Advances in Computer Science Research, volume 88
International Conference on Computer, Network, Communication and Information Systems (CNCI 2019)

A Land Cover Refined Classification Method Based on the Fusion of LiDAR Data and UAV Image

Lian Liu[1, a], Xing Liu[1, b], Junjiong Shi [1, c] and Anran Li[1, d]

[1]School of Civil Engineering, Chongqing University, Chongqing 400000, China;

[a]453178160@qq.com, [b]liuxingtm@cqu.edu.cn126.com, [c]17782240248@qq.com, [d]1426271062@qq.com

Abstract. Refined classification is one of the important research directions of remote sensing data classification. With the support of LiDAR data, the fusion of UAV image feature will be an effective mean to expand feature sets and improve classification category differentiation. After expanding the feature sets, there are high dimensional, massive data, small training samples and so on, which are important problems affecting the classification accuracy. Therefore, based on the DSM features obtained from LiDAR, this paper extracts the geometric features of the relevant space according to the neighborhood constraint rules, and proposes a SVM-DBN classification algorithm for small training samples after the fusion spectra, textures and other features. Using the small training samples classification of SVM, the algorithm predicts a large number of inexpensive unmarked samples on the basis of fewer training samples, uses the prediction results as prior-information training DBN and corrects the results of SVM prediction. The classification results show that the method has high overall precision and reliable results, which provide a new idea for the study of land cover refined classification.

Keywords: Refined Classification, Airbrone LiDAR, DSM, SVM-DBN.

1. Introduction

At present, single-source remote sensing data has become increasingly difficult to meet the needs of refined classification. For the classification of high-resolution images, due to the phenomenon of "homologous spectrum" and "foreign body spectrum", only rely on the spectrum is not enough to accurately distinguish different categories [1].

As a three-dimensional data, it is one of the effective means to improve the number of classification categories and ensure the accuracy of classification with the combination of LiDAR data and images. Fusion data, including multi-dimensional contents of information of spectral, spatial… [2], can improve the object classification number and class differentiation. With the support of point clouds elevation data, the method of refined classification in remote sensing image was proposed and achieved out [3]. In order to effectively improve the accuracy of urban land use classification, a classification method was proposed using high-resolution Digital Surface Model (DSM) and Digital Orthophoto Map (DOM) [4]. Multi-source data fusion has now become an indispensable technical mean in the process of classification [5,6].

In the current process of supervision classification, It take the Support Vector Machine (SVM) sorter with structural risk minimization, Deep Belief Network (NN) sorter with nonlinear mapping ability as the mainstream. The core idea of SVM is to projection to high-dimensional feature space through nonlinear mapping, seeks most optimal classification hyperplane as far as possible [7]. SVM has the ability of small train sample classification, but its kernel function and parameter optimization are important and difficult problem. A variety of optimization parameter methods are easy to fall into the local optimal solution, and there are common problems of time-consuming and complex calculation. The core of DBN lies in its nonlinear learning characteristics, which have the ability to fit and generalize large amounts of data, it also determines that the process of training data requires a lot of training samples. In classification, it would be a hassle to select training samples suitably in rich detail land cover images, and the selection of training samples is another key problem affecting the classification accuracy [8,9].

In this paper, the composite bands of UAV images and DSM generated by LiDAR are used as data source. Considering the small training samples, the SVM-DBN classification method was proposed and turned to be effective, which use SVM to classify a large number of inexpensive unmarked samples, through the statistical information of various classified categories of land cover in the feature space, as prior-information to train DBN parameters and correct their classification categories. Experiments show that, on the basis of the feature sets that combine surface spatial information with spectral and texture information, the method mentioned, which makes up for the deficiency of a single classifier, obtains better classification results.

2. the Information Extraction of Surface Spatial and Spectral

2.1 Description of Experimental Area.

The study area is Belgian port UAV image and airborne LiDAR data, and both types of data are registration. UAV image and LiDAR data as shown in Fig. 1.

a. UAV image b. LiDAR data

Fig. 1 Original data (500*500)

2.2 Second-order Statistics of Elevation-Level and Gray-Level.

Texture features are described by Gray-Level Co-occurrence Matrix (GLCM), which is a method to describe textures by using the characteristics of pixel grayscale space correlation [10].

By calculating the correlation statistics between pixels in a certain distance and direction in the image, the texture statistics are obtained on the assumption that the spatial distribution relationship between the pixel and the neighborhood one. In essence, it is a second-order statistical information based on spatial grayscale information.

The commonly used texture statistics are shown in the table 1. *f(i,j)* is the pixel value of row *i* and column *j*, μ is the mean of the pixels in the window and σ is the standard deviation of the pixels.

Table 1 Common Texture Statistics

Texture statistics	Calculation Formula
Homogeneity	$\sum_{i}\sum_{j} f(i,j) / 1+(i-j)^2$
Entropy	$-\sum_{i}\sum_{j} f(i,j)\log[f(i,j)]$
Second Moment	$\sum_{i}\sum_{j} (f(i,j))^2$
Contrast	$\sum_{i}\sum_{j} (i-j)^2 f(i,j)$
Dissimilarity	$\sum_{i}\sum_{j} \lvert i-j \rvert f(i,j)$

Correlation	$\frac{\sum_i \sum_j p_{ij}(i-\mu_i)(i-\mu_j)^2}{\sqrt{\sigma_i^2 \sigma_j^2}}$
Mean	$\frac{1}{n^2}\sum_i \sum_j f(i,j)$
Variance	$\frac{\sum_i \sum_j (f(i,j)-f(i,j))^2}{n^2-1}$

Texture is the local spatial information between adjacent pixels, which is the neighborhood space constraint of grayscale [11]. The surface texture is a local neighborhood feature that characterizes the change of elevation space, so there is also a certain spatial relationship between the elevation primitive separated by a certain distance. It not only reflects the distribution characteristics of the surface, but also reflects the location distribution characteristics between the same object. The Elevation-level co-occurrence Matrix (ELCM) based on the DSM can reflect the Characterization information of height on direction, adjacent interval, variation amplitude, etc.

The GLCM is defined by the grayscale-combined probability density of different position pixels in the neighborhood, while in the ELCM, the counted object becomes the elevation primitive, which is the elevation-combined probability density of pixels. According to the classification of different objects with similar spectral information, such as tree and grass, the degree of inter-category differentiation is increased after the elevation distribution characteristics are fused. The partial feature extraction results are shown in Fig. 2:

a. Contrast Based on ELCM | b. Second Moment Based on ELCM | c. Contrast Based on GLCM (r band) | d. Second Moment Based on GLCM (r band)

Fig. 2 The partial feature extraction results

2.3 General Statistical Characteristics.

The general statistical feature is obtained by the point operation of the original band, and for the surface features, the operation is based on DSM or DEM data. Surface and spectral are the essential characteristics of the features, which can be used to construct feature sets to distinguish different ground objects. For UAV images, the spectral value of each pixel represents the radiant brightness value of the object, while for surface feature, the value of each pixel represents the height of the object or the elevation.

The general statistical characteristics calculation formula is shown in the Table 2. In order to expand the feature sets, joined the vegetation features, height difference, LiDAR intensity.

Table 2 General Statistical

Other Characteristics	Calculation Formula
Homogeneity	$\sum_i \sum_j f(i,j) / 1+(i-j)^2$
Entropy	$-\sum_i \sum_j f(i,j)\log[f(i,j)]$

Second Moment	$\sum_i \sum_j (f(i,j))^2$
Contrast	$\sum_i \sum_j (i-j)^2 f(i,j)$

2.4 Construction of Feature Sets.

After extracting the representative features of UAV image and LiDAR data, the initial feature space is constructed, and the experimental window size is 3*3.

The feature sets of the study area are shown in Table 3:

Table 3 The Feature Sets

Characteristics	Spectral(I)	Surface Spatial(D)
Original data	I_1, I_2, I_3	D_1
Mean	I_4, I_5, I_6	D_2
Variance	I_7, I_8, I_9	D_3
Entropy	I_{10}, I_{11}, I_{12}	D_4
Skewness		D_5
ELCM and GLCM	$I_{13}, \ldots, I_{36}$	$D_6, \ldots, D_{13}$
other	VI (I_{37}), HD (D_{14}), LiDAR-Int (D_{15})	

3. SVM-DBN Classification Method

On the basis of fewer training samples, this method uses a large number of inexpensive unknown samples to correct the classification deviation of the tagged samples, which may be caused by insufficient representation by excavating the inherent structure information of the types to be classified in the unknown samples. The process is shown in Fig. 3

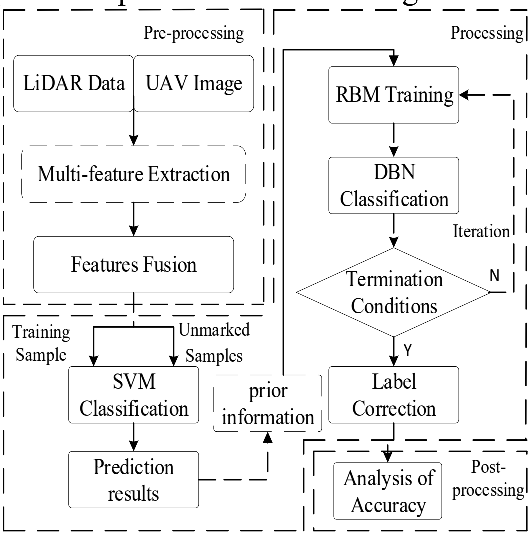


Fig. 3 Scheme of land cover refined classification

3.1 SVM Classification.

The main idea of SVM is to establish an optimal classification hyperplane as a decision surface, so that the isolation edges between different samples are maximized. The best position of the classification hyperplane is to give the maximum distance between the two types of samples, and the vectors closest to the taxonomic hyperplane can construct two edge hyperplane, which is called the support vector. Once the support vector that satisfies the optimal condition is obtained, other training data in a sense is useless [12].

Nonlinear mapping is introduced for nonlinear separable data: $\Phi : x_i \rightarrow \Phi(x_i)$, the original training data is mapped to the high-dimensional feature space so that it can be divided linearly. The kernel function calculation is introduced because the training data only needs to be calculated by the inner product and does not have to get the actual form of $\varphi(x)$. The common kernel functions are polynomial kernel function, radial base kernel function, sigmoid kernel function and so on. In addition to the kernel function parameters, the penalty coefficient *C* is also an important parameter. Find the maximum classification interval by adjusting *C* without having to make all training samples meet the correct classification. Because the SVM with RBF as kernel function can obtain satisfactory results in multidimensional data classification [13], it is selected as kernel function, at the same time, the training samples are less and the separation is better, and the penalty coefficient is set to 100.

3.2 DBN Classification.

A DBN can be defined as a stack of Restricted Boltzmann Machines (RBM) [14], in which each RBM layer communicates with both the previous and subsequent layers. The nodes of any single layer don't communicate with each other laterally [15]. A typical DBN consists of multilayer RBMs and a layer of Back Propagation (BP), the structure of which is as Fig. 4:

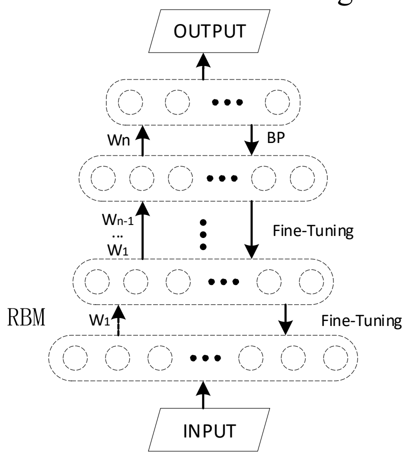


Fig. 4 The structure of DBN

The RBM contains the visible layer *v* and the hidden layer *h*, which has the input units $v=\{v_1,v_2...v_m\}$, the hidden units $h=\{h_1,h_2...h_n\}$, *m* and *n* are the number of visible and hidden elements respectively. *w* is the connection weight between the two layers, a_i and b_j are the bias value of the visible layer *i*-th neurons and the hidden layer *j*-th neurons respectively.

The training process of RBM is actually to find a probability distribution that can best produce training samples. In other words, to find a distribution so that the probability of training samples in this distribution is greatest. When a given unit state, the activation probability of the forward or back unit is satisfied with the condition Independent [16]. For RBM, its system energy function is:

$$E(v,h|\theta) = -\sum_{i=1}^{m} a_i v_i - \sum_{i=1}^{m} b_j h_j - \sum_{i=1}^{m}\sum_{j=1}^{n} v_i W_{ij} h_j \tag{1}$$

$\theta=\{W_{ij},b_i,c_j\}$ are parameters of RBM and training targets. Theoretically, when parameters are determined, the joint probability distribution *(v,h)* can be obtained:

$$P(v,h|\theta) = e^{-E(v,h|\theta)} \big/ Z(\theta) \tag{2}$$

$$Z(\theta) = \sum_{v,h} e^{-E(v,h|\theta)} \tag{3}$$

Z(θ) is normalized factor. In practical problems, the probability distribution of the observed data *v* is the marginal distribution $P(v|\theta)$ of the joint probability distribution $P(v,h|\theta)$, that is:

$$P(v|\theta) = \sum_h e^{-E(v,h|\theta)} \big/ Z(\theta) \tag{4}$$

According to conditional independence assumption, the probability of each layer neurons are:

$$P(h_j = 1|v,\theta) = \sigma\left(b_j + \sum_i v_i W_{ij}\right) \tag{5}$$

$$P(h_j = 0|v,\theta) = 1 - P(h_j = 1|v,\theta) \tag{6}$$

$$P(v_i = 1|h,\theta) = \sigma\left(a_i + \sum_j h_j W_{ij}\right) \tag{7}$$

$$P(v_i = 0|h,\theta) = 1 - P(v_i = 1|h,\theta) \tag{8}$$
$$\sigma(x) = (1 + e^{-x})^{-1} \tag{9}$$

After maximizing the logarithmic likelihood function of RBM on the training sets, the parameter θ^* of the fitting t samples are obtained, that is:

$$\theta^* = \underset{\theta}{argmax}\, L(\theta) = \underset{\theta}{argmax}\, P(v^t|\theta) \tag{10}$$

By using the stochastic gradient ascent algorithm and the contrastive divergence algorithm, the parameter update rules are obtained by k-th (usually k=1) Gibbs sampling [17]:

$$\Delta W = \varepsilon(P(h_1 = 1|v_1)v_1' - P(h_2 = 1|v_2)v_2') \tag{11}$$
$$\Delta b = \varepsilon(P(h_1 = 1|v_1) - P(h_2 = 1|v_2)) \tag{12}$$

BDN training usually requires a large number of marked samples to train unknown parameters, in contrast, SVM has more appropriate requirements for the number of marked samples. In this paper, by sampling a small number of sample points, based on SVM classification, the classification results as prior-information pre-training parameters of the DBN classifier, using a trained classifier for classification.

4. Experimental Analysis

As mentioned earlier, the experiment is divided into two stages, the first stage is feature sets construction. The second stage is based on SVM classification results as prior-information training DBN, to strengthen the classification of classifiers ability to predict.

4.1 Data Pre-processing.

If the pixel is $P(i,j)$, the dataset is $P_{i,j,q} \in \mathbb{R}^{m*n*(l+k)}$, where m, n is the image width and height, l is the feature dimension, k is the label. For visual interpretation of UAV images, the classification categories are determined as Table 4.

Table 4 Descriptions of Classification Categories

Category	Description
Tree	Independent shrubs or tall trees
Building1	Red brick spire house
Building2	Asphalt spire house
Building3	Asphalt flat-top houses
Road	Asphalt road
Grass	Artificial green space
River	River
Bare-land	Bare land

The dataset are divided into training samples and unknown samples, and the initial categories of samples to be classified are obtained by using SVM classification. For pixels, feature dimensions $X=\{I_1,I_2\cdots I_{37},D_1,D_2\cdots D_{15}\}$, the category label is $Y\in\{y_1,y_2\cdots y_k\}$,($k=8$), and the category labels of unknown samples are null. Each dimensions is standardized and the formula is as follows:

$$X = {x - x_{min}}/{x_{max} - x_{min}} \tag{13}$$

4.2 Classifier Training.

In order to verify the optimization function of the spatial constraint feature of the added DSM neighborhood in the classification, two sets of different feature sets based on SVM predictive initial label were designed experimentally. As feature sets $X_1=\{I_1,I_2,\ldots,I_{37},D_1,D_2,D_3\}$and feature sets $X_2=\{D_4,D_5,\ldots,D_{37}\}$, the Initial label Y_1^* and Y_2^* as the prior information of training neural network.

After the primary training is completed, the global error is obtained from the prediction result Y_1, Y_2 and the initial label sample set, and the repetition algorithm is not in the allowable range until the error value. The pre-trained parameters are assigned to the DBN network to predict the label of the samples.

Table 5 Descriptions of experimental samples

	SVM feature sets	Pre-label	DBN feature sets	Result
TEST 1	X	Y_1^*		Y_1^*
TEST 2	X	Y_1^*	X	Y_1
TEST 3	X_1	Y_2^*		Y_2^*
TEST 4	X_1	Y_2^*	X_1, X_2	Y_2

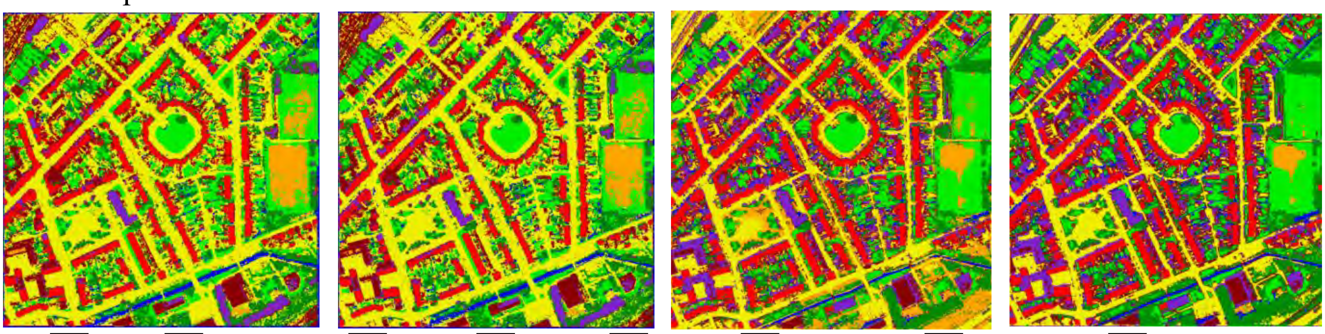
4.3 Results Analysis.

Based on different feature sets, this paper completes the refined classification results of four groups of urban features under different priori samples, and evaluates the accuracy of each classification result. Based on the classification results and visual interpretation results, the evaluation method is mainly confused matrix. Evaluation indexes include overall accuracy, kappa coefficient, user's accuracy and producer's accuracy. The results are shown in the Table 6.

Table 6 Three scheme comparing (%)

	TEST 1		TEST 2		TEST 3		TEST 4	
	Prod. Acc	User Acc.	Prod. Acc	User Acc.	Prod. Acc	User Acc.	Prod. Acc	User Acc.
Tree	47.38	68.53	45.12	65.21	90.19	90.57	93.76	91.63
Building1	86.27	94.57	84.41	97.48	98.18	99.56	98.69	99.63
Building2	63.22	44.66	70.89	42.59	91.11	71.12	93.56	70.7
Building3	72.91	56.73	71.9	62.22	47.12	73.32	47.12	76.58
Road	98.3	90.41	98.5	89.33	96.19	95.45	96.96	96.25
Grass	68.57	80.73	68.01	80.15	84.28	93.17	87.93	91
River	94.77	94.58	89.73	91.68	91.47	97.52	98.64	93.74
Bare-land	99.07	75.71	96.06	77.97	95.82	70.48	89.33	100
OA	79.15		78.51		90.61		92.16	
Kappa	0.7489		0.7416		0.8863		0.9049	

In TEST 1, due to the X high feature sets dimensions (52 dimensions), it is not ideal to rely on SVM classification results, and the OA is only 79.15%. TEST 2 shows that training DBN and classifying samples on the basis of too large prior-information error did not improve the classification results. Using X_1 for SVM classification in TEST 3, the OA has achieved 90.61%. In TEST 4, Using X_1 based on SVM to get the initial labels, and Using Y_2^* as prior information and adding spatial constraint features X_2, the OA reached 92.16%. TEST 4 shows that the classification accuracy can be improved effectively by using prior-information for classification after adding surface space detail information. Further analysis shows that after adding the spatial constraint feature X_2, the classification accuracy of each category increases, and the spectral characteristics of similar features, such as grass and trees, have improved.



Tree Building1 Road Grass River Building2 Building3 Bare-land

a. TEST 1 b. TEST 2 c. TEST 3 d.TEST 4

Fig. 5 The result of four kind of classification methods

In addition, to verify the effect of the number of priori samples on the classification results, the influence of the priori-samples with different ratios on the classification accuracy was randomly selected experimentally.

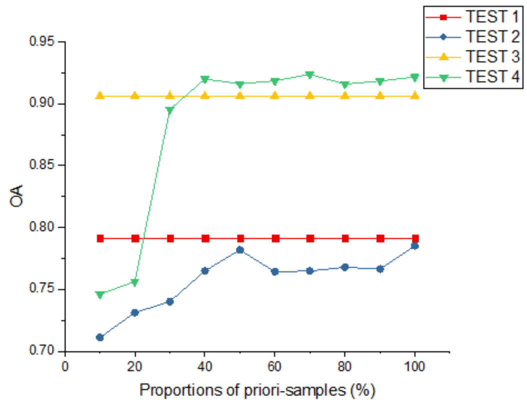


Fig. 6 The OA of classification based on different proportional priori-samples

TEST 1 and TEST 3 are the SVM classification results, which are plotted in the same diagram for reference, and two experiments are not extracted proportionally.

The experimental results show that when the proportion of priori samples is more than 30%, the overall accuracy is improved, so under the condition of strong time constraints, the initial labels can be optimized by randomly extracting more than 30% priori samples.

5. Conclusion

In the refined classification, for categories with similar spectral information, such as building2 and building3 or grass and tree, the separating degree of different categories can be enhanced by introducing second-order statistics of Elevation-Level and Gray-Level to improve classification accuracy

The optimal results are obtained under the condition of SVM limited information, which can easily solve the problems such as high dimensional features and small training samples. In the actual classification process, sampling large training data are often time-consuming and power-intensive. Pre-classification based on SVM by selecting less training samples, the results are not completely accurate, but it can be used as prior-information training DBN classifier to achieve self-learning and correction of sample labels, and improve the accuracy of classification.

References

[1]. ZHAI Tian-lin, JIN Gui, DENG Xiang-zheng, Research of WUHAN city land use classification method based on multi-source remote sensing image fusion. Resources and Environment in the Yangtze Basin, 2016, 25(10):1594-1602.

[2]. PENG Daifeng, ZHANG Yongjun, XIONG Xiaodaong, 3D Building change detecyion by combining LiDAR point clouds and aerial image[J]. Geomatics and Information Science of Wuhan University, 2015, 40(04):462-468.

[3]. Dong Baogeg, Ma Hongchao, Che Sen, Method of Land Cover Refinded Classification Supported by LiDAR Point Clouds[J]. Remote Sensing Technology and Application, 2016, 31(1):165-169.

[4]. Song X Y, Huang Y H, Dong D L, Urban land use classification from UAV remote sensing images based on digital surface model[J]. Journal of Geo-information Science, 2018, 20(5):703-711.

[5]. YAN Li, JIANG Yuwen, Fusion Classification of Laser Point Cloud and Airborne Image[J]. Bulletin of Surveying and Mapping, 2015(10):44-46+64.

[6]. XIE Rui, CHENG Xiaojun, GUAN Haiyan, Classification and Accuracy Analysis of LiDAR and Aerial Images[J].Journal of Tongji University(Natural Science),2013,41(04):607-613.

[7]. Christopher, M., Bishop. Pattern recognition and machine learning[M]. Springer, 2016.

[8]. Li, Fan Xu, Linlin Siva, Hyperspectral Image Classification With Limited Labeled Training Samples Using Enhanced Ensemble Learning and Conditional Random Fields[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2015, 8(6):2427-2438.

[9]. CUI Binge, MA Xiudan, XIE Xiaoyun, Hyperspectral image de-noising and classification with small training samples [J]. Journal of Remote Sensing, 2017, 21(05):728-738.

[10]. Haralick R M, Shanmugam K, Dinstein I,Texture Features for Image Classification. IEEE Transactions on Systems Man and Cybernetics. 1973.

[11]. Hou Q Q, Wang F, Yan L, Extraction of color image texture feature based on gray-level co-occurrence matrix [J]. Remote Sensing for Land and Resources, 2013, 25(4):26-32.

[12]. John A. Richards, Xiuping Jia, Remote Sensing Digital Image Analysis: An Introduction. Springer [M], 2012-9.

[13]. TIAN Yuan, TASHPOLAT.Tiyip, DING Jian-li, Land Cover Classification Using Remote Sensing Image Based on Support Vector Machines. Resources Science, 2008(08):1268-1274.

[14]. Hinton GE, Salakhutdinov RR, Reducing the dimensionality of data with neural networks. Science. 2006 Jul 28; 313(5786):504-7.

[15]. N.Le Roux, Y.Bengio, Representational power of Restricted Boltzman Machines and Deep Belif Networks. Neural Computation, 20(6):1631-1649, 2008.

[16]. Y.Freund, D.Haussler, Unsupervised learning of distributions on binary vectors using two layer networks. Technical Report UCSU-CRL-94-25, University of California, Santa Cruz, 1994.

[17]. Hinton G E. Training products of experts by minimizing contrastive divergence[M]. MIT Press, 2002.