Research on the technology of terrain classification and change detection based on deep learning

Xiao-nan Wang* 1, Qiang Zhang1, Feng-yu Wang1 and Yang Gao1

1Beijing Institute of Spacecraft System Engineering, postbox5142, Beijing, China
bcbb2004@126.com

Abstract. With the rapid development of urbanization in China, the change of land property is very rapid. The detection of land change is helpful to the government’s supervision and the development and protection of land resources. The traditional methods of land feature classification and change detection of remote sensing image generally adopt the methods of manual annotation and recognition. Due to the performance defects, they can not meet the requirements of remote sensing image business application. In this paper, the technology of satellite remote sensing image feature classification and change detection based on deep learning is proposed, and several demonstration cases of automatic classification and change detection system of buildings, roads, water bodies, forest land and other features based on high-resolution remote sensing image are formed. Relevant research results show that the technology can meet the requirements of remote sensing image classification and change detection, and the accuracy of the review is more than 90%. The method adopted can effectively solve the application needs of related fields.

1. Introduction

China is in the process of high-speed urbanization, the change of land attributes is very rapid, the detection of land change is conducive to the government’s supervision and policy-making, and is conducive to the development and protection of land and resources. And the classification of land features is conducive to the overall planning of urban and rural areas, so as to promote the process of urbanization.

The traditional remote sensing image classification and change detection generally use the methods of manual annotation and recognition, which is time-consuming and laborious. Although machine learning method can complete the function of remote sensing image feature labeling, there are still performance defects, which can not meet the business application requirements of remote sensing image. With the rapid development of deep learning technology, significant breakthroughs have been made in the fields of automatic recognition, labeling, change detection of remote sensing images, which makes it possible to meet the application requirements of automatic production of remote sensing image products.

In order to effectively improve the utilization rate of remote sensing image, automatically mine the deep information of remote sensing image, and provide high time effective data support for land monitoring, urban planning, surveying and mapping and other fields. In this paper, the technology of feature classification and change detection based on in-depth learning is proposed, and the high-resolution remote sensing image is used to make a demonstration application case for feature classification and change detection to verify the correctness and effectiveness of the technology. In order to reduce the labor cost, improve the automatic degree of remote sensing image information
extraction, and solve the contradiction between the lack of labor and the rapid increase of remote sensing data processing demand, the deep learning method is used to automatically interpret remote sensing image data.

2. Research on land classification module algorithm

Land classification needs to mark remote sensing image pixel by pixel. Through the statistics of the classification of each pixel in the whole map, we can get the aggregation status, distribution location data and other information of all kinds of features, so as to get the useful information of land use and land cover. Tradition machine learning method can complete the function of remote sensing image feature labeling, but there are still performance defects, which can not meet the requirements of remote sensing image intelligent products. In recent years, the deep convolution neural network, which has made a great breakthrough in image recognition, labeling and other fields, makes it possible to mass produce intelligent remote sensing image products [1].

Since 2012, deep convolution neural network (DCNN) has made great achievements in image detection, recognition, classification and other fields, and the accuracy and complexity are also constantly improving. But classic DCNN is more suitable for classification and regression, because the network converts the input image into a numerical factor, and the size of each dimension of the vector represents the probability that the return image belongs to a certain class. The traditional method is to extract pixel blocks one by one and then use convolution neural network (CNN) to classify them. The classification result is used as the classification of the central pixel of the pixel block, and then a classification result graph is obtained. The defects are: (1) the larger scale spatial information is ignored, resulting in information loss, which reduces the performance of the algorithm; (2) limited by the full connection layer, the size of the input image of DCNN is fixed. If the resolution of the image is different from that of the training sample, the extracted pixel block is difficult to reflect the nature of the corresponding pixel points, so it is difficult to classify them; (3) because the algorithm needs to traverse the whole image, the time consumed by the algorithm increases with the square level of image edge length, which wastes computing resources. In order to solve the above problems, this paper uses the neural network of Deeplab model to solve the problem of pixel annotation in remote sensing images [2].

2.1. Deeplab Model

The establishment of Deeplab model is based on Fully convolutional networks (FCN). FCN can input any size of image, and the output of each layer is also an image, which is conducive to the global understanding of the image. Ordinary CNN can only input images of a specified size, and the output of full connection layer is vector. The core idea of FCN is to replace the full connection layer of the traditional CNN network with a convolution layer whose convolution kernel is 1x1, so that all the neuron layers with parameters in the network are convolution layers, so as to allow the input of images of any size. If the output is a two-dimensional image, the spatial information of the input image will be well preserved, while CNN of image level classification completely loses the spatial information of the image.

The results of FCNet not only identify the target, but also show the shape, size and location of the target. Ordinary CNN can only give the category of image. Because FCNet does not need to traverse the whole image, it is much faster than the DCNN algorithm applied in scene annotation.

In this paper, the deeplab model is used to improve the accuracy of FCN by using the improved full convolution neural network and the full connection conditional random field algorithm. (CRF)
Except for CRF, the structure of deeplab is basically the same as that of FCNet, which is improved on the basis of VGG-Net: increase output heat map size. Because there are 32 times down sampling in FCN model, even if 100 pixels wide filling area is added around the image during input, the input is 500x500, and the size of the output heat map is only 16x16, so the spatial information of the image is seriously lost. In deeplab, the step size of the 4th and 5th lower sampling layer in FCN is changed from 2 to 1, and the width of filling operation is 1, which makes the lower sampling multiple of the whole model reduce to 8 times[3]. Thus, when the input is 500x500, the output is 65x65. This can reduce the loss of spatial information and improve the accuracy of the model.

2.2. RES-FCN Model

In order to further improve the expressiveness and detection performance of the network, this paper further studies the idea of ResNet. ResNet requires each layer of network to directly learn its corresponding mapping relationship, while in the residual network, each layer needs to learn its “residual” mapping. Through formula expression, if the corresponding ideal mapping relationship is $H(x)$, the residual mapping $F(x)$ that each layer of network needs to learn now is $H(x)-x$, and then the output of each layer of network is remapped to $F(x)+x$. This method comes from “shortcut mapping”. Through shortcut mapping, some of the data can directly reach the next layer in the process of each layer transmission, which can effectively alleviate the problem of gradient disappearance/explosion.

According to the idea of residual network, the original convolution layer of the original deeplab network is replaced by the 50 layer network structure of residual network, and the network structure is adjusted according to the characteristic of remote sensing image. In the experimental comparison, the residual network has the following performance:

(1) Convergence performance: after a certain number of layers are added to the residual network, the performance is improved to a certain extent. The residual network has a lower convergence loss and does not produce too much over fitting.

(2) Classification performance: as the number of layers of residual network is generally higher than that of previous models, and there is a residual structure as the support premise of extreme depth, the performance of residual network is generally higher than that of previous excellent models, and the performance of residual network increases with the increase of network layers[4].

The full convolution neural network model used in this paper is shown in table 1:

| Layer group name | Layer number | Type                  | Number of convolution kernels/Lower sampling step | Convolution Kernel size |
|------------------|--------------|-----------------------|--------------------------------------------------|-------------------------|
| conv1            | 1            | Convolution layer     | 64/2                                             | 3x3                     |
| maxpool          | -            | Maximum lower sampling layer | 1/2                                                 | 3x3                     |
The specific processing flow of image segmentation algorithm based on Res-FCN is as follows: firstly, the input remote sensing image gets the thermal map through Res-FCN network, that is, each layer of image represents a category, and the pixel value represents the confidence degree of the location belonging to the category. Then, the thermal graph is input into the all connected random field, and the spatial relation of the original image is used to post process the thermal graph, which can supplement the spatial relation for the segmentation result of neural network and improve the accuracy of segmentation. Finally, the result of conditional random field processing is selected by maximum value and bilinear interpolation, and the corresponding annotation result of the original image is obtained.

2.3. Adaptive joint loss function

At present, in the field of deep learning, the loss functions commonly used are Softmax loss and Euclidean loss. Softmax loss is often used in the field of image classification. It provides a smooth and differentiable operation to get the maximum value. It can be seen as an extension of logistic regression. Its function form is as follows:

\[
p_j = \frac{e^{o_j}}{\sum_k e^{o_k}} \quad (1)
\]

\[
L = -\sum_j y_j \log p_j \quad (2)
\]

Where \( o \) is the output of neural network, \( y \) is the label of sample, \( L \) is Softmax loss. For regression problems, the Euclidean distance loss function, Euclidean loss, is often used. Its function form is as follows:

\[
L = \frac{1}{2} \| y - o \|^2 \quad (3)
\]

As you can see, Softmax loss pays more attention to the relative size of the output value, that is, the maximum value. In this mode, the optimization direction of neural network is to increase the output of target class position and reduce the output of non target class position. Euclidean loss is more concerned with the absolute value of the output. In this mode, the optimization direction of neural network is to make the output and label as consistent as possible.

For image classification, obviously Softmax loss is more suitable, because if the output of neural network is correct, but the absolute value is larger, Euclidean loss will punish it, thus reducing the classification ability of neural network. FCN algorithm transforms image segmentation problem into pixel by pixel classification problem, and achieve good results. Considering that Euclidean loss can learn the continuity information of tag image, we propose Self-adaption joint loss function(SAJ loss) to integrate the two loss functions mentioned above. SAJ loss combines Softmax loss and Euclidean loss, so that the neural network can learn not only the spatial relationship of the input image, but also the spatial relationship of the label image. At the same time, the weight relationship between two is obtained through optimization learning, which is more conducive to training to get better network parameters. The equation of adaptive joint loss function is as follows:

\[
L_{SAJ} = -\sum_j y_j \log \left( \frac{e^{o_j}}{\sum_k e^{o_k}} \right) + \frac{1}{2} \| y - (W^T o + b) \|^2 \quad (4)
\]
Where \( y_i \) and \( o_j \) are the i-th elements of \( y \) and \( o \), \( W \) and \( b \) are learnable parameter matrices and biases, which can adaptively adjust the relationship between the two classical loss functions, the experimental results are shown in Fig2 and Fig3.

![Figure 2. Results of training with Softmax loss](image1)

![Figure 3. Results of training with SAJ loss](image2)

3. Research on Algorithm of land change detection module

The remote sensing data contains a large amount of information, but the change of ground features is always slow. For multitemporal remote sensing data, it contains a large number of redundant information, which is not conductive to the development of related monitoring work. In order to solve this problem, by introducing the change detection function, the intelligent extraction results of remote sensing images are processed twice to obtain more valuable change information\(^5\).

The function of land change detection requires high registration accuracy of the front and back phases, so it is necessary to input the remote sensing image extraction results after RPC information correction. Because the multitemporal remote sensing image is often not completely coincident, it is necessary to extract the coincident region from the input image according to its geographic information. After getting the overlapped area, according to the location of the overlapped area, the
results of intelligent extraction of time-phase remote sensing images before and after comparison are obtained. If it changes, the pixel is retained. In addition, in order to further improve the accuracy of change extraction, the input image needs to be corroded and expanded first, and the extraction results also need to be processed later. Through the connectivity domain analysis algorithm, the change patterns with an area of less than 2 meters are removed. The flow chart of this function is shown in Fig4.

Remote sensing image change detection is mainly to monitor the changes between two or more remote sensing images in the same geographical and different periods. The classic remote sensing image change detection method is based on the significance test. Let $I_1(x, y)$ and $I_2(x, y)$ represent the two remote sensing images before and after the change, let $D(x, y) = I_1(x, y) - I_2(x, y)$. In an ideal case, the region with $D(x, y)$ value of 0 is the region without change, and the region without $D(x, y)$ value is the region with change. Due to the interference of the equipment imaging noise, $D(x, y)$ is composed of two parts: the real change of the region and the noise. Generally, it is assumed that the noise obeys the Gaussian distribution, and its mean value and variance can be estimated from the difference graph. In most cases, the change area only accounts for a small part of the whole image, and change detection can be considered as a signal detection problem in noise. Assuming that $H_0$ represents that the pixel has not changed, and $H_1$ represents that the pixel (block) has changed, the probability density functions of the unchanged region and the changed region are $p(D(x, y)|H_0)$, $p(D(x, y)|H_1)$ respectively. Under the assumption that the noise is Gaussian distribution, the probability density function of the difference graph is:

$$p(D(x, y)|H_0) = \frac{1}{\sqrt{2\pi\sigma}}\exp \left( -\frac{(D(x, y) - \mu)^2}{2\sigma^2} \right)$$

The basic idea of significance test is to determine the threshold according to the probability distribution and false alarm rate, and judge whether the difference image belongs to the changing area or the non-changing area by the threshold. Given the false alarm probability $\alpha$, the:

$$P \left( \frac{D(x, y) - \mu}{\sigma} > z_{1-\alpha} \right) = \alpha.$$  \hspace{1cm} (6)

By calculating the exit limit $\tau_{1,2} = \mu \pm \sigma \times z_{1-\alpha} / \sqrt{2}$, you can get the final output:

$$Out(x, y) = \begin{cases} 0, & \tau_2 < D(x, y) < \tau_1 \\ 1, & \text{otherwise} \end{cases}$$

The region of 1 in the output image $Out(x, y)$ is the real changed region, and the region of 0 in the output image $Out(x, y)$ is the non changed region. However, this method is often affected by light, seasons and sensors, so it is difficult to obtain two images that fully meet the requirements. The deep learning method can overcome these effects, so we propose a change detection method based on deep learning and significance test. This method first marks the multi-temporal remote sensing image pixel by pixel, then subtracts the labeling results and detects the connected domain. Finally, it uses the significance test to screen the connected domain, and finally retains the larger connected domain, which is considered as the changed region. The flow chart of this method is shown in Fig5.
Figure 5. Change detection method based on deep learning and significance test

4. Test Result

4.1. Test results of land classification module

By inputting the original three channel remote sensing image data, output the corresponding building, water, road, forest area mask map, calculate the detection accuracy; by recording the land classification time, test whether it meets the requirement. Table 2 shows the accuracy results of various targets extraction in the land classification module, and table 3 shows the time consumption of various targets extraction in the land classification module. The test results show that the average accuracy is higher than 90%.

Table 2. Test results of extraction accuracy of various targets in land classification module

| Task                | Test set extraction accuracy |
|---------------------|-----------------------------|
| Building extraction | 0.9380                      |
| Water extraction    | 0.9089                      |
| Road extraction     | 0.8887                      |
| Woodland extraction | 0.9553                      |

Table 3. Test results of extraction time of various targets in land classification module

| Task               | Time consuming (s) |
|--------------------|--------------------|
| Specific target extraction | 73                 |

Taking buildings as an example, the extraction results of specific features in remote sensing image are shown in Fig6 and Fig 7.
4.2. Change Detection module test result

Through the time-phase pair of the original remote sensing image as the test case, the change detection module test is carried out for the, building, water, road, forest and other types of features, represented by the road. The front time phase diagram and the back time phase diagram of the change detection are shown in Fig 8 and Fig 9, respectively, with the yellow edge as new and the magenta edge as less.
Figure 9. Time phase diagram after road change detection

5. Conclusion
In this paper, the technical ability of satellite remote sensing image classification and change detection based on deep learning is established, and several demonstration cases based on high-resolution image are formed. Relevant research results show that the technology can meet the requirements of remote sensing image classification and change detection, with an average accuracy of more than 90%, and the processing results meet the accuracy requirements. The method used can effectively solve the application requirements of related fields, especially in the face of massive image processing and real-time monitoring applications, this research method has the advantages of high efficiency and high precision, and good versatility, which is convenient for replication and promotion to other regions.

Reference
[1] Tansey K, Chambers I, Anstee A, et al. Object-oriented classification of very high resolution airborne imagery for the extraction of hedgerows and field margin cover in agricultural areas[J]. Applied geography, 2009, 29(2): 145-157.
[2] Zhang L, Gui-song X, Wu T, et al. Deep Learning for remote sensing image understanding[J]. Journal of Sensors, 2015, 501: 173691.
[3] Krizhevsky A, Sutakever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// Advances in neural information processing systems. Lake Tahoe: Curran Associates, Inc., 2012: 1097-1105.
[4] Chen Y, Lin Z, Zhao X, et al. Deep learning-based classification of hyperspectral data[J]. IEEE Journal of Selected topics in applied earth observations and remote sensing, 2014, 7(6): 2094-2107.
[5] TAHMASEBI P, HEZARKHANI A A. A hybrid neural networks-fuzzy logic-genetic algorithm for grade estimation[J]. Computers & Geosciences, 2012, 42: 18-27.
[6] TARANTINO C, ADAMO M, LUCAS R, et al. Detection of Changes in Semi-natural Grasslands by Cross Correlation Analysis with WorldView-2 Images and New Landsat 8 Data[J]. Remote Sensing of Environment, 2016, 175: 65-72.