

## RECOGNITION OF SAR TARGET BASED ON MULTILAYER AUTO-ENCODER AND SNN

ZHIJUN SUN<sup>1,2,\*</sup>, LEI XUE<sup>1,2</sup> AND YANGMING XU<sup>1,2</sup>

<sup>1</sup>Electronic Engineering Institute  
No. 460, Huangshan Rd., Hefei 230037, P. R. China  
\*Corresponding author: robotman@126.com

<sup>2</sup>Key Laboratory of Electronic Restriction of Anhui Province  
No. 460, Huangshan Rd., Hefei 230037, P. R. China

Received November 2012; revised March 2013

**ABSTRACT.** *Automatic target recognition (ATR) of synthetic aperture radar (SAR) image is investigated. One feature extraction algorithm of SAR image based on multilayer auto-encoder is proposed. The method makes use of a probabilistic neural network, restricted Boltzmann machine (RBM), modeling probability distribution of environment. Through the formation of more expressive multilayer neural network, the deep learning model learns shared representation of the target and its shadow to reflect the target shape. Targets are classified automatically through two classification models. The experiment results based on the MSTAR verify effectiveness of the proposed algorithm.*

**Keywords:** SAR, Feature, Restricted Boltzmann machine, Multilayer auto-encoder, SNN

1. **Introduction.** SAR imaging mechanism results in serious speckle which is different from ordinary optical image. Therefore, most studies often explore only on the target itself with the subtraction of shadows and background. However, the shadow contour containing certain target shape information can improve recognition performance in a reasonable fusion manner.

The automatic target recognition from their radar signatures is an important and difficult problem that has attracted considerable research effort. It has also been a hot research spot of SAR image processing to recognize SAR target based on the target image and its shadow. A number of outstanding researches based on SAR image shadow emerge in the research community. The reference [1] recognizes the target with target profile and its shadow contour respectively. Then a tandem fusion is used to classify the test sample [1]. The literature [2] researches SAR target recognition based on the multi-view shadow. However, the amount of experiment data is limited [2]. HMM model is used in modeling target shadow information to recognize target automatically [3]. The reference [4] introduces the HRR and shadow information for target recognition [4].

At present, most of the researches deem target and shadows as independent features, while there is interaction between these attributes in the sense of the recognition [5]. Extracting the complementary features of joint representation of the two type data will contribute to the classification task. SAR shadow formation mechanism makes its shadow outline contain target's structure information, which makes it possible of SAR ATR based on targets and shadow contour. This paper introduces an acoustic recognition model, multi-layer auto-encoder to learn multi-mode representation [6]. The joint representation fusing target and shadow contour information will avoid the problem of feature redundant, or the dilemma of building complex mapping relationship on the original image directly.

**2. Deep Encoding Model.** When people listen to the voice, simultaneously observing the pronunciation mouth, even if the sound is very faint the content of the speech can be identifiable. The human brain extracts joint features which are more suitable for classification tasks. Based on this idea, the multi-layer auto-encoder has been successfully applied to robust speech features extraction in visual environment [6]. Such multilayer encoder is adopted to encode the target and shadow contour in the SAR image to obtain a fusion representation of target and shadow outline.

Concepts of deep learning derived from the study of artificial neural networks with multiple hidden layers. Multilayer Perception (MLP) is one deep structural model. The deep learning model constructs high-level representation (attributes or characteristics) by combining low-level feature to discover the input data's distributed representation.

Through the training of multi-layer nonlinear network structure, deep learning algorithm characterizes probability distribution of the input data and demonstrates the powerful ability to learn essential features. Hinton and Salakhutdinov proposed deep belief network (DBN) with an unsupervised greedy training algorithm [7] in 2006, which is supposed to resolve the optimization problems of deep network structure, and cause widespread concern in machine learning related research community [7-10]. The subsequent multilayer auto-encoder [11] proposed based on the idea is one of typical deep learning model, shown in Figure 1. Encoding part map two types of primary data to feature layer; decoding part reconstruct input samples from feature layer. Target real data's encoding and decoding parts utilize sparse Gaussian-Bernoulli RBM model, while other layers use sparse Bernoulli-Bernoulli RBM model.

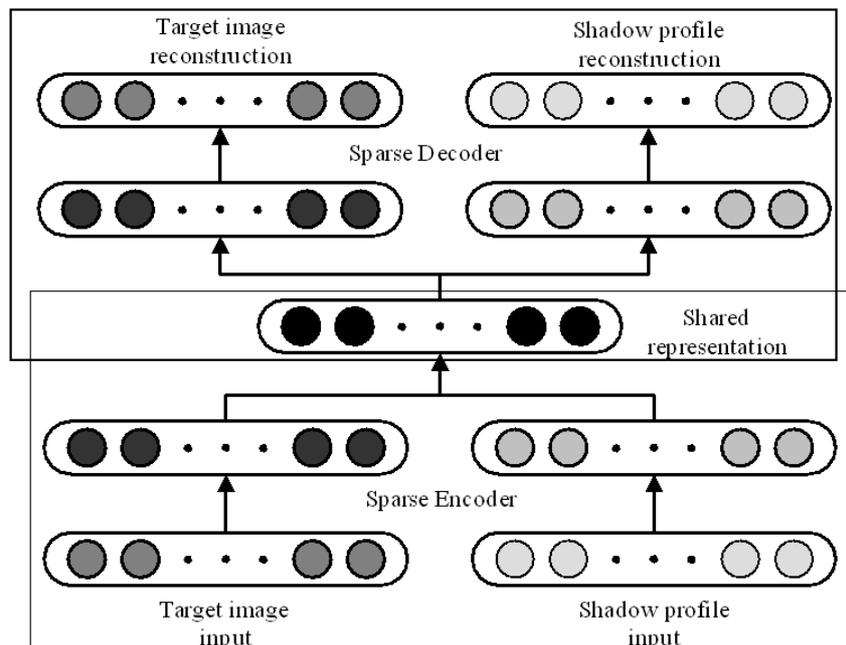


FIGURE 1. Bimodal deep auto-encoder model

The training of multi-layer auto-encoder is divided into two stages: pre-training and deep model fine-tune.

**2.1. Pre-training stage.** RBM is a typical neural network [12], as shown in Figure 2. To restrict sparsity of features learned, regularization factor  $S$  is introduced:

$$S = \lambda \sum_j \left( \rho - \frac{1}{m} \left( \sum_{k=1}^m E_{data} [h_j | v^k] \right) \right)^2 \quad (1)$$

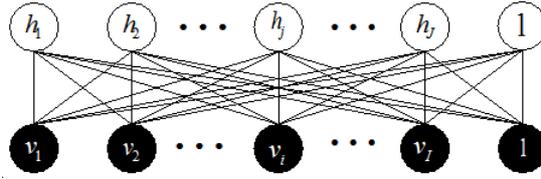


FIGURE 2. RBM model

Here,  $\{v^1, \dots, v^m\}$  represents the training set;  $\{h^1, \dots, h^n\}$  represents the hidden layer unit.  $\rho$  determines the sparsity of the hidden units;  $E_{data}$  denotes expectation over input data. Units in the visual layer and the hidden layer are interconnected. The training of hidden units is used to obtain high-level correlation. Compared with the traditional sigmoid belief network, RBM weights are relatively easy to train. Pre-training with unsupervised greedy layer-by-layer manner (Hinton called divergence contrast) is adopted to achieve generative weights. During the training process, the visual vector is mapped to hidden layer unit firstly; then the visual units are reconstructed by hidden units; these new visual units are mapped to the hidden units again, so that you can get further visual layer reconstruction. Such repeated steps are called Gibbs sampling. Associated differences between hidden layer and visual input units form the basis of weight update.

The joint distribution of RBM with visual units and hidden units in the conditions of given parameters can be represented with energy function:

$$p(v, h; \theta) = \exp(-E(v, h; \theta)) / Z \tag{2}$$

Here,  $Z = \sum_v \sum_h \exp(-E(v, h; \theta))$  is normalization factor. Marginal probability of the model given a visual vector is

$$p(v; \theta) = \sum_h \exp(-E(v, h; \theta)) / Z \tag{3}$$

The energy function for RBM of Bernoulli (visual)-Bernoulli (Hidden) is defined as

$$E(v, h; \theta) = - \sum_{i=1}^I \sum_{j=1}^J w_{ij} v_i h_j - \sum_{i=1}^I b_i v_i - \sum_{j=1}^J a_j h_j + S$$

Here  $w_{ij}$  represents symmetric weights of visual units and hidden units.  $b_i$  and  $a_j$  are the offset.  $I$  and  $J$  are the number of visual units and hidden units. The conditional probability can be calculated as follows:

$$\begin{aligned} p(h_j = 1|v; \theta) &= \delta \left( \sum_{i=1}^I w_{ij} v_i + a_j \right) \\ p(v_i = 1|h; \theta) &= \delta \left( \sum_{j=1}^J w_{ij} h_j + b_i \right) \end{aligned} \tag{4}$$

Here  $\delta(x) = 1/(1 + \exp(x))$ . Similarly, energy function for a Gaussian (visual)-Bernoulli (hidden) RBM can be rewritten as follows:

$$E(v, h; \theta) = - \sum_{i=1}^I \sum_{j=1}^J w_{ij} v_i h_j + \frac{1}{2} \sum_{i=1}^I (v_i - b_j)^2 - \sum_{j=1}^J a_j h_j + S \tag{5}$$

The corresponding conditional probability becomes:

$$p(h_j = 1|v; \theta) = \delta \left( \sum_{i=1}^I w_{ij} v_i + a_j \right) \tag{6a}$$

$$p(v_i = 1|h; \theta) = N\left(\sum_{j=1}^I w_{ij}h_j + b_i, 1\right) \quad (6b)$$

Here  $v_i$  satisfy Gaussian distribution with mean of  $\sum_{j=1}^J w_{ij}h_j + b_i$  and variance of 1. The Gauss-Bernoulli RBM can be used for modeling real target data. Bernoulli-Bernoulli RBM is used for modeling integer shadow data. With logarithmic likelihood probability gradient, we can deduce the RBM weight updating criterion:

$$\Delta w_{ij} = E_{data}(v_i h_j) - E_{mod\ el}(v_i h_j) \quad (7)$$

Here  $E_{data}(v_i h_j)$  is expectation over the observational training data.  $E_{mod\ el}(v_i h_j)$  is defined as expectations over the model. Unfortunately, it is difficult to calculate  $E_{mod\ el}(v_i h_j)$ . Comparison divergence algorithm similar to gradient is replaced with a full Gibbs sampling. Carefully training of RBM is the key to the successful application of deep learning. Table 1 provides a practical training guidance of RBM.

TABLE 1. RBM training algorithm

$\mathbf{x}$ represents training data; $\mathbf{h}$ represents hidden vector; $\varepsilon$ is learning rate for stochastic gradient descent method; $\mathbf{W}$ is weight matrix for RBM; $\mathbf{b}$ is input bias; $\mathbf{c}$ is output offset.
<ul style="list-style-type: none"> <li>• Step 1 sample <math>h</math> from <math>Q(h_{1i} = 1 x_1)</math> against Equation (6a).</li> </ul>
<ul style="list-style-type: none"> <li>• Step 2 sample <math>h</math> from <math>P(x_{2j} = 1 h_1)</math> against Equation (6b).</li> </ul>
<ul style="list-style-type: none"> <li>• Step 3 sample <math>h</math> from <math>Q(h_{2i} = 1 x_2)</math> against Equation (6a).</li> </ul>
<ul style="list-style-type: none"> <li>• Step 4 weights are updated.           <math display="block">W \leftarrow W + \varepsilon(h_1 x_1' - Q(h_2 = 1 x_2)x_2')</math> <math display="block">b \leftarrow b + \varepsilon(x_1 - x_2)</math> <math display="block">c \leftarrow c + \varepsilon(h_1 - Q(h_2 = 1 x_2))</math> </li> </ul>

**2.2. Multilayer encoder network fine tuning.** Multilayer encoder is one of the most successful deep learning frameworks. Being different from shallow model, it can exploit multi-model representation of target and shadow contour. Therefore, we consider pre-training one RBM for each model firstly. Then the output of the first hidden layer can be used as the input of the next new training layer. Essentially it is easier to achieve higher order correlation on the basis of representation of the first layer. A multilayer feature learning model is shown in Figure 3. Gauss-Bernoulli RBM is adopted in encoding and decoding target data, while Bernoulli-Bernoulli RBM is used to model shadow data.

After the model's construction, you need to fine-tune the network weights. Marginal Fisher criterion proposed in [13] is adopted to fine-tune network. Automatic encoder parameters  $\omega = [\omega_{enc}, \omega_{dec}]$ , here  $\omega_{enc}$  denotes encoding parameters, and  $\omega_{dec}$  denotes decoding parameters. The encoding part achieves the mapping from sample space to feature space. Optimal parameters should meet Equation (8):

$$\omega_{enc}^* = \arg \min_w J_{II} = \arg \min_{\omega_{enc}} \frac{\phi_{\omega_{enc}}(X)^T (D - W) \phi_{\omega_{enc}}(X)}{\phi_{\omega_{enc}}(X)^T (D^p - W^p) \phi_{\omega_{enc}}(X)} \quad (8)$$

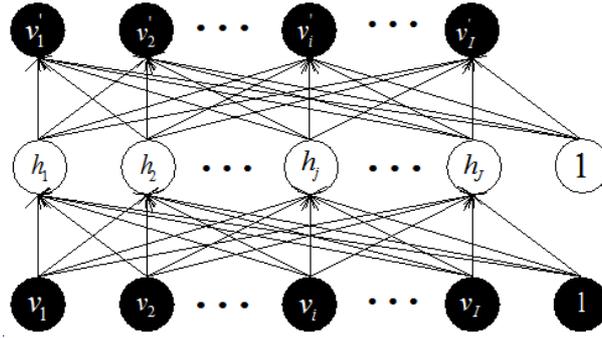


FIGURE 3. RBM encoding and decoding model

Here  $W_{ij} = \begin{cases} 1, & \text{if } i \in N_{k1}^+(j) \text{ or } j \in N_{k1}^+(i) \\ 0, & \text{else} \end{cases}$ ;  $D_{ii} = \sum_{j \neq i} W_{ij}, \forall i$

$N_{k1}^+(j)$  denotes nearest neighbor index of the sample set whose labels are the same to  $x_j$ .

Similarly,  $W_{ij}^p = \begin{cases} 1, & \text{if } (i, j) \in P_{k2}(c_i) \text{ or } (i, j) \in P_{k2}(c_j) \\ 0, & \text{else} \end{cases}$ ;  $D_{ii}^p = \sum_{j \neq i} W_{ij}^p, \forall i$

Here  $P_{k2}(c_i)$  are  $k_2$  nearest neighbor pairs in  $\{(i, j), i \in \pi_c, j \notin \pi_c\}$ .

Due to the powerful expressive ability of the deep network, the training often results in over-fitting. Then the generalization performance will be poor. So there is need to further introduce objective function regularization.

Extracted features should follow the internal structure of the data distribution. General access to the data distribution is to minimize the reconstruction error. Decoding part of the multilayer automatic encoding network naturally forms rebuilding function. Thus part of regularization is defined as follows:

$$R = \frac{1}{n} \sum_{i=1}^n \|x_i - \psi(\phi(x_i|\omega_{enc})|\omega_{dec})\|_2^2 \tag{9}$$

Here  $\|\cdot\|_2$  is the vector norm. It is also essential to have a smooth and simple model for better generalization performance. Weights and bias decay are chosen:

$$Dec = \|\omega_{enc}\|_2^2$$

Thus, the target function to optimize the feature turns into Formula (10):

$$E = J_{II} + \alpha R + \beta Dec \tag{10}$$

Here  $\alpha$  and  $\beta$  are regularization coefficients. Thus better feature can be extracted through the deep network with optimal solution to Formula (11).

$$W^* = \arg \min_W E = \arg \min_W J_{II} + \alpha R + \beta Dec \tag{11}$$

To be convenient for describing, we set  $s_c = \sum_i^n \sum_j^n (D_{i,j}^\omega - W_{i,j}^\omega) \|f_i - f_j\|_2^2 = \mathbf{1}_n^T (L^\omega \star \mathbf{Q}) \mathbf{1}_n$ .

Here  $L_{i,j}^\omega = D_{i,j}^\omega - W_{i,j}^\omega$ ;  $\mathbf{Q}_{i,j} = \|f_i - f_j\|_2^2$ ,  $\mathbf{1}_n$  represents  $n$ -dimensional vector of all the elements are 1.  $\star$  represents the operator of matrix dot product.

$$\frac{\partial s_c}{\partial f_i} = \sum_j (L_{i,j}^\omega + L_{j,i}^\omega) \cdot \frac{\partial \|f_i - f_j\|_2^2}{\partial f_i} = f_i \cdot \mathbf{1}_n^T \cdot (L^\omega + (L^\omega)^T)_i - F_n \cdot (L^\omega + (L^\omega)^T)_i \tag{12}$$

Similarly,  $\frac{\partial s_p}{\partial \mathbf{F}_n} = \mathbf{F}_n \cdot \left[ \text{Diag} \left( \mathbf{1}_n^T \cdot \left( \mathbf{L}_p^\omega + (\mathbf{L}_p^\omega)^T \right) \right) - \left( \mathbf{L}_p^\omega + (\mathbf{L}_p^\omega)^T \right) \right]$

$$\frac{\partial J_{II}}{\partial \mathbf{F}_n} = \frac{\partial \begin{pmatrix} s_c \\ s_p \end{pmatrix}}{\partial \mathbf{F}_n} = \frac{s_c \cdot \frac{\partial s_p}{\partial \mathbf{F}_n} - s_p \cdot \frac{\partial s_c}{\partial \mathbf{F}_n}}{(s_p)^2} \tag{13}$$

Derivative formula can be obtained from the coding layer to the input layer iteratively:

$$\frac{\partial J_{II}}{\partial \mathbf{W}} = \frac{\partial J_{II}}{\partial \mathbf{F}_n} \frac{\partial \mathbf{F}_n}{\partial \mathbf{W}}$$

For regularization factor,  $\frac{\partial R}{\partial \tilde{\mathbf{X}}_n} = \frac{2}{n}(\tilde{\mathbf{X}}_n - \tilde{\mathbf{X}}_n)$ ;  $\frac{\partial R}{\partial \mathbf{W}} = \frac{\partial R}{\partial \tilde{\mathbf{X}}_n} \frac{\partial \tilde{\mathbf{X}}_n}{\partial \mathbf{W}}$ .

Weight decay factor  $Dec$ ,  $\frac{\partial Dec}{\partial \mathbf{W}} = 2\mathbf{W}$ .

From the derivative formula before, we can achieve the training algorithm of network:

$$\frac{\partial E}{\partial \mathbf{W}} = \frac{\partial J_{II}}{\partial \mathbf{W}} + \alpha \frac{\partial R}{\partial \mathbf{W}} + \beta \frac{\partial Dec}{\partial \mathbf{W}} \tag{14}$$

According to Equation (10) of the objective function and Formula (14) of the gradient function, conjugate gradient method is adopted to obtain the corresponding solutions of minimizing the objective function E.

**3. Synergetic Neural Network.** Once obtaining the feature with the encoder, we need choose a classifier. An important theory of collaborative learning in the field of pattern recognition is: “The process of pattern recognition is the process of pattern formation” [14]. Pattern formation process includes the configuration of initial state, competition of the order parameter. Finally, the order parameters which win the competition control the entire system and make it into this particular order. In pattern recognition, this mechanism is used to restore original features.

Pattern recognition process corresponds to a dynamic process. Recognizing patterns of  $q(0)$  can be constructed in a dynamic process. After the intermediate state  $q(t)$  the  $q$  is mapped into the prototype  $q(0)$ . The process can be described by  $q(0) \rightarrow q(t) \rightarrow v_k$ . Haken network can be transformed as shown in Figure 4, which is similar to the three layer forward neural network. The middle order parameter corresponds to the dynamic equation of evolution.

$v_k$  must satisfy the normalization and zero mean condition:

$$\sum_{l=1}^N v_{kl} = 0, \quad \|v_k\|^2 = \left( \sum_{l=1}^N v_{kl}^2 \right)^{1/2} = 1$$

$\|v_k\|_2$  represents the Euclidean norm of  $v_k$ . The initial input vector  $q_0$  must also meet the same conditions, which can be achieved through the necessary preprocessing. The adjoint vector needs to satisfy:

$$(v_k^+, v_j) = v_k^+ v_j = \begin{cases} 1, & L_k == L_j \\ 0, & L_k \neq L_j \end{cases} \tag{15}$$

Here  $L_j$  represents label of  $j$ th sample.

Based on synergetic theory, pattern recognition process can be understood as the competition process of a number of order parameters.

**Definition 3.1.** *Order parameter can be defined as follows:*

$$\xi_k = v_k^+ q \tag{16}$$

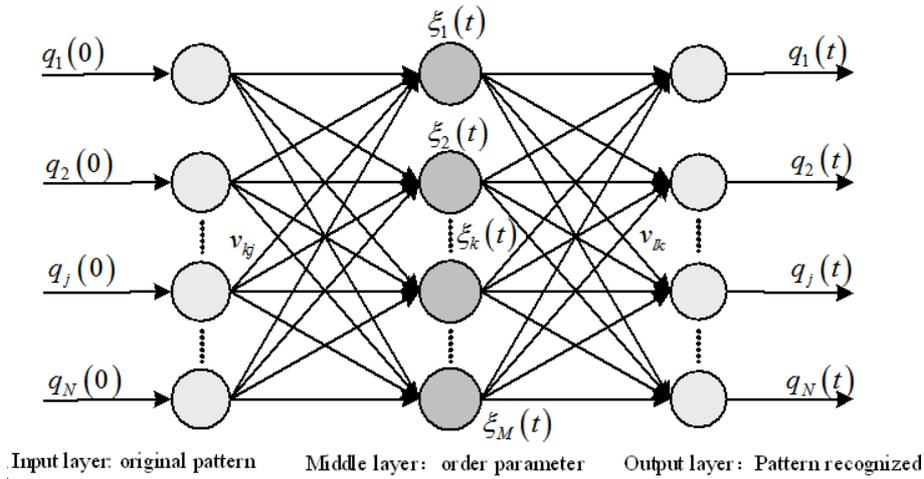


FIGURE 4. Structure of synergetic neural network

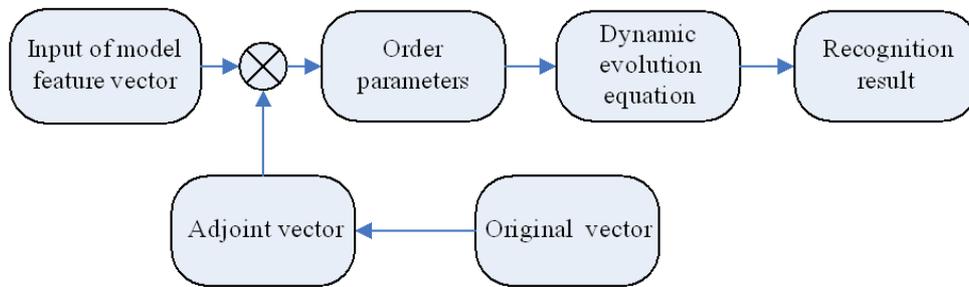


FIGURE 5. Synergetic pattern recognition process

The dynamic equation of order parameter is:

$$\begin{aligned} & \xi_k(n+1) - \xi_k(n) \\ &= \left( \lambda_k - C\xi_k^2(n) - (B+C) \sum_{L_{k'} \neq L_k} \xi_{k'}^2(n) + (B+C) \sum_{L_{k'} = L_k} \xi_{k'}^2(n) \right) \xi_k(n) \end{aligned} \quad (17)$$

Here  $L_k$  represents the label of  $k$ th sample. The first item in the bracket of Equation (17) represents their own excitation regulator; attention parameters  $\lambda_k$  adjust excitation intensity. The second item is self suppression regulator, which represents inhibition for its excessive growth. The third one is lateral inhibition, which represents mutual inhibition between representatives. The increase of order parameter of different label would produce inhibition. Final item represents lateral excitation, which means that any increase of the order parameter of the same label would produce excitation.

Operation process of synergetic neural network is shown in Figure 5.

#### 4. Simulation Experiment.

4.1. **Experimental data description.** The SAR ground stationary target dataset used in simulation experiment is provided by United States DARPA/AFRL MSTAR working group. Image size is 128\*128, resolution of which is 0.3m\*0.3m. The description of training samples is shown in Table 2. Table 3 gives the description of the test sample dataset.

TABLE 2. Training dataset description

Class	1°	2°	3°	4°	5°	6°	7°
T-72	232	116	78	58	47	39	34
BTR70	233	117	78	59	47	39	34
BMP2	233	117	78	59	47	39	34
Total	698	350	234	175	141	117	102

TABLE 3. Test dataset description

Class	T72-812	T72-S7	T72-132	BMP2-C21	BMP2-9563	BMP2-9566	BTR-C72	Total
size	195	191	196	196	195	196	196	1365

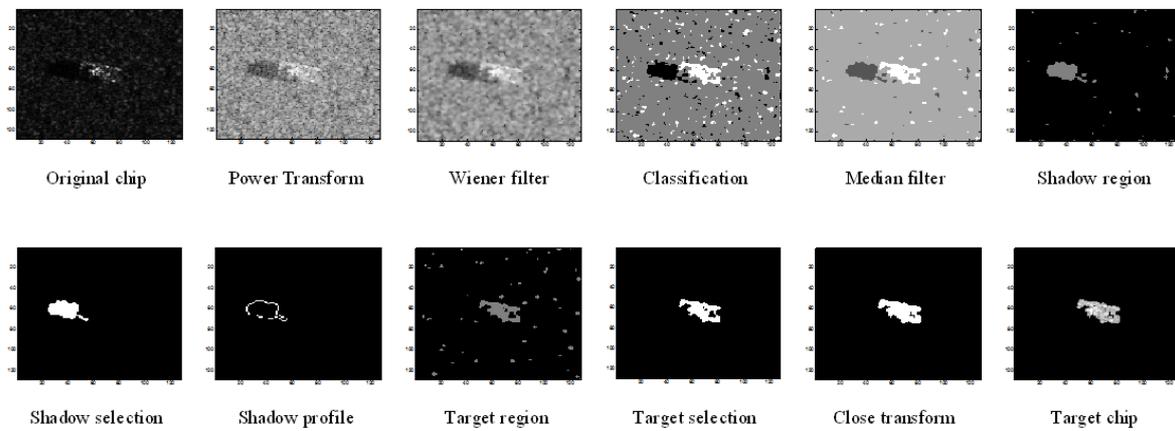


FIGURE 6. SAR target chip and shadow outline

**4.2. Image pretreatment.** The following segmentation technique is chosen with compromise on the algorithm complexity and exactness. First power transformation is used to convert the SAR image to be approximately Gaussian distribution. Wiener filtering method is adopted for denoising serious speckle noise. Image is marked by different label according to the statistics mean and variance. The image is then smoothed by median filter. Sobel operator is utilized to extract shadow boundary. The image pretreatment results are shown in Figure 6.

**4.3. Simulation setting.** The SAR contour segment algorithm shown in Figure 5. To further reduce the size of training images, the obtained image is cut, in which shadow image cutting range is  $(40\sim 89) \times (20\sim 69)$ ; target image segmentation pixel range is  $(40\sim 89) \times (40\sim 89)$ .

Then according to Table 1, the training step for RBM, two kinds of training data are modeled by RBM respectively. Sparse degree for shadow contour and the target image:  $\rho_{\text{shadow}} = \{0.05 \sim 0.3\}$ ;  $\rho_{\text{target}} = \{0.2 \sim 0.5\}$ ; the coefficient value of the degree in accordance with the different size of the training set in the above range. Iteration number of the training process: epoch = 2000; Bernoulli-Bernoulli RBM is used in modeling the shadow zone; the target gray scale image is modeled by the Gauss-Bernoulli RBM; dimensions of two first layer models are  $2500 \times 1000$ ; size of second layer SRBM model is  $2000 \times 1000$ .

After the pretraining of 3 RBM models as shown in Figure 3, the network can be further supervisedly fine-tuned as algorithm given in Section 2.2. The number of rounds for Supervised training is 200.

Once the training is completed, training samples and test samples are mapped from the sample space to the feature space. Then the classification of test data in the feature space can be operated through synergetic neural network. All the training samples are voting representative, different attention parameters of which are  $\lambda_k = \frac{N_{total}}{N_k}$ .  $N_{total}$  represents the size of training sample;  $N_k$  represents the size of samples belonging to the same category. The parameters for the evolution equation are  $B = 1$ ,  $C = 1$ . The iteration step number of Evolution is 200.

4.4. Experimental Result.

4.4.1. *Input data comparison.* Experiments results with the image of whole image (not segment), target, and the targets and shadow are shown in Table 4. The models of classification are NC (Nearest Center) and SNN respectively. Figure 7 shows an evolution curve. The results obtained by the simulation are shown in Table 4.

4.4.2. *Feature comparison.* To compare the result of different feature extracting algorithm, PCA [15] and KPCA feature analysis algorithms are used to extract principal components and the principal component of the kernel transformation. The largest heterogeneous distance feature extraction algorithm is proposed in [16]. We also give the results based KLDA feature extraction algorithm. Recognition performance of different algorithms utilizes the same nearest neighbor classifier for classification. The results are shown in Table 5.

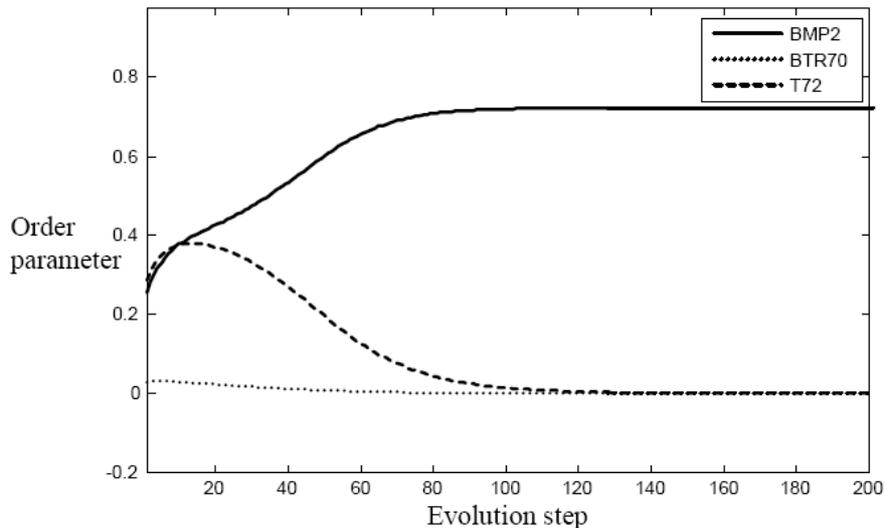


FIGURE 7. Dynamics evolution curve

TABLE 4. Rate comparison of different input data (%)

Size of training set	whole image (NC)	Target (NC)	Target and shadow (NC)	Target and shadow (SNN)
698	90.84	95.16	95.38	96.63
350	89.74	93.77	94.87	95.46
234	89.38	93.55	93.41	93.92
175	88.79	91.28	92.31	93.33
141	82.27	90.55	89.30	90.84
117	82.49	85.35	87.25	89.30
102	78.17	84.25	85.71	87.77

TABLE 5. Performance comparison of different feature extraction algorithm (%)

Size of training set	PCA <sup>[15]</sup>	KPCA <sup>[15]</sup>	KLDA <sup>[16]</sup>	MID <sup>[16]</sup>	Our results
698	94.65	95.38	95.46	98.17	95.16
350	94.21	94.43	94.36	96.19	93.77
234	90.62	91.06	89.30	94.29	93.55
175	88.35	88.86	86.37	90.40	91.28
141	82.49	82.93	83.30	85.20	90.55
117	81.32	82.27	84.18	80.59	85.35
102	82.49	83.00	84.69	77.29	84.25

TABLE 6. Recognition performance comparison of different classification (%)

Template matching <sup>[1]</sup>	SVM <sup>[14]</sup>	HMM <sup>[15]</sup>	Our results
84.93	90.99	94.00	96.63

4.4.3. *Comparative experiment results with shallow structure models.* The literature [1] fuse target and shadow outline using Fourier descriptors, maximum correlation method get the recognition results in a cascade manner. [17] recognized the target with SVM model directly on the image samples. [18] introduced HMM to model target variability. The results with different models are shown in Table 6.

4.5. **Analysis of experiment results.** To investigate the classification performance of the different input, the input data are described in Table 4, which also shows the comparison results. It can be seen the entire image have spots noise, affecting the accuracy of identification. After the image segmentation, classification performance on the pure target alone is improved significantly. Then the introduction of shadow information extracts robust feature, further improving the recognition accuracy.

Table 4 gives the experimental results show that the proposed algorithm in recognition performance is superior to most of other correlation algorithm. From the recognition results shown in Table 5, there is little difference between KLDA and KPCA on the recognition rate. The maximum heterogeneous distance algorithm obtains good results with large size of sample sets. However, the performance decline soon when the sample set size become small. The reason is the algorithm only considers different training samples, not taking into account the same sample. This further is validated that the feature extraction network can be used to enhance the recognition rate. As the training set size reduction can be found in the recognition result is reduced gradually, this is due to the decrease of the sample increases the recognition error of the system. Besides, we can see that when the training set size decrease seriously, the performance still keep high over others. This is because our methods not only use the labeled sample, all unlabeled sample are also used in the unsupervised experiment. At the same time, when the labeled sample is limited, the shadow information can further enhance the recognition rate.

Table 6 shows the recognition result of the shallow structure. This paper utilizes multi-layer automatic encoder for shared representation of the target and shadow, which could involve the complex mapping relationship. In contrast, the shallow algorithms or feature extraction, or “link” feature would have limitation in modeling the mapping function. The results verify the effectiveness of the proposed algorithm.

5. **Conclusion.** We combine the shadow contour and target image information to improve the recognition performance, and introduce marginal fisher criterion to regulate network weights. RBM and multilayer automatic encoder are used for feature extraction.

For the difficulty of obtaining the typical SAR image representative, we introduce multi-represent synergistic neural network algorithm. Besides introducing different representative attention parameters, corresponding multi-represent synergistic evolution dynamic equation is also given. Experimental results show that the proposed algorithm has certain advantage. The advantages of the proposed algorithm are derived from unsupervised pre-training process. The algorithm is applicable to semi-supervised learning. More training data would improve the performance. How to build a reasonable large size training dataset is one further research interest.

**Acknowledgment.** The authors gratefully acknowledge the helpful comments and suggestions of the reviewers and the editors, which have improved the presentation.

## REFERENCES

- [1] K. Yin, L. Jin, C. Li et al., An SAR ATR based on fusion of target contour and shadow contour, *Journal of Air Force Engineer University (Natural Science Edition)*, vol.12, no.1, pp.24-28, 2011.
- [2] L. Yang, W. Hao and D. Wang, SAR image recognition based on multi-aspect of shadow information, *Transactions of Nanjing University of Aeronautics & Astronautics*, vol.26, no.4, pp.320-326, 2009.
- [3] S. Papsion and R. M. Narayanan, Classification via the shadow region in SAR imagery, *IEEE Transactions on Aerospace and Electronic Systems*, vol.48, no.2, pp.969-979, 2012.
- [4] J. Cui, J. Gudnason and M. Brookes, Radar shadow and superresolution features for automatic recognition of MSTAR targets, *IEEE International Radar Conference*, Arlington, VA, USA, pp.534-539, 2005.
- [5] A. Jakulin, *Machine Learning Based on Attribute Interactions*, Ph.D. Thesis, University of Ljubljana, 2005.
- [6] J. Ngiam, A. Khosla, M. Kim et al., Multimodal deep learning, *Proc. of the 29th International Conference on Machine Learning*, University of Edinburgh, Scotland, pp.1-9, 2012.
- [7] G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, vol.313, pp.504-507, 2006.
- [8] G. E. Hinton, S. Osindero and Y. W. Teh, A fast learning algorithm for deep belief nets, *Neural Computation*, vol.18, no.7, pp.1527-1554, 2006.
- [9] Z. J. Sun, L. Xue, Y. M. Xu et al., Overview of deep learning, *Journal of Application Research of Computers*, vol.12, no.8, pp.2806-2810, 2012.
- [10] W. Yan and H. J. Cai, A simulation study of deep belief network combined with the self-organizing mechanism of adaptive resonance theory, *International Conference on Computational Intelligence and Software Engineering*, pp.1-4, 2010.
- [11] P. Vincent, H. Larochelle, Y. Bengio et al., Deep learning with denoising autoencoders, *Proc. of the 25th International Conference on Machine Learning*, pp.1096-1103, 2008.
- [12] H. Lee, C. Ekanadham and A. Y. Ng, Sparse deep belief net model for visual area V2, *Advances in Neural Information Processing Systems*, pp.873-880, 2008.
- [13] S. C. Yan, D. Xu, B. Y. Zhang et al., Graph embedding and extensions: A general framework for dimensionality reduction, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.1, pp.40-51, 2007.
- [14] S. P. Gou, L. C. Jiao and X. L. Xiao, Image recognition using synergetic neural networks based on immune clonal clustering, *Journal of Electronics & Information Technology*, vol.30, no.2, pp.263-266, 2008.
- [15] P. Han, R. B. Wu, Z. H. Wang et al., SAR Automatic target recognition based on KPCA criterion, *Journal of Electronics & Information Technology*, vol.25, no.10, pp.1297-1301, 2003.
- [16] B. Wang, Y. L. Hang, J. Y. Yang et al., A feature extraction method for SAR automatic target recognition based on maximum interclass distance, *Science China Technology Science*, vol.41, no.10, pp.1388-1392, 2011.
- [17] Q. Zhao and J. C. Principe, Support vector machines for SAR automatic target recognition, *IEEE Transactions on Aerospace and Electronic Systems*, vol.37, no.2, pp.643-654, 2001.
- [18] D. P. Kottke, P. D. Fiore, K. L. Brown et al., A design for HMM based SAR ATR, *Proc. of SPIE*, Orlando, FL, USA, vol.3370, pp.541-551, 1998.