Analysis and Recognition., Edinburgh, UK, 2003: 958.
[2] Zhang, S., Gong, Y.H., Wang, J.J. (2019) The Development of Deep Convolution Neural Network and Its Applications on Computer Vision. Chinese Journal of Computers., 42(03):453-482.
[3] Li, B.J., Liu, K., Gu, J.J., Jiang, W.Z. (2021) Review of Convolutional Neural Network. Computer Era., 40(6):1229-1251.
[4] Qiu, N.J., Wang, X.X., Wang, P., Zhou, S.C., Wang, Y.C. (2020) Research on Convolutional Neural Network Algorithm Combined with Transfer Learning Model. Combined with Transfer Learning Model., 56(05):43-48.
[5] Teng, W.X., Wen, X.R., Wang, N., Shi, H.H. (2019) Tree Species Classification and Mapping Based on Deep Transfer Learning with Unmanned Aerial Vehicle High Resolution Images. Laser & Optoelectronics Progress., 56(07):277-286.
[6] Zhang, W.T. (2019) Research on Parameter Transfer Learning Algorithm Based on BP Neural Network. Computer Knowledge and Technology., 15(04):189-191.
[7] Mnih, V. (2013) Machine learning for aerial image labelin. Canada:University of Toronto.