An Improved Remote Sensing Image Classification Method Based on DCNN

Bin Xu, Jiafei Zong, Ling Ding*, Wei Zhang and Guoguang Cao

School of Computer Science and Technology, Hubei University of Science and Technology, Xianning, China
Email: dn0715dn@163.com

Abstract. The popular deep learning method of the year established a recognition model to identify remote sensing images, which is a new method. Based on Google Inc.’s Inception-v3 convolutional neural network recognition model, this paper adopted a conception of migration learning to achieve the accurate automatic classification of remote sensing image scenes. Since the model already possesses good convolution parameters for image feature extraction, the idea of migration learning was used to retrain the model. To achieve the purpose of learning UC Merced Land Use Dataset related data types, the authors use the training set and test set to test separately. The learning model has a good learning effect and a high classification recognition rate.

1. Introduction

According to the differences of various remote sensing images, human beings can judge the target of remote sensing to a large extent and realize the classification of remote sensing images. However, people try to use the computer to automatically classify remote sensing images to reduce the consumption of human and material resources. The judgment of the image needs to be improved. With the continuous progress of astronomical science and aviation science, human beings gradually obtain the technology of remote sensing images with higher resolution. The contours of surface objects are more and more clear, the texture features are more and more complex, and the integrated spatial information is more and more abundant. Therefore, how to use these remote sensing images to obtain useful information is very important for the classification of remote sensing images. Remote sensing image processing is the main method to obtain remote sensing information. Remote sensing image classification is the feature extraction of various image information in remote sensing image processing. The relevant feature information is used to calculate the correlation probability. Finally, the feature type is discriminated.

2. Related Work

The original remote sensing image classification method is judged by visual observation. Although this method is simple and flexible, the accuracy is guaranteed. However, this method requires high professional knowledge of the visual survey personnel, which takes more time and is not applicable. Massive remote sensing images. Therefore, people have realized the traditional remote sensing image classification method through the spectral features and texture features of the image [1].

The traditional classification methods mainly include supervised minimum distance, Markov distance, unsupervised K-average classification [2], etc. But with the increasing resolution of remote sensing images, high-resolution remote sensing images have a lot of detailed information and the
complexity of the Earth's spectroscopy problem, the applicability of traditional classification methods is gradually decreasing. Therefore, new classification methods such as support vector machine, random forest classification, artificial neural network and other classifications [2] began to gradually use the remote sensing image classification. Tan et al. [3] realized hyperspectral remote sensing image classification based on the support vector machine. Zhang et al. [4] used random forest model to classify the forest vegetation of Landsat-8 remote sensing image. Wang et al. The method of BP neural network in remote sensing image classification processing is expounded, and the classification method based on BP neural network is studied [5]. However, these methods belong to the shallow structure model. The common feature of these shallow structure models shows that the original image information is less linear or nonlinear to achieve the classification purpose. For the multi-source and feature-sensitive remote sensing images, the shallow layer Structural models have become less and less applicable and are not automated enough.

3. Proposed Method

This paper adopted the migration learning method. Based on the Inception-v3 object recognition model parameters of Google’s deep convolutional neural network, all convolution parameters are preserved before the last fully connected layer of the model. The last part of the model is re-used with the data set. Training, design experiments, training and predictive classification of 21 types of remote sensing satellite images in a small number of scenes, the classification effect is better overall, and the time is shorter. However, the number of training sets is small and some image features are not obvious due to the initial iterations are too small. Some image classification results are not ideal, and the accuracy rate is low [6]. By optimizing the number of training sets and quality, the effective recognition ability of the model is improved, which also shows the good adaptability of the model in the field of remote sensing images. Besides, Google's Inception-v3 model is derived from a recognition model obtained by training a large number of training sets for various objects in life. The convolution feature extraction parameters have good adaptability and do not require related processing of remote sensing images. The imaging distance of the image, the requirements of the pixels, the like are also reduced, and the time consumption and the reduction of the computer resource consumption make it have considerable applicability.

3.1. Regression Model and Classification Model

The traditional BP neural network [7] has some problems: First, due to too many weights, the calculation is large and complicated; second, because there are too many weights, it is necessary to prepare a large number of samples first, and then it takes a long time. Go to the training using the training set for the model.

In the 1960s, neuroscientist Hubel et al. of Harvard Medical School proposed a new concept of receptive field [8] through the study of cat visual cortical cells, which was obtained through stimulation of different parts. The neural signal of the corresponding area of the cat reflects that this concept is very similar to the feature of the block extraction of convolutional nerves. The neurocognitive machine (Neocognitron) [9] proposed in the 1980s based on the concept of receptive field is considered to be the first implementation network of convolutional neural networks. Convolutional formula:

\[ g(f_1, f_2) = E(f_1(x, y) * f_2(x, y)) \]  

where \( g \) is the result of convolution extraction, \( f_1 \) is the original image input, \( f_2 \) is the parameter of the feature matrix, \((x, y)\) is the corresponding coordinate, the corresponding position image value is multiplied by the formula, and then the single nerve is The values of all the locations corresponding to the meta are summed to obtain a result as the value of the result of the feature extraction. A multi-layer filter is set in the convolutional neural network to reduce the situation of being too single, and the
The extraction effect is significantly improved. As shown in figure 1, feature extraction can be performed by different matrices to obtain different feature extraction results.

The connection between the convolutional layer and the pooling layer is generally followed by a pooling operation once the convolution is completed. The purpose of the pooling is to reduce the output size and reduce over-fitting. The more common pooling has max-pooling. For the sake of simplicity and convenience, assuming a 4\*4 window, the maximum pooling is to take the maximum parameters from each region. As shown in figure 2, the 4\*4 image is divided into the upper left, upper right, lower left, and lower right. Finally, the maximum values in the four regions are taken out respectively, which includes the pooling result, and a 4\*4 image is converted into 2\*2.

Another common way of pooling is mean-pooling, which extracts the average value for each region, as shown in figure 3. Also assume that there is a 4\*4 output image, press it to the top left, top right, bottom left, bottom right. Four parts are divided, and each area is averaged as the representative value of the area, such as 1,1,2,3 in the upper left area and the last output (1+1+2+3)/4 as the upper left area.

In general, convolutional neural networks have a convolutional layer function. For example, a low-level convolutional layer filter is used to detect simple features, such as edges and corners [10]. As the convolutional layer is improved, the image characteristics detected by the corresponding filter will increase the complexity, such as detecting the combination of the last filtering result, such as a
semicircle, a rectangle, and the like. In other words, the construction of the neural network is the repeated correction of various filter parameters. This process is also called the training process. Convolutional neural networks are such a feedback model that iteratively modifies a blank filter into a pattern that can detect a particular model.

3.2. Softmax Regression Model
The Softmax regression model is a model that converts the output into probability. Suppose now that digital image recognition, its input is a number between 0-9, our model may guess that the probability that a picture is a number 9 is 80%. The probability that the number 8 is 10%, the other numbers are less likely, but the overall probability adds up to 1. This is an application of the Softmax regression model, which can be used to assign probabilities to different objects. The mathematical formula is shown in equation (2), here $x_i$ is one of all possibilities, and $x_j$ corresponds to all outputs.

$$\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum \exp(x_j)}$$ (2)

3.3. Inception-V3 Classification Model
Inception-v3 is the third upgraded version of GoogleNet, an image recognition model developed by Google in 2012. It is a model for image information extraction. The model obtained after training is accurate and deep. Recognized, the greatest performance is the great use of dimensionality reduction processing, compared to the Alex convolutional neural network to greatly speed up the training. The final output to the ultimate pooling layer and the Softmax layer [11] converts the signal into probability and gets the output.

4. Experiment and Analysis

4.1. Experimental Data
The dataset used in this paper is UC Merced Land Use Dataset [12]. The dataset was collected in August 2010 in the United States. There are 21 classification scene remote sensing satellite image classification databases, each of which contains 100 image files in the format of tiff. Each image contains 256*256 pixels as shown in figure 4.

![Figure 4. UC merced land use dataset.](image)

4.2. Experimental Environment Construction
This article uses tensorflow1.9.0 version, Anaconda 3.2.0 (64-bit), encoding language python 3.6.2 operating system is win10, the setup environment is personal PC, the processor model is Inter(R) Core (TM)i5-8400 CPU @2.80GHz the running memory is 16G.
5. Experimental Results and Analysis

5.1. Model Retraining
In the process of retraining the model, this paper initially sets the number of iteration steps to 200, and takes 100 images of each class, which are numbered as the first 96 images as the training set, and the last 4 images as the test set. Before retraining the model, Google's deep learning model Inception-v3 model is obtained for free in github. In addition, retrain the Inception model is required by using the retrain.py interface, using the official correlation. The function is retraining and learning. At this time, the author repeatedly performs image training by writing windows batch processing files. The output obtained after training are output_graph.pb and output_labels.txt. The output_graph.pb holds the model file, the output_labels.txt is the labels corresponding to the classification. Then, the model obtained by retraining is used to analyze the correlation between the training set and the test set.

5.2. Training Set Classification Analysis
After the model has been retrained, it is necessary to use the images of the training to test the learning ability of the model to prove the applicability of the migration learning idea. Therefore, many different kinds of images are randomly selected from the training set image as the experimental model input to determine the accuracy and certain learning ability of the model. The selected pictures are shown in figure 5, and each category contains a random numbered picture.

![Figure 5. Random selection of training set.](image-url)

The results of the image test are expressed in the form of probability output. Each picture corresponds to 21 probabilities. When calculating all classifications, the category with the highest probability is automatically considered by the system as the category of the remote sensing image.

5.3. Test Set Classification Analysis
Testing with the training set above, in order to ensure that the model has good learning ability, testing with the test set that is not involved in training is to illustrate whether the model has generalization ability to achieve the purpose of application in the field of remote sensing image.

The test set selected in this paper is from 21 scenes in multiple times. Each scene has four images. A total of 84 images that have not participated in the test randomly is selected several untrained images as test sets, just like the training set. Each image model calculates the corresponding probability through the regression function, but this paper takes only the two largest probabilities as impressions. Select a result here, and select the picture as shown in figure 6.

![Figure 6. Test set image selection.](image-url)

The results of the test set show that there are no problems in image classification and some deviations. It can be seen that the model has a better recognition effect for the species with obvious features, such as agriculture, which has a higher recognition rate of 85%.
6. Conclusion
This paper firstly introduces the status quo of remote sensing image recognition classification. Since the traditional visual classification, statistical classification, and shallow model structure improve the imaging effect of remote sensing images, the recognition and classification effect of remote sensing images is more and more limited. One of the novel designs in this paper is that through the establishment of the remote sensing image deep learning model based on Google's Inception-v3, based on the principle of migration model, the model is retrained and realized in a short time with high precision on UC Merced Land Use Dataset. The effective identification of 21 types of remote sensing image types in the Land Use Dataset, the experimental results can be classified correctly for all classifications, most of the classification prediction probabilities are higher than 80%, and the model extracts features through convolutional neural networks, not Some artificial manipulation of the initial image is required, which reduces the possibility of artificially affecting the experimental results. It can also be explained that the robustness and generalization ability of the model is good, and the image size and resolution requirements of the training data are small.

Acknowledgements
This work is supported by the Innovation training program for college students: S201910927028, Ministry of Education Collaborative Innovation Project: 2020QT02.

References
[1] Liu Q, Jin H and Dong G 2019 Research on high resolution remote sensing image classification method for UAV Computer and Digital Engineering (003) 638-642+727.
[2] Gong W, Wang Z, Liang Y, Fan X and Hao J 2019 Research on high resolution remote sensing image classification based on convolution neural network IFIP Advances in Information and Communication Technology (545) 87-97.
[3] Gong M G, Yang H L and Zhang P Z 2017 Feature learning and change feature classification based on deep learning for ternary change detection in SAR images ISPRS Journal of Photogrammetry and Remote Sensing (129) 212-225.
[4] Zhang X Y, Li F R, Pei, et al. 2016 Classification of forest vegetation based on random forest model based on terrestrial satellite-8 remote sensing image Journal of Northeast Forestry University 44 (06) 53-57+74.
[5] Qureshi A S, Khan A, Zameer A, et al. 2017 Wind power prediction using deep neural network based meta regression and transfer learning Applied Soft Computing 58 742-755.
[6] Fadaeddini A, Eshghi M and Majidi B 2018 A deep residual neural network for low altitude remote sensing image classification 6th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS) (Kerman) pp 43-46.
[7] Ma W, Fan J, Li Q, et al. 2018 A raw milk service platform using BP neural network and fuzzy inference Information Processing in Agriculture 2214-3173.
[8] Turcsany D, Bargiela A and Maul T 2016 Local receptive field constrained deep networks Information Sciences 349-350 229-247.
[9] Zhang Y, Xia G, Wang J, et al. 2019 Multiple feature fully convolutional network for road extraction from high-resolution remote sensing image over mountainous areas IEEE Geoscience and Remote Sensing Letters (99) 1-5.
[10] Peng C, Li Y, Jiao L, et al. 2019 Densely based multi-scale and multi-modal fully convolutional networks for high-resolution remote-sensing image semantic segmentation IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (8) 2612-2626.
[11] Li L and Ying S C 2017 Implementation of softmax layer of convolutional neural networks based on FPGA Modern Computer (Professional Edition) (26) 21-24.
[12] Dataset Acquisition Source http://vision.ucmerced.edu/datasets.