Research on Surface Information Extraction Based on Deep Learning and Transfer Learning

Zhen Chen[1*], Yiyang Zheng[2]

[1]School of Earth Sciences and Resources, China University of Geosciences (Beijing), Beijing, China
[2]Mudanjiang Natural Resources Survey Center, China Geological Survey, Mudanjiang, China
Email: *chenzhen@cugb.edu.cn

Abstract

The land cover types in South China are varied, and the terrain is undulating, and the area of different land types is small, and the remote sensing monitoring work was difficult. In order to solve these problems, an automatic classification method based on transfer learning and convolutional neural network model was established in this paper, with a total classification accuracy of 98.1611%. This paper proposes a land use classification remote sensing method based on deep learning, which improved the automation level and monitoring accuracy of complex land surface remote sensing monitoring in South China, and it provided technical support for the land consolidation work in China.

Keywords

Land Classification, Convolution Neural Network, Transfer Learning

1. Introduction

There are many methods currently applied to remote sensing image classification, among which convolutional neural network is the one with better application effect. The convolutional neural network is a feed-forward neural network based on a deep layered structure of convolution operations. Convolutional neural network is the best among many deep learning algorithms. Convolutional neural network is a translation-invariant artificial neural network, which is characterized by the ability to achieve translation-invariant classification (Zhang et al., 1988).

Remote sensing land classification has always been an important content of remote sensing research. The technology of remote sensing land classification has also been greatly developed with the introduction of new technologies such

as machine learning and deep learning. In deep learning, convolutional neural network is a model more suitable for remote sensing land classification. Convolutional neural network is a feedforward network with a multi-layer structure, and it is a translation-invariant multi-layer neural network. At present, the application in the field of remote sensing is also expanding, and the classification accuracy has been improved compared with other traditional machine learning methods.

Transfer learning is a new machine learning method that uses existing prior knowledge to solve recognition problems in different related fields. Ballard et al. (2007) proposed a method to select potentially useful samples, construct the corresponding normalized feature space, and transform the classification problem of multispectral and hyperspectral remote sensing images into the problem of classifying the selected samples. Weng (2006) proposed that the diversity and uncertainty in the transfer learning process can be used to filter samples from the source sample area to the target sample area, so as to implement the classification of remote sensing images. In addition, transfer learning methods are widely used in large-scale remote sensing image classification applications where training samples are insufficient (Li, 1992). Literature (Wang et al., 2013) used the nonlinear feature method to extract features from samples from the source domain to the target domain, and verified its applicability to remote sensing image classification. Transfer learning is adapted to remote sensing image classification in large areas where training samples are insufficient.

Convolutional neural network (CNN) has a wide range of requirements for input data formats and can handle multi-dimensional data. For example: 1-dimensional CNN accepts 1 - 2 dimensional data; 1-dimensional data is generally time or spectrum data; 2-dimensional CNN accepts 2-3D data; 3D CNN accepts 4D data (Ng et al., 2018). Because CNN is widely used in remote sensing image processing, CNN is generally set to three-dimensional input data (three-channel input data) by default, that is, RGB channel.

CNN is similar to BP neural network, and CNN needs to normalize the input data. For example, use the Sigmoid kernel function to normalize the gray value of the remote sensing digital image distributed between [0, 255] to [0, 1] (Ng et al., 2018). After normalizing the input data, errors caused by learning and computing data of different dimensions are avoided.

Compared with the input layer and output layer, the components located in the center of the convolutional neural network structure are collectively called hidden layers, including convolutional layers, pooling layers, fully connected layers, Inception modules, residual blocks, etc.

The advantage of transfer learning is that the parameters of the trained model (pre-trained model) are moved to the new model to help the new model train. Transfer learning is an idea, not a method. Specifically, and it is the application of knowledge or patterns learned in one domain or task to different but related domains or problems. Prior to fine-tuning, the initial performance of the model

was higher; During the training process, the model was improved faster. After the training, the resulting model converged better.

The advantages of convolutional neural networks are more obvious when the input of the network is an image, so that the image can be directly used as the input of the network, avoiding the complicated process of feature extraction and data reconstruction in the traditional recognition algorithm, and having great advantages in the processing of two-dimensional images. The combination of transfer learning and convolutional neural network not only avoids too long training time, but also automatically realizes the feature extraction of two-dimensional images.

2. Overview of the Study Area

The coordinates of Dongguan and Shenzhen in Guangdong province, which are located in the coastal areas of southern China, which are typical areas of tropical monsoon climate with sufficient sunlight, warm climate, Sufficient rainfall is suitable for the growth of tropical crops, and it is located on the east side of the Pearl River Estuary, close to Shenzhen and Hong Kong, with superior transportation and geographical location, and belongs to the economically developed area in the south.

The northwestern part of Dongguan, Shenzhen and Shenzhen is a delta plain formed by the alluvial of the Dongjiang River, the northeastern part is close to the Dongjiang River, with gentle terrain and dense rivers, and the southwestern part is an alluvial plain along the Pearl River estuary. The geographic location of the study area (Shenzhen Guangming North Land Development and Complementary Land Use Project Area) is shown in Figure 1.

3. Data Preprocessing

The Shenzhen city of Guangdong province were taken as the study area. The Landsat 8 image of the study area (February 15, 2022) were used in the paper, which were preprocessed such as atmospheric correction, orthorectification, geometric correction, fusion, and coordinate transformation, and were visually interpreted. The remote sensing image of the study area is shown in Figure 2.

4. Research Methods

4.1. Convolutional Neural Network Structure

Convolutional neural network (CNN) is a special type of neural network which has the ability of hierarchical abstract representation so as to be suitable for computer image processing and speech recognition and other fields. The CNN consists of multi-layer interconnected neural networks, in which low-, mid- and high-level features of remote sensing images are extracted hierarchically. A typical CNN framework contains two main layers including convolutional layers and pooling layers, both of which are called the convolutional basis of convolutional neural networks (Chollet, 2017). Convolutional layers have filtering functions
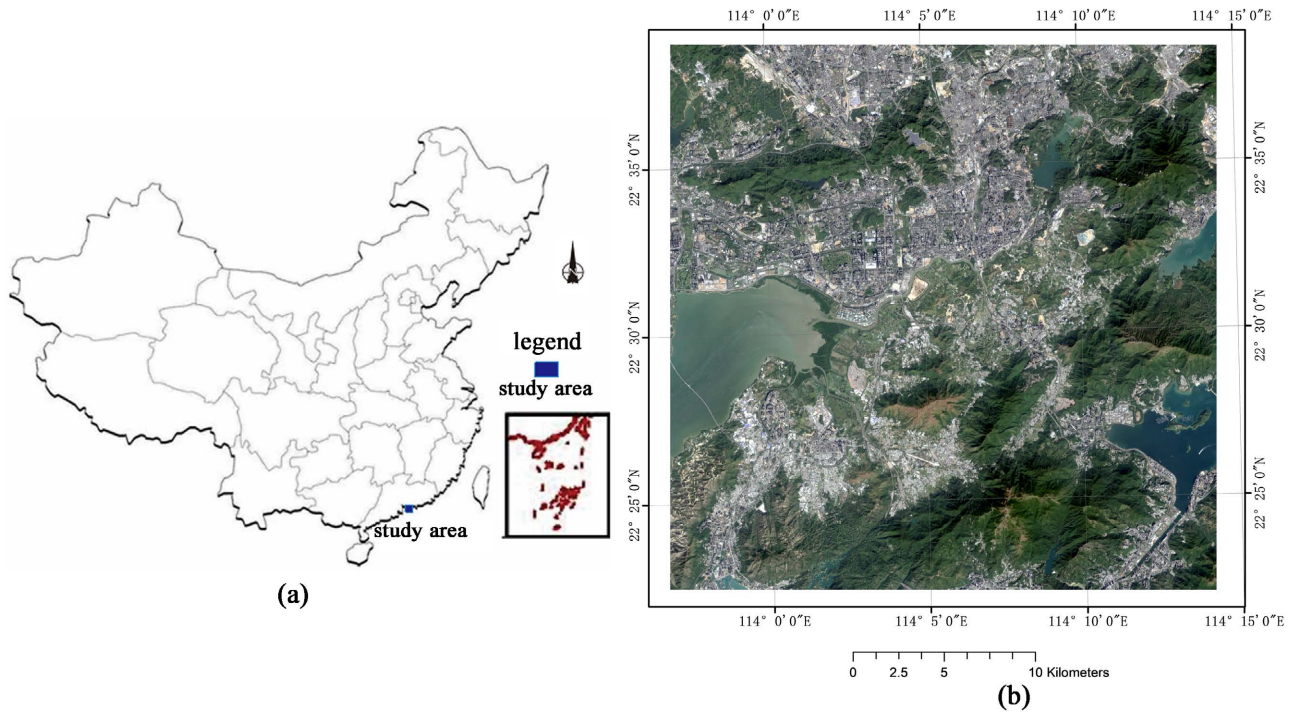
Figure 1. (a) Geographical location of the study area; (b) The remote sensing image of the study area.
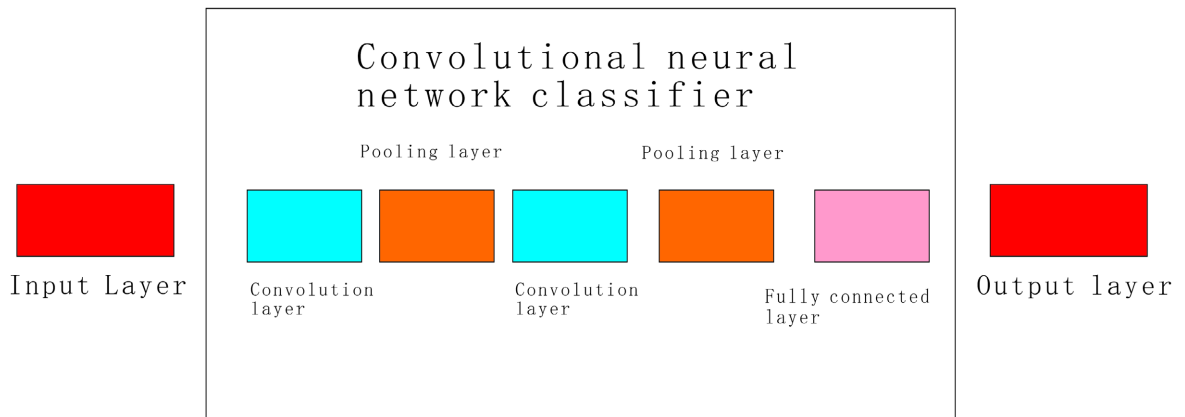

Figure 2. The structure of the convolutional neural network.

that can extract spatial features from images. Typically, the first convolutional layer extracts low-level features or small-scale local patterns, such as edges and corners, while the last convolutional layer extracts high-level features, such as image structure. This hierarchical structure allows CNN to improve the learning efficiency of spatially hierarchical patterns. In general, only convolutional neural network structures have convolutional and pooling layers. The convolution kernel in the convolutional neural network is similar to the neuron in the BP neural network, including weight coefficients. The role of the pooling layer is to filter the features to reduce the number of parameters. Taking the classic convolutional neural network model LeNet-5 as an example, the structure of the convolutional neural network is as follows: input-convolution layer-pooling layer-convolution

layer-pooling layer-fully connected layer-output. The structure diagram is as follows.

Convolutional layers are usually defined using two components: the kernel size (e.g., 3 × 3 or 5 × 5) and the number of filters. CNN uses a rectangular sliding window with a fixed size and a predetermined step size, and uses the dot product between the weight of the kernel and the input small area (i.e., the receptive field) to generate a convoluted feature map. The feature map is obtained by this simple convolution operation. The resulting new image is a visual representation of the extracted features. The reason why the convolutional layer can perform feature extraction on the input data, such as shape features (Fidler et al., 2006, 2008; Fidler & Leonardis, 2007) is that it contains multiple convolution kernels inside, and the convolution sum can be regarded as an element, which contains A weight coefficient and a bias value, similar to the neurons of the BP neural network. Such neurons (convolution kernels) in the convolutional neural network are connected to neurons (convolution kernels) that are relatively close to each other in the previous layer. The size of the convolution kernel is called "receptive field" (Gu et al., 2015). The convolution kernel scans the input data according to the step size, which is equivalent to performing matrix multiplication on the input data within the receptive field (convolution kernel) and finally adding the deviation value (Goodfellow et al., 2016).

$$Z^{l+1}(i,j) = \left[Z^l \otimes w^l\right](i,j) + b = \sum_{k=1}^{kl}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[Z_k^l\left(s_0 i + x, s_0 j + y\right)w_k^{l+1}(x,y)\right] + b$$

$$(i,j) \in \left\{0,1,\cdots,L_l+1\right\}, \quad L_l+1 = \frac{L_l + 2p - f}{s_0} + 1$$

(1)

The above formula actually calculates a cross-correlation and then superimposes the deviation value, b is the deviation amount, Z^l and Z^{l+1} is the convolution input and output feature map of the first layer $l+1$ (Qiu, 2018). The size of Z_{l+1} is indicated by L_{l+1}. The pixels of the feature map $Z(i,j)$ correspond to the number of channels K, and f, p and s_0 are the size of the convolution kernel, the number of filling layers and the step size (Goodfellow et al., 2016).

4.2. Migration Learning

When using ordinary neural network models for image classification in order to ensure the accuracy and stability of the classification results, it is generally assumed that: 1) training samples and test samples should obey different distributions; 2) the number of samples should be sufficient, in order to ensure large sample training and train an accurate classification model. However, the above two assumptions are very difficult to realize in actual work. The reasons may be: 1) Insufficient funds; 2) Time constraints; 3) Insufficient human resources; 4) It is very difficult to obtain some samples. So can you find a way to solve the problem of insufficient samples? This introduces the concept of transfer learning. The concept of transfer learning is a machine learning method that uses mature

prior knowledge in known fields to classify specific problems in different related application fields. Currently, transfer learning has been widely used in many related fields, such as image classification (Long et al., 2015), emotion classification (Blitzer et al., 2006), dialogue system (Mo et al., 2016) and urban computing (Wei et al., 2016). Building a convolutional neural network model requires a large number of calibration training samples, and the computational cost is high. Migration learning can transfer knowledge from existing data to facilitate learning tasks in new environments. Convolutional neural network models can be trained using existing typical datasets and then fine-tuned using samples from the study area in order to achieve the purpose of fully training the model. To this end, the clustering mechanism of a deep convolutional neural network (DCNN) is transformed into a land use classification task by using a transfer learning mechanism. Using the transfer learning mechanism, the features obtained during the training process of the DCNN-based linear elimination model can be transformed into land use information, and the land use remote sensing classification of complex surface areas can be completed.

Many scholars have carried out in-depth research on the algorithm of transfer learning in the past ten years. Based on the previous research results, the principle of transfer learning is to map the high-dimensional features of related fields to low-dimensional features (Li et al., 2018). The source domain features and target domain features obey similar distributions (Chen et al., 2015) in the low-dimensional feature space so as to achieve dimensionality reduction and distribution, and the model (classifier) trained by the samples in the source domain can be used to classify the target domain based on the above results.

In the CNN classification method based on migration learning, the CNN model can be pre-trained using typical terrain datasets, then the CNN model can be fine-tuned (fine-tuning its parameters) with samples from the research area, and finally the fine-tuned model can be used to classify the ground objects. The first CNN layer is similar to the Gabor filter, and the edge detector extracts color and texture features. Specifically, fine-tuning is used to adjust the parameters of the pre-trained network by utilizing samples from the study area. Girshick et al. (2014) showed that fine-tuning a pre-trained CNN can significantly improve classification performance. Zhao et al. fine-tuned AlexNet (Krizhevsky et al., 2012; Chatfield et al., 2014) and achieved better results. Xie et al. also used the fine-tuning pre-trained CNN model to achieve certain results in remote sensing classification work (Yue et al., 2015). Yue et al. (2015) also achieved certain results by using the fine-tuning method to classify hyperspectral images.

5. Results and Analysis

5.1. Sample Making

1) Run the matlab program, traverse and intercept samples of 32×32 size, and get all the 8 categories to be classified within the fixed-size neighborhood

(grassland, pond, greenhouse, land use, road, residential area, bare land, crops, etc. Object type) image samples, each image is saved in tif format, and the main file name is a natural number, starting from 1 and increasing sequentially.

2) Select the ROI (training area) of each type of surface object from ENVI, and record the row and column number of each pixel in the ROI in Excel. According to the row and column numbers, calculate the screenshot sample number in step 1, and select the corresponding screenshot sample. The samples are stored in folders of each category by category, and then the samples of each category are divided into training samples and testing samples, and finally two sets of training and testing samples are formed.

3) Make a sample tag file. Divide the selected samples into training samples and test samples (divided into 8 categories in total, training samples and test samples of 8 types of land features such as grassland, pond, greenhouse, land use, road, residential area, bare land, and crops), Use DOS batch commands and Python code to generate summary files and label files, and convert training and test sample sets and label files into train_leveldb and test_leveldb files in caffe format.

5.2. CNN Model Pre-Training

Multispectral satellite images can generally be expressed as $m \times n \times h$, that is, 3D tensors, where m and n represent the height and width of the image, respectively, and h corresponds to the number of channels. Convnets convolutional neural network models require 3D tensors as input. the paper uses the CNN_UCMerced-LandUse_Caffe_finetune dataset to pre-train CNN on the cifar10_full.prototxt model in the caffe framework.

5.3. Model Fine-Tuning Based on Transfer Learning

Fine-tuning a pre-trained network is the best solution when only a limited number of training samples are available in the study area. In this case, the parameters of all layers or top layers in the pertained network are fine-tuned, while freezing the first layer representing general features. Freezing refers to not updating the weights of a layer or set of layers during training. The advantage of this method is that the CNN_UCMerced-LandUse_Caffe_finetune dataset is used to train the CNN model, and then the parameters of the pre-trained model are fine-tuned using samples from the study area. Fine-tune the parameters of the trained model to make it more relevant to the dataset in the study area. It is effective to fine-tune those specific feature parameters bbecause the first layer in CNN encodes high-weight features, and the last layer encodes more specific features. Furthermore, fine-tuning all layers may lead to overfitting due to the large number of parameters (Ng et al., 2018). In this study, only the preceding parameters in the CNN model were fine-tuned.

The number of input channels of CNN is set to 3 in the paper, and the CNN_UCMerced-LandUse_Caffe_finetune dataset is used in the pre-training

stage to pre-train the CNN model. The parameters of the original deep architecture are maintained during fine-tuning. However, a learning rate of 0.001 and a decay rate of 0.004 are chosen for full training and fine-tuning experiments. For fine-tuning and full training, the number of iterations and snapshots (output a fine-tuned model after fine-tuning a certain number of times) are set to 8000 and 1000, respectively. All of the above functions are implemented using the Caffe function library (Abadi et al., 2016).

5.4. Convolutional Neural Network Classification Based on Transfer Learning

The Python language is used to write codes to classify the images of the study area based on the above-mentioned fine-tuned model, and finally the classified raster images are obtained. Finally, the matlab program is used to assign projections to the classified raster images. The map of classification is shown in Figure 3. The classification maps of various ground object types is shown in Figure 4.
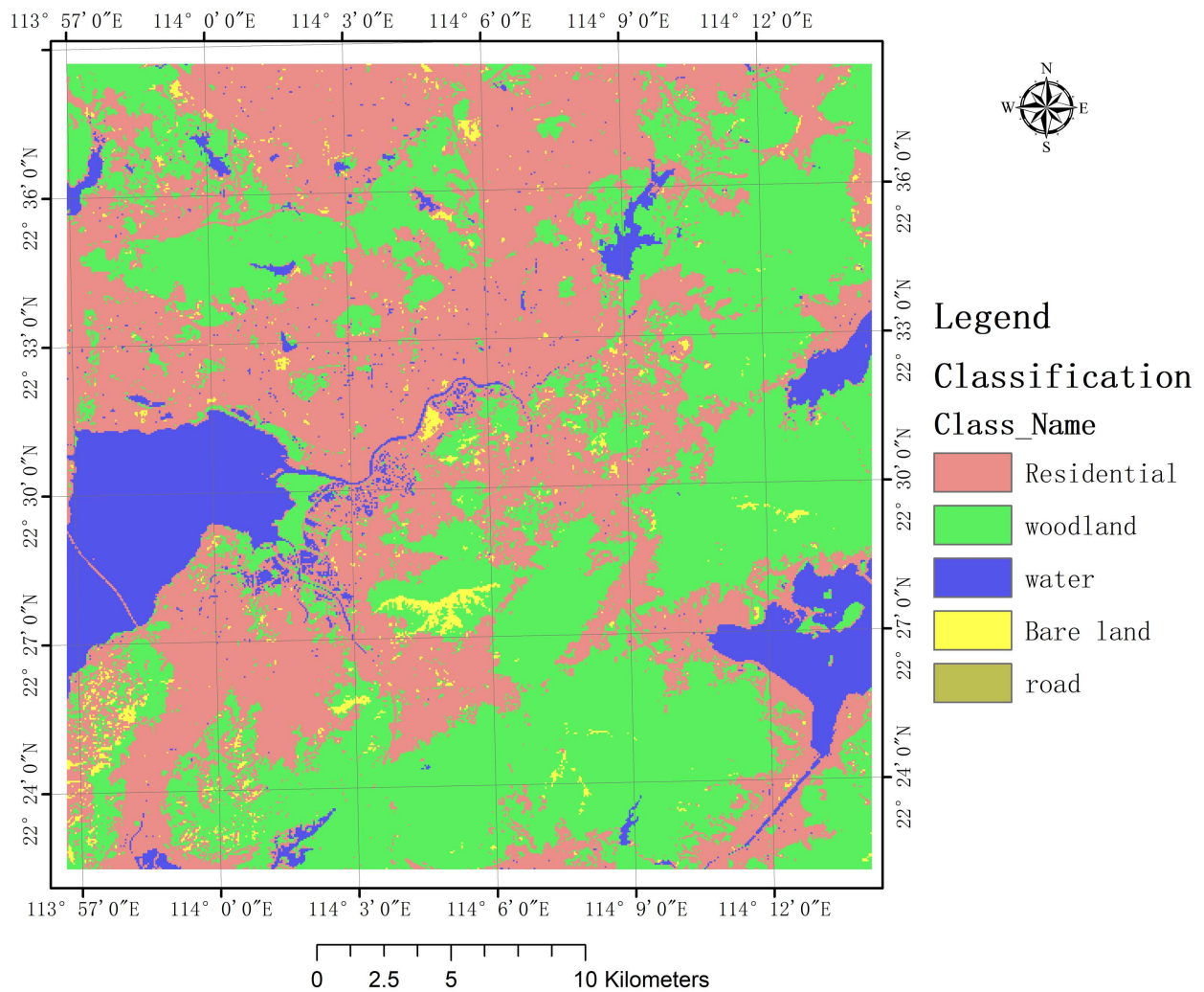

Figure 3. The map of classification.

(1) Residential

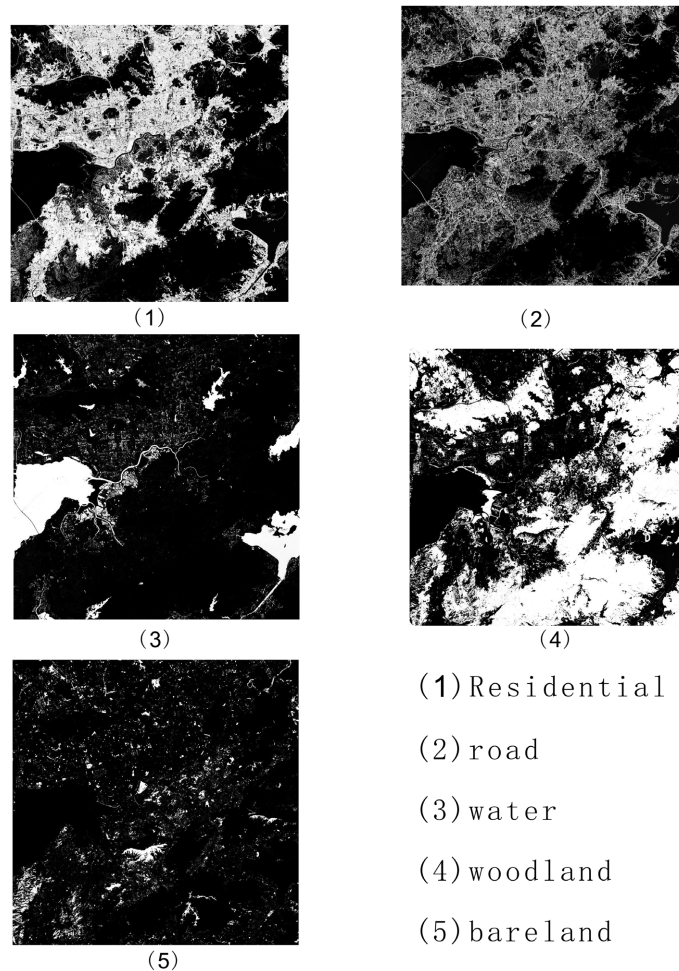(2) road

(3) water

(4) woodland

(5) bareland

Figure 4. The classification maps of various ground object types.

5.5. Accuracy Evaluation Based on Migration Learning and Convolutional Neural Network Classification

After the above-mentioned convolutional neural network based on transfer learning is used to classify land use in complex surface areas, the section compares the accuracy of classification methods based on transfer learning, convolutional neural network and support vector machine. The overall accuracy and Kappa coefficient are used to quantitatively evaluate the classification accuracy. The overall accuracy representation is calculated by dividing the number of correctly classified cells by the total number of cells in the confusion matrix. The Kappa coefficient determines the degree of agreement between the reference data and the classification map (Liu et al., 2018). The above results are shown in Table 1.

Table 1 shows that the overall accuracy of the convolutional neural network classification method based on migration learning is 90.3588%, and the Kappa coefficient is 0.8897; Table 2 shows that the overall accuracy of the support vector machine classification method is 84.83%, and the Kappa coefficient is 0.7954. This proves that the accuracy of classification based on transfer learning and

Table 1. The convolving neural network classification confusion matrix accuracy evaluation based on transfer learning.

Overall Accuracy = 90.3588%, Kappa Coefficient = 0.8897

Class	Residential	woodland	water	Bare land	road
Unclassified	1373	0	3	0	0
Residential	1373	0	3	0	1
woodland	0	1926	0	0	0
water	0	0	1196	0	0
Bare land	0	0	0	1231	0
road	0	0	0	0	1009
Total	2746	1926	1202	1231	1010

Table 2. Accuracy evaluation of SVM classification confusion matrix.

Overall Accuracy = 84.83%, Kappa Coefficient = 0.7954

Class	Residential	woodland	water	Bare land	road
Unclassified	2842	26	2784	0	107
Residential	57,778	2400	4764	0	621
woodland	192	1492	0	0	0
water	1088	34	21,418	0	0
Bare land	1	0	0	16,814	5
road	1974	0	29	0	17,358
Total	63,875	3952	28,995	16,814	18,091

convolutional neural network is relatively high, and it is more suitable for remote sensing land use classification in complex land surface areas in South China.

6. Conclusion

The Landsat 8 image was used as the data source in the paper. The Shenzhen city of Guangdong province were taken as the study area, to carry out research on the application of land use classification technology in complex surface areas. The transfer learning method was applied to transfer the prior knowledge of the source domain to the target domain, and the CNN_UCMerced-LandUse_Caffe_finetune dataset was used to pre-train the convolutional neural network model, and then the sample set in the research area was used to fine-tune the pre-trained model, which solve the commonly existing problem of insufficient training samples of the convolutional neural network model and improves the efficiency and accuracy of model training.

This paper compares the accuracy of transfer learning-based deep learning convolutional neural network classification and support vector machine classification methods. The comparison results show that the overall accuracy of the con-

volutional neural network classification method based on transfer learning is 90.3588%, and the Kappa coefficient is 0.8897; the overall accuracy of the support vector machine classification method is 84.83%, and the Kappa coefficient is 0.7954. The research proves the high accuracy of classification based on migration learning and convolutional neural network.

In summary, the land use classification method based on transfer learning and convolutional neural network based on automatic feature selection is more suitable for land use classification by remote sensing in complex land surface areas in South China, and can improve the automation level and extraction accuracy of land use monitoring information extraction by remote sensing.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., & Devin, M. (2016). Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv: 1603.04467

Ballard, A. H. (2007). Rosette Constellations of Earth Satellites. *IEEE Transactions on Aerospace and Electronic Systems, AES-16,* 656-673. https://doi.org/10.1109/TAES.1980.308932

Blitzer, J., McDonald, R., & Pereira, F. (2006). Domain Adaptation with Structural Correspondence Learning. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* (pp. 120-128). Association for Computational Linguistics. https://doi.org/10.3115/1610075.1610094

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014). Return of the Devil in the Details: Delving Deep into Convolutional Nets. In *Proceedings of the British Machine Vision Conference 2014.* BMVA Press.

Chen, N. C., Zhang, W. J., & Wang, X. L. (2015). Performance Evaluation and Methodology for Sensor Observation Services. *Bulletin of Surveying and Mapping, No. 4,* 61-64.

Chollet, F. (2017). *Deep Learning with Python.* Manning Publications Co.

Fidler, S., & Leonardis, A. (2007). Towards Scalable Representations of Object Categories: Learning a Hierarchy of Parts. In *2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE.

Fidler, S., Berginc, G., & Leonardis, A. (2006). Hierarchical Statistical Learning of Generic Parts of Object Structure. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 182-189). IEEE. https://doi.org/10.1109/CVPR.2006.134

Fidler, S., Boben, M., & Leonardis, A. (2008). Similarity-Based Cross-Layered Hierarchical Representation for Object Categorization. In *2008 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-8). IEEE. https://doi.org/10.1109/CVPR.2008.4587409

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 580-587). IEEE. https://doi.org/10.1109/CVPR.2014.81

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* (Vol. 1, pp. 326-366). MIT Press.

Gu, J., Wang, Z., Kuen, J., Ma, L., Shahroudy, A., Shuai, B., Liu, T., Wang, X., Wang, L., Wang, G., & Cai, J. (2015). Recent Advances in Convolutional Neural Networks. arXiv: 1512.07108

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems* (pp. 1097-1105).

Li, D. Y. (1992). The Determination of the Orbit of Satellite along the Outer Edge of the Ground Cover Zone. *Chinese Space Science and Technology, No. 3,* 19-26.

Li, H. L., Hu, X. J., & Guo, H. (2018). Land Use/Land Cover Classification Supported by Transfer Learning. *Bulletin of Surveying and Mapping, No. 9,* 50-54.

Liu, Y., Fan, B., Wang, L., Bai, J., Xiang, S., & Pan, C. (2018). Semantic Labeling in Very High Resolution Images via a Self-Cascaded Convolutional Neural Network. *ISPRS Journal of Photogrammetry and Remote Sensing, 145,* 78-95. https://doi.org/10.1016/j.isprsjprs.2017.12.007

Long, M., Cao, Y., Wang, J., & Jordan, M. (2015). Learning Transferable Features with Deep Adaptation Networks. *Proceedings of the 32nd International Conference on International Conference on Machine Learning, 37,* 97-105.

Mo, K., Li, S., Zhang, Y., Li, J., & Yang, Q. (2016). Personalizing a Dialogue System with Transfer Learning. arXiv: 1610.02891

Ng, A., Kian, K., & Younes, B. (2018). *Convolutional Neural Networks Course (Deep-Learning.AI).*

Qiu, X. P. (2018). *Chapter 5. Convolutional Neural Networks. Neural Networks, Deep Learning.*

Wang, W., Zheng, Z., Li, P. F. et al. (2013). Measuring the Observation Capability of a Mission-Oriented Imaging Satellite Sensor. *Journal of Wuhan University (Information Science Edition), 38,* 1480-1483.

Wei, Y., Zheng, Y., & Yang, Q. (2016). Transfer Knowledge between Cities. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1905-1914). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939830

Weng, H. H. (2006). *Analysis and Simulation of Ground Coverage by Remote Sensing Satellite.* University of Information Engineering.

Yue, J., Zhao, W., Mao, S., & Liu, H. (2015). Spectral-Spatial Classification of Hyperspectral Images Using Deep Convolutional Neural Networks. *Remote Sensing Letters, 6,* 468-477. https://doi.org/10.1080/2150704X.2015.1047045

Zhang, W., Tanida, J., Itoh, K., & Ichioka, Y. (1988). Shift Invariant Pattern Recognition Neural Network and Its Optical Architecture. *Proceedings of Annual Conference of the Japan Society of Applied Physics, 6p-M-14,* 734.