

# Change Detection of Remote Sensing Image Based on Deep Neural Networks

Yan Chu, Guo Cao\* and Hassan Hayat

School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing Jiangsu 210094, China

\*Corresponding author

**Abstract**—How to improve the quality of difference image (DI) for change detection task is an important issue in remote sensing images. This paper propose a new DI creation method based on deep neural networks. Deep belief network (DBN) which is an important model in deep learning is applied, and back propagation (BP) algorithm is improved according to change detection task in our method. The modified BP algorithm tunes DBN to increase the difference on changed areas and decrease the difference on unchanged areas in DI. The change detection result is obtained through clustering analysis of DI. The experiment result shows that **the proposed method can avoid the radiometric correction procedures for change detection, enhance the difference in change areas remarkably, and suppress the noise effectively compared with traditional method of DI creation methods. The change detection result is more accurate, and the Kappa values increase 25.93% averagely.**

**Keywords**—back propagation; deep belief networks; remote sensing image; difference image; change detection

## I. INTRODUCTION

To detect the change in remote sensing images is a technology that determines the ground surface changes based on the images acquired at different time phases. **It is a key technology for surveying the land, assessing disasters and monitoring the environment.** At present, two kinds of detection methods are available:

The first one is Post-Classification Comparison[1] that requires respective classification of two images and compare their classification results, so as to realize the detection of changed and unchanged areas. The problem of such method is to require a high accuracy of classification.

The second method is comparative analysis[2] that constructs DI(Difference Image) between remote sensing images with different time phases and then detect the change through analysis, therefore, it is also known as DI analysis method. In this method, the quality (strong distinguish ability) of DI and its construction has a significant impact on the final results.

The traditional methods to construct DI is: Take the radiometric correction[3,4] of two images at different time phases and then construct DI by using original spectral differencing [5] and rationing [6], texture rationing [7], or using PCA transform [2], or difference image fusion [8], or wavelet transform. However, during the actual detection of change, the traditional methods rely heavily on the quality of radiometric correction. For example: if the unchanged area is

not properly radically corrected then the traditional method can barely construct quality DI.

In recent years, Deep learning [9-11] has become a new and popular research focus in the field of machine learning and has shown an excellent performance and great potential in the domain of classification. One breakthrough of deep learning is the model and fast learning algorithm of "Restricted Boltzmann Machine" (RBM) by Hinton et al.[12-14], and successfully applied in the image classifications task.

**In neural network, Back propagation (BP) algorithm [15-16] is an important tuning algorithm that is featured with solid theoretical basis, rigorous derivation process, strong commonality, and wide application in the training of multi-layer neural network.** This article applies DBN models and BP into the construction of DI. Multi-temporal remote sensing data requires no more radiometric correction. The two original remote sensing images are input directly into the neural network model, and obviously distinguished DI are constructed through tuning instructed by BP algorithm. Experiment shows better results than the traditional methods.

## II. DEEP NEURAL NETWORKS

### A. Restricted Boltzmann Machine

1) *Structure and Principle*: Restricted Boltzmann machine (RBM) is a variant of Boltzmann machine. Restricted means that the units must form a bipartite graph: a pair of units from two different layers (visible layer and hidden layer). It has symmetrical connection weights, while the units inside one layer have no weight connection.

The principle of RBM is to map the visible layer into hidden layer which shall be subsequently used to reconstruct the visible layer and adjust the parameters in RBM so as to minimize the reconstruction deviation.

RBM contains  $m$  visible units  $v=(v_1,...,v_m)$  to accept the data and  $n$  hidden units  $h=(h_1,...,h_n)$  for output. It also maps  $v$  as the weight matrix  $W$  of  $h$ . The energy function of model is defined as

$$E(v, h) = -h'Wv - a'v - b'h \quad (1)$$

$a$  is the deviation vector of the visible layer, while  $b$  is the deviation vector of the hidden layer. The value of each unit is mapped between 0-1 by sigmoid function, reflecting the

activation degree of the unit. The hidden units and visible units are subject to the following joint distribution:

$$P(v, h) = \frac{1}{Z} e^{-E(v, h)} \quad (2)$$

$$Z = \sum_{all} e^{-E(v, h)} \quad (3)$$

The activation value of hidden units are calculated from the visible units by:

$$P(h_j = 1/v) = \text{sigm}(b_j + \sum_{i=1}^m w_{i,j} v_i) \quad (4)$$

The activation value of hidden units reversely reconstructing visible units is

$$P(v_i = 1/h) = \text{sigm}(a_i + \sum_{j=1}^n w_{i,j} h_j) \quad (5)$$

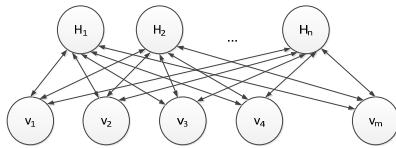


FIGURE I. RBM STRUCTURE

### B. Deep Belief Network

RBM is a single-layered neural network. In order to extract the in-depth features of images, it is required to construct a neural network with deep structure. Deep belief network (DBN) is a deep structure formed by stacking RBM, where the output of the previous layer of RBM serves out the input of the next layer of RBM.

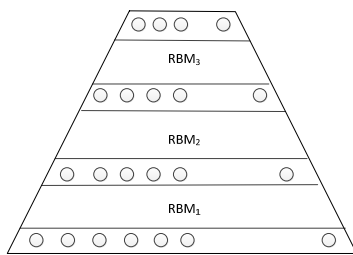


FIGURE II. DBN STRUCTURE

The training of DBN is to insatiably train RBM layer by layer, which means the first layer of RBM shall be sufficiently trained using CD<sup>[14]</sup>(contrastive divergence) training algorithm, and then its parameters shall be fixed, and the status of its hidden layer units shall be used as the input of the second layer of RBM. After the sufficient training for the second layer of RBM, it shall be stacked above the first layer of RBM, and then after above mentioned steps shall be repeated from the bottom to the top so as to train all layers of RBM.

## III. CHANGE DETECTION METHOD BASED ON NEURAL NETWORK

### A. Change Detection Framework

The framework consists of two DBN models (DBN<sub>1</sub> and DBN<sub>2</sub>) to learn the features of the remote sensing images I<sub>1</sub> and I<sub>2</sub>. Take the each pixel value with its neighborhood values reside in g\*g (we take g=7) window. The information of all pixels inside I<sub>1</sub> constructed the group S<sub>1</sub> to train DBN<sub>1</sub>, while the pixel points inside I<sub>2</sub> constructed group S<sub>2</sub> to train DBN<sub>2</sub>. The features of different kinds of pixel points can be preliminarily extracted from trained DBN.

Select the corresponding elements s<sub>1</sub> and s<sub>2</sub> from group S<sub>1</sub> and S<sub>2</sub> respectively. (Where s<sub>1</sub> and s<sub>2</sub> represent the same position in I<sub>1</sub> and I<sub>2</sub> respectively) s<sub>1</sub> and s<sub>2</sub> are the inputs for the DBN<sub>1</sub> and DBN<sub>2</sub> and the output is f<sub>1</sub> and f<sub>2</sub>.

The distance between f<sub>1</sub> & f<sub>2</sub> is:

$$\text{Definition 1: } d(f_1, f_2) = \sum_{i=1}^n |f_{1i} - f_{2i}|$$

Select some prominent changed area manually to construct the changed training sample C, select some prominent unchanged area manually to construct the unchanged training sample NC, then apply BP tuning DBN<sub>1</sub> and DBN<sub>2</sub>. The purpose of tuning DBN<sub>1</sub> and DBN<sub>2</sub> is to decrease the distance (mentioned in Definition1) between the output of two DBNs for C, and increase the distance for NC. To achieve this, Guide the other DBN with the output of one DBN by turns. After several times of tuning, the output of two DBNs reach the standard, then use the distance to create the DI image D, the final change detection result is gotten through PCA-kmeans clustering analysis.

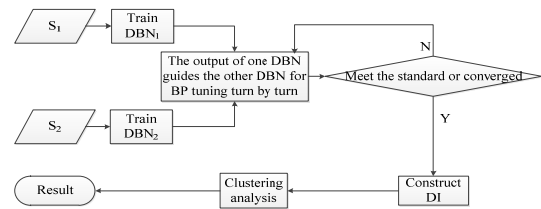


FIGURE III. FLOWCHART OF PROPOSED METHOD

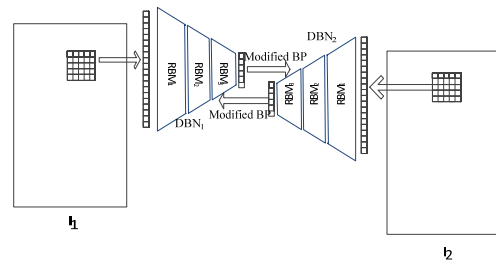


FIGURE IV. FRAMEWORK OF PROPOSED METHOD

### B. Modified BP

BP is a supervised learning algorithm of neural networks. For training sample  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$ , The learning procedure of BP is "tuning the parameter of neural network to make the output closer to y".

To describe the modified BP, we represent the parameter of neural network as follows; the neural network has  $n$  layers with sigmoid function  $f(x)$ :

Definition 2  $z^{(l+1)} = W^{(l)}a^{(l)} + b^{(l)}$

Definition 3  $a^{(l+1)} = f(z^{(l+1)})$

The cost function of BP defines as:

Definition 4  $J(W, b; x, y) = \frac{1}{2} \|a^{(n)} - y\|^2$

The following steps are followed to modify the BP algorithm.

We initialize the parameters of neural network through CD from input  $x$  to output layer  $y$  and then compute the  $a^{(0)}$  (Definition.3) and  $z^{(0)}$  (Definition.2) of each layer. For output layer (layer  $n$ ): (• represents the element wise multiplication)

$$\delta^{(n)} = \frac{\partial J}{\partial z^{(n)}} = -(y - a^{(n)}) \bullet f'(z^{(n)}) \quad (6)$$

Reversely compute  $\delta^{(l)}$  of each layer, for  $l=n-1, n-2, \dots, 2$ :

$$\delta^{(l)} = ((W^{(l)})^T \delta^{(l+1)}) \bullet f'(z^{(l)}) \quad (7)$$

Compute partial derivative of each layer:

$$\nabla_{W^{(l)}} J = \delta^{(l+1)} (a^{(l)})^T \quad (8)$$

$$\nabla_{b^{(l)}} J = \delta^{(l+1)} \quad (9)$$

We labeled the changed and unchanged C & NC respectively samples for training. The purpose of training is to decrease the distance between the outputs of DBN<sub>1</sub> and DBN<sub>2</sub> for NC and increase the distance between the outputs of two DBN for C.

For the NC training samples, updated the parameter positively and for the C training sample, updated the parameter negatively.

For the training samples of the same class, we tried to decrease the variance of the output of two DBN.

For unchanged training sample  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(nc)}, y^{(nc)})\}$

$$a_i = \frac{d_i - d_{\min}}{d_{\max} - d_{\min}} \quad (i = 1, \dots, nc) \quad (10)$$

For changed training sample  $\{(x^{(1)}, y^{(1)}), \dots, (x^{(c)}, y^{(c)})\}$

$$a_j = \frac{d_{\max} - d_j}{d_{\max} - d_{\min}} \quad (j = 1, \dots, c) \quad (11)$$

Update parameters in the modified BP are as follows:

- For all layers:  $\Delta W^{(l)} := 0, \Delta b^{(l)} := 0$

- For  $i=(1, \dots, nc)$ , compute  $\nabla_{W^{(l)}} J, \nabla_{b^{(l)}} J, a_i$

$$\Delta W^{(l)} := \Delta W^{(l)} + a_i \nabla_{W^{(l)}} J \quad (12)$$

$$\Delta b^{(l)} := \Delta b^{(l)} + a_i \nabla_{b^{(l)}} J \quad (13)$$

- For  $j=(1, \dots, c)$ , compute  $\nabla_{W^{(l)}} J, \nabla_{b^{(l)}} J, a_j$

$$\Delta W^{(l)} := \Delta W^{(l)} - a_j \nabla_{W^{(l)}} J \quad (14)$$

$$\Delta b^{(l)} := \Delta b^{(l)} - a_j \nabla_{b^{(l)}} J \quad (15)$$

- Update the parameter: Where  $\alpha$  is learning rate and  $\lambda$  is the Weight decay parameter.

$$W^{(l)} = W^{(l)} - \alpha \left[ \frac{1}{c + nc} \Delta W^{(l)} \right] + \lambda W^{(l)} \quad (16)$$

$$b^{(l)} = b^{(l)} - \alpha \left[ \frac{1}{c + nc} \Delta b^{(l)} \right] \quad (17)$$

- Repeat the procedure until the results are converged and meet the standard.

### C. Clustering analysis

The method of clustering analysis uses the PCA-Kmeans<sup>[2]</sup> algorithm. Take the non-overlapped  $h \times h$  (we take  $h=5$ ) window from DI and then map to the feature space through PCA, then apply Kmeans clustering to detect changed and unchanged groups. Finally the changed group converted into the changed map. We use this method to get the result of change detection and evaluate the quality of DI.

## IV. EXPERIMENT

### A. Introduction

The experiment compared the traditional method of DI which is based on radiometric correction and our method which is based on neural network. **Kappa<sup>[17]</sup> is used for evaluate the accuracy of result**

**We selected some labeled area for both methods manually.**

In radiometric correction, we selected some unchanged area as the basis and in our method we selected some changed and unchanged area for training sample. In our method; the strategy of selecting the area is to select all kinds of changed and unchanged areas present in an image. In traditional method, there are two kinds of strategy to select the unchanged area for radiometric correction:

**1<sup>st</sup> Type** : Select the part of typical unchanged area.

**2<sup>nd</sup> Type**: Select all the unchanged area based on the ground-truth.

In the experiment, we applied & compared these two methods.

In traditional method, after the radiometric correction, AD (Absolute Distance) and ED (Euclidian Distance) are important

and widely use methods to create the DI<sup>[18]</sup>. We compared these two methods.

Our method builds the neural network for each image. The size of the window is 7\*7 with RGB channels, so the dimension of the input data for neural network is 147(7\*7\*3). To find out the effect of layers in neural network, we designed two neural networks and compare their results.

a) Neural Network with 3 layers (147-300-250-200)

b) Neural Network with 5 layers (147-300-250-200-170-150).

### B. Image Infomation

The images are gotten from the Google Earth with RGB channels. We choose three images at different resolution. We record the shooting date, latitude and longitude, resolution of each image, which are as follows:

TABLE I. IMAGE INFORMATION

Group	Position	Date		Pixel	Resolution(m/pixel)
		Image 1	Image 2		
1	N: 41°58'05.41" E: 121°03'11.46"	2013.05.16	2015.09.12	651*521	0.57
2	N: 31°53'23.37" E: 119°03'07.92"	2014.07.23	2010.10.04	701*420	0.9
3	N: 32°13'57.32" E: 118°35'57.08"	2011.02.05	2014.12.16	701*561	1.5

### C. Results

#### 1) Experiment 1:

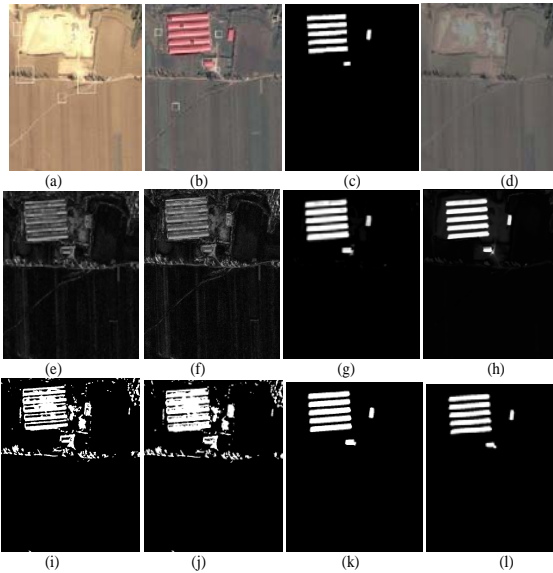


FIGURE V. (A)RS IMAGE1 (B)RS IMAGE2 (C)GROUND TRUTH (D)1<sup>ST</sup> TYPE OF RADIOMETRIC CORRECTION OF "A" (E)DI BASED ON AD WITH 1<sup>ST</sup> TYPE OF RADIOMETRIC CORRECTION (F)DI BASED ON ED WITH 2<sup>ND</sup> TYPE OF RADIOMETRIC CORRECTION (G)DI BASED ON 3 LAYERS NEURAL NETWORK (H)DI BASED ON 5 LAYERS NEURAL NETWORK (I)-(L)CLUSTERING ANALYSIS OF "E-H"

In this experiment, we wanted to detect the red building that represents the change. The white windows in (a) is the selected

unchanged areas for radiometric correction. Traditional method can't suppress the influence of trees, so there are many noises in the final result.

The black and white windows in image (b) are the changed and unchanged training samples respectively for the neural network. The DI (m) and (n) constructed by the 3 and 5 layers of neural networks respectively and suppress all kinds of noise with highlighted the change area.

The experiment results are as follows:

TABLE II. TRADITIONAL METHOD

Criteria	1 <sup>st</sup> Type		2 <sup>nd</sup> Type		Average
	AD	ED	AD	ED	
FP	16130	15152	17293	16197	16193
FN	4390	1143	4363	1230	11126
Kappa	54.08%	65.12%	52.67%	63.37%	58.81%

TABLE III. NEURAL NETWORK METHOD

Criteria	3 layers	5 layers	Average
FP	2802	1336	2069
FN	211	1140	676
Kappa	90.57%	91.68%	91.13%

From table II we concluded that 1<sup>st</sup> Type of radiometric correction is slightly better than the 2<sup>nd</sup> Type and ED is better than AD. From table III we concluded that 5 layers neural network is slightly better than 3 layers with Kappa increased 1.11%. According to the above mention statistics, we found that the average performance of neural network method is much better than traditional method with Kappa increase 32.3%.

#### 2) Experiment 2:

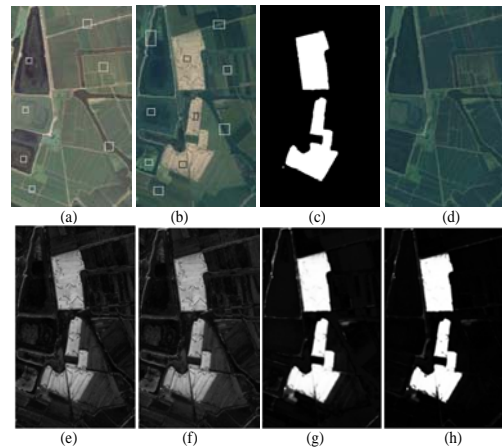


FIGURE VI. (A)RS IMAGE1 (B)RS IMAGE2 (C)GROUND TRUTH (D)1<sup>ST</sup> TYPE OF RADIOMETRIC CORRECTION OF "A" (E)DI BASED ON AD WITH 1<sup>ST</sup> TYPE OF RADIOMETRIC CORRECTION (F)DI BASED ON ED WITH 1<sup>ST</sup> TYPE OF RADIOMETRIC CORRECTION (G)DI BASED ON 3 LAYERS NEURAL NETWORK (H)DI BASED ON 5 LAYERS NEURAL NETWORK

Because the 2<sup>nd</sup> Type of radiometric correction produced the bad result, so we ignored it. The comparison of these two



methods indicated that the neural network method highlighted the change area better with high contrast.

The results of experiment 2 are as follows:

TABLE IV. TRADITIONAL METHOD(1<sup>ST</sup> TYPE)

Criteria	AD	ED	Average
FP	136	112	124
FN	3496	3397	3446
Kappa	94.65%	94.83%	94.74%

TABLE V. NEURAL NETWORK METHOD

Criteria	3 layers	5 layers	Average
FP	60	17	38
FN	836	2240	1538
Kappa	98.72%	96.71%	97.71%

According to table IV and V, we found that the average performance of neural network method is slightly better than traditional method with Kappa increase 2.97%.

### 3) Experiment 3:

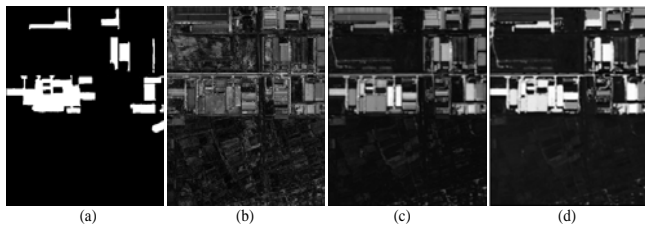


FIGURE VII. (A)GROUND TRUTH (B)DI BASED ON AD WITH 1<sup>ST</sup> TYPE OF RADIOMETRIC CORRECTION(C)DI BASED ON 3 LAYERS NEURAL NETWORK(D)DI BASED ON 5 LAYERS NEURAL NETWORK

The results of experiment 3 are as follows:

TABLE VI. TRADITIONAL METHOD

Criteria	1 <sup>st</sup> Type		2 <sup>nd</sup> Type		Average
	AD	ED	AD	ED	
FP	28422	29655	27438	25133	27662
FN	25039	24709	26905	26987	25910
Kappa	32.75%	32.65%	29.68%	30.89%	31.49%

TABLE VII. NEURAL NETWORK METHOD

Criteria	3 layers	5 layers	Average
FP	17202	4622	10912
FN	9556	8427	8992
Kappa	67.69%	82.34%	75.02%

According to table VII, we found that the 5 layers neural network performed better than 3 layers with kappa increased 14.65% .According to table VI and VII, we found that the average performance of neural network method is much better than traditional method with Kappa increase 42.53%.

## V. CONCLUSION & FUTURE WORK

This paper has proposed a novel method to calculate DI based on DBN and BP. The method also avoids the radiometric correction to calculate the DI. The results of the experiment show that the proposed method performed much better than traditional method with high contrast in DI and have more accurate results.

Our method labeled the typical area manually for training but how to label the typical area automatically will be a direction for future research. Besides this if RS image has many different kinds of changes and un-change areas then how to improve the ability of neural network to better recognize these?

## ACKNOWLEDGMENT

The authors acknowledge the support to this research from The National Natural Science Fund of China(Grant No. 61371168).

## REFERENCES

- [1] Radke R J, Andra S, Al-Kofahi O, et al. Image change detection algorithms: a systematic survey.[J]. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2005, 14(3):294-307.
- [2] Celik T. Unsupervised Change Detection in Satellite Images Using Principal Component Analysis and -Means Clustering[J]. IEEE Geoscience & Remote Sensing Letters, 2009, 6(4):772-776.
- [3] Chen H, Ma L, Qi B G, et al. Relative radiometric correction for remote sensing images based on Vivado HLS[C]// Iet International Radar Conference. 2015.
- [4] Yan W Y, Shaker A. Radiometric Correction and Normalization of Airborne LiDAR Intensity Data for Improving Land-Cover Classification[J]. IEEE Transactions on Geoscience & Remote Sensing, 2014, 52(12):7658-7673
- [5] T. Sohl, "Change analysis in the United Arab Emirates: an investigation of techniques", Photogrammetric Engineering and Remote Sensing, 65,1999, pp. 475-484.
- [6] Tang P Q, Yang J Y, Zhang C, et al. An Object-oriented Post-classification Remote Sensing Change Detection after the Pixel Ratio[J]. Remote Sensing Information, 2010, 32(1):69-72.
- [7] Li S, Hua H. Automatic recognition of landslides based on change detection[C]// International Symposium on Photoelectronic Detection and Imaging. International Society for Optics and Photonics, 2009:27-31.
- [8] Wei L, Zhong Y, Zhang L. Adaptive change method of remote sensing image fusion[J]. Journal of Remote Sensing, 2010.
- [9] Chen B, Polatkan G, Sapiro G, et al. Deep Learning with Hierarchical Convolutional Factor Analysis[J]. IEEE Transactions on Software Engineering, 2013, 35(8):1887-901.
- [10] Arel I, Rose D C, Karnowski T P. Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier][J]. IEEE Computational Intelligence Magazine, 2010, 5(4):13-18.
- [11] Bengio Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-127.
- [12] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks.[J]. Science, 2006, 313(5786):504-507.
- [13] Hinton G E. Learning multiple layers of representation[J]. Trends in Cognitive Sciences, 2007, 11(10):428-34.
- [14] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets.[J]. Neural Computation, 1960, 18(7):1527-54.
- [15] Hecht-Nielsen R. Theory of the backpropagation neural network[J]. Neural Networks, 1988, 1(1):65-93.
- [16] Lari-Najafi H, Nasiruddin M, Samad T. Effect of initial weights on back-propagation and its variations[C]// IEEE International Conference

- on Systems, Man and Cybernetics, 1989. Conference Proceedings. IEEE, 1989:218-219 vol.1.
- [17] Rosenfield G. A Coefficient of Agreement as a Measure of Thematic Classification Accuracy.[J]. Photogrammetric Engineering & Remote Sensing, 1986, 52(2):223-227.
- [18] Du P, Liu S, Gamba P, et al. Fusion of difference images for change detection in urban areas[C]// Urban Remote Sensing Event. IEEE, 2011:165-168.