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

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 hotspot, 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 spatial 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...
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 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 \( 5 \times 5 \) 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, suppose that there is $N$ arbitrary samples $(X_j, t_j)$, where $X_j = [x_{j1}, x_{j2}, \ldots, x_{jm}]^T \in \mathbb{R}^m, t_j = [t_{j1}, t_{j2}, \ldots, t_{jn}]^T \in \mathbb{R}^n$.

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

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

where $g(x)$ is an activation function, $W_i = [w_{i1}, w_{i2}, \ldots, w_{im}]^T$ is input weight, $\beta_i$ 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 $\beta_i$, $W_i$, and $b_i$, which can be regarded as

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

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 $\hat{W}_i$, $\hat{b}_i$, and $\hat{\beta}_i$ that makes

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

where $i = 1, 2, \ldots, L$; 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 $\beta$ 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

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

where $n$ is the number of layers for the convolution layer, $W_{ij}$ is the convolution kernel, $\phi_j$ 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

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

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.
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 $5 \times 5$ 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 $5 \times 5$, 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 8G memory. In this experiment, the dataset are from MSTA database. There
are 10 kinds of target images including 2S1, BMP_2, BRDM_2, BTR_60, BTR_70, D_7, T_62, T_72, ZII_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
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.
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.

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.

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2SI</td>
<td>85.36</td>
<td>98.18</td>
<td>95.62</td>
<td>89.95</td>
<td>90.65</td>
<td>94.53</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>BMP_2</td>
<td>84.48</td>
<td>98.98</td>
<td>98.46</td>
<td>94.54</td>
<td>92.5</td>
<td>95.97</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>BRDM_2</td>
<td>90.11</td>
<td>99.27</td>
<td>95.26</td>
<td>92.54</td>
<td>93.34</td>
<td>85.04</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>BTR_60</td>
<td>96.53</td>
<td>96.41</td>
<td>100.00</td>
<td>95.68</td>
<td>98.95</td>
<td>97.46</td>
<td>99.99</td>
<td>100.00</td>
</tr>
<tr>
<td>BTR_70</td>
<td>98.33</td>
<td>99.49</td>
<td>99.49</td>
<td>85.66</td>
<td>99.79</td>
<td>91.33</td>
<td>97.66</td>
<td>100.00</td>
</tr>
<tr>
<td>D_7</td>
<td>98.32</td>
<td>99.27</td>
<td>100.00</td>
<td>90.85</td>
<td>98.12</td>
<td>98.18</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>T_62</td>
<td>90.54</td>
<td>99.64</td>
<td>98.17</td>
<td>89.99</td>
<td>90.52</td>
<td>91.21</td>
<td>99.50</td>
<td>100.00</td>
</tr>
<tr>
<td>T_72</td>
<td>94.59</td>
<td>100.00</td>
<td>100.00</td>
<td>96.79</td>
<td>91.38</td>
<td>98.18</td>
<td>99.50</td>
<td>100.00</td>
</tr>
<tr>
<td>ZIL_13_1</td>
<td>96.35</td>
<td>99.64</td>
<td>99.49</td>
<td>94.68</td>
<td>94.65</td>
<td>93.80</td>
<td>99.00</td>
<td>100.00</td>
</tr>
<tr>
<td>ZUS_23_4</td>
<td>92.3</td>
<td>99.64</td>
<td>99.64</td>
<td>93.65</td>
<td>94.62</td>
<td>98.18</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Average</td>
<td>92.70</td>
<td>99.13</td>
<td>98.56</td>
<td>92.43</td>
<td>94.48</td>
<td>94.38</td>
<td>99.57</td>
<td>100.00</td>
</tr>
</tbody>
</table>

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, 201701D121012, and 201701D221121), and the Shanxi Scholarship Council of China (Grant No. 2016-088).

References


