Hyperspectral Image Recognition Based on Artificial Neural Network

Feng Liang, Hanhu Liu*, Xiao Wang, Yanyan Liu

ABSTRACT

This paper aims to reduce the dimensions of the data in remote sensing hyperspectral image (HIS), and solve the information redundancy caused by the numerous bands in the image. To this end, the neural network sensitivity analysis (NNSA) was introduced to simplify the dimensionality reduction process. Meanwhile, the convolutional neural network (CNN) was adopted as the classification algorithm for the HIS, seeking to prevent the complex data reconstruction of feature extraction and classification. Then, the proposed method was contrasted with several other classification methods in several experiments. The results show that the proposed method outperformed the contrast plans in classification accuracy. Thus, the artificial neural network (ANN) is good at reducing the dimensions of remote sensing HSI and the CNN is a reliable classification tool. The research findings shed new light on remote sensing image processing and other related operations.

Key Words: Artificial Neural Network (ANN), Hyperspectral Image (HIS), Convolutional Neural Network (CNN)

DOI Number: 10.14704/nq.2018.16.5.1244 **NeuroQuantology 2018; 16(5):699-705**

Introduction

Inspired by the neural network in human brain, the artificial neural network (ANN) is a computing model of the adaptive, nonlinear and adaptive network system involving multiple interconnected neurons (Herb, 1949; Liu, 2006; Hu *et al.,* 2016; Kadri and Mouss, 2017). The ANN processes, learns and stores information similarly to the human brain. Below is a brief review of the key moments in the history of the ANN.

The first computing model for neural network, the McCulloch-Pitts (MP) model, came into being in 1943 (McCulloch, 1943). From late 1950s to early 1960s, Rosenblatt (Rosenblatt, 1958; Rosenblatt, 1962) combined the learning function and the MP model into the single layer perceptron model, marking the start of practical research into the neural network. Nevertheless, the single layer model cannot solve linearly inseparable problems. The defect was overcome in 1986 when Rumelhart *et al*., proposed an error backpropagation (BP) training algorithm, which targets multilayer feedforward network containing a BP network. In 2006, Hinton highlighted the importance of deep learning, setting off a new upsurge of ANN research.

At present, the deep learning algorithm has already outperformed the traditional methods in the classification of remote sensing images (Salakhutdinov *et al.,* 2012), prediction of traffic flow (Vincent *et al.,* 2008) and blind assessment of image quality (Vinrent *et al.,* 2010). For the classification of remote sensing images, most of the existing studies focus on the classification of multispectral remote sensing data. However, there is little report on the hyperspectral image (HSI), a multidimensional information image acquired by hyperspectral remote sensing (Landgrebe, 2002). The high-

Corresponding author: Hanhu Liu
Address: Key Laboratory of Geoscience Spatial Information Technology, Ministry of Land and Resources, Chengdu University of Technology, Chengdu 610059, China
e-mail ✉ liuhanhu@qq.com
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
Received: 21 February 2018 ; **Accepted:** 27 March 2018

dimensional spectral data are the attribute that differentiates the HSI from the traditional multispectral data.

Considering the multidimensional features of the HSI, this paper explores the data reduction and HSI object recognition using the ANN, an imitator of the thinking mode of human brain. The purpose is to provide a new solution to the classification of the HSI.

Methods

Data reduction based on ANN

Remote sensing HSI often carries numerous redundant information, due to its high spectral resolution, multiple spectral bands and huge data size. The redundancy can be solved by dimensionality reduction. There are two common ways to reduce the dimensionality of the HSI: feature extraction and band selection. By feature extraction, the spectral bands are recombined and optimized to reduce the number of dimensions. However, this method lacks sufficient classification accuracy, owing to the destruction of spectrum features and the loss of spectral information. By contrast, band selection can avoid the Hughes phenomenon through band reduction. Therefore, band selection was adopted for HSI dimensionality reduction.

Besides band selection, the neural network sensitivity analysis (NNSA) was also adopted for our research. The analysis method quantifies the impact of input attributes on the output of the model, ranks the input attributes by the degree of impact, and eliminates those low impact ones. For dimensionality reduction, the NNSA was combined with BP neural network (BPNN) classifier, seeking to quantify the effect of each band on classification accuracy and determine the most impactful band combinations (Wang *et al.,* 2016; Fichera *et al.,* 2017).

NNSA

In the NNSA, the model is expressed as $y = f(x_1, x_2, \ldots x_n)$, where x_i is the i-th attribute value. Each attribute of the model has a range of possible values. The model predicts how much each attribute affects the output value of the model. The effect is also known as the sensitivity coefficient. In other words, the sensitivity coefficient quantifies the effect of attribute xi on output value y. Then, the sensitivity coefficients are ranked in descending order, and the bottom ones are eliminated, making it possible to reduce the dimensions.

Band selection

As mentioned above, the NNSA was combined with the BPNN classifier to select the bands that contribute most to the classification accuracy, laying the basis for HSI dimension reduction. Here, the band selection is implemented in the following steps:

(1) Data pre-processing

Due to the strong correlation between the bands in the original remote sensing HSI, it is very important to identify the bands with loose connections as the inputs. To this end, all the bands in the original image were divided into a number of subspaces for band selection. The data portioning was realized by the adaptive subspace decomposition based on correlation filtering (Zhang *et al.,* 2000).

First, the correlation coefficient $R_{i,j} \in [-1, 1]$ between two bands was computed, with i and j being the i-th band and the j-th band, respectively.

The absolute value of the coefficient is positively correlated with the relevance between the two target bands, i.e. the closer it gets to 0, the weaker the relevance. The correlation coefficient can be expressed as:

$$R_{i,j} = \frac{E\{(x_i - u_i)(x_j - u_j)\}}{\sqrt{E\{(x_i - u_i)^2\}}\sqrt{E\{(x_j - u_j)^2\}}} \quad (1)$$

Where u_i and u_j are mean greyscale of the two target bands, that is, the mean values of x_i and x_j, respectively; $E()$ is the mathematical expectation of the function in brackets, i.e. the mean number of sampling points in each band. Let the total number (dimensions) of the HSI be N. Then, the correlation coefficient matrix$R$ can be obtained as:

$$R = \begin{bmatrix} 1 & \cdots & R_{1n} \\ \vdots & \ddots & \vdots \\ R_{n1} & \cdots & 1 \end{bmatrix} \quad (2)$$

Where the diagonal is the autocorrelation coefficient whose value is 1. In light of the pre-set threshold T, all the continuous bands satisfying $R_{i,j} \geq T$ were combined into a new subspace. The size of T can be adjusted to change the number of bands and that subspaces in each space. In the space, the value of T is positively correlated with the number of subspaces, and negatively with the number of bands.

The partitioned feature of correlation between HSI bands was quantified by the

correlation coefficient matrix *R*, laying the basis for the division of strong continuous band in subspace. Then, the composition of the bands was selected in each subspace to reduce the band correlation.

(2) Sensitivity analysis and dimensionality reduction

The activation function in the network was employed to find the partial derivative and disclose the effect of input data on output value. Then, the Ruck sensitivity analysis, a method based on partial derivative sensitivity, was introduced to reduce the dimensions of the HSI (Ruck *et al.,* 1990).

Suppose there is an N-L-1 feed forward neural network, with N being the number of input layer neurons, and L being the hidden layer neurons. Let the activation function of the hidden layer be a hyperbolic tangent function $f(x) = (e^x-e^{-x})/(e^x+e^{-x})$, and that of the output layer be a linear function. Then, the sensitivity coefficient of input attribute x_i to output value y can be expressed as:

$$s_i = \frac{\partial y}{\partial x_i} = f'(net_k)\sum_{j=1}^{L} w_{ij}\, v_{j1} f'(net_j) \quad (3)$$

Where $w_{i,j}$ and v_{j1} are the connection weights of input layer neuron i to hidden layer neuron j, and that of hidden layer neuron j to output layer neuron k=1 respectively; f'(net_j) and f'(net_k) are the activation function of hidden layer neuron j, and the partial derivative function of the activation function of output layer neuron k=1, respectively. Since the activation function of the output layer be a linear function and f'(net_k)=1, formula (3) can be rewritten as:

$$s_i=\partial y/\partial x_i=\Sigma_{j=1}^{L} w_{ij}v_{j1}(1+f_j(x))(1-f_j(x)) \quad (4)$$

where f'(net_j)=(1+ f$_j$(x))(1-f$_j$(x))f'(net_k)=1.

The Ruck sensitivity analysis above only discusses the effect of input attribute on output value of a single sample. A comprehensive evaluation function is needed to synthesize the sensitivity analysis results of all samples. Hence, the measurement system analysis (MSA) metrics (Zurada *et al.,* 1997) were chosen to be the comprehensive evaluation function.

It is assumed that S_{ik} is the sensitivity coefficient of input variable i to output variable Y_k(k=1) for all samples, and s_{ik}^t is the sensitivity coefficient of input variable i to output variable Y_k(k=1) for the t-th sample. Thus, the comprehensive evaluation function can be expressed as:

$$S_{ik} = \sqrt{\frac{\sum_{t=1}^{n}(s_{ik}^t)^2}{n}} \quad (5)$$

Where n is the total number of samples; S_{ik} is a nonnegative parameter that sorts out the input bands by the sensitivity to output values, that is, the degree of impact of input variables on output results.

Through Ruck sensitivity analysis, the HSI dimensionality was reduced in the following steps:

(1) Pre-process the original remote sensing HSI, eliminate the interference bands, select the preliminary objects, perform subspace division, and determine the BPNN topology.

(2) Optimize the weight threshold of the BPNN.

(3) Rank the bands by the sensitivity coefficient, remove those of low sensitivity, and find the band combinations that affect the classification accuracy.

(4) Classify the selected band combinations by a convolutional neural network (CNN) classifier, aiming to verify the effect of dimensionality reduction.

CNN-based classification

The CNN is an integration between the ANN and the structure of nervous system in human brain. With deep learning ability, the bio-inspired network has unique advantages in speech recognition and image processing, especially in visual image processing. First proposed by Fukushima in 1980, the CNN has been improved by Yan Lecun in 1998 (Seide *et al.,* 2011), generalized by Seven Behnke and simplified by Patrice Simard in 2003 (Cho *et al.,* 2014), and further enhanced by Dan Ciresan in 2011 and 2012. So far, the CNN has achieved desirable results in the experiments on image databases like MNIST, NORB, HWDB and CIFAR10 (Lee *et al.,* 2003).

In this paper, the CNN is adopted to introduce feature learning into multilayer neural network through weight sharing and reconfiguration of network structure. Meanwhile, the network was subject to the classification training by the BP algorithm. The CNN contains 3 types of layers: the convolutional layer, pooling layer (optional) and fully-connected layer. Each layer consists of multiple 2D planes, each of which involves many independent neurons.

Every pixel on the convolutional layer is a neuron. The neuron is connected to a local area in the upper layer, and can thus be viewed as a local feature detector. Each neuron can extract many kinds of visual features, namely, the primary directions, the turning points, and the local connections that hinder the training of network parameters (Hubel *et al.*, 1962).

The convolutional layer can be expressed as:

$$x_j^l = f x_j^l \left(\Sigma_{i \in M_j} x_j^{l-1} \cdot k_{ij}^l + b_j^l \right) \quad (6)$$

Where l is the number of convolutional layers; k is a convolution kernel; b is the bias; f is the activation function; M_j is an input feature map of the previous layer.

The convolutional layer is usually followed by a pooling layer, which samples the feature image of the convolutional to reduce the resolution. The pooling layer can be expressed as:

$$x_j^l = f x_j^l (\beta_j^l down(x_j^{l-1}) + b_j^l) \quad (7)$$

Where *down* () is the pooling function; β is the pooling coefficient; b is the bias; f is the activation function.

The fully-connected layer classifies the features of acquired higher-order invariants. After the information is transformed and computed in this layer, the forward propagation of learning is completed, and the final results are outputted by the output layer.

During the classification of remote sensing HSI, the CNN was trained in two phases (Fukushima, 1980):

(1) Forward propagation

The sample (X, Y_p) was extracted from the sample set, and X was inputted to the network. Then, the information was transferred step-by-step from the input layer to the output layer. Finally, the actual output was exported from the output layer. The entire process can be described as:

$$O_p = F_n(\ldots(F_2(F_1(\mathrm{XW}_1)W_2)\ldots)W_n) \quad (8)$$

(2) BP

The BP is a phase of error propagation. The deviation of the actual output O_p and the ideal output Yp was calculated, and the weight matrix was adjusted by error minimization method. The BP process can be described as:

$$E_p = 1/2(\Sigma_j(y_{pj} - O_{pj})^2) \quad (9)$$

The successful application of the CNN in the classification of remote sensing HSI requires two hypotheses (Hubel, 1962): First, each neuron has only a few inputs, such that the gradient is spread across as many layers as possible; Second, the hierarchical local connection structure boasts strong prior knowledge and fits in with computer vision tasks; the gradient-based optimization algorithm can achieve good learning effects if the parameters of the entire network are in the right place.

Experiment

Study area and data source

The study area is about 15km south of Goldfield, the seat of Esmeralda County, Nevada. Most of the rocks in the area are exposed, including Cambrian sedimentary rocks and metamorphic sedimentary rocks, Tertiary volcanic rocks and Quaternary alluvial rocks (Abrams *et al.,* 1980). To verify the accuracy of rock identification, the rock distribution map of the study area was compiled based on the data in USGS Circular 1995 (Figure 1).

The experimental data were extracted from the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). The imaging took place on June 19, 1997 at the spatial resolution of 20m. As shown in Figure 1, the image shows the radiant energy data measured by the unit of uW/(cm²·nm·sr).

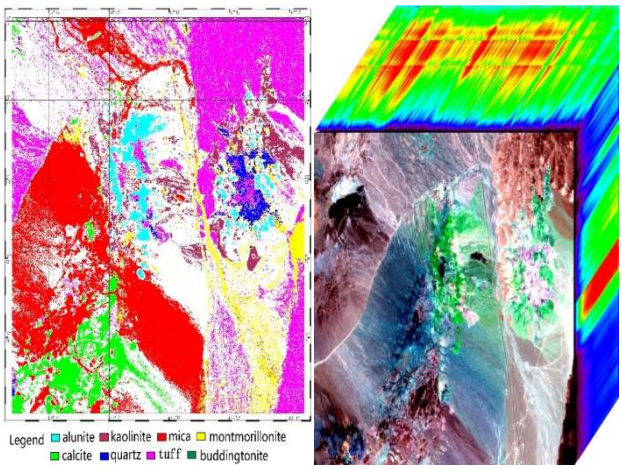

Figure 1. Rock distribution map (left) and AVIRIS image (right)

Experimental design

Only one output layer neuron was designed according to the precondition of Ruck sensitivity

analysis. Considering the multiple objects to be classified, the output variables were in the form of real numbers rather than binary figures. The output neuron could be classified into several categories. All band combinations were collected for the same band in the same subspace. The correlation coefficient between each two bands was identified by the adaptive subspace decomposition based on correlation filtering.

Because of the limited space, the author only presented the grayscale diagram of the AVIRIS correlation coefficient matrix. As shown in Figure 2, the correlation coefficients are relatively high in the bright spots, and peak at the brightest spots; those along the diagonal lines are valued as 1. It can be seen that the HSI carries obvious block features. Thus, the adaptive subspace decomposition can reasonably group the bands, and speed up the processing of hyperspectral data.

To sharpen the HSI grouping features, 6 subspaces were set up after the decomposition. The band sets in each subspace are listed in Table 1.

Next, the CNN classification algorithm was contrasted with the support vector machine (SVM) classification algorithm (Cannistraro *et al*., 2016). The experimental groups are given in Table 2.

Table 1. Subspaces and band sets

	Band range	Subspace Dimension
Subspace 1	3-44	42
Subspace 2	45-101	57
Subspace 3	102-107, 114-153	46
Subspace 4	170-200	31
Subspace 5	201—210	10
Subspace 6	211-221	11

Table 2. The experimental groups

Experimental group	Experimental content
E1	Without support for sensitivity analysis, direct SVM classification
E2	Use sensitivity analysis and then SVM classification
E3	Without support for sensitivity analysis, direct CNN classification
E4	Use sensitivity analysis then CNN classification

Results and Discussion

Classification

The pixel purity index (PPI) of the bands selected through the NNSA were calculated and the N-D scatter plot of the regions of interest were prepared. Since scatter of all bands rotate timely in an N-dimensional space, the geometry of the scattered points was used to pinpoint the spectral endmember of the target image (Figure 3). Here, three bands are discussed considering the restrictions of 3D space and visual effects. Then, the type of the spectrum endmember was determined against the JPL spectral library.

Four classification experiments were performed on the experimental data, each of which uses a specific classification method. The four classification methods are denoted as E1, E2, E3 and E4 (Table 2), in which E4 is our classification plan. The results of each experiment are displayed in Figure 4.

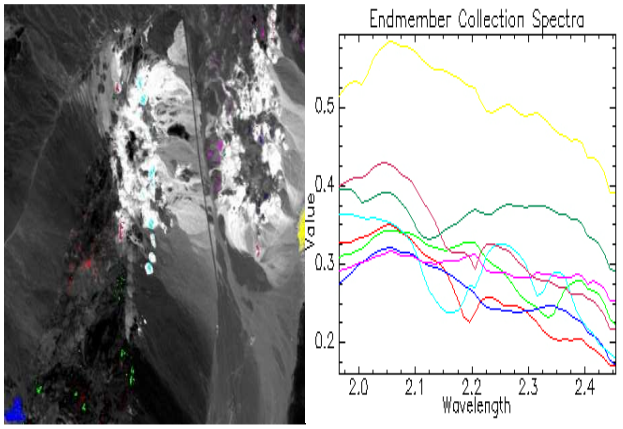

Figure 3. The pixels (left) and the corresponding endmember spectrum (right)

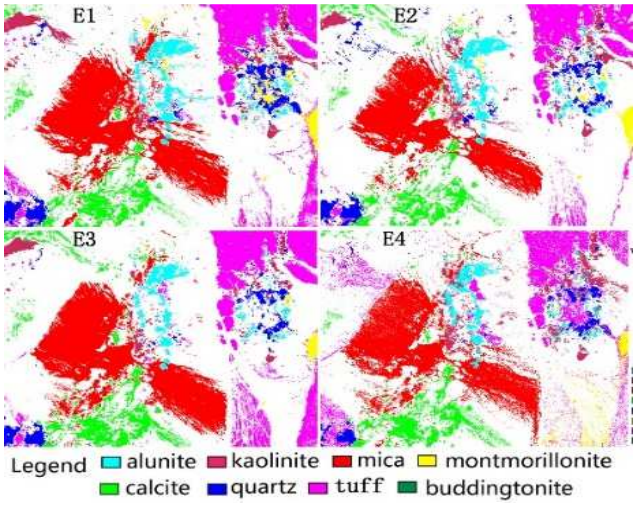

Figure 4. The experimental results

Accuracy analysis

Figure 4 lists the classification accuracy of each classification plan (E1, E2, E3 and E4) in each type of mineral. The minerals include alunite (A), kaolinite (B), mica (C), calcite (D), quartz (E), montmorillonite (F), tuff (G) and graphite (H).

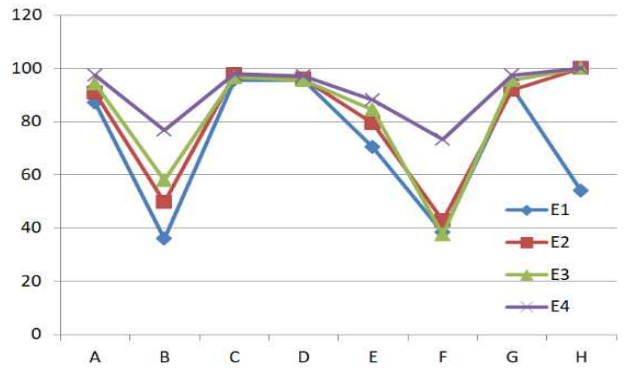

Figure 4. The accuracy of each classification method on each type of mineral

As can be seen from Figure 4, the classification method E4 outperformed all the other three methods in the recognition accuracy on every type of mineral. On the contrary, the classification method E1 was the least accurate one among all contrastive methods. Plans E2 and E3 were slightly better than E1 in terms of accuracy. The results fully demonstrate the effect of the proposed method E4, indicating that the ANN is good at reducing the dimensions of remote sensing HSI and the CNN is a reliable classification tool. This conclusion is supported by the overall classification accuracy and Kappa coefficient.

Table 3. The overall classification accuracy of all classification methods

	E1	E2	E3	E4
Kappa coefficient	0.84	0.87	0.90	0.91
Overall Accuracy	88.8	90.2	93.6	94.3

Conclusions

Based on the basic principle of sensitivity analysis, this paper applies the NNSA method to reduce the dimensions of remote sensing HSI. To minimize the correlation between input attributes, the adaptive subspace decomposition was adopted to select proper band combinations. Meanwhile, the sensitivity coefficient was computed rapidly by sensitivity analysis based on partial derivative, and the bands of small coefficients were eliminated for dimensionality reduction. The resultant image had much fewer dimensions than the original one. On this basis, the CNN was introduced in details and applied to the classification of remote sensing HSI. Through contrast experiments, it is learned that the CNN-based classification plan was much more accurate than the SVM classification algorithm and other contrast plans. The results prove that ANN is good at reducing the dimensions of remote sensing HSI and the CNN is a reliable classification tool. The research findings shed new light on remote sensing image processing and other related operations.

References

Cannistraro M, Lorenzini E. The applications of the new technologies “e-sensing” in hospitals, International Journal of Heat and Technology 2016; 34(4): 551-57.

Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. 2014.

Fichera A, Pagano A. (2017). A neural tool for the prediction of the experimental dynamics of two-phase flows, International Journal of Heat and Technology 2017; 35(2): 235-42.

Fukushima K, Miyake S. A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in poition. Biological Cybernetics 1980; 36: 193-202.

Herb DO. The organization of behavior: a neuropsychological theory. New York; John Wiley, 1949: 100-36.

Hinton U, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science 2006; 313 (5786): 504- 07.

Hu LJ, Tang L, Pan Q, Song H, Wen PG. Research and analysis of PI control strategy based on neural network in power grid, Mathematical Modelling of Engineering Problems 2016; 3(1): 25-28.

Hubel DH, Wiesel TN. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. Journal of Physiology 1962; 160: 106-54.

Kadri O, Mouss LH. Identification and detection of the process fault in a cement rotary kiln by extreme learning machine and ant colony optimization, Academic Journal of Manufacturing Engineering 2017; 15(2): 43-50.

Landgrebe D. Hyperspectral image analysis. IEEE Signal Processing Magazine 2002; 19(1): 17-28.

Lee T, Mumford D. Hierarchical Baycsian lnference in the Visual Cortex. Journal of Optical Society of America 2003; 20: 1434-48.

Liu M. Discrete-time delayed standard neural network model and its application. Science in China Series F: Information Sciences 2006; 49(2):137-54.

McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. The Bulletin of Mathematical Biophysics 1943; 5(4):115-33.

Rosenblatt F. Principles of Neurodinamics: Preceptron and Theory of Brain Mechanisms. Washington, USA: Spartan Books, 1962.

Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 1958; 65(6): 386-408.

Ruck DW, Rogers SK, Kabrisky M. Feature selection using a multilayer perceptrons. Journal of Neural Network Computing 1990; 2(2): 40-48.

Rumelhart DE, Hinton U, Williams RJ. Learning representations by back-propagatig errors. Nature 1986; 323(6088): 533-36.

Salakhutdinov R, Hinton U. An efficient learning procedure for deep Boltzmann machines. Neural Computation 2012; 24(8): 1967-2006.

Seide F, Liu YD. Conversational speech transcription using context-dependent deep neural networks//Internatio-nal Speech Communication Association. Annual Conference. 12th 2011. Red Hook, NY: Curran Associates, 2011: 437-40.

Vincent P, Larochelle H, Bengio Y. Extracting and composing robust features with denoising autoencoders// Proceedings of the 25th international Conference on Machine Learning. Helsinki, Finland 2008: 1096-1103

Vinrent P, Larorhelle H, Bengio Y. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. Journal of Machine Learning Research 2010; 11(12): 3371-3408

Wang TC., Xie YZ, Yan H. Research of multi sensor information fusion technology based on extension neural network, Mathematical Modelling of Engineering Problems 2016; 3(3): 129-34.

Zhang JP, Zhang Y, Zou B. Fusion classification of hyperspectral image based on adaptive subspace decomposition. IEEE International Conference on Image Processing, Vancouver, BC, Canada 2000; 3: 472-75.

Zurada JM, Malinowski A, Usui S. Perturbation method for deleting redundant inputs of perceptron networks. Neurocomputing 1997; 14(2): 177-93.