Feature Extraction of Radar Echo Image Based on Improved Convolutional Neural Network

Zhang Chenhao$^{1, *}$, Li Hongquan$^{1}$, Feng Xun$^{1}$ and Tang Jingmian$^{1}$

$^{1}$ Air Force Early Warning Academy, Wuhan, China

*1996zhangchenhao@163.com

Abstract. Feature extraction of radar echo image is an important part of identification and recognition of air targets. In recent years, with the rapid development of deep learning, new solutions are provided for it. In this paper, the convolution neural network (CNN) is applied to feature extraction. Based on the classic CNN model, Adam is used to update the model parameters, dropout is used to prevent overfitting, and an improved CNN model is constructed. Then, the radar echo image data set is used to train the model, so as to extract target features and classify them. Simulation results show that the accuracy of the improved model is 99%, and the training speed is greatly improved. Different from the traditional extraction method which relies on manual experience, the improved CNN can improve the efficiency of feature extraction of radar echo image and lay a solid foundation for further research and identification of air targets.

1. Introduction

In the military field, air defense is very important for national security, and it is necessary to track every target entering the airspace in detail. Radar is a clairvoyant to guard airspace, and can monitor air targets. Based on oscillograms of the radar screen, the commander can get the batch, quantity and other information of the air targets. However, for a long time, the features of the radar echo image have been mainly extracted by manual recognition, which obviously does not conform to the rapidly changing the form of war. At the same time, it is subject to more subjective influences and the error rate of feature extraction is higher. How to extract the radar echo image features autonomously and accurately is an urgent problem to be solved at present. In recent years, the rise of artificial intelligence has provided a new way for extracting radar echo image features. Due to the long-term early warning missions of air defense, the radar has accumulated a large number of echo images. Through the multi-layer processing in deep learning, from the initial feature to the global feature, the radar echo image features are finally obtained.

Aiming at the problem of radar image feature extraction, some scholars have conducted some researches by different methods. Zhang Hanhua intercepts valid echo data, removes unnecessary data interference, and defines a series of echo parameters to extract echo features[1]. But the parameters defined by this method may not completely include all features. Shao Yun used fuzzy max-min neural network to determine the sortie of the target according to the waveform expansion. This algorithm is a feed-forward global connection network, which will lead to a higher complexity of the neural network, and involves many parameters, which reduces the training efficiency of the network[2]. When Wang Jiabao used CNN to identify synthetic aperture radar image targets, he combined multiple basic operating units in the CNN into different modules, and constructed a hierarchical structure network...
that combines feature extraction and classifier training[3]. In that paper, the stochastic gradient descent (SGD) method is used to train the network, and the direction of the gradient does not point to the direction of the minimum value, which will reduce the speed of parameter update.

To sum up, using deep learning can solve the problems related to the feature extraction of echo images, especially the CNN has a significant advantage in the field of image recognition. Considering that the determination of parameters is an important part in training neural network model, this paper will improve the CNN to solve the problems of slow parameter optimization speed, long training time and over-fitting, so as to improve the efficiency of extracting radar echo image features.

2. Analysis of feature extraction

The radar echo image is a visualization of the signal received by the radar on the screen. By extracting the features of the radar echo image, the target can be identified and recognized. At present, A scope, P scope and 3D scope are commonly used in radar. Among them, the A scope is a synchronous oscilloscope; the P scope is a radial circular scan display; the 3D scope can show the echo in three dimensions.

However, when identifying and recognizing air targets manually, it is mainly based on the echo image features of A scope. For aircrafts, the echo in A scope has a large amplitude, wide width and bright wave color. At the same time, the peaks of one aircraft are stable, while the peaks of two aircrafts and above are beating. For meteorological targets, the echo of the A scope is weak, the peaks are acuminate, the pulsation are irregular, and the waves are not patterned.

CNN is a deep feed-forward neural network with local connection and weight sharing[4]. As one of the representative algorithms of deep learning, it is widely used in computer vision, natural language processing and other fields. The main feature of CNN is that it has a convolution operation[5]. In the convolutional layer, the weight of each neuron connecting data window is fixed, and each neuron only focuses on one feature. Neurons are filters in image processing, and each filter in the convolutional layer has its own image feature to focus on. All these neurons together can be regarded as a set of feature extractors for the entire image.

3. Improved CNN model

Generally, CNN mainly includes convolutional layers, activation functions, pooling layers, fully connected layers, and objective functions. This paper considers the characteristics of radar echo images, recombines the above types of layers, uses Adam to update parameters, and Dropout prevents overfitting, so as to build an improved CNN model.

3.1. Basic framework of model

3.1.1. Convolutional layer. The convolution layer mainly uses the convolution kernel to perform convolution operation on the radar echo image to extract its features. For three-dimensional image data such as radar echo images, in addition to height and length, processing channels are also required. When there are multiple feature maps in the channel, each feature map is first convolved with the convolution kernel, and then the convolution operation results of all feature maps are added to obtain the output of the convolution layer. If the input of convolutional layer \( l \) is \( x^l \in \mathbb{R}^{H^{l} \times W^{l} \times D^{l}} \), the convolution kernel of this layer is \( f^{l} \in \mathbb{R}^{H^{l} \times W^{l} \times D^{l}} \), and \( b \) is the bias, then the formal convolution operation can be expressed as:

\[
y_{i^l, j^l}^{l+1} = \sum_{i=0}^{H^{l}} \sum_{j=0}^{W^{l}} \sum_{d=0}^{D^{l}} f_{i,j,d}^{l+1} \odot x_{i^l,i,j,d}^{l+1} + b
\]  

(1)

Where \((i^{l+1}, j^{l+1})\) is the position coordinate of the convolution result, which satisfies:

\[
0 \leq i^{l+1} < H^{l+1} \\
0 \leq j^{l+1} < W^{l+1}
\]  

(2)
3.1.2. **Pooling layer.** The pooling layer can further compress the feature map output by the convolution layer, so that the model can remove redundant information, further extract features, and reduce the number of parameters. Max pooling and mean pooling are two common methods. Max pooling is to take the maximum value of the target area, and mean pooling is to average the target area. Since mean pooling mainly preserves the background information of the image, and Max pooling mainly preserves the texture information, this paper will use Max pooling. If the pooling kernel of layer $l$ is $P^{l} \in \mathbb{R}^{H \times W \times D}$, Max pooling can be expressed as:

$$y_{i^{l+1}, j^{l+1}, d^{l}} = \max_{0 \leq i \leq H^{l+1}, 0 \leq j \leq W^{l+1}} x_{i+i^{l}+H, j+j^{l}+W, d+d^{l}}$$

(3)

where $0 \leq i^{l+1} < H^{l+1}$, $0 \leq j^{l+1} < W^{l+1}$, $0 \leq d < D^{l}$.

3.1.3. **Activation function.** The activation function is a simulation of the characteristics of biological neurons, that is, there is a threshold for neurons. When the accumulation of signals exceeds this threshold, the neuron will be activated; otherwise, the neuron is in a state of inhibition. And the activation function can increase the non-linearity of CNN. At present, the common activation functions are sigmoid function and ReLU (Rectified Linear Unit) function. Their expressions are as follows:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

(4)

$$\text{rectified}(x) = \max\{0, x\} = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}$$

(5)

![Figure 1. Sigmoid function and its gradient](image1)

![Figure 2. ReLU function and its gradient](image2)

By comparing figure 1 and figure 2, we can find that the Sigmoid function is all compressed to 1 when $x \geq 5$, and all compressed to 0 when $x \leq -5$, causing a saturation effect, so that the gradient of the function tends to 0 when $x \geq 5$ or $x \leq -5$, and the parameters cannot be updated to train the network model[6]. For the ReLU function, the gradient is 0 when $x < 0$, and the gradient is 1 when $x \geq 0$. In this part, the gradient saturation effect is completely eliminated. In addition, compared with the Sigmoid function, the ReLU function is beneficial to the convergence, and the convergence speed is about 6 times faster[7]. Therefore, the ReLU function is selected as the activation function in this paper.

3.1.4. **Fully connected layer.** The fully connected layer plays the role of a classifier in CNN, that is, after extracting features of radar echo image through the convolutional layer, activation function layer and pooling layer, the full-connection layer combines the extracted features nonlinearly to obtain the output and realize the recognition and classification. Each node of the full connected layer is connected to all nodes of the previous layer to integrate the extracted features. Because of the full connection, the parameters of the fully connected layer are also the most.
3.1.5. Output layer. For the classification problem of radar echo image feature extraction, the Softmax function needs to be used as the activation function in the output stage of CNN. The Softmax function can regularize the input values so that the sum of their output values is 1. Its expression is:

$$\text{softmax}(x) = \frac{\exp(x_i)}{\sum_{i=1}^{n} \exp(x_i)}$$

where \(n\) is the number of nodes in the previous layer.

The output layer also includes a loss function that measures the error between the predicted value and the true value. In CNN, when looking for the optimal parameters, including weights and biases, it is necessary to find parameters that make the loss function as small as possible, and generally the mean square error and the cross-entropy error are used. The mean square error is usually used as the loss function in the regression problems, while the classification problems involve the discrete One-Hot vector, and the cross entropy error is usually used as the loss function. Therefore, the cross entropy error is used as the loss function when extracting the radar echo image features, and the expression is:

$$E = - \sum_{k} t_k \log y_k$$

where \(y_k\) is the output of CNN, and \(t_k\) is the label of the correct solution.

Firstly, the radar echo image data processed by convolution layer, activation function layer and pooling layer are connected into one column. Secondly, because the neural network is a kind of supervised learning, during model training, the model is trained according to the training samples, so as to obtain the weight of the fully connected layer. Finally, when using this model for result recognition, weighted summation is performed according to the weights obtained through training and the results calculated through the convolution layer, activation function layer, and pooling layer to obtain the scores of each result. The structure of CNN model is shown in figure 3.

![Figure 3. The structure of CNN model](image)

3.2. Model optimization

3.2.1. Parameters updating. The purpose of the training phase of CNN is to obtain the parameters that minimize the loss function. But because the number of parameters of a neural network is large, it is a complicated task to optimize it. Stochastic gradient descent (SGD) is the most common parameter optimization method. It samples randomly and uniformly, and then updates the parameters along the gradient direction. After several iterations, it gradually approaches the optimal parameters. However, SGD has a lot of noise, so that the parameters are not updated in the direction of overall optimization every iteration, which may lead to local optimization. In addition, the scaling of the gradient in each dimension is the same when updating the parameters, which may reduce the training speed.

In order to solve the problems existing in SGD, parameter update methods such as Momentum and AdaGrad are proposed. Momentum simulates the inertia of an object in motion. When updating parameters, the optimized direction is retained to some extent. Meanwhile, fine-tuning is carried out based on the gradient of the current batch to determine the direction of parameter updating. This method can increase stability, improve training speed, and reduce the probability of local optimization. For the problem of using the same update rate for all parameters, AdaGrad can adaptively adjust the
learning rate for each parameter. In this paper, Adam method is used, combining the advantages of Momentum and AdaGrad, to achieve efficient search in the parameter space, improve training speed. Besides, after bias correction, the learning rate of each iteration has a certain range, making the parameters relatively stable. The mathematical expression of Adam is as follows:

$$v_t = \beta_1 v_{t-1} + (1 - \beta_1) \frac{\partial L}{\partial W}$$

$$s_t = \beta_2 s_{t-1} + (1 - \beta_2) \left( \frac{\partial L}{\partial W} \right)^2$$

$$W_t = W_{t-1} - \eta \frac{v_t}{\sqrt{s_t} + \epsilon}$$

where formula (9) calculates the estimation of the first-order moment, and formula (10) calculates the estimation of the second-order distance. $\eta$, $\beta_1$, $\beta_2$ and $\epsilon$ are the hyperparameters.

3.2.2. Preventing overfitting. CNN can only fit training data, but cannot fit other data well[8]. This phenomenon is overfitting. The main reason for this is that the training data is too small for the sample to represent the predetermined classification rules, or the network model becomes complex because of too many parameters. Weight decay is a common method to prevent overfitting, which mainly punishes larger weights in the process of training the model. This method can prevent overfitting by simple means, but when the CNN model is more complicated, it cannot effectively prevent overfitting. Therefore, Dropout is chosen in this paper. This method will randomly delete some hidden layer neurons every time data is transferred during training. During the test, although all neuron signals are transmitted, the output of each neuron needs to be multiplied by the deletion ratio.

4. Simulation

4.1. Preparation for simulation
In this paper, the Python programming language with Tensorflow is used for simulation to realize the extraction of radar echo image features. The radar echo images of this experiment consist of one aircraft in a batch and two aircrafts in a batch. Each type includes 1000 images as training and test datasets, and uses their category names as classification labels.

4.2. Simulation process
First of all, since this paper is about feature extraction of two types of images, the fully connected layer of the network needs to be set to 2 outputs, that is, 2 neurons. Secondly, according to the needs of the task and the size of the echo image data set the hyperparameters. Then, use the training data set to train the improved CNN model to make the error converge. Finally, the test data is input into the trained model for verification.

4.3. Analysis of simulation results
Figure 4 and 5 show the changes in accuracy and loss value with the increase of iterations in the training process before and after the model improvement. It can be found from the figures that the model before improvement starts to converge at about 35,000 iterations, while the improved model starts to converge at about 2,000 iterations. The performance of the improved CNN model is greatly improved, and the accuracy rate is 99%, which is quite high.
Figure 4. Changes in loss and accuracy before model improvement

Figure 5. Changes in loss and accuracy after model improvement

5. Conclusion
In this paper, the improved CNN model is applied to feature extraction of radar echo image, which greatly reduces the training time of the model and improves the accuracy to 99%. It confirms that CNN has a powerful feature extraction capability and lays a theoretical and practical foundation for the identification and recognition of air targets.

Reference
[1] Zhang H H, Wang W, Jiang W D and Chen Z P 2003 J. J. Natl. Univ. Def. Technol. 25 38-41+52
[2] Shao Y, Wang H L, Zhang H J and Chen H 2015 J. Shipb. Electron. Countermeas. 38 62-65+69
[3] Wang J B, Li Y, Zhang G N, Miao Z, Li H and Xu W G 2017 J. Appl. Res. Comput. 34 1597-1600
[4] Zhou F Y, Jin L P and Dong J 2017 J. Chin. J. Comput. 40 1229-1251
[5] Lecun Y, Boser B, Denker J S, Henderson D, Howard R E and Hubbard W 1989 J. Neural Comput. 1 541-551
[6] Hochreiter S 1998 J. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. 6 107-116
[7] Krizhevsky A, Sutskever I and Hinton G 2012 J. Adv. Neural Inf. Process. Syst. 25
[8] Yang Z Z, Kuang N, Fan L and Kang B 2018 J. J. Signal Process. 34 1474-89