

Research Article

A Method Combining CNN and ELM for Feature Extraction and Classification of SAR Image

Peng Wang , Xiaomin Zhang, and Yan Hao

Department of Mathematics, North University of China, Taiyuan, Shanxi 030051, China

Correspondence should be addressed to Peng Wang; wpmath@nuc.edu.cn

Received 25 January 2019; Revised 4 October 2019; Accepted 31 October 2019; Published 28 November 2019

Academic Editor: Eduard Llobet

Copyright © 2019 Peng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the large number of Sigmoid activation function derivation in the traditional convolution neural network (CNN), it is difficult to solve the question of the low efficiency of extracting the feature of Synthetic Aperture Radar (SAR) images. The Sigmoid activation function in the CNN is improved to be a rectified linear unit (ReLU) activation function, and the classifier is modified by the Extreme Learning Machine (ELM). Finally, in this CNN model, the improved CNN works as the feature extractor and ELM performs as a recognizer. A SAR image recognition algorithm based on the CNN-ELM algorithm is proposed by combining the CNN and the ELM algorithm. The experiment is conducted on the Moving and Stationary Target Acquisition and Recognition (MSTAR) database which contains 10 kinds of target images. The experiment result shows that the algorithm can realize the sparsity of the network, alleviate the overfitting problem, and speed up the convergence speed of the network. It is worth mentioning that the running time of this experiment is very short. Compared with other experiment on the same database, it indicates that this experiment has generated a higher recognition rate. The accuracy of the SAR image recognition is 100%.

1. Introduction

Synthetic Aperture Radar (SAR) is an important mean to obtain information, and it is widely used in geological survey, topographic mapping, utilization of marine resources, and so on. The reason why it is worthy to explore in the future is due to its vast practical applications.

SAR image recognition always is seemed as research hotpot, in the process of obtaining information, while feature extraction is one key factor in the success of an image target recognition system. Mishra and Mulgrew firstly put forward the application of principal component analysis (PCA) to SAR image classification. Experiments on Moving and Stationary Target Acquisition and Recognition (MSTAR) database show that the classifier based on PCA is better than the bias classifier based on Gaussian model under the condition of limited training data [1]. Knee et al. proposed an automatically classifying method using image partitioning and sparse representation-based feature vector generation for SAR image [2]. Wang et al. proposed a complementary spa-

tial pyramid coding (CSPC) approach in the framework of spatial pyramid matching. Both the coding coefficients and coding residuals are explored to develop more discriminative and robust features for representing SAR images [3]. Chen et al. proposed the deep convolutional networks (DCN) for target classification of SAR images, which can achieve an average accuracy of 99% on classification of ten-class targets of MSTAR database [4]. Ding et al. proposed the method combining of global and local filters (CGLF) for SAR target recognition under the standard operating condition and various extended operating conditions, which can be seen as a standard of comparison [5].

Convolutional neural network (CNN) is a class of feedforward neural networks including convolution computation and deep structure. It is one of the representative algorithms of deep learning [6]. CNN has been widely used in many fields. Particularly in the field of recognition, CNN is used for handwritten digit recognition [7, 8], speech recognition [9, 10], facial expression recognition [11, 12], human face recognition [13, 14], refrigerator fruit

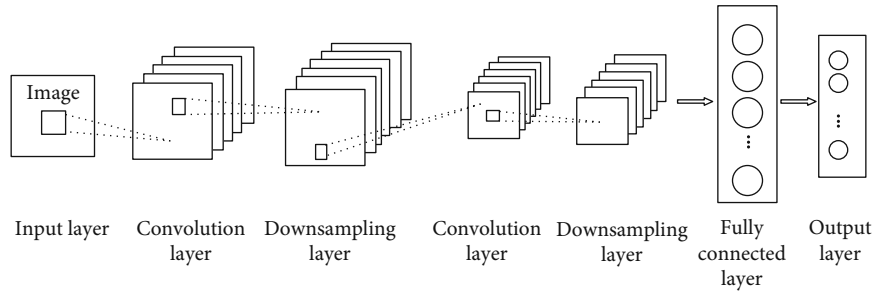


FIGURE 1: Structure of CNN.

and vegetable recognition [15], verification code recognition [16], traffic sign classification [17] and recognition [18], and so on. In the field of image recognition, the images can be directly made the input of CNN, which reduces the complexity of the experiment. Furthermore, the image information can be passed through the forward propagation to the convolution layer and the downsampling layer by CNN. At the same time, it can be handled in the different network layers, which avoids the extraction process of the complex features in the traditional algorithms. The most basic features of the images can be accessed by the neuron in the local sensing domain of CNN. CNN still remains a high degree of invariance in extracting complex features of images, regardless of shift, scaling, deformation, rotation, or other forms of deformation of the image [19]. Cireşan et al. built multicolumn deep neural networks to apply to Mixed National Institute of Standards and Technology database (MNIST) handwriting database, which has a great recognition effect (recognition rate is less than 0.3%) [20]. Therefore, the CNN is a better image feature extractor.

In recent years, CNN has also been successfully applied to radar image recognition. Du et al. proposed a displacement and rotation insensitive deep CNN model which was trained by augmented dataset [21]. Peng et al. proposed a new CNN architecture with spatial pyramid pooling (SPP) which can build a high hierarchy of feature map by dividing the convolved feature maps from finer to coarser levels to aggregate local features of SAR images [22]. Krizhevsky et al. conducted a CNN model to apply to ImageNet large scale visual recognition challenge (ILSVRC-2012) and obtained some good results [23]. Cho and Park proposed CNN architecture using aggregated features and fully connected layers, which the accuracy recognizing the 10 classes of military targets on MSTAR dataset is 94.38% [24].

Generally speaking, the last layer of CNN can be regarded as a linear classifier, but it is not the optimal classifier. The commonly used optimal classifier is support vector machine (SVM) and its improved algorithms. In 2006, Huang and LeCun combined convolutional networks (CN) and SVM algorithm to identify targets and generated a high recognition rate [25]. In 2012 under the combining algorithms, the experiment on handwritten digit was conducted; the accuracy can be up to 99.81% [7]. Particularly in reference [26], CNN works as a trainable feature extractor and SVM performs as a recognizer, the classification

accuracy 98.49% is gained on the Kennedy Space Center (KSC) dataset and 99.45% is gained on the Pavia University Scene (PU) dataset.

However, compared with the traditional SVM, back propagation (BP) and other classification algorithms, the Extreme Learning Machine not only has a fast training speed, less adjustment parameters, but also has short running time and high training precision [27]. A new SAR image recognition method based on the CNN-ELM algorithm is proposed in this paper. The process of the method is as follows: firstly, ReLU function is used in CNN, instead of Sigmoid function; secondly, the image features are extracted; finally, the last layer of CNN will be replaced with ELM in order to recognize these images. And the method is characterized by a high recognition rate and short running time.

2. CNN and ELM

2.1. Convolution Neural Network. In recent years, because machine learning does not need to change the topological structure of the images, it is very popular in image recognition. Convolution neural network (CNN) is not only one of the deep learning [28] but also one of the artificial neural networks, which mainly is used in the fields of speech analysis [29] and image recognition [30].

The structure of a traditional CNN model is shown in Figure 1. There are five layers in the CNN model. The input layer is a matrix of the normalized pattern with size $S \times S$. The feature map connects inputs with its previous layer. It means that the features obtained by convolution layer are used as input of pooling layer. All the neurons in one feature map share the same kernel and connecting weights (known as the sharing weights in [31]). For example, with a kernel size of 5 and a subsampling ratio of 2, each feature map layer reduces the feature size from the previous feature size S to $(S - 4)/2$.

There are three unique structural characteristics in the CNN model: local sensing domain, weight sharing [31], and downsampling. The local perception domain is the single neuron of each layer. It is the neurons of every layer which is only relative to the neuron in a certain domain (generally the neurons in the rectangular area are 5×5) in the network input layer. Due to the unique structural characteristics, the structural features of the input image can be extracted by each neuron. Weight sharing can greatly reduce the training parameters of the network and the number of training

samples. Downsampling is an effective feature of extracting images, which can make the model have a good antinoise capability and greatly reduce the feature dimension of images. The CNN model is divided into input layer, hidden layer, and output layer. There are two hidden layers: convolution layer (extracting feature) and downsampling layer (selecting the optimizational feature).

2.2. Extreme Learning Machine. The Extreme Learning Machine (ELM) algorithm is a new fast learning algorithm. It does not need to adjust during the training process, but only needs to set the number of the hidden layer neurons, and finds the optimal solution. Compared with the traditional classification algorithm such as CNN-SVM [7], the advantages of fast learning speed, strong generalization ability, and few adjustment parameters have been found. For a single hidden layer neural network, supposing that there is N arbitrary samples (X_j, t_j) , where $X_j = [x_{j1}, x_{j2}, \dots, x_{jm}]^T \in \mathbb{R}^n$, $t_j = [t_{j1}, t_{j2}, \dots, t_{jm}]^T \in \mathbb{R}^m$.

If there are L hidden layer nodes, the single hidden layer neural network can be expressed as

$$o_j = \sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i), \quad j = 1, 2, \dots, N, \quad (1)$$

where $g(x)$ is an activation function, $W_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is input weight, β_i is output weight, b_i is bias, $W_i \cdot X_j$ is the inner product of W_i and X_j . The learning aim of the single hidden layer neural network is to minimize the error of output. It can be considered as

$$\min \sum_{j=1}^N \|o_j - t_j\|, \quad (2)$$

that means, there are existences β_i , W_i , and b_i , which can be regarded as

$$\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) = t_j, \quad j = 1, 2, \dots, N. \quad (3)$$

It can be expressed as a matrix

$$H\beta = T. \quad (4)$$

In order to train the single hidden layer neural network, there is a hope to get \widehat{W}_i , \widehat{b}_i , and $\widehat{\beta}_i$ that makes

$$\|H(\widehat{W}_i, \widehat{b}_i) \widehat{\beta}_i - T\| = \min \|H(W_i, b_i) \beta_i - T\|, \quad (5)$$

where $i = 1, 2, \dots, L$; it is equivalent to the minimum loss function

$$\min E = \min \sum_{j=1}^N \left(\sum_{i=1}^L \beta_i g(W_i \cdot X_j + b_i) - t_j \right)^2. \quad (6)$$

There is no need to adjust parameters in the ELM algorithm. Once the input weights W_i and hidden layer bias b_i are randomly determined, the output matrixes H and β of the hidden layer are uniquely determined.

3. Recognition of SAR Images Based on Improved CNN-ELM

In the convolution layer of Figure 1, the characteristic graph (map) is convoluted by convolution kernel, and the map of the convolution layer is outputted by the convolution structure through the activation function. The convolution layer and the downsampling layer alternately appear, and each output map of the convolution layer is related to the input map. Generally, the output of the convolution layer is

$$y_j^n = f \left(\sum_{i \in M_j} y_j^{n-1} * W_{ij} + \phi_j \right), \quad (7)$$

where n is the number of layers for the convolution layer, W_{ij} is the convolution kernel, ϕ_j is the bias, and M_j is the input map. $f(\cdot)$ is the activation function. The standard CNN activation function is the Sigmoid function which can be expressed as [32]

$$f(x) = (1 + \exp(-x))^{-1}. \quad (8)$$

The output range of Sigmoid function is (0,1) in the process of adjusting the weights; it is positive proportions between the change of weight value and the output of the upper layer; when part of it tends to zero, the reduction or no adjustment of the weight adjustment will increase the training time. From Figure 2(a), it can be seen that its derivative curve, in turn, looks like a bowl, which easily causes the problem of gradient dispersion.

Therefore, it is improved to ReLU function [23], where it is an unsaturated nonlinear function; it is easy to derive and realize the sparsity of the network. From Figure 2(b), the output of some neurons is 0, which reduces the dependence among the parameters, relieves the overfitting problem, and can transmit the gradient to the front network very well in reverse propagation. Meanwhile, it can reduce the problems which are caused by gradient dispersion and speed up the convergence speed of the network. The formula of the ReLU function is

$$f(x) = \max(x, 0) = \max \left(\sum_{i \in M_j} y_i^{n-1} * W_{ij} + \phi_j, 0 \right). \quad (9)$$

In this paper, the improved CNN is used to extract the image features which are used as the input of the ELM algorithm to get the recognition accuracy. Therefore, a SAR image recognition algorithm based on improved CNN-ELM is proposed, and the structure of proposed algorithm can be seen in Figure 3.

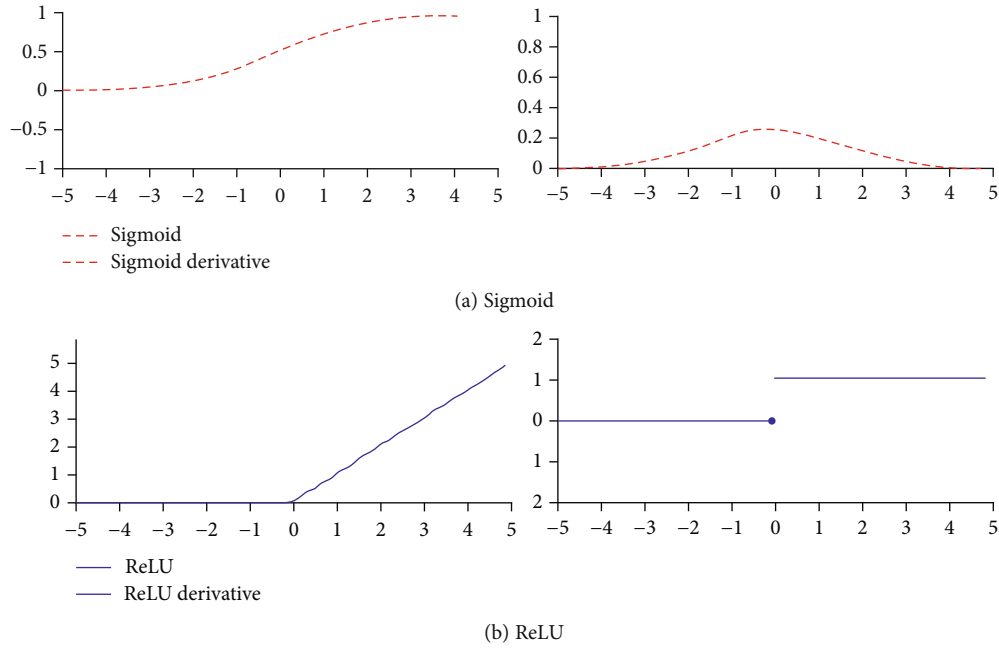


FIGURE 2: Sigmoid, ReLU function, and its derivative curve.

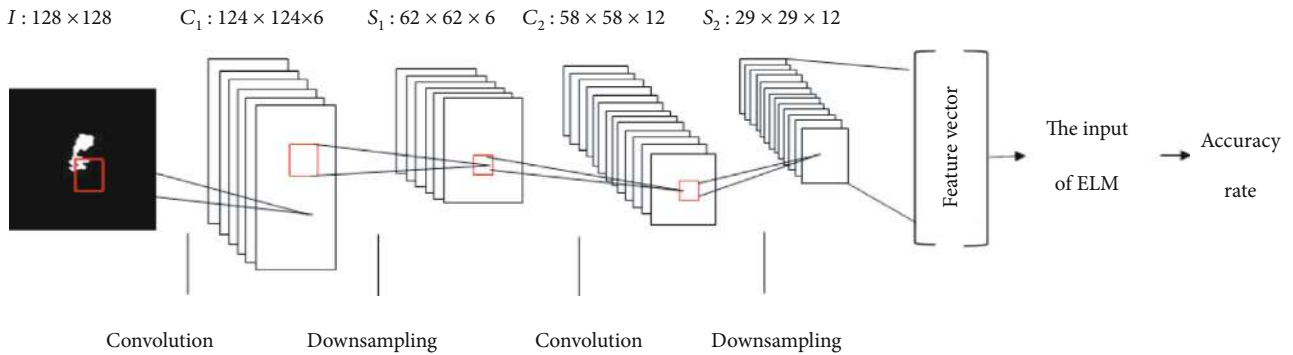


FIGURE 3: Structure of the improved CNN-ELM for SAR image recognition.

The algorithm steps are shown in the following.

- (1) Input of CNN: the SAR gray-scale images after denoising, segmentation, edge detection, and other operations
- (2) The first convolution layer: the SAR gray-scale images are convoluted by 6 filters whose size are 5×5 and bias and then activated by ReLU function to get 6 feature maps
- (3) The first downsampling layer: the mean pooling is used in each map which is obtained from the previous step. In this step, the length of the mean pooling is 2. The aim of this step is to get the structure information of the image
- (4) The second convolution layer: these maps are convoluted by bias and 12 filters whose size are 5×5 , then they are activated by ReLU activation function to get 12 maps
- (5) The second downsampling layer: the mean pooling is used in each map which is obtained from the previous step. In this step, the length of the mean pooling is 2. The aim of this step is to get the feature vector of the image
- (6) ELM algorithm recognition: the feature vector obtained by step (5) is used as the input of the ELM algorithm, and the ELM train function and training sets are used to create and train ELM; next, the trained parameters and ELM predict function are used to test the test sets; finally, the recognition accuracy of the training sets and the test sets will be obtained

4. Simulation Experiment and Analysis

4.1. Simulation Experiment. All experiments were performed on a MATLAB 2016b with i7-8700 CPU and 8 G memory. In this experiment, the dataset are from MSTA database. There



FIGURE 4: Target shapes and SAR image shape.

are 10 kinds of target images including 2S1, BMP_2, BRDM_2, BTR_60, BTR_70, D_7, T_62, T_72, ZIL_13_1, and ZUS_23_4; their shapes and SAR image shapes will be shown in Figure 4. Each target of SAR images is observed in all directions. Generally, the SAR images taken from 15° of depression angle are used to train the proposed CNN architecture,

and the SAR images taken from 17° of depression angle are used to evaluate the performance of the proposed method.

Comparing with the optical image, the SAR image contains a lot of noise; the SAR image will be pretreated including denoising, segmentation, and edge detection. These images will be shown in Figure 5. Then, the SAR images

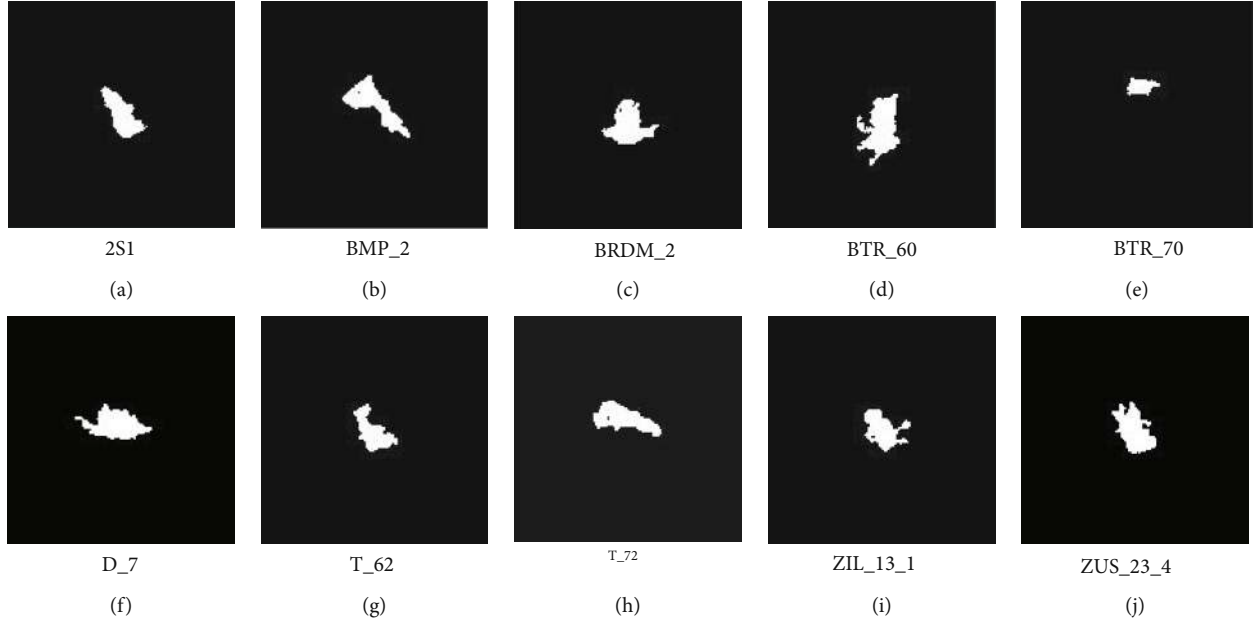


FIGURE 5: 10 kinds of tank SAR images after denoising, segmentation, edge detection, and other pretreated images under 15° of depression angle.

TABLE 1: Training and testing set of MSTAR database.

MSTAR database	Training set	Testing set
2S1	299	274
BMP_2	233	196
BRDM_2	296	274
BTR_60	256	195
BTR_70	233	196
D_7	299	274
T_62	299	273
T_72	232	196
ZIL_13_1	299	274
ZUS_23_4	299	274
Total	2747	2426

which are pretreated and taken from 15° of depression angle are used as the training set; meanwhile, the SAR images taken from 17° of depression angle are used as the test set. The details can be seen in Table 1, the total size of the training set is 2747, and the size of the test set is 2426. Next to it, the improved CNN is used to extract the feature of the training set and the test set, and the acquired feature vectors are extracted. These vectors are seemed as the input of the Extreme Learning Machine (ELM). The recognition accuracy which is obtained through the ELM algorithm to identify and classify is compared with other experimental algorithms (the algorithm in the literature [7, 25]); the performance of the proposed algorithm is significantly improved.

4.2. Experimental Result Analysis. The calculation method of accuracy is the 10 kinds of SAR images (2S1, BMP_2, BRDM_2, BTR_60, BTR_70, D_7, T_62, T_72, ZIL_13_1,

ZUS_23_4) of training set and test set which are, respectively, labeled with 1,2,...,10 labels as original labels. The ELM algorithm is used to get their labels. If the number of labels tested is the same as that of the original labels, then add 1 and finally, divide the number of each label by the total number of the labels to get the accuracy. The time of feature extraction using CNN is about 1.2 seconds, and the time of ELM recognition is about 0.15 seconds, so the total time is about 1.35 seconds.

In order to test the performance of the proposed CNN-ELM algorithm for SAR image recognition, comparisons are made with principal component analysis (PCA) [1], deep convolutional networks (DCN) [4], sparse representation classification (SRC) [2], complementary spatial pyramid coding (CSPC) [3], combination convolutional nets and support vector machine (CN-SVM) [25], and CNN-SVM [7] methods, respectively. The results of the experiment are shown in Table 2.

In Table 2, from the classification and recognition of every kind of SAR images, it is obvious that the accuracy rate of CNN-ELM is significantly higher than other methods; the average accuracy rate of CNN-ELM is 100%, next comes CNN-SVM (99.57%), DCN (99.13%), CGLF (98.56%), CSPC (94.48%), CN-SVM (94.38%), SRC (92.70%), and PCA (92.43%). Particularly, the detailed comparison with similar CN-SVM [25] and CNN-SVM [7] is as follows.

Compared with CN-SVM, the accuracy rate of CNN-ELM in the type 2S1 is 5.47% higher, in the type BMP_2 is 4.03% higher, in the type BRDM_2 is 14.96% higher, in the type BTR_60 is 2.54% higher, in the type BTR_70 is 8.67% higher, in the type D_7, T_72, and ZUS_23_4 is 1.82% higher, in the type T_62 is 8.79% higher, and in the type ZIL_13_1 is 6.2% higher, respectively.

TABLE 2: Recognition and contrast results (%) of different algorithms for SAR images.

Accuracy rate	SRC [2]	DCN [4]	CGLF [5]	PCA [1]	CSPC [3]	CNSVM [25]	CNN- SVM [7]	CNNELM
2S1	85.36	98.18	95.62	89.95	90.65	94.53	100.00	100.00
BMP_2	84.48	98.98	98.46	94.54	92.5	95.97	100.00	100.00
BRDM_2	90.11	99.27	95.26	92.54	93.34	85.04	100.00	100.00
BTR_60	96.53	96.41	100.00	95.68	98.95	97.46	99.99	100.00
BTR_70	98.33	99.49	99.49	85.66	99.79	91.33	97.66	100.00
D_7	98.32	99.27	100.00	90.85	98.12	98.18	100.00	100.00
T_62	90.54	99.64	98.17	89.99	90.52	91.21	99.50	100.00
T_72	94.59	100.00	100.00	96.79	91.38	98.18	99.50	100.00
ZIL_13_1	96.35	99.64	99.49	94.68	94.65	93.80	99.00	100.00
ZUS_23_4	92.3	99.64	99.64	93.65	94.62	98.18	100.00	100.00
Average	92.70	99.13	98.56	92.43	94.48	94.38	99.57	100.00

Compared with CNN-SVM, the classification accuracy rate of these five SAR images (BTR_60, BTR_70, T_62, T_72, and ZIL_13_1) has been improved. Separately, the accuracy rate of CNN-ELM in the type BTR_60 is 0.01% higher, in the type BTR_70 is 2.34% higher, in the type T_62 and T_72 is 0.5% higher, and in the type ZIL131 is 1% higher, respectively. On the whole, the recognition accuracy rate of the improved CNN-ELM algorithm is 5.62% higher than CN-SVM, and the accuracy rate is also 0.43% higher than CNN-SVM. The experiment time is very short, which shows that the algorithm has a very strong feasibility and can be further applied to the classification and recognition of other objects.

5. Conclusion

In this paper, the CNN is improved through improving the activation function; then, the improved CNN is used to extract the depth features of 10 kinds of SAR gray images which are under the processes of denoising, segmentation, and edge detection. Following this process, these SAR images are classified by the ELM algorithm combined with the features extracted from CNN. The experimental result shows that the combination of CNN and ELM has a high accuracy rate and short time for the recognition of SAR images, which demonstrates the effectiveness of the algorithm and can be further applied to the classification and recognition of other objects, especially for some images whose features are not obvious.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The research project is supported by the National Nature Science Foundation of China (Grant No. 61127008), the Natural Science Foundation of Shanxi Province (Grant Nos. 201801D121026, 201701 D121012, and 201701D221121), and the Shanxi Scholarship Council of China (Grant No. 2016-088).

References

- [1] A. K. Mishra and B. Mulgrew, "Radar signal classification using PCA-based features," in *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, pp. 1104–1107, Toulouse, France, May 2006.
- [2] P. Knee, J. J. Thiagarajan, K. N. Ramamurthy, and A. Spanias, "SAR target classification using sparse representations and spatial pyramids," in *IEEE RadarCon (RADAR)*, pp. 294–298, Kansas City, MO, USA, May 2011.
- [3] S. Wang, L. Jiao, S. Yang, and H. Liu, "SAR image target recognition via complementary spatial pyramid coding," *Neurocomputing*, vol. 196, no. C, pp. 125–132, 2016.
- [4] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4806–4817, 2016.
- [5] B. Ding and G. Wen, "Combination of global and local filters for robust SAR target recognition under various extended operating conditions," *Information Sciences*, vol. 476, pp. 48–63, 2019.
- [6] J. Gu, Z. Wang, J. Kuen et al., "Recent advances in convolutional neural networks," *Pattern Recognition*, vol. 77, pp. 354–377, 2018.
- [7] X.-X. Niu and C. Y. Suen, "A novel hybrid CNN-SVM classifier for recognizing handwritten digits," *Pattern Recognition*, vol. 45, no. 4, pp. 1318–1325, 2012.
- [8] H.-h. Zhao and H. Liu, "Multiple classifiers fusion and CNN feature extraction for handwritten digits recognition," *Granular Computing*, vol. 4, pp. 1–8, 2019.
- [9] S. Singhal, V. Passricha, P. Sharma, and R. K. Aggarwal, "Multi-level region-of-interest CNNs for end to end speech

- recognition,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 11, pp. 4615–4624, 2019.
- [10] T. Tan, Y. Qian, H. Hu, Y. Zhou, W. Ding, and K. Yu, “Adaptive very deep convolutional residual network for noise robust speech recognition,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 8, pp. 1393–1405, 2018.
 - [11] S. Xie and H. Hu, “Facial expression recognition with FRR-CNN,” *Electronics Letters*, vol. 53, no. 4, pp. 235–237, 2017.
 - [12] J. Li, D. Zhang, J. Zhang et al., “Facial expression recognition with faster R-CNN,” *Procedia Computer Science*, vol. 107, no. C, pp. 135–140, 2017.
 - [13] S. Kang, J. Lee, K. Bong, C. Kim, Y. Kim, and H.-J. Yoo, “Low-power scalable 3-D face frontalization processor for CNN-based face recognition in mobile devices,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, no. 4, pp. 873–883, 2018.
 - [14] C. Ding and D. Tao, “Trunk-branch ensemble convolutional neural networks for video-based face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002–1014, 2018.
 - [15] B. Park, A. Shin, S. K. Park et al., “Ecological study for refrigerator use, salt, vegetable, and fruit intakes, and gastric cancer,” *Cancer Causes & Control*, vol. 22, no. 11, pp. 1497–1502, 2011.
 - [16] L. G. Hafemann, R. Sabourin, and L. E. S. D. Oliveira, “Analyzing features learned for offline signature verification using deep CNNs,” in *2016 23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, December 2016.
 - [17] D. Cireşan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-column deep neural network for traffic sign classification,” *Neural Networks*, vol. 32, pp. 333–338, 2012.
 - [18] X. Mao, S. Hijazi, R. Casas, P. Kaul, and C. Rowen, “Hierarchical CNN for traffic sign recognition,” in *2016 IEEE Intelligent Vehicles Symposium (IV)*, Gothenburg, Sweden, June 2016.
 - [19] I. Arel, D. C. Rose, and T. P. Karnowski, “Deep machine Learning - A new frontier in artificial intelligence research [Research Frontier],” *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13–18, 2010.
 - [20] D. Cireşan, U. Meier, and J. Schmidhuber, “Multi-column deep neural networks for image classification,” 2012, <http://arxiv.org/abs/1202.2745>.
 - [21] K. Du, Y. Deng, R. Wang, T. Zhao, and N. Li, “SAR ATR based on displacement- and rotation-insensitive CNN,” *Remote Sensing Letters*, vol. 7, no. 9, pp. 895–904, 2016.
 - [22] L. Peng, X. Liu, M. Liu, L. Dong, M. Hui, and Y. Zhao, “SAR target recognition and posture estimation using spatial pyramid pooling within CNN,” in *2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology*, p. 106200W, Beijing, China, January 2018.
 - [23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, pp. 1097–1105, Lake Tahoe, NV, USA, December 2012.
 - [24] J. H. Cho and C. G. Park, “Additional feature CNN based automatic target recognition in SAR image,” in *2017 Fourth Asian Conference on Defence Technology - Japan (ACDT)*, pp. 1–4, Tokyo, Japan, December 2017.
 - [25] F.-J. Huang and Y. LeCun, “Large-scale learning with SVM and convolutional nets for generic object categorization,” in *Proceedings. Computer Vision and Pattern Recognition Conference (CVPR’06)*, New York, June 2006.
 - [26] J. Leng, L. Tao, B. Gang, Q. Dong, and D. Han, “Cube-CNN-SVM: a novel hyperspectral image classification method,” in *2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI)*, San Jose, CA, USA, November 2016.
 - [27] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, “Extreme learning machine: a new learning scheme of feedforward neural networks,” in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541)*, pp. 985–990, Budapest, Hungary, July 2004.
 - [28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
 - [29] S. Sukittanon, A. C. Surendran, J. C. Platt, and C. J. Burges, “Convolutional networks for speech detection,” in *Eighth international conference on spoken language processing*, Jeju Island, Korea, October 2004.
 - [30] Y.-N. Chen, C.-C. Han, C.-T. Wang, B.-S. Jeng, and K.-C. Fan, “A CNN-based face detector with a simple feature map and a coarse-to-fine classifier,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 3, pp. 1–12, 2009.
 - [31] F. Lauer, C. Y. Suen, and G. Bloch, “A trainable feature extractor for handwritten digit recognition,” *Pattern Recognition*, vol. 40, no. 6, pp. 1816–1824, 2007.
 - [32] J. Han and C. Moraga, “The influence of the sigmoid function parameters on the speed of backpropagation learning,” in *International Workshop on Artificial Neural Networks*, pp. 195–201, Springer, 1995.

