Journal of Electronic Testing (2022) 38:77–89
https://doi.org/10.1007/s10836-022-05985-1

A New Neural Network Based on CNN for EMIS Identification

Ying-chun Xiao[1,2] · Feng Zhu[1] · Sheng-xian Zhuang[1] · Yang Yang[1]

Received: 15 November 2021 / Accepted: 13 February 2022 / Published online: 8 March 2022

Abstract
Electromagnetic interference sources (EMIS) must be identified in order to locate them promptly. Because representative features of EMIS broadband signals are difficult to extract, we propose a new identification method based on convolutional neural network (CNN) to extract EMIS deep features from spectrum signals and increase recognition accuracy. To achieve noise reduction, we added a noise reduction layer (NRL) to the network, which uses background noise data as the weight to determine its correlation with the input data. Furthermore, a new loss function based on intra-class and inter-class relative distances is presented, which is paired with the Softmax loss function to make the network converge fast and consistently. Experiments on three data sets are used to validate the created method's overall performance. Simulated results demonstrate that the suggested method can effectively extract the deep features of the EMIS signal, enhance signal classification speed and accuracy, and achieve 100% accuracy on our data set.

Keywords Electromagnetic interference sources (EMIS) · Convolutional neural network (CNN) · Noise reduction · Extraction of deep features · Loss function

1 Introduction

In recent years, with the widespread application of radio techniques, the electromagnetic environment has become more and more complex, and various electromagnetic emission sources (EMS) have appeared [20]. For example, narrowband signals produced by inter-modulation of broadcast signals, stray radiation caused by frequency drift of wireless communication equipment, high-frequency radiation signal produced by aging and discharge of electrical equipment, and broadband signal produced by signal jammers and the Pantograph arc of electrified railway, among others. If these signals interfere with the normal operation of nearby electrical equipment, they can be called an electromagnetic interference source (EMIS). For example, when the EMS is near the airport, it will affect the normal operation of communication and navigation equipment [3, 7]. The current locating technology of EMIS cannot achieve precise positioning and can only locate to a range.The current location accuracy of EMIS needs to be improved [25]. This can be narrowed down by identification. The spectrum peak of the narrow-band interference signal is obvious, the direction is clear, and it is easier to identify. The frequency spectrum of broadband interference signals is extremely wide, and the spikes have certain random characteristics, which makes them difficult to identify. So, effectively identifying the type of EMIS is very important for its quick location and radio monitoring.

Many conventional signal classification methods, including k-nearest neighbor classification, decision tree and template matching, rely on algorithms that are computationally intensive and require manual analysis of features [6, 30, 32]. For many years, the identification of EMS has been basically based on the extraction of signal features and pattern recognition methods, where the signal can be in the time domain or the frequency domain. In 2001, Antonini and Orlandi used wavelet packet decomposition and statistical methods to classify the time-domain signals of radiation sources [1]. However, the signal is a random sequence with white noise and a signal-to-noise ratio of 1, which has not been applied in practice. Liet et al. applied wavelet transform to the feature extraction of radiation signals and, combined with cluster analysis, analyzed the conduction radiation problem 14. Empirical mode decomposition (EMD) and its improvement methods are also widely used to analyze the

Responsible Editor: T. Xia

✉ Feng Zhu
zhufeng@swjtu.cn

[1] Southwest Jiaotong University, Chengdu 611756, China

[2] Lanzhou City University, Lanzhou 730070, China

characteristics of EMS signals [14, 31]. By combining frequency characteristics with intrinsic mode function (IMF), one can extract the amplitude ratio of the main EMS, thereby identifying the main emission sources. But others with less interference cannot be identified. The artificial neural network (ANN) can improve the recognition accuracy [23]. Support vector machines (SVM) are used to improve the recognition accuracy of the EMS signal, transforming spatial information from 3D data sets to 1D vectors to simplify the model, and the recognition accuracy is as high as 99.9%, but when the signal-to-noise ratio increases, the recognition accuracy is reduced significantly [22]. With the development of artificial intelligence and machine learning, deep learning methods have promoted the research of radio signal classification [11, 27, 33], there are more researches on radar signal recognition [2, 12, 13, 21, 34]. But, few researches on the identification of EMIS. Among them, the convolutional neural network (CNN) can reduce the computational complexity and suppress over-fitting 1712. The convolution operation is very beneficial for processing images and one-dimensional sequence data and has been widely used. Many classic networks have used convolutional operation, such as LeNet 3, AlexNet 24, VGG 36, ResNet [29, 36], Inception 25 and other networks. In 31, the CNN network classifies radio signals without extracting any features or other preprocessing on the original signals, but allows the network to learn the original time series features on high-dimensional data directly, and the training time is long. Literature 19 uses ResNet to classify time-series radio signals and finds that its training time is less, but its accuracy is not significantly improved.

Two problems need to be solved in order to effectively identify a single EMIS. One is to extract deep features based on simple EMIS signals to improve the recognition performance, and the other is to reduce the complexity of the recognition algorithm and shorten the training time. We propose a new CNN based on only the frequency domain signal of EMIS to identify it accurately and quickly.

The remainder of this article is organized as follows. Section 2 briefly introduces the background of the EMIS signal. Section 4 introduces the new CNN proposed in this paper. Our method is validated and analyzed in Sect. 4.1. Section 4.2 is conclusion and future directions.

2 Background

The electromagnetic environment is composed of different EMS signals and is becoming more and more complex in the time domain, frequency domain, space domain, and energy. When the frequency of the EMS signal is equal to or very close to the working frequency of the sensitive equipment, and the amplitude reaches a certain value, it will affect the normal operation of the relevant sensitive equipment. So, frequency and amplitude are important parameters of the EMIS signal. Generally, the types of electromagnetic interference (EMI) can be summarized as single-frequency interference, multiple-frequency interference, narrow-band interference, partial-band interference, and full-band interference.

Single-frequency interference is a single-frequency sine wave interference, and its frequency spectrum is a spectral line. Multiple-frequency interference is the superposition of several single-frequency interferences, and its frequency spectrum has multiple spectral lines. There are more interference frequency components compared with single-frequency interference.Narrow-band interference refers to interference with signals in a relatively narrow frequency band. Full-band interference refers to interference with signals of all working frequencies. Partial-band interference refers to interference with signals in individual frequency bands. These frequency bands can be adjacent or non-adjacent.

Single-frequency interference, multiple-frequency interference, and narrow-band interference have obvious spectral peaks, and the interference frequency or frequency band is clear and easy to identify. Partial-band interference and full-band interference signals have a wide frequency range, and the spikes are random, most of which are impulsive non-Gaussian noise. For example, the impulsive noise caused by printers, microwave ovens, fluorescent lamps, elevators, and other equipment in daily life has more "spikes" 4. It is more difficult to identify such signals, and deeper features need to be extracted.

Although it is difficult to explain the physical meaning of the feature parameters extracted by the deep neural network (DNN), the DNN can independently learn a variety of features from the simple input data. We proposed a new identification network based on the CNN. It self-learns based on only EMIS frequency domain signals, extracts features, and identifies interference sources. The overall architecture is shown in Fig. 1. It includes a noise reduction layer (NRL), a deep feature extraction sub-network and an improved Softmax (Im-Softmax) classification layer, whose details are given in Sect. 4.

3 The New Netwok Based on CNN for EMIS Identification

3.1 Noise Reduction Processing

The EMIS frequency domain test data includes not only the frequency components of the interference source, but also the frequency components of other noises in the environment. The spectrum data of the same EMIS signal under different test backgrounds may be very different, so the background noise

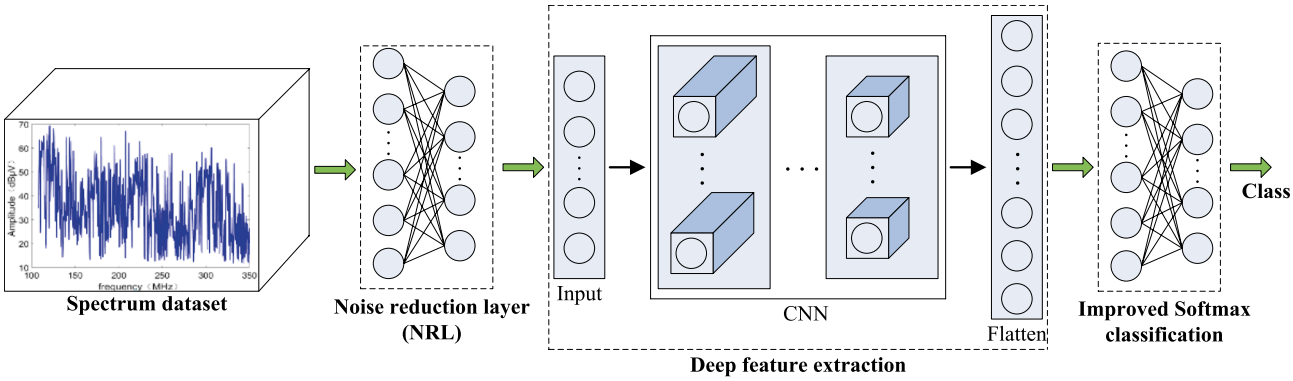

Fig. 1 Overall architecture of the proposed new network for EMIS Identification

has a great influence on the test data of the EMIS signal. The background noise data is a spectral vector of the electromagnetic environment in a frequency range. The Correlation operation can measure the similarity of two sequence signals. The correlation calculation result between EMIS test data and the background noise data of its test environment is greater than that with other background noise. Furthermore, the correlation calculation results of different EMIS test data and the same background noise can highlight the difference between interference source signals. So, we add a NRL in front of the convolutional layer and calculate the correlation between the EMIS spectrum data and the background noise data. In a sense, correlation calculation can reduce the influence of background noise on the EMIS signal, making the input data more reliable. The NRL also non-linearly normalizes the data after the correlation calculation through the activation function and uses the normalized data as the input data of the deep feature extraction sub-network. So, the NRL is a fully connected layer whose weight is the background noise data. Its structure is shown in Fig. 2.

In Fig. 2, X as the input data of NRL is the spectrum data sequence of the EMIS. The total length of the sequence is m, so $X \in R^{1\times m}$. W_j as the weight of the NRL, is the background noise data, and $W_j \in R^{1\times m}$. Vector I is the output sequence after noise reduction, and the length is $k \cdot m$. Because the *sigmoid* function can convert data into a number between 0 and 1, we use the *sigmoid* function as the activation function of NRL, which are:

$$h_{ji} = x_i \cdot W_j(i) \tag{1}$$

$$I(n) = Sigmoid\ (h(n)) \tag{2}$$

where x_i is the i-th data in the X vector, $0 \le i \le m$. $W_j(i)$ is the i-th data in the W_j vector,$0 \le j \le k$. All h_{ji} together make up vector h, $h(n)$ is the n-th data in the h, where $0 \le n \le k \cdot m$, and k represents the number of backgrounds. It can be seen that when the number of backgrounds increases, the dimension of $I(n)$ increases exponentially. Therefore, we hope to test the different EMIS in the same environment as much as possible. However, in actual application, the EMIS is generally located in the same environmental, so there is no need to think about this issue too much.

3.2 Deep Feature Extraction

The traditional CNN (T-CNN) is composed of several convolutional layers (Conv), ReLU activation functions, and pooling layers (Pooling). As shown in Fig. 3, we proposed a deep feature extraction sub-network based on one-dimensional CNN and an improved classification layer. The feature extraction sub-network consists of six convolutional layers, three pooling layers, one flattening layer,

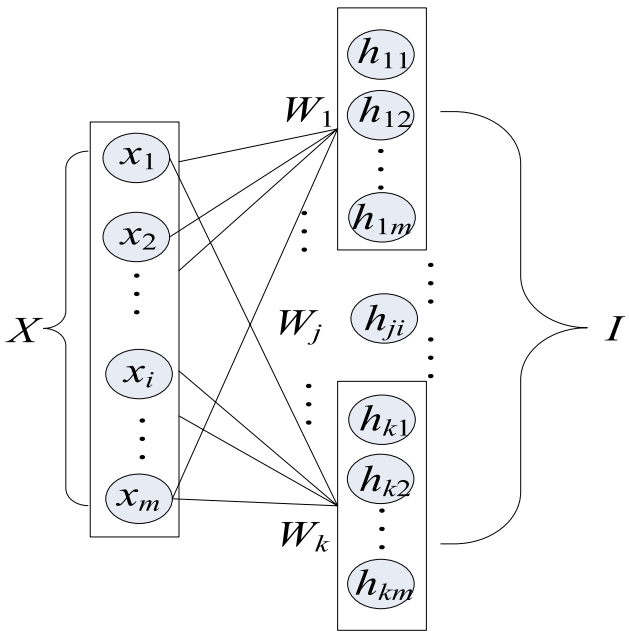

Fig. 2 NRL structure

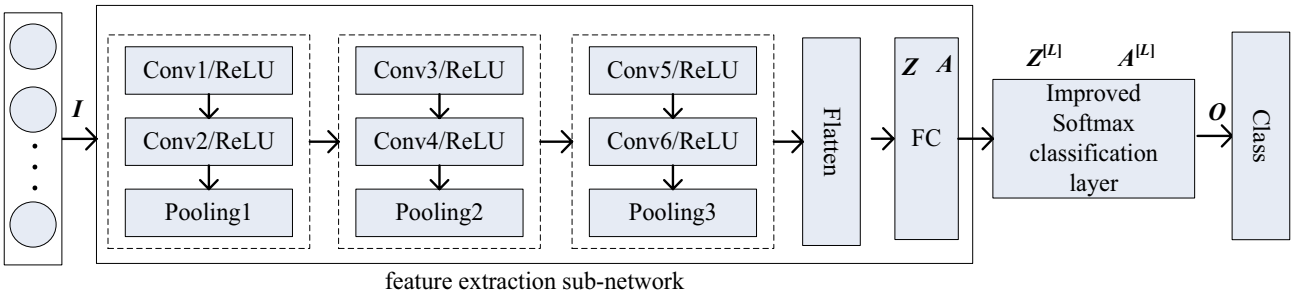

Fig. 3 The structure of feature extraction sub-network and classification layer

and one fully connected (FC) layer. The output features are used as the input to the classification layer.

In the convolutional layer, the convolution kernel uses a fixed kernel size, a fixed stride, and performs convolution operations with the input data. Different features can be obtained by using different convolution kernels. Suppose the dimension of the input sequence I is $(1, N_i)$, the convolution kernel is K, and its dimension is $(1, N_k)$, the result of the j-th convolution operation is 38:

$$C(j) = \sum_{n=1}^{N_k} I(j+n) \times K(n) \tag{3}$$

where, $1 \le j \le N_i - N_k + 1$. Assuming that the stride is s, the data length after the convolution operation is $\left[\left(N_i - N_k\right)/s\right] + 1$.

As with T-CNN, the ReLU function is used as an activation function in the convolutional layer. The input data is batch processed 5 before the convolution operation to speed up the training process, avoid excessive reliance on the initial value, and suppress over-fitting, especially for relatively small data sets.

The pooling layer down-samples the data and processes it sparsely to reduce the amount of computation. Commonly, the pooling methods include maximum pooling and average pooling. For the spectrum data of EMIS, the maximum value can better reflect the interference characteristics. We adopt the maximum pooling method to avoid the ambiguity problem of average pooling.

The feature vector after convolution has a larger dimension, which will affect the classification effect. The FC layer is used to refit the feature vector to reduce the feature dimension while retaining feature information as much as possible. In addition, to reduce over-fitting of training data, it can be decided whether to use Alpha Dropout 10 to regularize the FC layer according to the experimental situation. The FC layer includes linear operations and nonlinear operations. Assume that the FC layer is the q-th layer of the network, the input $A^{[q-1]}$ is the output of the q-1 layer, and the weight is $W^{[q]}$. The ReLU function as the activation function, then the output $A^{[q]}$ is,

$$A^{[q]} = \mathrm{ReLU}\left(A^{[q-1]} \cdot W^{[q]} + b^{[q]}\right) \tag{4}$$

$A^{[q]}$ is called a "deep feature." In order to evaluate the features more intuitively, we use t-distribution stochastic neighbor embedding (t-SNE) 27 to visualize them. The t-SNE uses two-dimensional feature maps to represent multidimensional features. Although the process of dimension reduction will lose information and have certain errors, two-dimensional visualization can more intuitively judge whether these deep features are discriminatory.

3.3 Improved Softmax Classification Layer

The classification layer of T-CNN is a FC layer based on the Softmax function, which is a multiple-class linear regression method that realizes multiple-type classification. If the Softmax classification layer is the l-th layer of the network, the input is $A^{[l-1]}$, the output is $A^{[l]}$, and the weight and bias are $W^{[l]}$ and $b^{[l]}$ respectively, then,

$$= W^{[l]} \cdot A^{[l-1]} + b^{[l]} \tag{5}$$

The activation process of the Softmax function is as follows:

Let $t_i = e^{Z_i^{[l]}}$, $1 \le i \le C$, where C is the number of types. The i-th data of $A^{[l]}$ is defined as 18,

$$A_i^{[l]} = \frac{t_i}{\sum_{i=1}^{C} t_i} \tag{6}$$

If O_i is the classification label at the output of the network, and $O_i \in \{0, 1\}$. $\hat{O}_i$ is the predicted value of O_i obtained by training the network. So $\hat{O} = A^{[l]}$, the maximum value in $A^{[l]}$ is highlighted, other values are significantly suppressed. In order to calculate the gradient of the error, the Softmax loss function is defined. That is

$$L_{Softmax} = -\sum_{k=1}^{C} O_k \log \widehat{O}_k = -logA_i = -log\frac{e^{(W_i \cdot A + b_i)}}{\sum_{k=1}^{C} e^{(W_k \cdot A + b_k)}} \tag{7}$$

where W_i and b_i represent the weight and bias of the FC layer corresponding to the i-th class. It can be seen that the output value of the FC layer determines the class to which the feature belongs. From Eq. (6), it is clear that $L_{Softmax}$ mainly increases the inter-class distance and does not take into account the intra-class distance.

Using the above Softmax loss function can solve many classification tasks. However, in order to make the classification of small samples faster and more accurate, the features of intra-class are required to be more compact, while the features of inter-class are more scattered. Therefore, the distribution of features needs to be optimized.

We have improved the Softmax classification layer, which is described as follows:

As the layer number increases, the activation distribution of the hidden layer gradually deviates. Standardizing the features can avoid the disappearance of the neuron node gradient during back propagation. We divide the feature vectors into multiple small batches of feature vectors, which are the input to the softmax layer. Let one small batch feature vector be $A_m = \{a_1, a_{2,} \cdots a_m\}$. After normalization, it is $\hat{A}_m = \{\hat{a}_1, \hat{a}_{2,} \cdots \hat{a}_m\}$. Then

$$\widehat{a}_i = \frac{a_i - \mu}{\sqrt{\sigma^2 + \xi}} \tag{8}$$

where ξ is a very small constant, used to avoid a zero denominator. μ and σ^2 are,

$$\mu = \frac{1}{m}\sum_{i=1}^{m} a_i \tag{9}$$

$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m} \left(a_i - \mu\right)^2 \tag{10}$$

The calculation of the FC layer is shown in Eqs. (5) and (6).

Calculate the class center value,c_{y_i}, of each class in each batch, then calculate the distance between each sample and the class center. The average intra-class distance is,

$$D_{intra-class} = \frac{1}{M}\sum_{i=1}^{M} \|a_i - c_{y_i}\|^2 \tag{11}$$

where $||$ means two-norm operator, c_{y_i} represents the center value of the y_i class, and M is the number of samples in the batch.

Calculate the center value of all samples in each batch, the average inter-class distance is,

$$D_{inter-class} = \frac{1}{C}\sum_{y_i=1}^{C} \|c_{y_i} - c_{all}\|^2 \tag{12}$$

where C is the number of classes.

The relative distance between the intra-class and the inter-class is defined as

$$D_{\text{re}} = \frac{D_{intra-class}}{D_{inter-class}} = \frac{\sum_{i=1}^{M} \|a_i - c_{y_i}\|^2}{MC\sum_{y_i=1}^{C} \|c_{y_i} - c_{all}\|^2 + \eta} \tag{13}$$

where η is a very small constant, used to avoid a zero denominator, and set to 1e-6. It can be seen that when $D_{\text{re}} << 1$ and close to 0, the intra-class distance is much smaller than the inter-class distance, and a better feature distribution can be obtained. We define the joint loss function as

$$J = L_{Softmax} + \lambda D_{re} \tag{14}$$

where λ is a constant used to balance the two loss functions. We set various values for the parameter λ when training the network. Choose the value of λ that gives the best network performance. After many trials, its value is finally chosen to be 1 in this article. We call J the improved Softmax (Im-Softmax) loss function, apply it to the classification layer. Through continuous training of network parameters to minimize the J.

3.4 Back Propagation

The back propagation of the network calculates the derivative of the loss function with each parameter. The optimization algorithm is used to update the parameters until the function converges and minimizes the loss value.

The derivative of D_{re} with respect to a_i is,

$$\frac{\partial D_{re}}{\partial a_i} = \frac{2\sum_{i=1}^{M}\left(a_i - c_{y_i}\right)}{MC\sum_{y_i}^{C}\|c_{y_i} - c_{all}\|^2 + \eta} \tag{15}$$

The derivative of D_{re} with respect to c_{y_i} is,

$$\frac{\partial D_{re}}{\partial c_{y_i}} = \frac{2\sum_{i=1}^{M}\left(c_{y_i} - a_i\right)}{MC\sum_{i=1}^{C}\|c_{y_i} - c_{all}\|^2 + \eta} - \frac{2\sum_{i=1}^{M}\|a_i - c_{y_i}\|^2 \cdot \sum_{i=1}^{C}\left(c_{y_i} - c_{all}\right)}{MC\cdot\left[\sum_{i=1}^{C}\|c_{y_i} - c_{all}\|^2 + \eta\right]^2} \tag{16}$$

The derivative of D_{re} with respect to c_{all} is,

$$\frac{\partial D_{re}}{\partial c_{all}} = \frac{2MC\sum_{i=1}^{C}\left(c_{all} - c_{y_i}\right)\cdot\sum_{i=1}^{M}\|a_i - c_{y_i}\|^2}{\left[MC\sum_{i=1}^{C}\|c_{y_i} - c_{all}\|^2 + \eta\right]^2} \tag{17}$$

The gradient of the convolutional layer is different. The forward propagation of the l-1 convolutional layer can be expressed as

$$\begin{cases} z^{[l-1]} = w^{[l-1]} \cdot a^{[l-2]} + b^{[l-1]} \\ a^{[l-1]} = f\left(z^{[l-1]}\right) = ReLU\left(z^{[l-1]}\right) \end{cases} \tag{18}$$

Let the error of the l-th layer is $\delta^{[l]}$, the error of the l-1 layer is,

$$\begin{aligned} \delta^{[l-1]} &= \frac{\partial J}{\partial z^{[l-1]}} = \frac{\partial J}{\partial z^{[l]}} \cdot \frac{\partial z^{[l]}}{\partial a^{[l-1]}} \cdot \frac{\partial a^{[l-1]}}{\partial z^{[l-1]}} \\ &= \delta^{[l]} \cdot \frac{\partial z^{[l]}}{\partial a^{[l-1]}} \cdot f'\left(z^{[l-1]}\right) \end{aligned} \tag{19}$$

where $\frac{\partial z^{[l]}}{\partial a^{[l-1]}}$ means that the convolution kernel is flipped 180° left and right. Then the gradient of w and b in the convolutional layer are respectively

$$\frac{\partial J}{\partial w^{[l]}} = a^{[l-1]} * \delta^{[l]} \tag{20}$$

$$\frac{\partial J}{\partial b^{[l]}} = \delta^{[l]} \tag{21}$$

The Adam 11 is used to update the parameters. If t is epoch and the initial value is 0, then $\partial J_t\left(W, b, c_{y_i}, c_{all}\right)$ is the gradient of the loss function in the t-th epoch.

The first moments before and after the deviation correction of the l-th layer at the t-th epoch are

$$v_t^{[l]} = \beta_1 v_{t-1}^{[l]} + \left(1 - \beta_1\right) \cdot \partial J_t\left(W, b, c_{y_i}, c_{all}\right) \tag{22}$$

$$v_t^{[l]} = \frac{v_t^{[l]}}{1 - \left(\beta_1\right)^t} \tag{23}$$

The second moments before and after the deviation correction of the l-th layer at the t-th epoch are

$$s_t^{[l]} = \beta_2 s_{t-1}^{[l]} + \left(1 - \beta_2\right) \cdot \left[\partial J_t\left(W, b, c_{y_i}, c_{all}\right)\right]^2 \tag{24}$$

$$s'^{[l]}_t = \frac{s_t^{[l]}}{1 - \left(\beta_2\right)^t} \tag{25}$$

Updated parameters are,

$$\begin{cases} W_t^{[l]} = W_{t-1}^{[l]} - \alpha \frac{v'^{[l]}_t}{\sqrt{s'^{[l]}_t} + \varepsilon} \\ b_t^{[l]} = b_{t-1}^{[l]} - \alpha \frac{v'^{[l]}_t}{\sqrt{s'^{[l]}_t} + \varepsilon} \\ c_{y_i t}^{[l]} = c_{y_i t-1}^{[l]} - \alpha \frac{v'^{[l]}_t}{\sqrt{s'^{[l]}_t} + \varepsilon} \\ c_{all t}^{[l]} = c_{all t-1}^{[l]} - \alpha \frac{v'^{[l]}_t}{\sqrt{s'^{[l]}_t} + \varepsilon} \end{cases} \tag{26}$$

where α is learning rate and set to 0.01. The parameters β_1 and β_2 are usually 0.9 and 0.999, respectively. ε is a very small constant to avoid the value of the denominator is be 0, generally set to 1e-8.

The new network proposed in this paper uses NRL and the Im-Softmax function, which can be called the NRL-CNN + Im-Softmax network. The weights of the FC layer need to be processed to reduce the network training time. The last convolutional layer gives multiple uniformly distributed class centers, that is, uniformly distributed sample points. We replace the weights of the last convolutional layer with the class center weights. When training the network, only the weights of the previous layer need to be updated to shorten the training time. At the end of training, in order to get a better classification, the weight of the last FC layer can be selectively fine-tuned according to different data sets.

4 Experiment and Result Analysis

We obtained spectral data of interference sources through laboratory experiments. Based on the actual spectrum data of the six EMIS and background noise in two environments, the NRL-CNN + Im-Softmax is verified on different data sets. The experiments are based on TensorFlow and are completed on the CPU.

4.1 Measurement Data

Taking the electromagnetic environment of one Chinese airport as an example, the working frequency of the communication and navigation equipment is mainly concentrated in the 108 MHz–2000 MHz range. We collected 3 types of aging power supplies and 2 types of signal jammers as EMIS. At the same time, the articulated neutral section (ANS) arc of the electrified railway is another type of EMIS. We tested spectrum data of their electromagnetic emissions (EM) in the laboratory. We marked the background noise of the laboratory as class 0, the ANS arc, the signal jammer-PBQ2X, the signal jammer-PBQ6X, the aging power supply for civil lighting (APS for CL), the aging power supply for airport runway lights (APS for ARL), and the aging power supply for airport lighting (APS for AL) as classes 1, 2, 3, 4, 5, 6, respectively.

The method and equipment for collecting spectrum data for EMIS meet the requirements of EN50121-2 20, and the test distance is 10 m. The EMI spectrum analyzer (N9010B EXA), cables and corresponding receiving antennas are used to test the EM. Figure 4 is the schematic diagram of the test.

In order to verify the method in this article, three data sets were collected, and experiments were performed on different data sets.

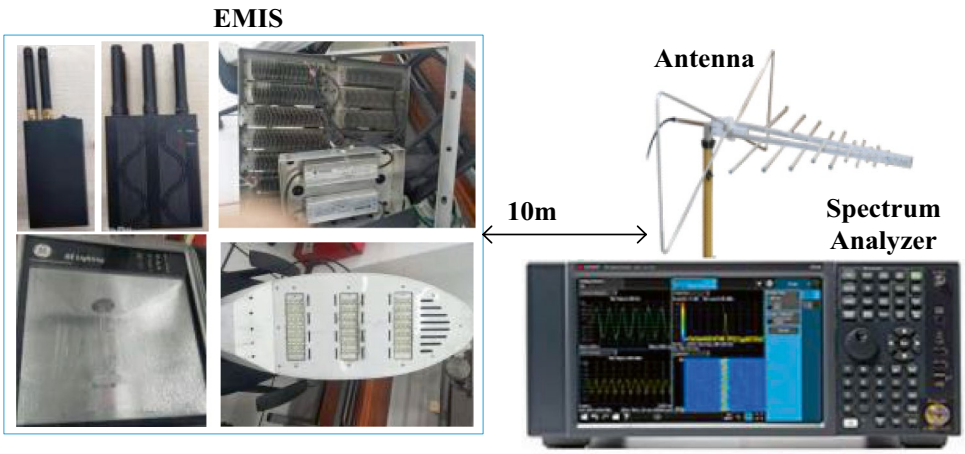

Fig. 4 Schematic diagram of the test

Dataset 1 contains background noise as well as the EM of each EMIS in six frequency bands (108–148 MHz, 108–350 MHz, 310–350 MHz, 962–2000 MHz, 1010–1050 MHz, and 1070–1110 MHz).The test was repeated 20 times for each frequency band, and a total of 960 samples were obtained. Each class has 120 samples. There are 96 samples in the training set and 24 samples in the test set, and each sample has 1001 data points.

Dataset 2 contains the background noise and the EM of each EMIS in the three frequency bands of 108–148 MHz, 108–350 MHz, and 310–350 MHz. A total of 480 samples,

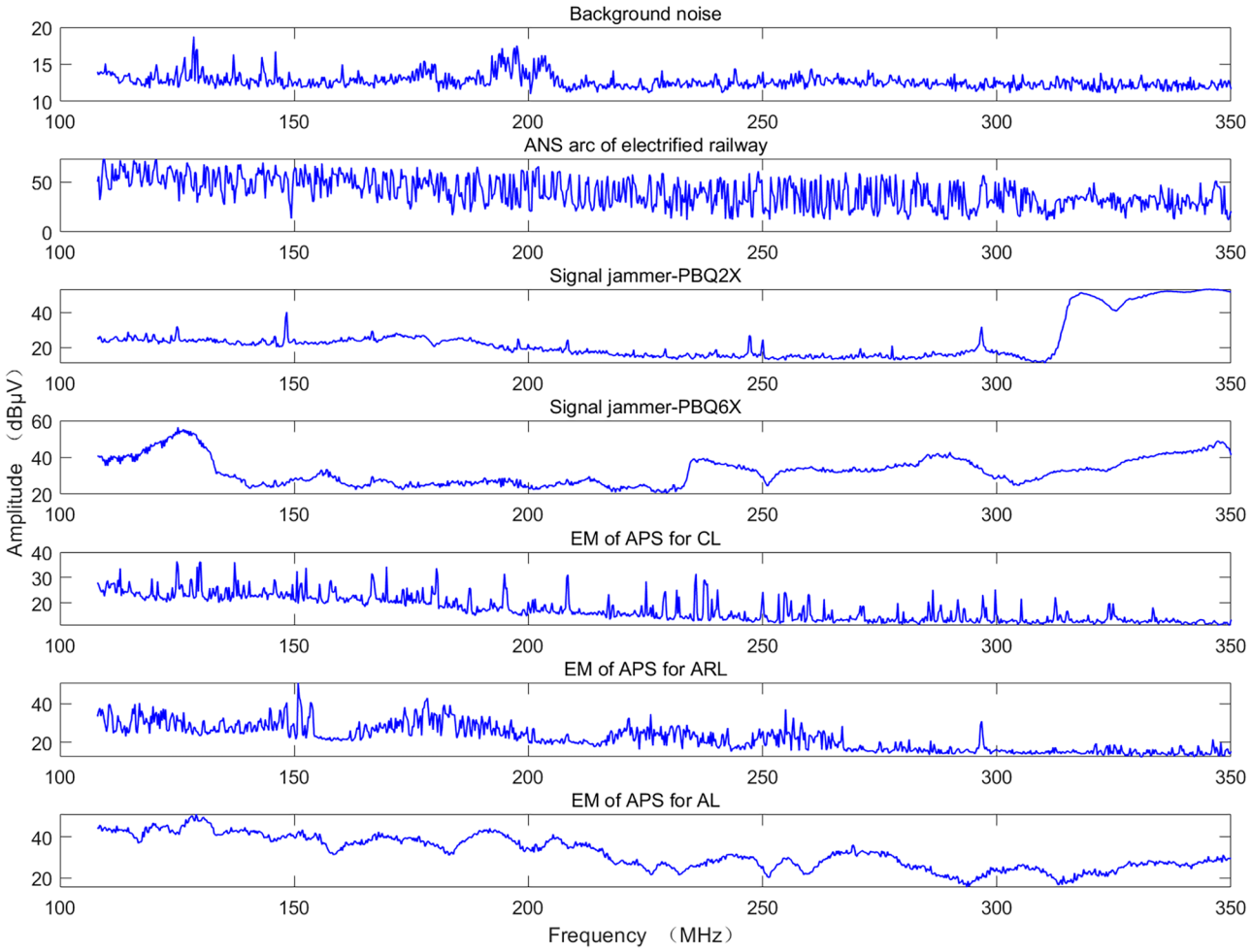

Fig. 5 One sample of each class in the 108–350 MHz range

60 in each class. There are 48 samples in the training set and 12 samples in the test set.

Dataset 3 contains the background noise and the EM of each EMIS in the 108–350 MHz range. A total of 160 samples, 20 in each class. There are 16 in the training set and 4 samples in the test set.

For example, one sample of each class in the 108–350 MHz range is shown in Fig. 5. The signal amplitudes of various EMIS are greater than the background noise, and some classes are similar in some frequencies. It should be noted that the test background of the ANS arc signal is different from the test background of other EMIS.

4.2 Network Parameters

The NRL-CNN + Im-Softmax network in this article is composed of a NRL, 6 convolutional layers, 3 maximum pooling layers, 1 flattening layer, 1 FC layer, 1 batch normalization layer, and an Im-Softmax classification layer. The network parameters are shown in Table 1.

Table 1 The network parameters

Layer	Related parameters	Output Shape	Parameters#
NRL	1001	(2002, 1)	0
Conv1	filters = 16 size = 3 strides = 1	(2000, 16)	64
Conv2	filters = 16 size = 3 strides = 1	(1998, 16)	784
Max_Pooling1	filters = 16 size = 3 strides = 3	(666, 16)	0
Conv3	filters = 64 size = 3 strides = 1	(664, 64)	3136
Conv4	filters = 64 size = 3 strides = 1	(662, 64)	12,352
Max_Pooling2	filters = 64 size = 3 strides = 3	(220, 64)	0
Conv5	filters = 64 size = 3 strides = 1	(218, 64)	12352
Conv6	filters = 64 size = 3 strides = 1	(216, 64)	12352
Max_Pooling3	filters = 64 size = 3 strides = 3	(72, 64)	0
Flatten	None	4608	0
FC	100	100	460900
Batch Normalization	batch size = 64	100	400
Im-Softmax	7	7	707

During the experiment, it is found that the dropout regularization method has no effect on the research content of this article. The Batch Normalization layer is introduced to solve the problem of the disappearance of the gradient of the neuron node during back propagation. The Adam optimization method is used and β_1=0.9, β_2=0.999. The parameters that need to be initialized in the network use the Xavier method, and the learning rate is 0.001. The number of neurons in the FC layer is set to 100, and the number of neurons in the Im-Softmax classification layer is set to 7. The epochs are set to 100 when training the network. There are a total of 503,047 parameters, of which 502,847 parameters need to be trained.

4.3 Experimental Results

We trained and tested the T-CNN, the NRL-CNN, and the NRL-CNN + Im-Softmax on three data sets. The loss function of T-CNN and NRL-CNN is Softmax, and the NRL-CNN + Im-Softmax network uses the Im-Softmax loss function. This section compares and analyzes the loss function, feature distribution, network training time, and accuracy obtained by different methods on different data sets.

The training loss function and test loss function of the three methods on the three data sets are shown in Fig. 6. It can be seen that the loss functions of the three methods have large fluctuations on Dataset1. This is because Dataset1 is composed of data in multiple different frequency bands, and the emission signal at the high frequency of the EMIS is small and similar to the background noise. The test loss of the T-CNN is larger when it reaches a stable value, and reduced after adding NRL, which is reduced by 54.85%, 90.77% and 90.55% on the three data sets, but requires more epochs. The test loss of the NRL-CNN + Im-Softmax is close to 0, which has fewer epochs compared with the NRL-CNN. That is, the new network proposed in this paper can update the network parameters in the shortest time and make the loss function reach the desired minimum in the shortest time.

The training accuracy and training time of the three networks on the three data sets are shown in Table 2. The training accuracy reached 100%, indicating that our data set is reliable. T-CNN has the least training time, and NRL-CNN has the most training time. The training time of the NRL-CNN + Im-Softmax and T-CNN are almost the same. This proves that the method proposed in this paper can obtain a stable and fast convergence loss function, thereby improving the accuracy.

Use the t-SNE algorithm to visualize the features obtained by the three networks. The two-dimensional

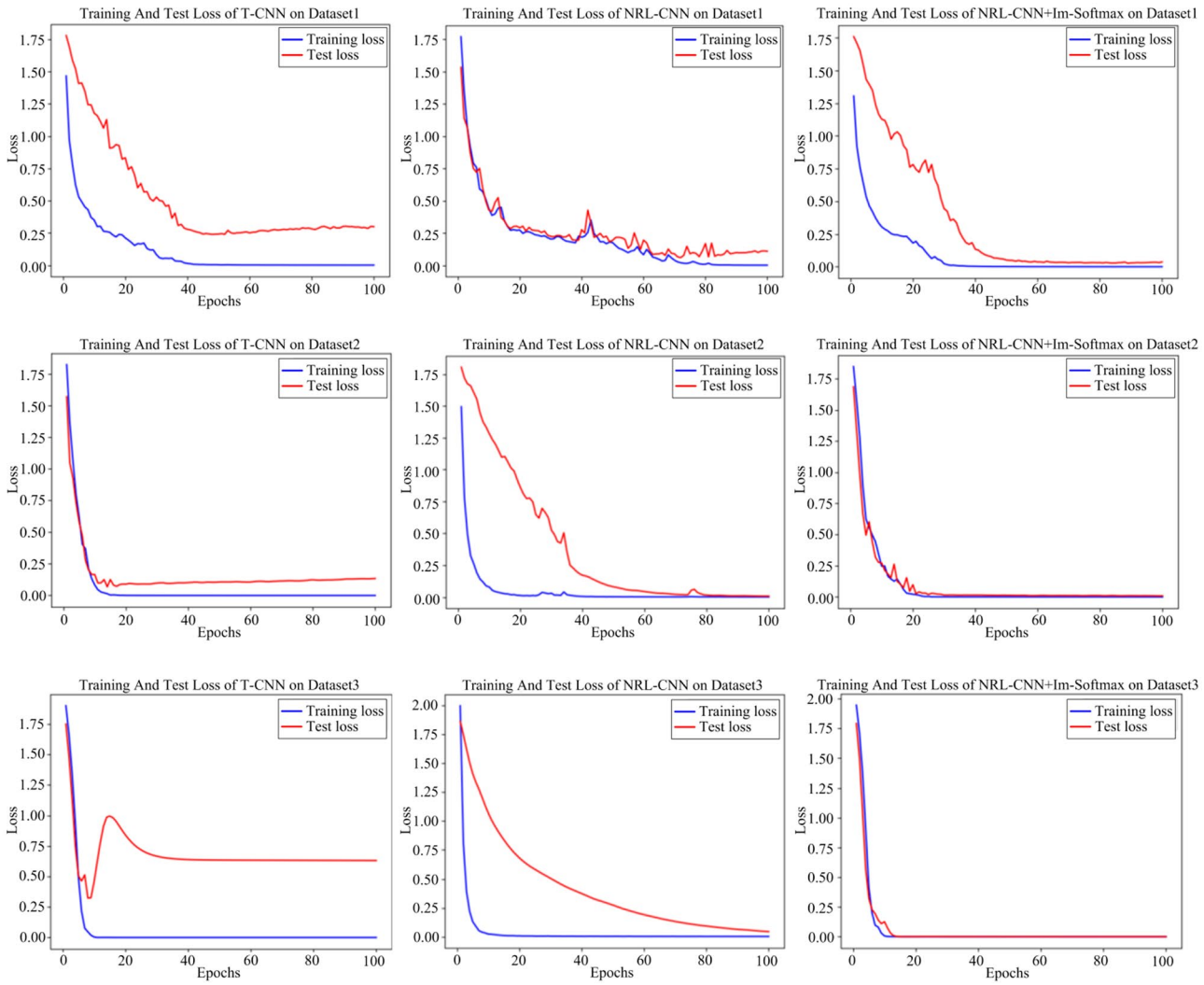

Fig. 6 The training loss and test loss of the three methods on the three data sets

feature distribution of the three methods on the three data sets is shown in Fig. 7. It can be seen that the ANS arc features extracted by T-CNN are quite different from other EMIS in the feature distribution. Only individual samples can be seen on Dataset1 and Dataset2. This is because the test background of the ANS arc is different from other EMIS, and the T-CNN method does not take into account the different background noises. After adding NRL, the relative distance between inter-class and intra-class was increased, and the feature distribution improved. After adding NRL, the relative distance between inter-class and intra-class was increased, and the feature distribution improved. On the same dataset, it can be seen that when the NRL-CNN classification layer only uses the Softmax loss function, the distance between intra-classes is not reduced, and there is no obvious difference in the relative distance between intra-class and inter-class. When the Im-Softmax loss function is used in NRL-CNN + Im-Softmax,

Table 2 Training accuracy and training time

Networks	Training Accuracy			Training Time (s)		
	Dataset1	Dataset2	Dataset3	Dataset1	Dataset2	Dataset3
T-CNN	100%	100%	100%	56.488	29.102	10.633
NRL-CNN	100%	100%	100%	103.766	52.450	18.053
NRL-CNN + Im-Softmax	100%	100%	100%	57.138	30.051	11.146

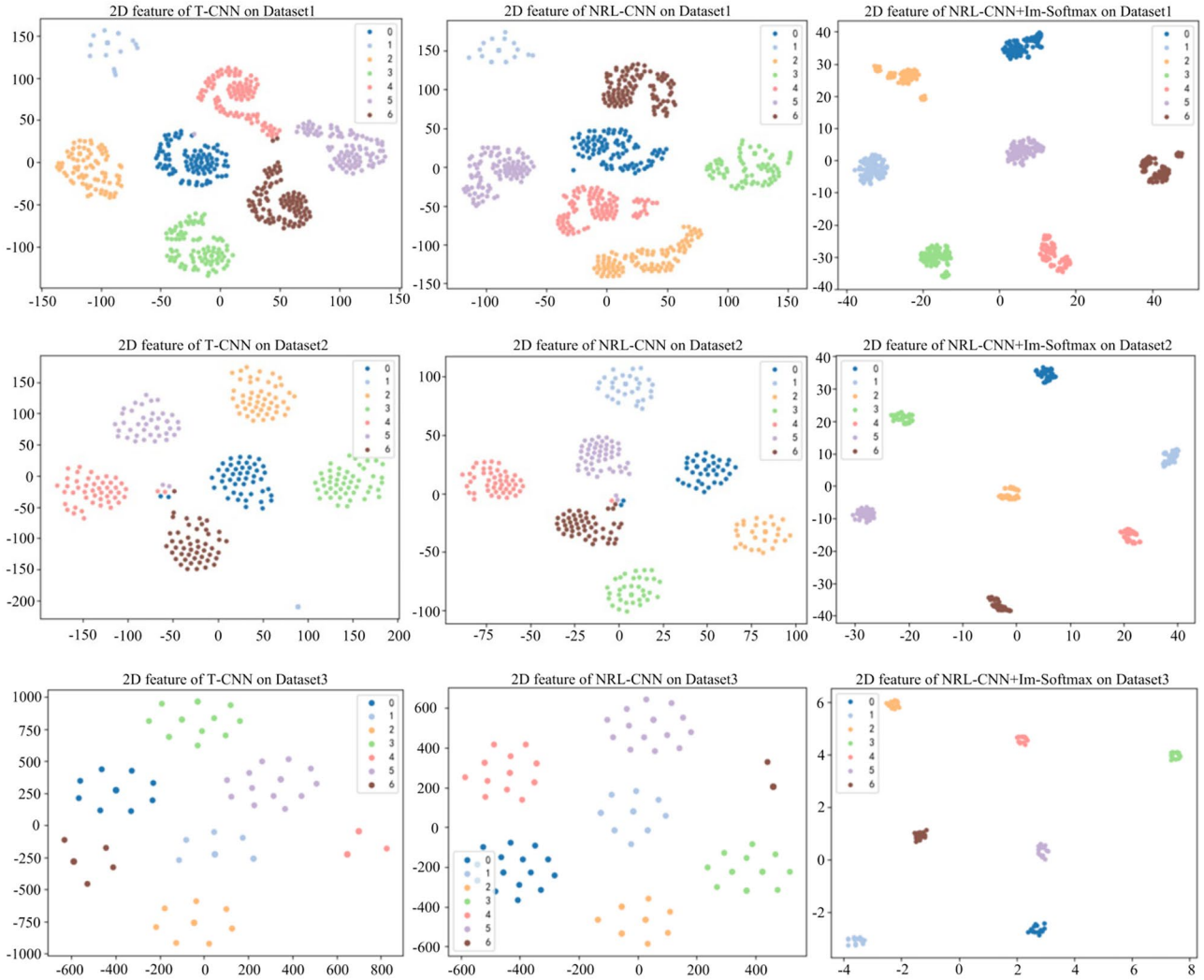

Fig. 7 The feature distribution of three methods on the three data sets

the distance between the sample and the center of the class is reduced. The average distance within the intra-class is reduced by 0.9 times, and the relative distance between the inter-class and intra-class is very obvious. That is, the improved classification layer can obtain better feature distributions on three data sets of different sizes, which shows that the new network proposed in this paper can also extract the features of small samples well.

Test the three trained networks based on the test set, and get the test accuracy of each network, as shown in Table 3. The confusion matrix of the three methods on the three test sets is shown in Fig. 8.

From Fig. 8, when T-CNN is used on Dataset1, the classes are easily confused and difficult to identify. The NRL-CNN + Im-Softmax can identify confusing classes, but only the class 0 and class 5 are misidentified with each other. On Dataset2 and Dataset3, T-CNN and NRL-CNN are easy to confuse class 0, class 1 and class 5. NRL-CNN + Im-Softmax can identify them well.

From Table 3, T-CNN has the lowest test accuracy, mainly because the EM signal of some EMIS in the high

Table 3 Test accuracy

Networks	Test Accuracy		
	Dataset1	Dataset2	Dataset3
T-CNN	91.67%	97.62%	96.43%
NRL-CNN	97.02%	98.81%	96.43%
NRL-CNN + Im-Softmax	98.21%	100%	100%

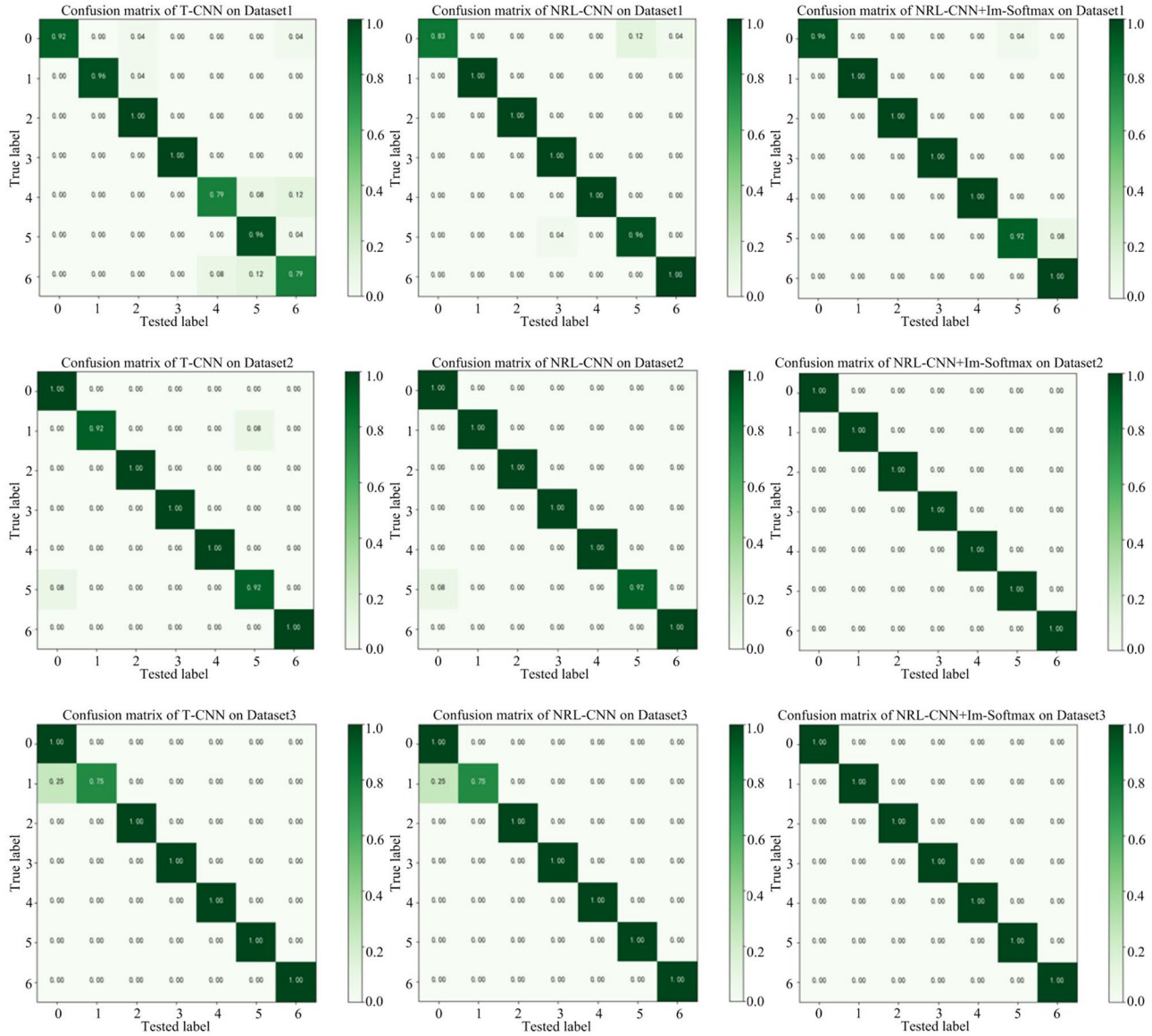

Fig. 8 The Confusion matrix of three methods on the three data sets

frequency range is weak, similar to background noise, easy to confuse, and deep features are not easy to extract. After adding NRL, the accuracy can be improved on both Dataset1 and Dataset2, but it has no effect on the smallest Dataset3. After using the Im-Softmax loss function, the deep features of various EMIS signals can be obtained, thereby improving the test accuracy. It can be seen that the method proposed in this article also has good classification ability on small sample data sets.

5 Conclusion and Future Directions

We proposed a new NRL-CNN + Im-Softmax network based on the traditional CNN to identify the EMIS. The frequency spectrum of the EMIS is used as data sets. The comprehensive performance of NRL-CNN + Im-Softmax is verified through experiments of 3 networks on 3 data sets. The results show that calculating the correlation between the test data of each EMIS and its background noise can reduce noise, improve input data, thereby reducing the test loss. The test loss on the three data sets has dropped by 78.72% on average. The NRL-CNN + Im-Softmax can extract the deep features of EMIS signals, shorten the network training time, and the test accuracy can even reach 100%. The NRL-CNN + Im-Softmax is universal, not limited by the size of the data set, and can get good identification results on small sample data set. The NRL-CNN + Im-Softmax has good comprehensive performance and provides a reliable technical means for EMIS identification. In addition, the method in this paper

can only identify a single EMIS and know the number of EMIS in advance. In future work, we will collect more EMIS information to study the identification methods that can identify multiple EMIS at the same time [3–5, 8–10, 15–19, 24, 26, 28, 35, 37].

Acknowledgements This work is supported by National Key R&D Program of China (No. 2018YFC0809500).

Data Availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of Interest I certify that there is no actual or potential conflicts of interest in relation to this article.

References

1. Antonini G, Orlandi A (2001) Wavelet packet-based EMI signal processing and source identification. IEEE Trans Electromagn Compat 43(2):140–148. https://doi.org/10.1109/15.925533
2. Bu K, He Y, Jing X, Han J (2020) Adversarial Transfer Learning for Deep Learning Based Automatic Modulation Classification. IEEE Signal Process Lett 27:880–884. https://doi.org/10.1109/LSP.2020.2991875
3. Elsaadouny M, Barowski J, Rolfes I (2020) "Extracting the Features of the Shallowly Buried Objects using LeNet Convolutional Network," in Proceedings of 14th European Conference on Antennas and Propagation (EuCAP), Copenhagen, Denmark. 1–4. https://doi.org/10.23919/EuCAP.48036.2020.9135701
4. Gao R, Li Z, Li HS, Ai B (2015) A Spectrum Sensing Method via Absolute Value Cumulating with Laplacian Noise. Wireless Pers Commun 83(2):1387–1404. https://doi.org/10.1007/s11277-015-2457-4
5. Goodfellow I, Bengio Y, Courville A, Bach F (2016) Deep Learning. MIT Press, Cambridge, MA
6. Hong Z et al (2019) Electromagnetic Pattern Extraction and Grouping for Near-Field Scanning of Integrated Circuits by PCA and K-Means Approaches. IEEE Trans Electromagn Compat 61(6):1811–1822. https://doi.org/10.1109/TEMC.2018.2890026
7. Jiao Y, Han J, Xu B, Xiao M, Shen B, Sun H (2021) "Research on Domain Entity Extraction in Civil Aviation Safety," in Proceedings IEEE 3rd International Conference on Civil Aviation Safety and Information Technology (ICCASIT). 384–388. https://doi.org/10.1109/ICCASIT53235.2021.9633439
8. Kingma D, Ba J (2014) Adam: A method for stochastic optimization. Computer Science 12–22. http://arxiv.org/abs/1412.6980
9. Klambauer G, Unterthiner T, Mayr A, Hochreiter S (2017) Self normalizing neural networks. Adv Neural Inform Proc Sys 30:972–981. http://arxiv.org/abs/1706.02515
10. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://doi.org/10.1145/3065386
11. Kulin M, Kazaz T, Moerman I, De Poorter E (2018) End-to-End Learning From Spectrum Data: A Deep Learning Approach for Wireless Signal Identification in Spectrum Monitoring Applications. IEEE Access 6:18484–18501. https://doi.org/10.1109/ACCESS.2018.2818794
12. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86(11):2278–2324. https://doi.org/10.1109/5.726791
13. Li D, Yang R, Li X, Zhu S (2020) Radar Signal Modulation Recognition Based on Deep Joint Learning. IEEE Access 8:48515–48528. https://doi.org/10.1109/ACCESS.2020.2978875
14. Li H, Wang C, Zhao D (2015) An improved EMD and its applications to find the basis functions of EMI signals. Mathematical Problems in Eng 2015(150127). https://doi.org/10.1155/2015/150127
15. Luo Y, Wong Y, Kankanhalli M, Zhao Q (2020) G-Softmax: Improving Intraclass Compactness and Interclass Separability of Features. IEEE Transactions on Neural Networks and Learning Systems 31(2):685–699. https://doi.org/10.1109/TNNLS.2019.2909737
16. Nikdel S, Noh S, Shortle J (2021) "Common Cause Failure Analysis for Aviation Safety Assessment Models," in Proceedings IEEE/AIAA 40th Digital Avionics Systems Conference (DASC). 1–10. https://doi.org/10.1109/DASC52595.2021.9594402
17. O'Shea TJ, Roy T, Clancy TC (2018) Over-the-Air Deep Learning Based Radio Signal Classification. IEEE Journal of Selected Topics in Signal Processing 12(1):168–179. https://doi.org/10.1109/JSTSP.2018.2797022
18. Railway applications – Electromagnetic compatibility – Part 2: Emission of the whole railway system to the outside world, EN50121–2
19. Santamaría-Vázquez E, Martínez-Cagigal V, Vaquerizo-Villar F, Hornero R (2020) EEG-Inception: A Novel Deep Convolutional Neural Network for Assistive ERP-Basraed Bin-Computer Interfaces. IEEE Trans Neural Syst Rehabil Eng 28(12):2773–2782. https://doi.org/10.1109/TNSRE.2020.3048106
20. Sevgi L (2003) Complex electromagnetic problems and numerical simulation approaches. Wiley, Piscataway, New Jersey
21. Sheeny M, Wallace A, Wang S (2020) 300 GHz radar object recognition based on deep neural networks and transfer learning. IET Radar Sonar Navig 14(10):1483–1493. https://doi.org/10.1049/iet-rsn.2019.0601
22. Shi D, Gao Y (2013) A method of identifying electromagnetic radiation sources by using support vector machines. China Communications 10(7):36–43. https://doi.org/10.1109/CC.2013.6570798
23. Shieh CS, Lin CT (2002) A vector neural network for emitter identification. IEEE Trans Antennas Propag 50(8):1120–1127. https://doi.org/10.1109/TAP.2002.801387
24. Sun S, Zhang T, Li Q, Wang J, Zhang W, Wen Z, Tang Y (2021) Fault Diagnosis of Conventional Circuit Breaker Contact System Based on Time-Frequency Analysis and Improved AlexNet. IEEE Trans Instrum Meas 70:1–12. https://doi.org/10.1109/TIM.2020.3045798
25. Tian Y, Tatematsu A, Tanabe K, Miyajima K (2014) Development of Locating System of Pulsed Electromagnetic Interference Source Based on Advanced TDOA Estimation Method. IEEE Trans Electr Compatibility 56(6):1326–1334. https://doi.org/10.1109/TEMC.2014.2315438
26. Van Der Maaten L, Hinton G (2008) Visualizing data using t-SNE. J Mach Learn Res 86:2579–2605
27. Wang F, Gong W, Liu J, Wu K (2020) Channel Selective Activity Recognition with WiFi: A Deep Learning Approach Exploring Wideband Information. IEEE Transactions on Network Science and Engineering 7(1):181–192. https://doi.org/10.1109/TNSE.2018.2825144
28. West NE, O'Shea T (2017) "Deep architectures for modulation recognition," in Proceedings of 2017 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN), Baltimore, MD, USA. 1–6. https://doi.org/10.1109/DySPAN.2017.7920754

29. Wickramasinghe CS, Marino DL, Manic M (2021) ResNet Autoencoders for Unsupervised Feature Learning From High-Dimensional Data: Deep Models Resistant to Performance Degradation. IEEE Access 9:40511–40520. https://doi.org/10.1109/ACCESS.2021.3064819
30. Willson GB (1990) "Radar classification using a neural network," in Proceedings of Applications of Artificial Neural Networks. Bellingham, WA: SPIE. 200–210
31. Xiao C, Zhao T (2017) Identification Method of EMI Sources Based on Measured Single-Channel Signal and its Application in Aviation Secondary Power Source Design. IEEE Trans Electromagn Compat 59(2):439–446. https://doi.org/10.1109/TEMC.2016.2612704
32. Yang J, Yang JY (2003) Why can LDA be performed in PCA transformed space? Pattern Recognition. 36(2):563–566. https://doi.org/10.1016/S0031-3203(02)00048-1
33. Ye X, Zhu Q (2019) Class-Incremental Learning Based on Feature Extraction of CNN With Optimized Softmax and One-Class Classifiers. IEEE Access 7:42024–42031. https://doi.org/10.1109/ACCESS.2019.2904614
34. Zhang F, Hu C, Yin Q, Li W, Li H, Hong W (2017) Multi-Aspect-Aware Bidirectional LSTM Networks for Synthetic Aperture Radar Target Recognition. IEEE Access 5:26880–26891. https://doi.org/10.1109/ACCESS.2017.2773363
35. Zhou Y, Chang H, Lu Y, Lu X, Zhou R (2021) Improving the Performance of VGG Through Different Granularity Feature Combinations. IEEE Access 9:26208–26220. https://doi.org/10.1109/ACCESS.2020.3031908
36. Zhao M, Zhong S, Fu X, Tang B, Pecht M (2020) Deep Residual Shrinkage Networks for Fault Diagnosis. IEEE Trans Industr Inf 16(7):4681–4690. https://doi.org/10.1109/TII.2019.2943898
37. Zhao X, Wen Z, Pan X, Ye W, Bermak A (2019) Mixture Gases Classification Based on Multi-Label One-Dimensional Deep Convolutional Neural Network. IEEE Access 7:12630–12637. https://doi.org/10.1109/ACCESS.2019.2892754

Yingchun Xiao was born in Gansu Province, China, in 1990. She received the B.S. degree in electronic information science and technology from Lanzhou University of Technology, Lanzhou, China, in 2012, and is currently working toward the Ph.D. degree in electrical engineering at Southwest Jiaotong University, Chengdu, China. At the same time she is a lecturer at Lanzhou City College. Her research interests include electromagnetic environment test and evaluation, electromagnetic compatibility analysis and design, and identification and location of electromagnetic interference sources.

Feng Zhu received the Ph.D. degree in railway traction electrification and automation from Southwest Jiaotong University, Sichuan, China, in 1997. He is currently a Full Professor in the School of Electrical Engineering, Southwest Jiaotong University. His current research interests include locomotive over-voltage and grounding technology, electromagnetic theory and numerical analysis of electromagnetic field, and electromagnetic compatibility analysis and design.

Shengxian Zhuang is currently working with the School of Electrical Engineering at Southwest Jiaotong University as a professor. He got his M.S. and Ph.D. degrees, respectively, at Southwest Jiaotong University and the University of Electronic Science and Technology of China in 1991 and 1999. From 1999 to 2003, he did postdoctoral research at Zhejiang University and Linkoping University of Sweden. He was a visiting professor at Paderborn University in Germany in 2005 and at the University of Leeds, UK in 2017. His research interests include power conversion for sustainable energies, motor control and drive systems, power electronics and systems integration, modeling, diagnosis, and suppression of electromagnetic interference of power electronic converters.

Yang Yang was born in Shanxi, China on April 19, 1989. She received her bachelors degree in measurement and control technology and instrumentation from Shaanxi University of Science and Technology in 2011 and her masters degree in control theory and control engineering from Northwestern Polytechnical University in 2014. She is currently working toward a Ph.D. degree in electrical engineering at Southwest Jiaotong University, Chengdu, China. Her research interests include electromagnetic environment testing and evaluation, electromagnetic compatibility analysis and design, and electromagnetic compatibility problems in the field of railway power supply and rail transit.