Indian Journal of Geo Marine Sciences
Vol. 48 (11), November 2019, pp. 1822-1832

Underwater target recognition method based on t-SNE and stacked nonnegative constrained denoising autoencoder

Yuechao Chen*, Xiaonan Xu, Bin Zhou & Hengheng Quan

[1]Science and Technology on Sonar Laboratory, Hangzhou Applied Acoustics Research Institute, Hangzhou, China

*[E-mail: chenyc_715@163.com]

Received 02 May 2018; revised 12 June 2018

Underwater targets recognition is a difficult task due to the specific attributes of underwater target radiated noises, low signal to noise ratio and so on. In this paper, the input data optimization method and recognition model were researched. The underwater target radiated noise spectrum was chosen as the original feature. The t-distributed stochastic neighbor embedding (t-SNE) algorithm was used to reduce the dimensionality of the original spectrum segments divided by frequency. The optimal features can be obtained by analyzing the separability. Then the stacked nonnegative constrained denoising autoencoder (SNDAE) model was established to recognize the optimal features. The experimental signal spectra were processed by above methods. The results show that the recognition accuracy of SNDAE is higher than that of other contrastive methods. And the frequency of input band with the highest recognition accuracy is approximately the same as that with the best separability based on t-SNE, indicating that the above method can improve the recognition accuracy and efficiency.

[**Keywords***:* Feature optimize; Stacked nonnegative constrained denoising autoencoder; t-distributed stochastic neighbor embedding; Underwater target radiated noise]

Introduction

Underwater target recognition, as one of the key technologies to promote the intelligence of underwater acoustic equipment, is an important research direction of underwater acoustic signal processing. The core research contents of underwater target recognition are to extract and express the features of underwater target radiated noises. The traditional underwater target recognition is usually realized by separability feature extracting based on some signal processing methods and feature identifying based on classifier. Many scholars have researched underwater target recognition. Then a lot of separability feature extracting means and classifier design methods have been put forward[1-5]. However, the composition of underwater acoustic signals is very complex, and the features with obvious invariance and differentiation are often the comprehensive results of a variety of original features according to contribution and relevance. So the feature construction and selection need great skills. An important way to improve the utilization degree of target information hiding in signal is to increase the feature number or dimensions. But this will lead to the increase of information redundancy. Then the complexity of classifier calculation will increase and the recognition efficiency will decrease. If the nonlinear calculation capacity of the classifier is insufficient, the phenomenon of overfitting or underfitting may occur, and then the classifier performance will go down. Therefore, the underwater target radiated noise features for recognition should be optimized and classification model with strong computational performance needs to be built to realize underwater target recognition.

The t-distributed Stochastic Neighbor Embedding (t-SNE) is a nonlinear dimensionality reduction method based on probability analysis, which can recover low dimensional manifold structure from high-dimensional sampling data and realize visual projection[6]. Compared with principal component analysis (PCA) and other traditional methods, it has stronger nonlinear data processing ability. At present, many researchers have applied t-SNE algorithm to data dimensionality reduction and visualization analysis in the field of computer vision, medical diagnosis and so on[7,8,9]. Therefore, it is a good choice to introduce this algorithm into separability analysis and optimization for underwater target signal features.

The deep learning is a hot research direction in artificial intelligence field now. It was first put forward by Hinton G[10] in 2006. The structure of deep learning is based on the superposition of multiple hidden layers. The feature extraction ability is improved through layer by layer learning and the process is similar to human brain's cognition of things. Although each hidden layer usually only uses relatively simple nonlinear transformation, the combination of all hidden layers can produce very complex nonlinear transformation. So deep learning has very powerful computing ability. Since proposed, the deep learning has attracted wide attention. Not only new theoretical algorithms are continuing to be introduced[11-15] but also the practical applications such as image recognition, behavior recognition, speech signal processing, human brain simulation and so on are increasing[16-20]. The applications of deep learning in the field of underwater target recognition were relatively few and only simple researches based on basic deep learning algorithms were conducted[21,22,23]. Compared with image and speech signals, the underwater target radiated signal usually has lower signal to noise ratio and the signal composition is more complex. So it is necessary to establish an appropriate deep learning model and optimize this model according to the signal features for the purpose of achieving better results.

In this paper, the underwater target recognition method was studied based on t-SNE and deep learning algorithm. Firstly, the dimensionality of underwater target radiated signal spectrum was reduced based on t-SNE algorithm, and the features for recognition can be optimized by analyzing the separability of dimensionality reduction results. Then the stacked nonnegative constrained denoising autoencoder (SNDAE) was built according to the input signal characteristics, and the optimized features can be recognized by SNDAE. Finally, underwater target experimental signals were recognized by the proposed signal processing method, and the effectiveness is proved.

Methods

t-SNE algorithm

t-SNE is a manifold learning algorithm improved by stochastic neighbor embedding (SNE)[24]. The SNE transforms high dimensional Euclidean distance into conditional probability which represents the similarity between data points. The closer the distance between data points, the higher the conditional probability. When the data are mapped to low-dimensional space, the data points should also reflect the same similarity. The conditional probability between the corresponding data points is also constructed in low-dimensional space and the projection of high-dimensional data to low-dimensional data is realized by reducing the similarity error between the high-dimensional space condition probabilities and the low-dimensional space conditional probabilities.

As the SNE algorithm establishes the similarity between high-dimensional data based on asymmetric conditional probabilities, the relationship between the data cannot be fully expressed and there also exists the phenomenon of data congestion caused by the small space volume. The t-SNE algorithm improves the above disadvantages. It adopts a symmetric model and the conditional probabilities between the data points in SNE model are replaced by joint probabilities. Set and conform to the Gaussian probability distribution centered on . Then the conditional probability can be respectively expressed as:

$$p_{j|i} = \frac{e^{\left(-\|x_i - x_j\|^2 / 2\sigma_i^2\right)}}{\sum_{k \neq i} e^{\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)}} \quad \ldots (1)$$

$$p_{i|j} = \frac{e^{\left(-\|x_j - x_i\|^2 / 2\sigma_i^2\right)}}{\sum_{k \neq j} e^{\left(-\|x_j - x_k\|^2 / 2\sigma_i^2\right)}} \quad \ldots (2)$$

where σ_i is the variance of Gaussian function. The local symmetry joint probability distribution between high-dimensional data can be expressed as:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2} \quad \ldots (3)$$

In the low-dimensional space, the t-SNE takes the t-distribution whose freedom is one degree as a kernel function to generate the joint probability distribution between the embedded data points. As a typical heavy-tailed distribution, t distribution can make the distance between the embedded data larger than that generated by Gaussian kernel function, especially for the distance between the non-similar data points, which can effectively alleviate disadvantages of low dimensional manifold extrusion caused by using

Gaussian kernel function to construct the random neighborhood probability. Set y_j and y_i as the data points in low-dimensional space corresponding to x_j and x_i respectively, then the joint probability distribution q_{ij} can be expressed as:

$$q_{ij} = \frac{\left(1+\left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{k\neq l}\left(1+\left\|y_k - y_l\right\|^2\right)^{-1}} \quad \ldots (4)$$

The Kullback-Leibler divergence (KL divergence) is used as the cost function for difference measure between $p_{j|i}$ and $q_{j|i}$. The KL divergence of all points is minimized based on the gradient descent method to obtain the best embedded data. The objective cost function $J_{\text{t-SNE}}$ and the gradient C can be respectively expressed as:

$$J_{\text{t-SNE}} = KL\left(P\,\|\,Q\right) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad \ldots (5)$$

$$C = \frac{\partial J_{\text{t-SNE}}}{\partial y_i} = 4\sum_j \left(p_{ij} - q_{ij}\right)\left(1+\left\|y_i - y_j\right\|^2\right)^{-1}\left(y_i - y_j\right) \quad \ldots (6)$$

Stacked Nonnegative Constrained Denoising AutoEncoder

Stacked Denoising AutoEncoder

The autoencoder (AE) is a multi-layer feed forward network with structural symmetry for middle layer. The training objective is to reconstruct the input signal from the target expression. The AE can be used to learn identity mapping and extract unsupervised features. The structure of a single AE can be divided into encoder and decoder.

The denoising autoencoder (DAE) has the same structure as the traditional AE, except that it adds a certain amount of noises to the input data, and its learning goal is to reconstruct a pure input from the contaminated input, thereby the robustness of output is increased. Set $\boldsymbol{x}$ as original input without noise, the DAE adds noises to original input data through a random mapping transformation $\boldsymbol{x} \sim q_D\left(\boldsymbol{x}' \mid \boldsymbol{x}\right)$. Then the partially corrupted data $\boldsymbol{x}'$ can be obtained, and D is the data set. The encoder completes the mapping transformation from the input vector $\boldsymbol{x}$ to the output representation $\boldsymbol{y}$, and the process can be expressed as:

$$\boldsymbol{y} = f_\theta\left(\boldsymbol{x}\right) = s\left(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b}\right) \quad \ldots (7)$$

where θ is the model parameter and $\theta = \left\{\boldsymbol{W}, \boldsymbol{b}\right\}$, $\boldsymbol{W}$ is the weight matrix whose dimension is $d' \times d$, $\boldsymbol{b}$ is the offset vector whose dimension is d'.

The decoder maps the output representation $\boldsymbol{y}$ back into the input space and reconstructs the vector $\boldsymbol{z}$. The process can be expressed as:

$$\boldsymbol{z} = g_{\theta'}\left(\boldsymbol{y}\right) = s\left(\boldsymbol{W}'\boldsymbol{y} + \boldsymbol{b}'\right) \quad \ldots (8)$$

where θ' is the model parameter and $\theta' = \left\{\boldsymbol{W}', \boldsymbol{b}'\right\}$, $\boldsymbol{W}'$ is the weight matrix whose dimension is $d \times d'$, $\boldsymbol{b}'$ is the offset vector whose dimension is d. The objective function is to minimize the reconstruction error between $\boldsymbol{z}$ and $\boldsymbol{x}$, which can be expressed as:

$$\arg_{\theta,\theta'} \min J\left(\boldsymbol{x}, \boldsymbol{z}\right) \quad \ldots (9)$$

where J is a specific cost function, common with Mean Squared Error, Cross Entropy and so on. When some DAEs are overlapped layer by layer, that is, the output of last DAE is the input of the current DAE, the stacked denoising autoencoder (SDAE) can be obtained. Fig. 1 shows the block diagram of SDAE. The training of SDAE can be divided into two stages which are pre-training and fine-tuning. In the stage of

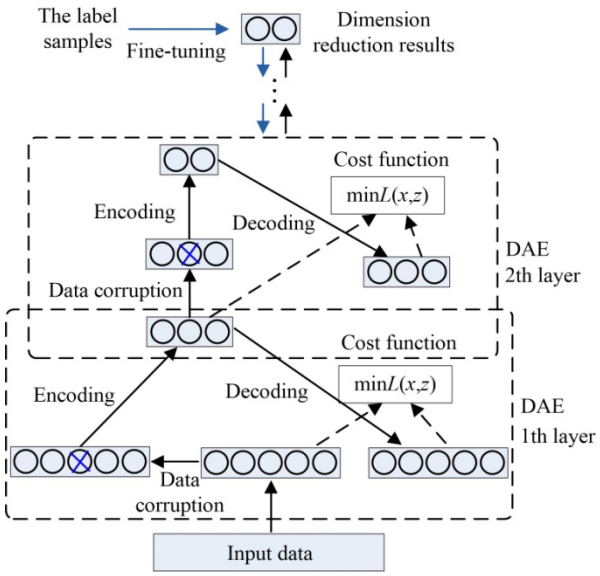

Fig. 1 — Block diagram of SDAE

pre-training, the unsupervised training is conducted for SDAE so that the model structure parameters of SDAE can achieve approximate optimum in the whole. Then the supervised reverse fine-tuning is conducted based on data with label to further optimize the network of SDAE.

In order to prevent over-fitting in the course of training, weight decay term $J_{\text{weight}}(\theta)$ can be added into the cost function, usually expressed as:

$$J_{\text{weight}}(\theta)=\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^{(l)}\right)^2 \quad \ldots (10)$$

where λ is the coefficient of weight decay term, n_l is the hidden layer number, s_l is the neuron node number of layer l, $W_{ji}^{(l)}$ is the element value of the weight matrix $\boldsymbol{W}$ with subscript of j and i.

At the same time, in order to further optimize the model performance, the sparse constrained term can be added into the cost function, usually expressed as:

$$J_{\text{sparse}}(\theta)=\beta\sum_{i=1}^{s}\text{KL}\left(\rho\,\|\,\hat{\rho}_j\right) \quad \ldots (11)$$

where β is the coefficient of sparse constrained term, ρ is the sparse parameter, $\hat{\rho}_j$ is the average activity of hidden neuron node j, s is the number of hidden neuron node number, $\text{KL}\left(\rho\,\|\,\hat{\rho}_j\right)$ is the KL divergence between $\hat{\rho}_j$ and ρ, and it can be expressed as:

$$\text{KL}\left(\rho\,\|\,\hat{\rho}_j\right)=\rho\log\frac{\rho}{\hat{\rho}_j}+(1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j} \quad \ldots (12)$$

The final cost function of SDAE can be expressed as:

$$J_{\text{SDAE}}=J(\boldsymbol{W},\boldsymbol{b})+\beta\sum_{j}^{s}\text{KL}\left(\rho\,\|\,\hat{\rho}_j\right)+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}\left(W_{ji}^{(l)}\right)^2 \quad \ldots (13)$$

Stacked Nonnegative Constrained Denoising AutoEncoder

As the input of deep learning model for target recognition is signal spectrum, and then the multi-sample input data is a non-negative matrix. With reference to non-negative matrix factorization (NMF), when the dimensionality reduction is conducted for the input data matrix, the results should also be non-negative. That is the overall performance of the results is superimposed by local features, then the nature of the input object can be better reflected. A viable option is to improve the weight decay term thus making the output data close to non-negative. The decay rate of negative numbers and nonnegative numbers in the weight matrix can be adjusted, and the weighting term is set to L2 regularization, then the following cost function can be obtained.

$$J_{\text{SNDAE}}=J(\boldsymbol{W},\boldsymbol{b})+\beta\sum_{j}^{s}\text{KL}\left(\rho\,\|\,\hat{\rho}_j\right)+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}f\left(W_{ji}^{(l)}\right) \quad \ldots (14)$$

$$f\left(W_{ji}^{(l)}\right)=\begin{cases}\gamma_1 W_{ji}^2 & W_{ji}<0\\ \gamma_2 W_{ji}^2 & W_{ji}\geq 0\end{cases} \quad \ldots (15)$$

where $\gamma_1>0$, $\gamma_2>0$. When setting $\gamma_1>\gamma_2$, the effect of negative weight coefficient on the cost function is greater than that of positive weight coefficient. Unlike the nonnegative constraint in literature 25, the proposed method can adjust the decay rate of the weight decay term by changing negative weight coefficient γ_1 and positive weight coefficient γ_2. Thus the nonnegativity of the weight matrix can be improved. If γ_1 and γ_2 are set appropriately, the iterative process of cost function can be effectively improved.

Based on the gradient descent method to minimize the cost function, the following weight update method can be obtained.

$$W_{ji}^{(l)}=W_{ji}^{(l)}-\alpha\left[\frac{\partial}{\partial W_{ji}^{(l)}}J(\mathbf{W},\mathbf{b})+\beta\frac{\partial}{\partial W_{ji}^{(l)}}\sum_{j}^{s}\text{KL}\left(\rho\,\|\,\hat{\rho}_j\right)+\lambda y\left(W_{ji}^{(l)}\right)\right] \quad \ldots (16)$$

$$y\left(W_{ji}^{(l)}\right)=\begin{cases}\gamma_1 W_{ji}^{(l)} & W_{ji}<0\\ \gamma_2 W_{ji}^{(l)} & W_{ji}\geq 0\end{cases} \quad \ldots (17)$$

In practice, the weight matrices are updated by applying batch gradient descent method. The process is as follows. Set $\boldsymbol{W}^{(l)}$ as the weight matrix of layer l. $\Delta\boldsymbol{W}^{(l)}$ is a matrix. The dimension of $\Delta\boldsymbol{W}^{(l)}$ is the same as that of $\boldsymbol{W}^{(l)}$, and the initial values of all

layers are set to 0. The number of samples per batch is m.

For all samples, the partial derivative of each node is calculated in turn using the back-propagation algorithm. The calculation method is can be expressed as:

$$\nabla_{W_k^{(l)}} J_{\text{SNDAE}} = \frac{\partial}{\partial \boldsymbol{W}_k^{(l)}} J(\mathbf{W},\mathbf{b}) + \beta \frac{\partial}{\partial \boldsymbol{W}_k^{(l)}} \sum_j^s \text{KL}\left(\rho \| \hat{\rho}_j\right) \quad \ldots (18)$$

where k is the sample sequence number. Next $\nabla_{W_k^{(l)}} J_{\text{SNDAE}}$ is added to $\Delta \boldsymbol{W}^{(l)}$, that is:

$$\Delta \boldsymbol{W}^{(l)} = \Delta \boldsymbol{W}^{(l)} + \nabla_{W_k^{(l)}} J_{\text{SNDAE}} \quad \ldots (19)$$

Then the weight matrix can be updated, and the update result can be expressed as:

$$\boldsymbol{W}^{(l)} = \boldsymbol{W}^{(l)} - \alpha \left[\frac{1}{m} \Delta \boldsymbol{W}^{(l)} + \lambda y\left(\boldsymbol{W}^{(l)}\right) \right] \quad \ldots (20)$$

Fig. 2 shows the asymmetric weight decay curve when $\gamma_1 = 1$ and $\gamma_2 = 0.5$.

Signal processing flow

When the above methods are comprehensively applied, the signal processing frame can be obtained, as shown in Fig. 3. Firstly, the underwater target radiated noises are pre-processed, and the purpose is to transform the signal into a processing domain in which the signal separability can appear. In this paper, the signal spectrum is generated in the pre-processed stage for further processing. The spectrum of the target signal is segmented by frequency. Then the t-SNE algorithm is used to reduce the dimensionality of signal spectra of different frequency bands. The dimensionality reduction results are visualized and analyzed to obtain optimal signal band with the best separability. At last, the signal spectrum of the preferred frequency band is recognized based on SNDAE model and the final recognition result can be obtained.

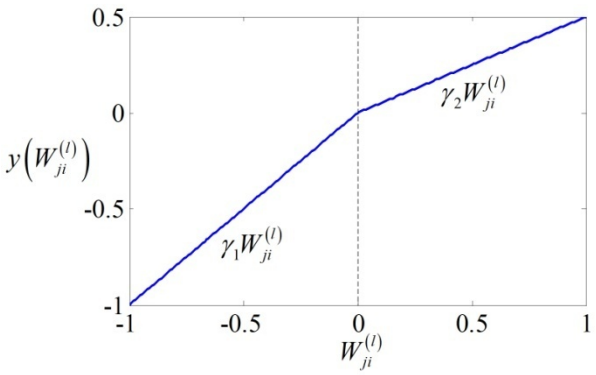

Fig. 2 — Asymmetric weight decay curve

Results and Discussion

Process for two types of underwater target experimental signals

Spectrum characteristics of underwater target radiated signals

Selecting the appropriate input data is one of the key factors to realize the effective learning and recognition. Usually, the signal to noise ratio of the time domain underwater acoustic data is low. Even through complex nonlinear transformation, it is difficult to extract feature with good separability. In frequency domain, the low frequency spectrum usually contains rich target characteristics and can be used as the input data of deep learning model.

Two types of experimental underwater target radiated signal were analyzed. The sampling rate of the experiment was 5000 Hz. The number of sampling points was 32768 for each target sample, and the total number of samples were 5400. Fig. 4 shows partial signal fragments of every target with random selection. It can be seen that the signal to noise ratio of each type of target signal is very low and the signal composition is complex. Fig. 5 shows the spectrum of each target signal in Fig. 4. It can be seen that the spectra are also complex, but there are some differences from the above spectra, reflecting that the spectra contain separability information.

The low frequency signal spectra of some bands were projected in two-dimension based on t-SNE

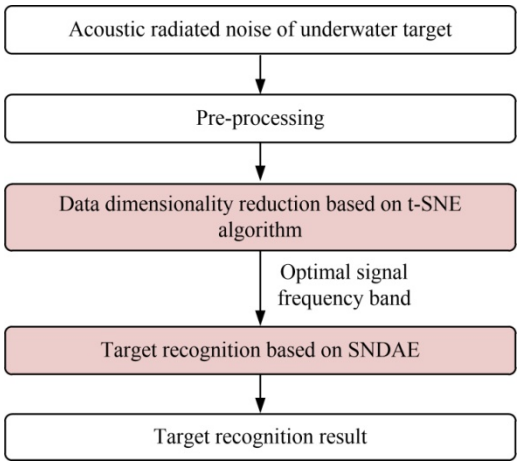

Fig. 3 — Signal processing frame

algorithm to realize visualization, and the principal component analysis (PCA), locality preserving projection (LPP) and stochastic neighbor embedding (SNE) were used as contrastive methods. The results are shown in Fig. 6. The components of target 1 have been almost overlapped on the components of target 2 from the results of PCA and LPP. The results cannot reflect the separability of different targets. The projection results of t-SNE and SNE can reflect some separability between target 1 and target 2, but SNE results still have a certain coincidence degree among different targets and the t-SNE results have the best separability.

Analysis of dimensionality reduction results of different frequency bands

The separability characteristics of the spectrum on different frequencies were studied. The spectrum with frequency range from 10 to 1000 Hz was segmented by frequency. The dimensionality of each spectrum segment was reduced and the results were visualized based on t-SNE algorithm, as shown in Fig. 7. It can be seen that as the frequency increases, the separability of the spectral segments exhibits a downward trend in general. The spectrum segments with frequency range from 10 to 150 Hz have the best separability. There exists large class spacing between the various objectives on the whole. The spectrum bands of 150H-250 Hz also have relatively strong separability. However, the class spacing between different targets has been reduced. The overlap degree gradually increases when the frequency is more than 250 Hz. For the spectrum bands whose initial frequency is higher than 500 Hz, the separability degree between different targets becomes very weak. Based on the above analysis, when the target signal is to be recognized, the spectrum whose end frequency is lower than 500 Hz can be mainly used. In particular, the components below 250 Hz may be important to consider.

Analysis of recognition results of two types of targets

The SNDAE model was used to recognize the above two types of underwater target radiated noises,

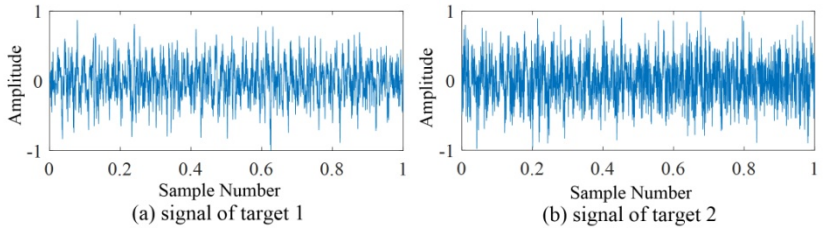

Fig. 4 — Two types of underwater target radiated signals

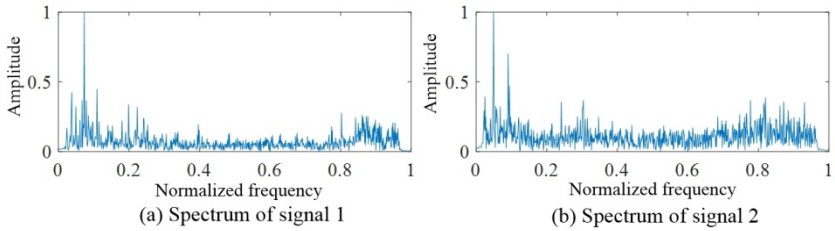

Fig. 5 — Two types of underwater target radiated signal spectra

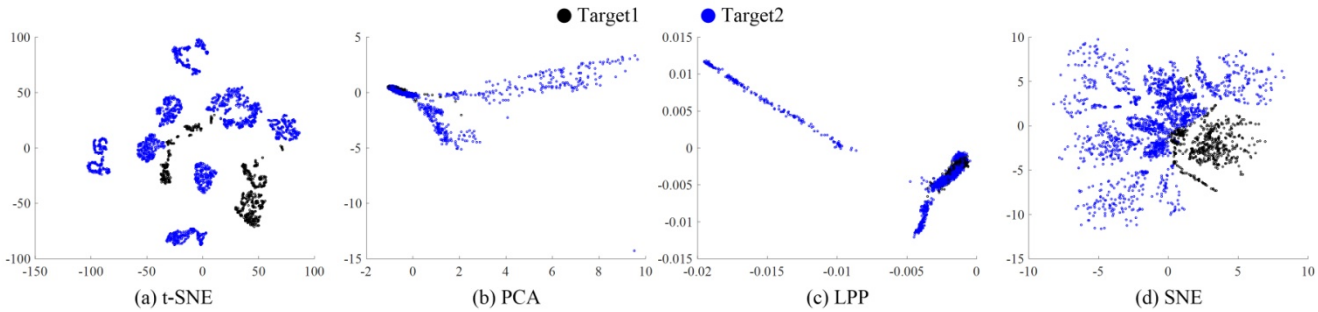

Fig. 6 — Projection results of different dimensionality reduction methods

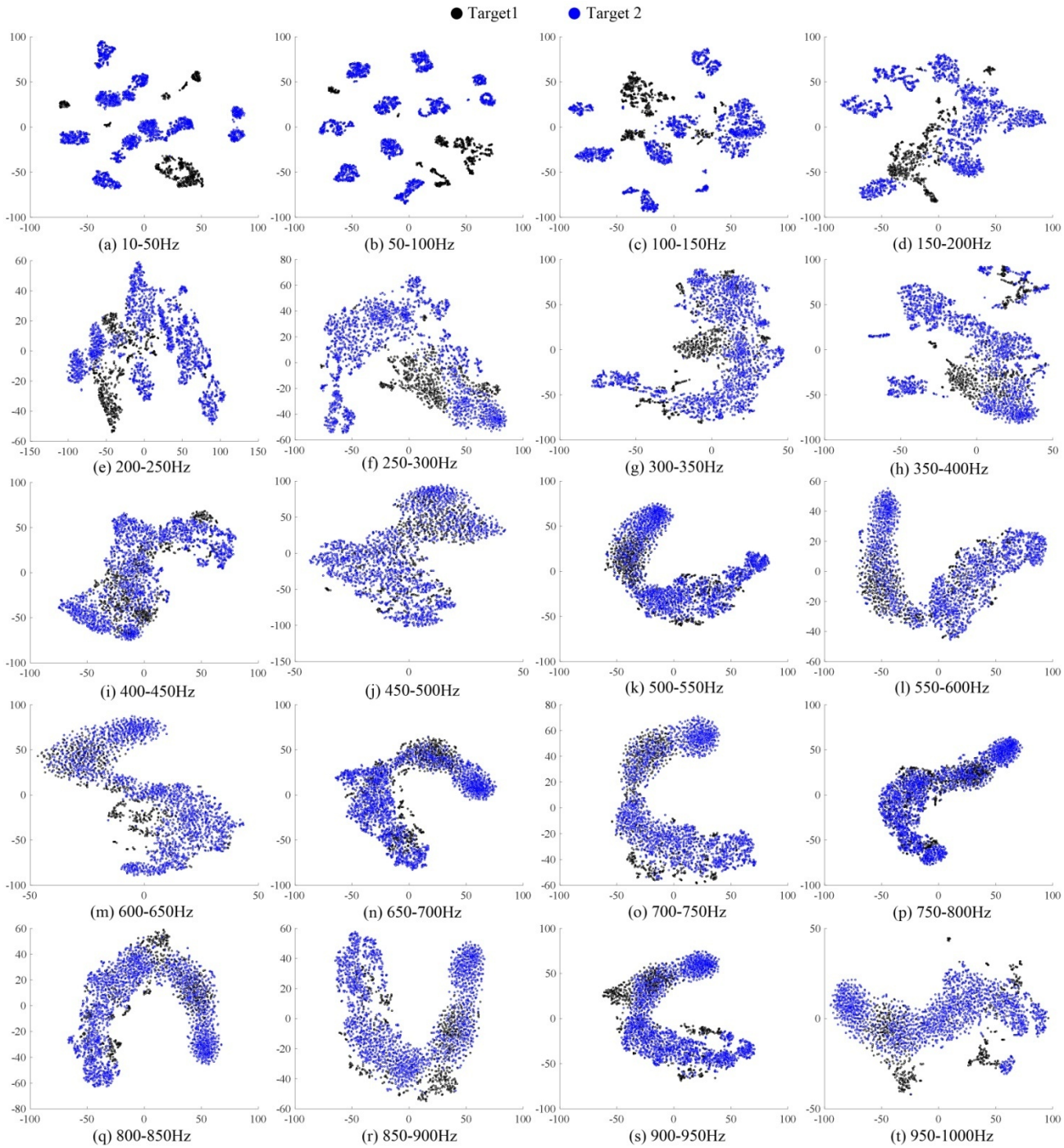

Fig. 7 —Spectrum dimensionality reduction results of two types of targets

and the recognition accuracy was analyzed. At the same time, SDAE and Support vector machine (SVM) were used for contrast recognition.

The spectrum whose frequency is below 250 Hz has relatively good separability from the analysis results in above section. Then the low frequency bands should be contained in the spectrum to be recognized. Therefore, the initial frequency of every spectrum band was set as 10 Hz and the end frequency was gradually increased from 50 Hz to 1000 Hz. Considering the low proportion of samples with label in actual underwater acoustic data, the number of samples with label was set as five percent of the total training samples. To a certain extent, it

can also reflect the advantages of unsupervised pre-training of autoencoder.

The number of hidden layers of SNDAE as well as that of SDAE was set to 2 respectively. The reference nodes were all set to 500 and 50 respectively, and it can be changed according to the actual situation. The weight coefficient γ_1 and γ_2 were set as 1 and 0.5 respectively, and the Softmax classifier was used as the top output layer. The Gaussian radial basis function (RBF) was used as the kernel function of SVM.

Table 1 shows the recognition accuracy of different frequency bands based on above recognition methods. As the initial frequencies of all spectrum bands for recognition are the same, the end frequency is used as the abscissa, and the curve of recognition accuracy changed with end frequency can be obtained, as shown in Fig. 8.

Then the recognition accuracy was analyzed. The recognition accuracy of each method shows a tendency to rise first and then decreases as the end frequency increases.

For SNDAE and SDAE, when the end frequency is raised from 50 Hz to 100 Hz, the recognition accuracy increases substantially, and the recognition accuracy fluctuates within a wide range when the end frequency is between 100-500 Hz, but all above spectrum bands have high recognition accuracy. When the end frequency is between 150-250 Hz, the recognition accuracy is the best and the recognition accuracy of SNDAE is better than that of SDAE on the whole. According to the analysis results based on t-SNE, the separable components mainly exist in the underwater target signal spectrum band whose frequency is below 250 Hz. And some separable components also exist in the spectrum band whose frequency is between 250-500 Hz. These results are in good agreement with the frequency bands when the above deep learning models have the best recognition accuracy. In addition, the above model still has a relatively high recognition accuracy when the end frequency is high, indicating that the above deep learning models have strong data analysis ability and can extract the separable components from the highly redundant data to realize target recognition. So the deep learning models have good data tolerance.

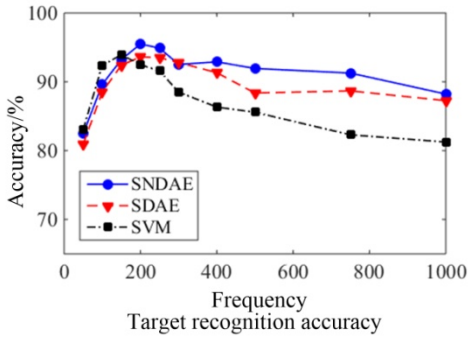

Fig. 8 — Curves of recognition accuracy with end frequency

For SVM, the recognition accuracy is also greatly improved when the end frequency is increased from 50 Hz to 100 Hz, and the best recognition accuracy appears when the end frequency is 150 Hz. This is due to the fact that SVM constructs the classification surface by mapping data from low dimensional space to high dimensional space, and its advantage is to solve the classification problem when the sample number is small and sample dimensions are low. Then, with the end frequency increasing, the recognition accuracy has a monotonically decreasing trend. When the end frequency is large, the recognition accuracy of SVM was significantly lower than that of SNDAE and SDAE. So the SVM cannot take full advantage of the spectrum band whose frequency is larger than 150 Hz and the data tolerance is lower than that of deep learning models.

In general, the recognition accuracy of SNDAE is better than that of SDAE and SVM whether on the whole or for the best frequency bands.

Process for multi types of underwater target experimental signals

Signal feature analysis

The experimental data are beam space signal of four types of ships. The sampling rate is also 5000 Hz. The segments of the spectrum of the signal were

Table 1 — Recognition accuracy of two types of targets

Frequency/Hz	Recognition accuracy/%			Frequency/Hz	Recognition accuracy/%		
	SNDAE	SDAE	SVM		SNDAE	SDAE	SVM
10-50	82.46	80.93	83.08	10-300	92.52	92.81	88.50
10-100	89.63	88.54	92.42	10-400	92.90	91.32	86.33
10-150	93.24	92.41	93.92	10-500	91.94	88.37	85.58
10-200	95.49	93.63	92.50	100-750	91.27	88.68	82.30
10-250	94.90	93.50	91.58	10-1000	88.22	87.26	81.25

processed based on t-SNE algorithm and the results were shown in Fig. 9.

The spectra have relatively high separability when the frequency is below 200 Hz and the components whose frequency is higher than 500 Hz have little separability. So only two dimensionality reduction results whose initial frequency exceeds 500 Hz are listed. The results are similar to that of targets with two types in Fig. 7.

Analysis of recognition results of four types of target

The SNDAE, SDAE and SVM were applied to recognize the four types of targets. The model parameters were consistent with the previous settings in Section 3.1.3. A total of 4000 samples were used as training data and 2000 samples as test data.

Table 2 shows the recognition accuracy and recognition program run time of different frequency bands. The recognition program runtime has been normalized for ease of comparison. Fig. 10 shows the curve of recognition accuracy changed with end frequency as well as the curve of recognition program run time changed with end frequency.

The recognition accuracy was analyzed. As the end frequency increase, the recognition accuracy of each method increases first and then decreases. The best

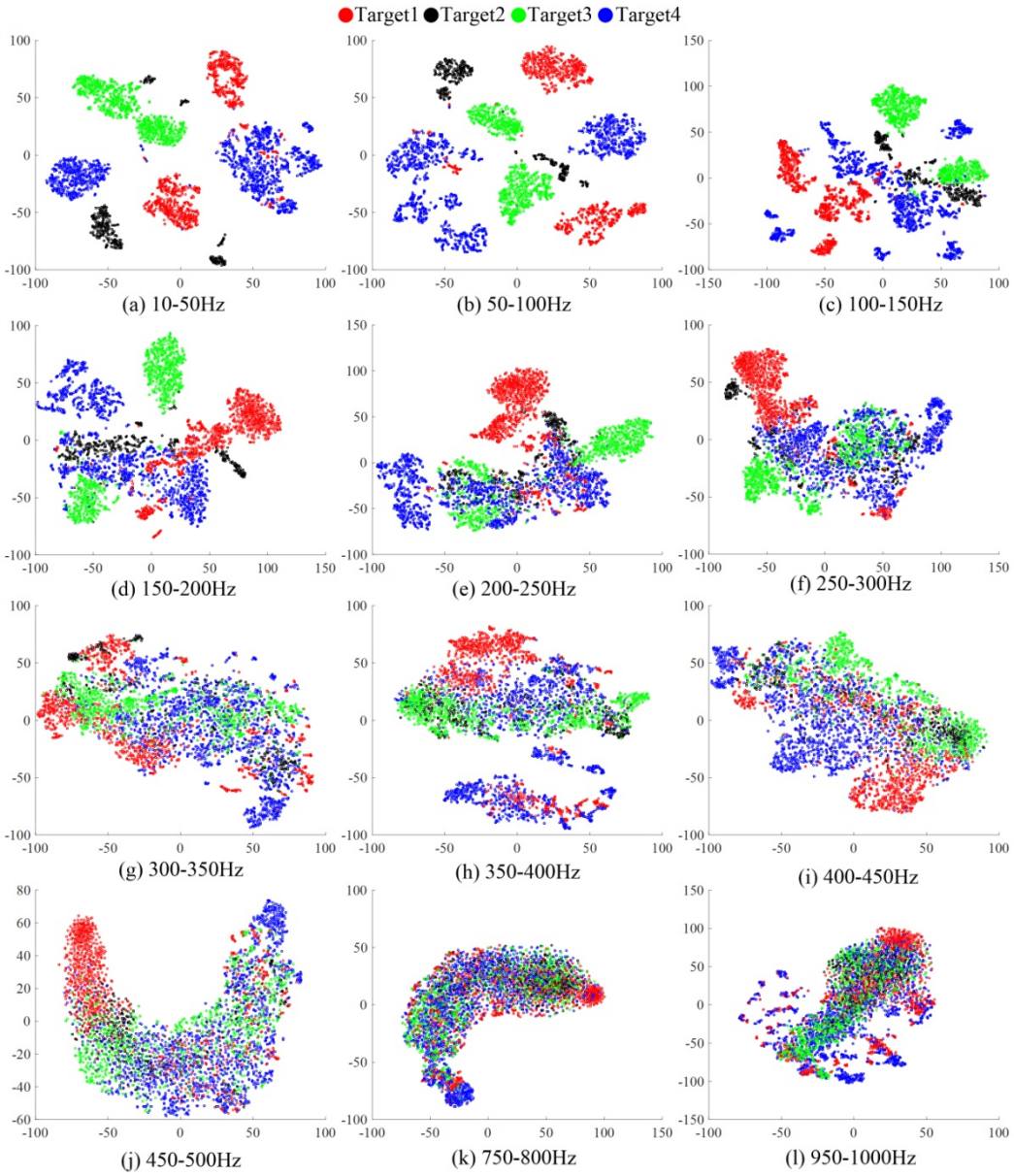

Fig. 9 —Spectrum dimensionality reduction results of multi types

Table 2 — Recognition accuracy and recognition program run time of four types of targets

Frequency/Hz	Recognition accuracy/%			Program run time		
	SNDAE	SDAE	SVM	SNDAE	SDAE	SVM
10-50	79.50	77.30	82.65	0.0098	0.0098	0.0264
10-100	92.70	89.80	92.95	0.0250	0.0258	0.0597
10-150	93.90	91.85	92.90	0.0367	0.0403	0.0965
10-200	94.35	90.85	90.25	0.1024	0.1072	0.1367
10-250	92.35	92.85	88.85	0.1340	0.1353	0.1707
10-300	93.45	92.25	85.65	0.1613	0.1675	0.2171
10-400	91.30	90.05	83.25	0.2216	0.2282	0.3135
10-500	92.55	90.65	82.85	0.2724	0.2827	0.4219
10-1000	89.85	85.50	68.70	1	1	1

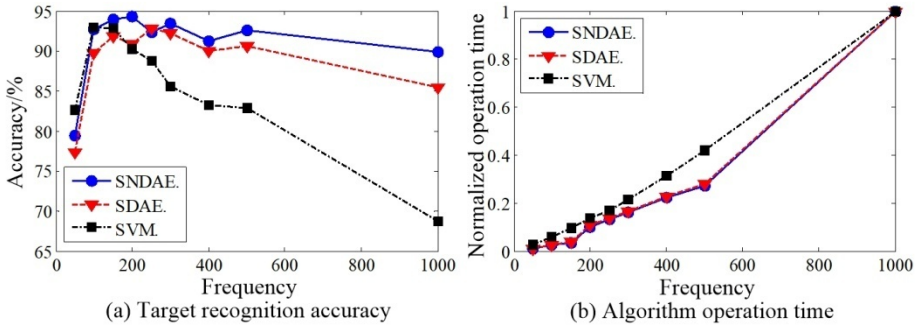

Fig. 10 — Curves of recognition accuracy and program run time changed with end frequency

recognition accuracy is obtained when the end frequency is between 150-300 Hz. The recognition accuracy of SNDAE is the best overall.

Then the recognition program efficiency was analyzed. The program run time of each method is gradually increased with the rise of the data dimension. Take SNDAE as an example, when it has the best recognition accuracy (frequency band of 10-200 Hz), the program run time only accounts for 13.67 % of that of broadband spectrum (frequency band of 10-1000 Hz). Therefore, the accuracy and efficiency of recognition can be effectively improved by selecting the optimal input data based on t-SNE algorithm and achieving target recognition based on SNDAE model.

Conclusion

The underwater target recognition was studied. The t-SNE was used to reduce the dimensionality of the target radiated noise spectrum segment divided by frequency for the purpose of analyzing separability. Then the SNDAE was established to recognize optimal data. The experimental data processing results show that the spectrum separability decreases with the increase of frequency. The recognition accuracy of SNDAE is higher than that of SDAE and SVM, and the optimal input data band approximately corresponds to the result of t-SNE. The program run time based on optimal data is significantly lower than that based on original data.

Acknowledgement

This study was funded by the National Natural Science Foundation of China [Grant No. 61701450, 61701449].

References

1 Azimi-Sadjadi, M. R., Yao, D., Huang, D., Dobeck, G. J., Underwater target classification using wavelet packet and neural networks. *IEEE. T. Neural. Networ.*, 11(2000), 784-794.

2 Li, D. H., Azimi-Sadjadi, M. R., Robinson, M., Comparison of different classification algorithms for underwater target discrimination. *IEEE. T. Neural. Networ.*, 15(2004), 189-194.

3 Cexus, J. C., Boudraa, A. O., Teager-Huang analysis applied to sonar target recognition. *Int. J. Signal Process*, 1(2004), 23-27.

4 Qu, H., Allen, J. S., Syrmos, V. L., Underwater target recognition using time-frequency analysis and elliptical fuzzy clustering classification. *Processings of the 28th International Conference on Ocean, Offshore and Arctic Engineering*, Honolulu, 2009, 725-733.

5 Wang, S. G., Zeng, X. Y., Robust underwater noise targets classification using auditory inspired time-frequency analysis. *Appl. Acoust.*, 78(2014), 68-76.

6 Maaten, L. V. D., Hinton, G., Visualizing data using t-SNE. *J. Mach. Learn. Res*, 9(2008), 2579-2605.

7 Yi, J. Z., Mao, X., Xue, Y. L., Compare, A., Facial expression recognition based on t-SNE and adaboostM2. *Int. J. Fuzzy Log. Intell. Syst*, 13(2013), 315-323.

8 Jamieson, A. R., Giger, M. L., Drukker, K., Li, H., Yuan, Y., Bhooshan, N., Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. *Med. Phys.*, 37(2009), 339-351.

9 Cheng, J., Liu, H. J., Wang, F., Li, H. S., Zhu, C., Silhouette analysis for human action recognition based on supervised temporal t-SNE and incremental learning. *IEEE. T. Image. Process.*, 24(2015), 3203-3217.

10 Hinton, G., Salakhutdinov, R. R., Reducing the dimensionality of data with neural networks. *Science*, 313(2006), 504-507.

11 LeCun, Y., Bengio, Y., Hinton, G., Deep learning. *Nature*, 521(2015), 436-444.

12 Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P. A., Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.*, 11(2010), 3371-3408.

13 Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. R., Improving neural networks by preventing co-adaptation of feature detectors. *Comp. Sci.*, 3(2012), 212-223.

14 He, K. M., Zhang, X. Y., Ren, S. Q., Sun J., Deep residual learning for image recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016, 770-778.

15 Vincent, P., Larochelle, H., Bengio, Y., Manzagol, P. A., Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning*, Helsinki, 2008, 1096-1103.

16 Schirrmeister, R. T., Springenberg, J. T., Ldj, F., Glasstetter, M., Eggensperger, K., Tangermann, M., et al,. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum. Brain. Mapp.*, 38(2017), 5391-5420.

17 Yamins, D. L. K., Dicarlo, J. J., Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.*, 19(2016), 356–365.

18 Ciresan, D., Meier, U., Masci, J., Schmidhuber, J., Multi-column deep neural network for traffic sign classification. *Neural Networks*, 32(2012), 333-338.

19 Ji, S. W., Xu, W., Yang, M., Yu, K., 3D convolutional neural networks for human action recognition. *IEEE. T. Pattern. Anal.*, 35(2013), 221-231.

20 Byeon, Y. H., Kwak, K. C., Facial expression recognition using 3D convolutional neural network. *Int. J. Adv. Comput. Sci. Appl*, 5(2014), 107-112.

21 Kamal, S., Mohammed, S. K., Pillai, P. R. S., Supriya, M. H., Deep learning architectures for underwater target recognition. *Ocean Electronics (SYMPOL)*, Kochi, 2013, 48-54.

22 Chen, Y. C, Xu, X. N., The research of underwater target recognition method based on deep learning, *IEEE International Conference on Signal Processing, Communications and Computing*, Xiamen, 2017.

23 Li, C., Zhou, Q. M., Han, X., Yin, J. W., Shao, M. Q., Underwater non-cooperative communication signal recognition with deep learning. *J. Acoust. Soc. Am.*, 142(2017), 2732-2732.

24 Hinton, G., Roweis, S., Stochastic neighbor embedding. *Advances in Neural Information Processing Systems*, 41(2002), 833-840.

25 Hosseini-Asl, E., Zurada, J. M., Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints. *IEEE. T. Neur. Net. Lear.*, 27(2016), 2486-2498.