A New Convolutional Network Structure for Power Quality Disturbance Identification and Classification in Micro-Grids

RENXI GONG AND **TAOYU RUAN**

School of Electrical Engineering, Guangxi University, Nanning 530004, China

Corresponding author: Renxi Gong (rxgong@gxu.edu.cn)

ABSTRACT Aiming at the problems of a low convergence speed, low accuracy and poor generalization ability of traditional power disturbance identification and classification methods, a new deep convolutional network structure is presented, and a power quality disturbance identification and classification method for microgrids based on the new network structure is proposed. The network consists of a five-layer one-dimensional modified Inception-residual network (ResNet) (1D-MIR) and a three-layer full-connection tier, which is a deep convolutional network. The idea of the method can be described as follows: First, power disturbance signals are expressed by an n-dimensional unit vector, and a database of these power disturbance signals is established. Second, the disturbance signals in the database are randomly sampled, and the power quality disturbances are calibrated with the n-dimensional unit vector to form both data and test sets. Finally, the gradient descent method and the adaptive moment estimation method (Adam) are adopted to train and optimize the network, respectively, and the trained and optimized network is applied to power quality disturbance identification and classification. A large number of experiments has been conducted, and the obtained results show that the constructed network can quickly extract the characteristics of the various disturbance signals, including single and composite disturbances, and identify and classify them. A comparison of the results obtained by the proposed method with those obtained by several other methods reveals that the proposed method attains a higher accuracy, higher convergence speed and stronger generalization ability.

INDEX TERMS New network structure, power quality disturbance detection, deep convolution neural network, deep learning.

I. INTRODUCTION

The application of a microgrid is an effective technical approach for the large-scale application of distributed generations, an effective way to realize an active distribution network, and an effective means to implement the transition from the traditional grid to the smart grid. However, due to the utilization of a large number of power electronic devices in a microgrid, a large number of harmonic signals are injected into the grid, resulting in power quality deterioration problems such as voltage waveform distortion, fluctuation, flickering and three-phase unbalance [1]. This will pose a serious threat to the security, stability and economic operation of the power system, as well as to the surrounding electrical environment. Therefore, microgrid disturbance signal identification and classification is of great significance.

In terms of identification, with the development of computer simulation technology, many advanced signal characteristic extraction methods, such as fast Fourier transform (FFT), wavelet transform (WT) and Hilbert-Huang transform (HHT), have emerged. However, there are certain respective shortcomings to all these methods that are difficult to overcome. For example, FFT requires that the signal must be a stationary periodic signal and must satisfy the sampling theorem; otherwise, the analysis results will exhibit spectrum aliasing and the fence effect, while WT suffers from the difficulty in choosing the wavelet basis, and HHT faces the problems of the endpoint flying wing and modal

The associate editor coordinating the review of this manuscript and approving it for publication was Canbing Li.

confusion [2]. Numerous studies have shown that the convolutional structures in convolutional neural networks have strong characteristic learning and expression abilities. The sharing of the convolutional kernel parameters in the hidden layer and the interconnection sparsity between the layers enable convolutional neural networks to learn grid-like topology characteristics such as pixels and audio with less computational effort, and the learning effect is stable, while additional characteristic engineering for the data is unnecessary. The structure has a strong self-learning ability and natural adaptability to one-dimensional signals, which can overcome the limitations of previous characteristic extraction methods. Against this background, scholars have pioneered the application of convolutional neural networks with traditional structures to the characteristic extraction of electric energy disturbance signals [3]–[7]. However, the convolutional network structure, which simply superimposes the convolution layer and then connects one or more fully connected layers, has a poor learning ability. As a result, the convergence speed of the network is slow, and the generalization ability is poor.

Many scholars have found that increasing the depth of the network structure is the most effective method for enhancing the network learning ability at present. However, this will lead to an excessive computation amount and overfitting for traditional convolution network structures, such as LeNet and AlexNet. In many convolutional network models, the direct connection of the residual network (ResNet) enables it to considerably increase the depth without an excessive computational effort, thus improving the accuracy. In the internal residual block, the problem of gradient disappearance due to the increased depth in the deep neural network is relieved by a jump connection, which has the advantage of easy optimization, while the inception structure has the advantage of a more efficient use of the computing resources, and regarding the structure, more characteristics can be extracted for the same computation amount, and the training effect can be improved. A method based on the combination of these two network structures has been previously proposed for image classification, and good results have been achieved [8]–[11].

In terms of classification, with the improvement of the computing power, power quality classification methods based on pattern recognition are constantly proposed, e.g., methods based on neural networks [12], random forest networks [13], wavelet neural networks [14], support vector machines [15], [16], k-nearest neighbors [17], and decision trees. These methods have partly solved the problems inherent to the previous methods such that the accuracy of power quality disturbance classification has been considerably improved. However, a large computation amount is still required. Thus, there is still room for improvement of the accuracy.

In this paper, a new method for power quality disturbance identification and classification is proposed by constructing a new deep convolutional network structure. The database established according to the power disturbance signal calibrated by an n-dimensional unit vector is first adopted to train

TABLE 1. Power quality standard signal and disturbance signal model.

Type	Signal model	Parameters
X_0	$v(t) = U_m \sin(\omega_0 t)$	$U_m=311V,\ \omega_0=314(rad/s)$
X_1	$v(t) = U_m \sin(\omega_0 t) + U_n \sin(A_n \omega_0 t + \varphi_n)$	$A_n =2N+1(N \in [0,15])$
X_2	$v(t) = U_m \sin(\omega_0 t) + \alpha(t)U_n \sin(A_n \omega_0 t + \varphi_n)$	$U_n = 10\sim20\ V$, $\alpha(t) = Me^{-\frac{t}{\tau}},\ \tau =0\sim0.05\ s$
X_3	$v(t) = U_m(1 + \alpha(u(t - t_1) - u(-t_2)))\sin(\omega_0 t)$	$0.1 \leq \alpha \leq 0.9$, $0.5T \leq t2 - t1 \leq 30T$
X_4	$v(t) = U_m(1 - \alpha(u(t - t_1) - u(-t_2)))\sin(\omega_0 t)$	$0.1 \leq \alpha \leq 0.9$, $0.5T \leq t2 - t1 \leq 30T$
X_5	$v(t) = U_m(1 - (1 - \alpha)(u(t - t_1) - u(-t_2)))\sin(\omega_0 t)$	$0.9 \leq \alpha \leq 1$, $0.5T \leq t2 - t1 \leq 30T$
X_6	$v(t) = [1 + \alpha(t)]U_m \sin(\omega_0 t + \varphi)$	$\tau = 1\ ms\sim3ms$
X_7	$v(t) = [1 + a\sin(b\omega t)]\sin(\omega t)$	$0.05 < a < 0.2, 0.1 < b < 0.5$

the deep convolution network consisting of a five-layer 1D-MIR structure and a three-layer full-connection layer. Then, the network trained and optimized by the gradient descent method and adaptive moment estimation method (Adam) is applied to power quality disturbance identification and classification, thus improving the classification accuracy and speed.

The remainder of this paper is organized as follows. Section II provides a detailed description of the construction method and process of the proposed new convolutional network structure and its application in power quality disturbance identification and classification, where the establishment of the database is contained in Section II. A. The construction of the deep convolutional network with a 1D-MIR structure is expounded in Section II. B, and the training process is described in Section II. C. In Section III, the simulation results obtained with the proposed method are reported. In Section IV, representative practical cases of single and composite disturbances are analyzed, which illustrates the validity of the classification method. Finally, the Conclusion Section summarizes the main outcomes of this paper.

II. DESCRIPTION OF THE METHOD

A. ESTABLISHMENT OF THE DATABASE

According to IEEE standards, seven mathematical models of a single power quality voltage disturbance signal are established: harmonics (X_1), transient oscillation (X_2), transient swell (X_3), transient sag (X_4), interruption (X_5), transient pulse (X_6) and fluctuation (X_7). As indicated in TABLE 1, the fundamental wave frequency f_0 is 50 Hz, where T is the power frequency period, $u(t)$ is the unit step function, and U is the voltage amplitude. According to the actual situation of the power grid, considering harmonics below 30 times, the frequency range of the transient oscillation is 800∼1800 Hz.

Python is adopted to generate the disturbance signals listed in TABLE 1, where the sampling frequency is $f_s = 10$ kHz, and 25 fundamental wave cycles, namely, at 500 ms, are realized, with 5000 sampling points. The waveforms corresponding to TABLE 1 are shown in Fig. 1.

B. BUILDING OF THE DEEP CONVOLUTIONAL NETWORK WITH A 1D-MIR STRUCTURE

The convolutional network consists of three parts, which are used for the extraction of the characteristics, classification of the characteristics and optimization of the
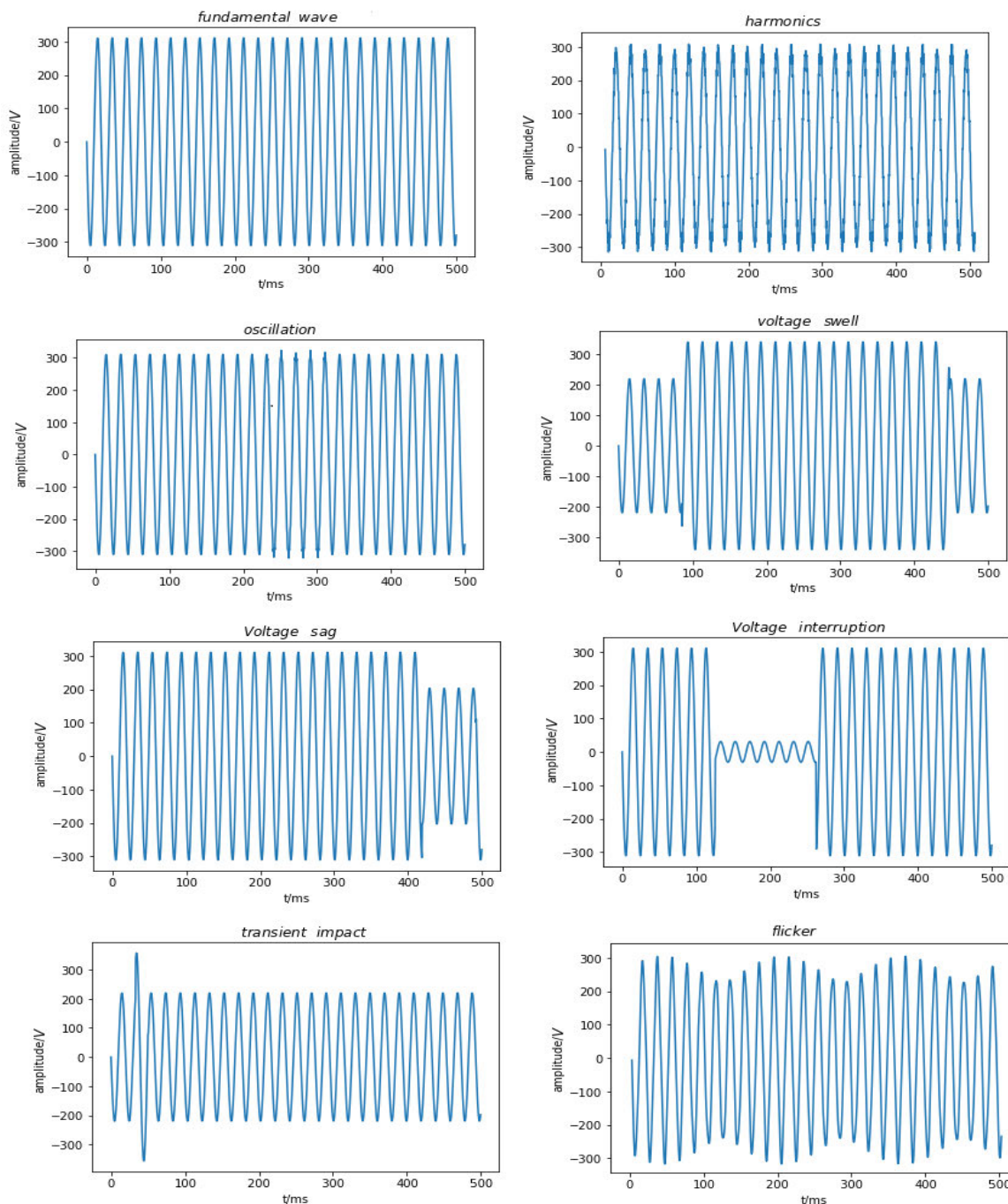
FIGURE 1. Eight kinds of waveforms.

output, and contains eight layers altogether. The building of the deep convolutional network can be described as follows:

The extraction of the characteristics relies on five 1D-MIR modules. Each of them consists of four branches, three of which consist of a subnetwork with two convolution layers in each branch, and the remaining branch is a shortcut, as shown in Fig. 2.

The three branches are named Branch_0, Branch_1 and Branch_2. Their functions are to extract the subtle

characteristics of signals and reduce the computational load of the network. Among them, all the first convolution layers are convolution layers with 1×1 convolution cores and 1 channel, and the second layers are convolution layers with a 1×3 convolution core and 32 channels, 1×5 convolution core and 16 channels, and 1×7 convolution core and 8 channels for Branch_0, Branch_1 and Branch_2, respectively, as indicated in TABLE 2. Because the dimensions of the output tensors of the three branches are inconsistent, the output of each branch should be filled symmetrically.
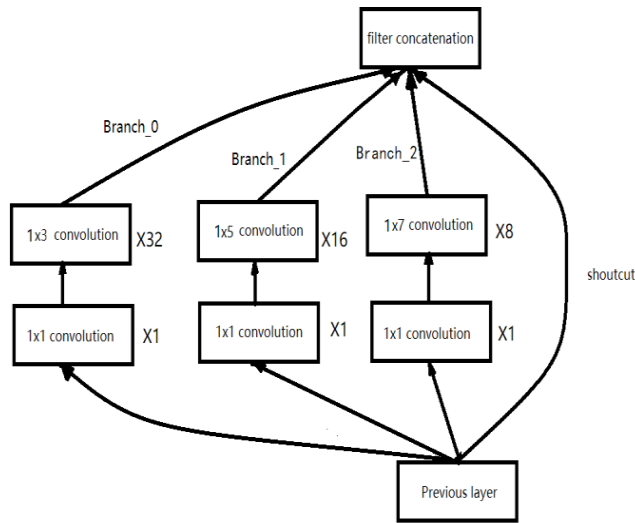
FIGURE 2. Modified Inception-ResNet (1D-MIR) structure.

TABLE 2. 1D-MIR module specification parameters.

branch	specification	quantity	stride	filling
Branch_0	1×1	1	1	Valid
	3×1	32	1	Valid
Branch_1	1×1	1	1	Valid
	5×1	16	1	Valid
Branch_2	1×1	1	1	Valid
	7×1	8	1	Valid

First, the outputs of the three branches are jointed to form a characteristic vector, and the vector is then made to multiply the output of the directly connected channel (also called the shortcut), while the multiplication results are then input to the next layer. The shortcut directly passes the input information to the output and protects the integrity of the information.

The classification part consists of three layers comprising the full-connection layer and the three layers of the dropout layer, which are alternately connected.

First, in all the fully connected layers, each node is connected with a node of the last layer. Next, the local information with category discrimination is integrated into a new characteristic vector. Finally, a new dropout layer is connected, and the weights at the nodes with less local characteristic information are set to 0, as shown in Fig. 3.

The optimization part of the output consists of an output layer and an optimization layer. In the output layer, the softmax function (also called the normalized exponential function) is utilized as the loss function. First, the probabilities of the output types in the classification part are individually calculated by the loss function. Next, the types of the power
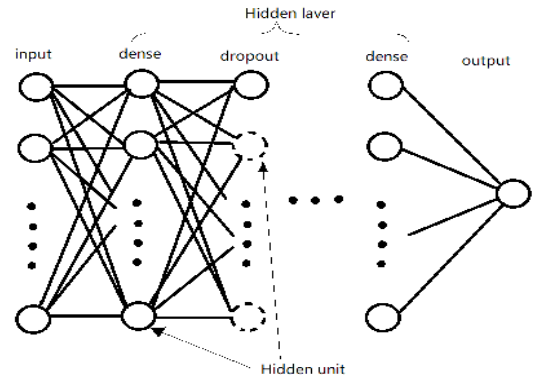

FIGURE 3. The structure of the classification part.

disturbance signals are determined according to the probabilities and are compared to the labels calibrated in advance to determine the ratio of the incorrect samples to the overall samples. Thereafter, the loss is output. In the optimization layer, the Adam function is first utilized to conduct first- and second-order moment estimations of the loss function of the output layer. Finally, the learning speed is dynamically adjusted according to the estimation results, thus regulating the learning steps of the back propagation process and optimizing the loss. The final built network structure is shown in Fig. 4.

C. TRAINING PROCESS

The training process of the 1D-MIR deep convolutional network on the power quality disturbance signal database is divided into two processes: the forward propagation process and back propagation process.

1) FORWARD PROPAGATION PROCESS
Step 1:

Python was adopted to generate the disturbance signals as listed in TABLE 1. Two thousand random samples are generated for each of the fundamental wave signals and the seven types of disturbance signals, totaling 16000 samples. For each type of disturbance, 1600 samples are selected to form the training set and 400 samples to form the test set. To ensure that the simulation signal is as close as possible to the actual power system disturbance signals, the begin-end times to each type of disturbance and random variation amplitude are set within a certain range. To improve the generalization ability of the network, data augmentation is required, which is achieved by randomly scrambling the different kinds of samples in the process for intercepting the disturbance signals [18].

To facilitate training, each disturbance type mentioned above is represented by a vector. That is, X_0 is represented by [1,0,0,0,0,0,0,0,0], X_1 by [0,1,0,0,0,0,0,0], X_2 by [0,0,1,0,0,0,0,0], X_3 by [0,0,0,1,0,0,0,0], X_4 by [0,0,0,0,1,0,0,0], X_5 by [0,0,0,0,0,1,0,0], X_6 by [0,0,0,0,0,0,1,0], and X_7 by [0,0,0,0,0,0,0,1].
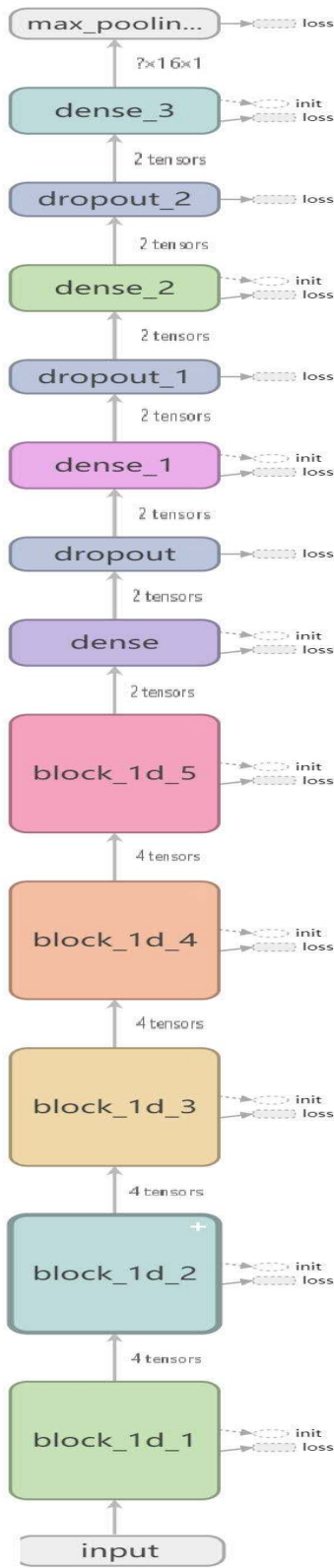
FIGURE 4. Overall structure of the modified Inception-ResNet (1D-MIR) deep convolution network.

The 16000 samples and corresponding labels are divided into 100 batches, and the batch size is 160, while the size of the sample labels is $1 \times$ the class number $\times 1$, where the class number is 8. The sample is written as a vector, as defined in Eq. (1):

$$X_i = [x_{i0}, x_{i1}, \ldots \ldots x_{im}]^T \tag{1}$$

In Eq. (1), $i = 1, 2, \ldots, n$, where n is the batch number of the samples, and m is the length of the signals collected.

The sample size of each batch is expressed as:

$$X_{batch\_size} = [X_1, X_2, \ldots \ldots X_n] \tag{2}$$

At the same time, at the moment the program starts, each network node is initialized, and the weights and biases are randomly assigned.

Step 2:

An input constructor (next batch) is constructed by calling the iterator of Python's bottom implementation. The samples and labels are fed into the network in batches to train the weights and biases of the characteristic extraction part.

A total of i characteristic vectors are obtained from the batch samples intercepted by the iterator under the action of the 5 1D-MIR modules.

The characteristic vector is output by the n-th 1D-MIR module:

$$y^{(n)} = \sum_1^l f\{f\{\ldots f\{X_{batchsize} \times \left(k_{n_{a0}} * k_{n_{a1}}\right.$$
$$\left. + k_{n_{b0}} * k_{n_{b1}} + k_{n_{c0}} * k_{n_{c1}}\right) * k_n \times X_{batchsize}\}\}\} \tag{3}$$

In Eq. (3), $y^{(n)}$ is the characteristic vector of the output of the n-th layer; $k_{n_{an}}$, $k_{n_{bn}}$ and $k_{n_{cn}}$ are the convolution kernels of the three branches of the n-th layer; k_n is the downsampling convolution kernel of the n-th layer; and "$*$" is the convolution operation, $n \in [0,1,2,3,4]$.

Step 3:

The output of the characteristic extraction part is input to the classification part to train the weights of the full-connection layer. The specific operation is as follows:

i. All the weights are aggregated into a weight matrix (W) in the order of the hidden units:

$$W = \{w[1], w[2], w[3], \ldots \ldots, w[q]\} \tag{4}$$

The same is true for the biases B:

$$B = \{b[1], b[2], b[3], \ldots \ldots b[q]\} \tag{5}$$

where q is the length of the input vector, i.e., the number of data points input into the full-connection layer.

Then, the weight matrix W and bias matrix B are initialized. The initialization process is as follows: the matrices are assigned by the random variables generated by the Gaussian process and then multiplied by $\sqrt{\frac{2}{q^{(l-1)}}}$.

ii. Each characteristic vector in each batch is input into the full-connection layer in the form of a vector space, and then, under the action of the full-connection layer, it is transformed into a simple characteristic vector, which can be expressed as:

$$z^{(1)} = \sum y_n w_i^{(1)} + b^{(1)} \tag{6}$$

iii. The outputs of the full-connection layer are input into the dropout layer, where the inactivation rate is in the range of [0,1], and the rate is set to 0.8. Then, a sparse vector is obtained by the weights of the partial nodes:

$$r_i^{(2)} \sim Bernoulli\,(p) \tag{7}$$
$$\tilde{z}^{(2)} = r_i^{(2)} * z^{(1)} \tag{8}$$

where $r_n^{(2)}$ is the inactivation factor randomly extracted from the Bernoulli distribution, $z^{(1)}$ is the output of the first fully connected layer, and $\tilde{z}^{(2)}$ is the final sparse vector. The following relationship can be obtained by repeating the same process three times.

$$x_{so} = f(w_i^{(6)} \tilde{z}^{(6)}) \tag{9}$$

The above is the final characteristic vector of the power disturbance signals.

Step 4:

The final characteristic vector is input into the softmax layer, and the vector is normalized with the softmax function. Then, the vector is mapped to the interval of (0,1), and an eight-dimensional probability vector is obtained. The estimated probability value of sample x for each category i is denoted as $p(i|x)$, which can be expressed as:

$$p(i|x) = \begin{bmatrix} \dfrac{e^{x_{os}^{[0]}}}{\sum_i x_{so}} \\ \dfrac{e^{x_{os}^{[1]}}}{\sum_i x_{so}} \\ \vdots \\ \dfrac{e^{x_{os}^{[8]}}}{\sum_i x_{so}} \end{bmatrix} \tag{10}$$

where i is the index of vector x_{so}. At the corresponding positions of the eight vector elements, the closer the element value approaches 1, the more likely it belongs to the perturbation type of the corresponding position of the value. The network classifies the sample as the one with the highest probability.

Step 5:

The error rate is calculated by comparing the judgment result of the network to the corresponding label. At the same time, the difference between the output and the sample label, which is called the loss value, is calculated node by node via the loss function, as expressed in Eq. (11):

$$L(a^{[L]}, y) = -[y log_{10}(a) + (1 - y) log_{10}(1 - a)] \tag{11}$$

where a is the input of each layer, and y is the basic truth label value of the sample.

Then, the cost function is calculated according to Eq. (12):

$$J\,(w, b) = \frac{1}{m} \sum_{i=1}^{m} L(a^{[i]}, y^{[i]}) \tag{12}$$

2) BACK PROPAGATION PROCESS
Step 1:

The derivatives of the weight (w) and bias (b) are determined for the loss functions obtained in the forward propagation process:

$$\frac{d}{dw} J\left(a^{[i]}, y^{[i]}\right) = \frac{1}{m} \sum_{i=1}^{m} \frac{d}{dw} L\left(a^{[i]}, y^{[i]}\right) \tag{13}$$

$$\frac{d}{db} J\left(a^{[i]}, y^{[i]}\right) = \frac{1}{m} \sum_{i=1}^{m} \frac{d}{db} L\left(a^{[i]}, y^{[i]}\right) \tag{14}$$

Step 2:

According to the chain rule, the derivatives of the loss function with respect to the weights and biases of the layer are determined for the output $y^{[i]}$ of each level to calculate the minimum value of the loss function. The vectored back propagation equation can be expressed as follows:

$$dZ^{[l]} = da^{[l]} * g^{[l]\prime}(Z^{[l]}) \tag{15}$$
$$dw^{[l]} = dZ^{[l]} * a^{[l-1]} \tag{16}$$
$$db^{[l]} = dZ^{[l]} \tag{17}$$
$$da^{[l-1]} = w^{[l]T} * dZ^{[l]} \tag{18}$$
$$dZ^{[l]} = w^{[l+1]T} dZ^{[l+1]} * g^{[l]\prime}(Z^{[l]}) \tag{19}$$

In the iteration process, the Adam optimization algorithm is adopted to reduce the calculation amount.

3) CONFIGURATIONS OF THE PARAMETERS AND ENVIRONMENT IN THE TRAINING PROCESS

i. The configuration of the parameters: First, the initial learning rate, the exponential decay rate of the first-order moment estimation β_1, and the exponential decay rate of the second-order moment estimation β_2 are set to 0.01, 0.9, and 0.999, respectively.

ii. The configuration of the hardware environment: Intel (R) Core (TM) i7-8759 CPU @ 2.20-2.21 GHz, 8 GB memory; Nvidia Quadro M4000 GPU, 8 GB memory.

iii. The configuration of the software environment: Windows 10 64-bit, CUDA Toolkit 9.0, CUDNN V9.0, Python 3.7.2, Tensor-GPU 1.13.1. model. The training and testing of the network are accelerated by the GPU.

III. RESULTS AND ANALYSIS

The key performance indicators mainly include the classification accuracy, computation cost, convergence speed, etc. for evaluating the quality (good and bad) of an algorithm for the identification and classification of the power quality disturbance. In this paper, these indicators are also adopted.

The classification accuracy is defined as the ratio of the correct number of samples to the total number of samples and is calculated according to Eq. (20).

$$accuracy = \frac{Correct\ number\ of\ classification}{Total\ number\ of\ samples} \tag{20}$$

The parameter quantity of the network to be trained (the number of weights) is determined by the network structure

shown in Fig. 2 and calculated by Eq. (21).

$$p_l = (K \times G + 1) \times C \qquad (21)$$

where p_l is the number of parameters in layer l, $K \times G$ is the number of convolution kernels of the layer, and C is the number of convolution kernels or channels.

The calculation quantity is also determined by the network structure and calculated by Eq. (22).

$$c = O \times (C - 1) \times C \times \left(\frac{H - K + P_h}{S} + 1 \right)$$
$$\times \left(\frac{W - G + P_W}{S} + 1 \right) \times p_l \qquad (22)$$

where O is the number of output characteristic graphs, H and K are the height and width of the input characteristic graphs, respectively, S is the step length, and P_h and P_w are the pixels filled along the height and width directions, respectively.

According to Eqs. (20), (21) and (22), the relationships between the loss value and accuracy of the network and the number of iterations on the training and testing sets can be determined, which are shown in Figs. 5, 6, 7 and 8.
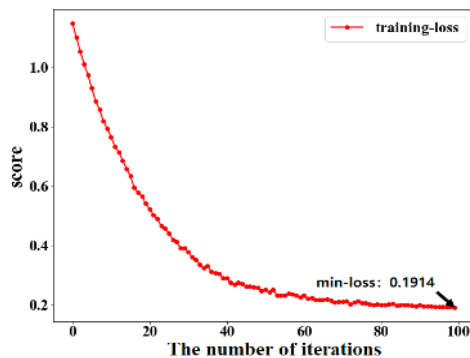
FIGURE 5. Loss value vs. the number of iterations on the training set. for data set I.
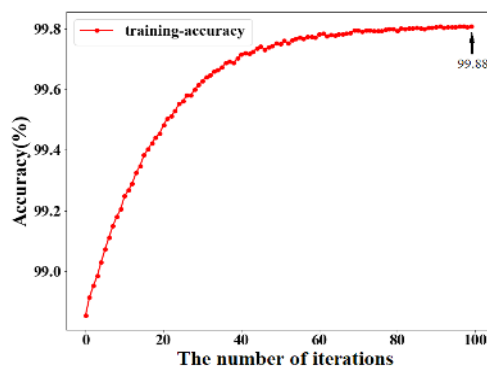
FIGURE 6. Accuracy vs. the number of iterations on the training set for data set I.

Figs. 5 and 6 show that after fewer than 100 iterations, the loss value and accuracy of the network structure are 0.1914 and 99.8%, respectively, on the training set. As the
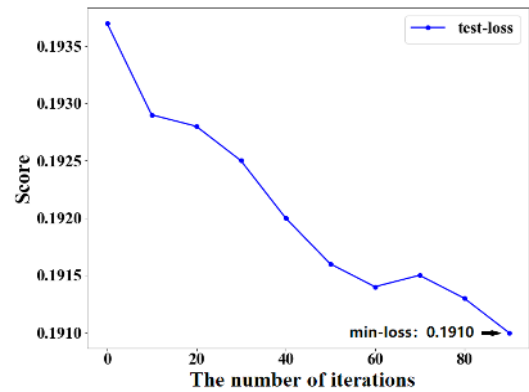
FIGURE 7. Loss value vs. the number of iterations on the testing set for data set I.
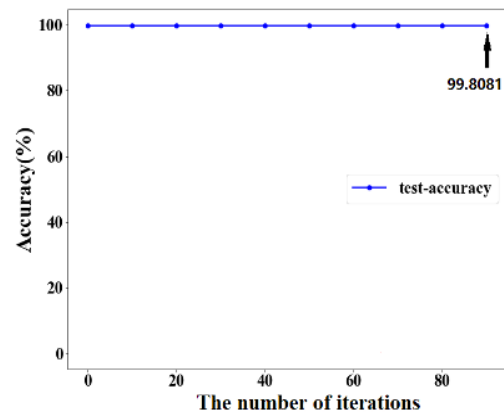

FIGURE 8. Accuracy vs. the number of iterations on the testing set for data set I.

number of iterations increases, the gap between the classification results and the predetermined results decreases, namely, the loss value decreases, and the statistical accuracy gradually increases.

As shown in Figs. 7 and 8, it is evident that the maximum accuracy of the network on the testing set is as high as 99.8%, and the loss value is further reduced to 0.1910. Through evaluation of the neural network model via 10-fold cross-validation, the average accuracy and loss value of the network in the simulation experiment are obtained, which are 98.5% and 0.1945, respectively.

The curve of the accuracy vs. the number of iterations is almost flat, implying that the dependence of the accuracy on the number of iterations is not notable. Therefore, the number of iterations on the testing set can be decreased to reduce the overall computation amount.

To examine the advantage of the network built over others, a test of 3 technical indicators (training time (times), total number of iterations, and number of iterations at convergence) for characterizing the convergence speed of a network is performed and 2 technical indicators (calculation quantity (times) and the number of parameters) are calculated for characterizing the computation cost by Eqs. (21) and (22),

respectively, for several kinds of networks (including the networks built in this paper and [3], [4], [19]) for the identification and classification of power quality disturbances.

The results obtained are summarized in TABLES 3 and 4. TABLE 3 lists the testing results of the training time, the number of iterations, etc. TABLE 4 summarizes the testing results of the calculation quantity and the number of parameters trained.

TABLE 3. Comparison of convergence rate.

CNN	Training time (min)	Total number of iterations (times)	Number of iterations at convergence (times)	Number of samples at convergence
LeNet-5[α*]	45	150	30	3200
AlexNet[β*]	50	1200	50	2000
Net-256[γ*]	50	5000	500	1400
1D-MIR[δ*]	6	100	80	1024

Note: α*: Ref. [3]; β*: Ref. [19]; γ*: Ref. [20]; δ*: the proposed network

TABLE 4. Comparison of the training cost.

CNN	Calculation quantity (times)	Number of parameters
LeNet-5	3474800	6610
AlexNet	7200000000	60000000
Net-256	738949088	28352
1D-MIR	321410	1494

TABLES 3 and 4 indicate that the training cost of the network built for the algorithm is much lower than that of the others, and its convergence speed is also much higher than that of the others.

To verify the generalization ability of the network, 64 types of disturbance samples are considered, including 8 single disturbances and 56 composite disturbances. The number of samples is 32000, and they are divided into two training sets.

Figs. 6, 9 and 10 show the accuracy variation obtained from the three training sets (data sets I, II and III, respectively). The figures reveal that the generalization accuracies of ID-MIR are 99.88%, 99.80% and 99.61% on the three data sets (I, II and III, respectively) whose size and type are different, but these values are very close.

Figs. 7 and 8 show the performance of the network on the test set.

The generalization precisions obtained by the 4 different kinds of networks for these 64 types of different disturbances are listed and compared in TABLE 5, whose calculation methods are the same as those used in [20], [21].

The classification accuracies for the 8 types of single disturbances are summarized in TABLE 6, and those for the 56 types of composite disturbances are listed in TABLE 7, where the accuracy rate is calculated with Eq. (20), and the number of correct or incorrect number of samples is obtained by screening the total samples.

In addition, the amount of data utilized in the network training is smaller than that utilized in the other networks.

TABLE 5. Generalization precision contrast.

CNN	Hyper parameter						Accuracy rate (%)		
	layer	n	size	s	p	action	I	II	III
leNet-5	C1	20	5×5	1	0	C			
	C2\C3	100	5×5	1	0	C	93.01	90.10	89.10
	C4	7	4×6	1	0	C			
	S1\S2\S3	1	2×2	2	0	max			
AlexNet	C1	3	11×11	4	0	C			
	C2	48	5×5	4	0	C			
	S2	1	2×2	1	0	max			
	C3	128	3×3	4	0	C	90.10	89.00	87.50
	C4\C5	192	3×3	4	0	C			
	S5	1	1×1	1	0	max			
	N1	256	6×6						
Net-256	C1	96	11×11	1	0	C			
	C2	256	7×7	4	0	C			
	C3	128	3×3	2	0	C			
	S1\S2\S3	1	2×2	2	0	max	96.05	85.50	85.01
	N1	128	4×4	0	0				
	N2	512	4×4	0	0				
	N3	64	4×4	0	0				
1D-MIR	Refer to the second item in section 2 of this article						99.88	99.80	99.61

Note: In the hyper parameter column, C_i (i =1, 2, 3, 4, 5) is the convolution layer; S_j (j=1, 2, 3) is the pooling layer; N_k (j=1, 2, 3) is the fully connected layer; s is the stride; n is the number of convolution kernels; max is the maximum pooling; C is the convolution action; p (padding) is the filling.

TABLE 6. Accuracy of the classification algorithm for a single disturbances.

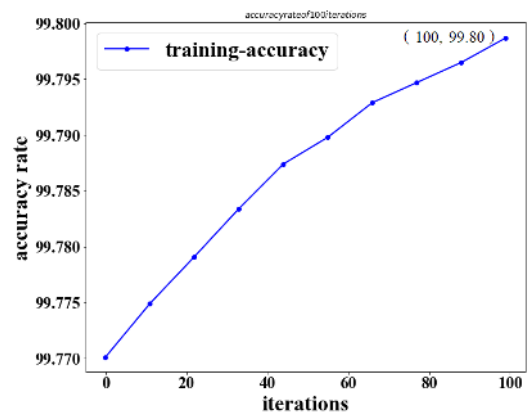| Order	Sample Type	Recognition accuracy(%)			
		The original signal	30 dB	40 dB	50 dB
1	X_0	100	99.62	99.90	99.15
2	X_1	98.80	98.58	98.80	98.33
3	X_2	97.60	98.17	98.00	97.49
4	X_3	97.43	97.43	97.43	97.11
5	X_4	98.75	98.75	98.75	98.22
6	X_5	98.50	98.52	98.50	98.27
7	X_6	99.10	98.69	97.60	97.60
8	X_7	98.22	98.22	98.22	98.19
9	Average	98.50	98.50	98.40	98.045

FIGURE 9. Accuracy vs. the number of iterations on the training set for data set II.

All these findings fully prove that the network has a strong generalization ability.

Moreover, to test the performance of the anti-interference ability of the method, three Gaussian white noises with

TABLE 7. Accuracy of the classification algorithm for the composite disturbances.

Order	Sample Type	Recognition accuracy(%)			
		The original signal	30 dB	40 dB	50 dB
1	X_1+X_3	98.58	98.23	97.90	97.90
2	X_1+X_4	97.59	96.89	96.50	96.48
3	X_1+X_5	98.61	98.60	98.60	98.60
4	X_1+X_6	98.70	98.68	98.69	98.68
5	X_1+X_7	98.55	97.85	97.80	97.80
6	X_2+X_3	99.01	98.99	98.97	98.95
7	X_2+X_4	99.05	99.12	98.96	98.90
8	X_2+X_5	98.99	98.89	98.85	98.80
9	X_2+X_6	99.59	98.85	98.85	98.80
10	X_2+X_7	98.56	98.45	98.45	98.45
11	X_3+X_4	99.90	99.85	99.85	99.85
12	X_3+X_5	98.99	98.98	98.97	98.97
13	X_3+X_6	98.99	98.95	98.95	98.95
14	X_3+X_7	99.05	99.03	99.03	99.03
15	X_4+X_5	99.13	99.10	99.08	99.08
16	X_4+X_6	99.01	98.99	98.98	98.98
17	X_4+X_7	99.50	99.49	99.49	99.49
18	X_5+X_6	99.80	99.79	99.79	99.79
19	X_5+X_7	99.80	99.80	99.79	99.79
20	X_6+X_7	98.87	98.75	98.75	98.75
21	$X_1+X_3+X_4$	98.66	98.65	98.64	98.64
22	$X_1+X_3+X_5$	99.01	99.01	99.00	99.00
23	$X_1+X_3+X_6$	99.77	99.75	99.74	99.74
24	$X_1+X_3+X_7$	98.80	98.45	98.40	98.08
25	$X_1+X_4+X_5$	99.58	99.55	99.55	99.55
26	$X_1+X_4+X_6$	99.80	99.80	99.79	99.78
27	$X_1+X_4+X_7$	97.20	96.98	96.97	96.97
28	$X_1+X_5+X_6$	97.35	97.34	97.34	97.33
29	$X_1+X_5+X_7$	98.10	98.09	98.07	98.07
30	$X_2+X_3+X_4$	99.20	99.15	99.14	99.14
31	$X_2+X_3+X_5$	98.75	98.74	98.74	98.74
32	$X_2+X_3+X_6$	98.99	98.98	98.98	98.98
33	$X_2+X_3+X_7$	98.50	98.47	97.90	97.82
34	$X_2+X_4+X_5$	97.99	97.99	97.95	97.95
35	$X_2+X_4+X_6$	98.01	98.00	98.00	98.00
36	$X_2+X_4+X_7$	95.98	95.91	95.90	95.90
37	$X_2+X_5+X_6$	97.00	96.91	96.90	96.90
38	$X_2+X_5+X_7$	98.03	98.02	98.02	98.02
39	$X_2+X_6+X_7$	96.99	96.99	96.98	96.98
40	$X_3+X_4+X_5$	99.82	99.82	99.82	99.81
41	$X_3+X_4+X_6$	99.50	99.49	99.49	99.49
42	$X_3+X_4+X_7$	99.57	99.56	99.56	99.54
43	$X_3+X_5+X_6$	99.63	99.63	99.62	99.62
44	$X_3+X_5+X_7$	99.43	99.42	99.41	99.40
45	$X_4+X_5+X_6$	99.58	99.58	99.58	99.57
46	$X_4+X_5+X_7$	98.58	98.57	98.57	98.57
47	$X_5+X_6+X_7$	98.77	98.76	98.76	98.76
48	$X_1+X_3+X_4+X_5$	97.56	97.54	97.54	97.54
49	$X_1+X_3+X_4+X_6$	97.55	97.55	97.55	97.53
50	$X_1+X_3+X_4+X_7$	97.56	97.56	97.55	97.55
51	$X_1+X_3+X_5+X_6$	98.38	98.37	98.37	98.37
52	$X_1+X_3+X_5+X_7$	98.24	98.23	98.23	98.23
53	$X_1+X_4+X_5+X_6$	97.86	97.85	97.85	97.85
54	$X_1+X_3+X_4+X_5+X_6$	96.78	96.77	96.76	96.76
56	$X_1+X_3+X_4+X_6+X_7$	96.79	96.77	96.77	96.77
57	Average	98.61	98.54	98.50	98.49

FIGURE 10. Accuracy vs. the number of iterations on the training set for data set III.

TABLE 8. Comparison of the accuracy obtained by the various kinds of commonly used classification methods.

Method	Accuracy (%)		
	30 dB	40 dB	50 dB
Short-time Fourier transform	93.68	93.68	90.93
S transform	96.63	96.63	91.20
Discrete wavelet transform	95.92	95.92	92.92
Hilbert yellow transformation	96.12	96.12	93.34
Method proposed in this paper	98.52	98.45	98.27

common methods, the proposed method attains a much higher accuracy rate.

IV. ANALYSIS OF THE EXAMPLES

To verify the effectiveness of the method, a simulation of the various single-disturbance types listed in TABLE 1 and several composite disturbance types is performed by utilizing Python, and an identification and classification is performed by the method proposed in this paper. The results obtained show that this method can clearly illustrate the start and end times when disturbances occur and the various characteristic information of the disturbances, thus realizing the identification and classification of the disturbance types. In the following, examples of single and composite disturbances are given, which illustrate the validity of the classification method.

A. STEADY-STATE HARMONIC SIGNAL
Given a Voltage Signal:

$$v(t)$$
$$= 220\sqrt{2}\sin(\omega_0 t) + 56\sqrt{2}\sin(3\omega_0 t) + 45\sqrt{2}\sin(5\omega_0 t)$$
$$+ 29\sqrt{2}\sin(7\omega_0 t) + 16\sqrt{2}\sin(11\omega_0 t) + 8\sqrt{2}\sin(13\omega_0 t).$$

The waveform of the given signal is shown in Fig. 12, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 11. The information in the figure shows that the steady-state harmonic signal occurs in the interval $t \in [0,500]$ ms, which is consistent with the information contained in the

signal-to-noise ratios (SNRs) of 30 dB, 40 dB and 50 dB are added to the disturbance signals to simulate the actual situation of noise pollution, and the classification accuracy rates are also listed in TABLES 6 and 7. It can be observed that they are still very high.

TABLE 8 lists the comparison results of the accuracy obtained by the various types of commonly used classification methods. The table indicates that compared to the
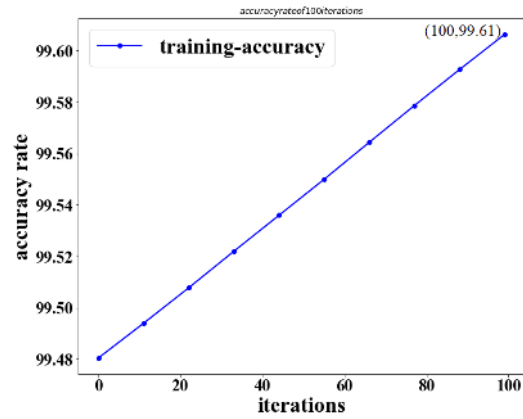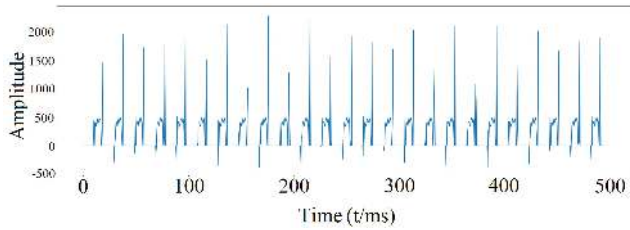
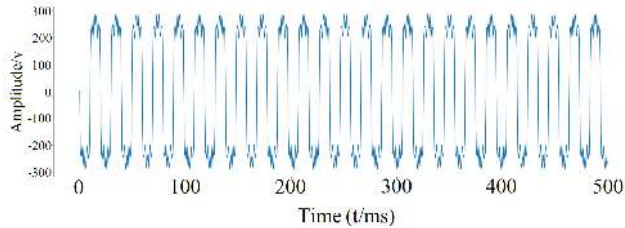**FIGURE 11.** The extracted characteristic waveform.

FIGURE 12. The signal waveform in the time domain.

given signal and reflects the characteristics of this type of disturbance.

B. VOLTAGE SWELL
Given a Voltage Signal:

$$v(t) = 220\sqrt{2}\,(1 + 0.8\,(u(t - t_1)$$
$$-u(-t_2)))\sin(\omega_0 t)\,,\,(0.5T \le t2 - t1 \le 30T).$$

The waveform of the given signal is shown in Fig. 14, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 13. The information in the figure shows that voltage swell occurs in the interval $t \in [349,385]$ ms, which is consistent with the information provided by the given signal and reflects the characteristics of this type of disturbance.
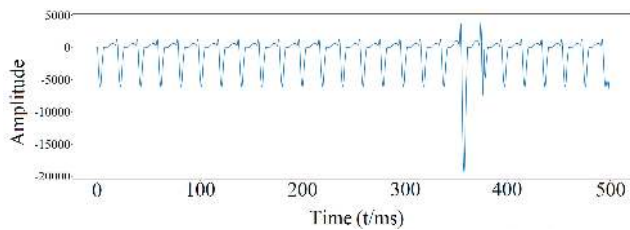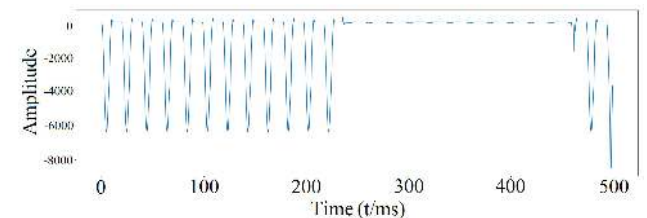

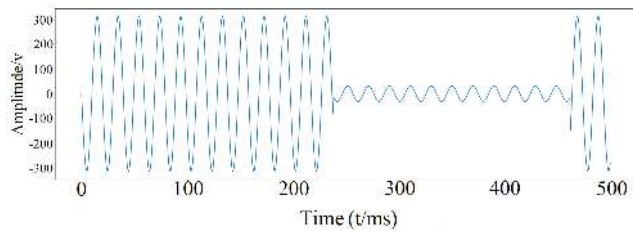**FIGURE 13.** The extracted characteristic waveform.
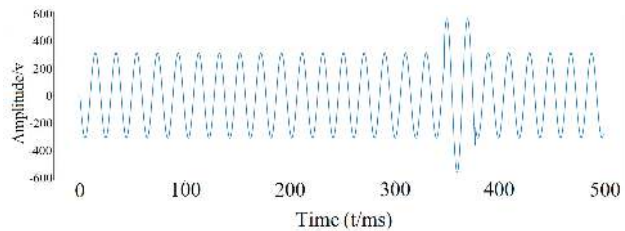
FIGURE 14. The signal waveform in the time domain.

C. VOLTAGE SAG
Given a Voltage Signal:

$$v(t) = 220\sqrt{2}\,(1 - 0.4\,(u(t - t_1) - u(-t_2)))\sin(\omega_0 t)\,,$$
$$(0.5T \le t_2 - t_1 \le 30T).$$

The waveform of the given signal is shown in Fig. 16, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 15. The information in the figure shows that voltage sag occurs in the interval $t \in [386,490]$ ms, which is consistent with the information in the given signal and reflects the characteristics of this type of disturbance.
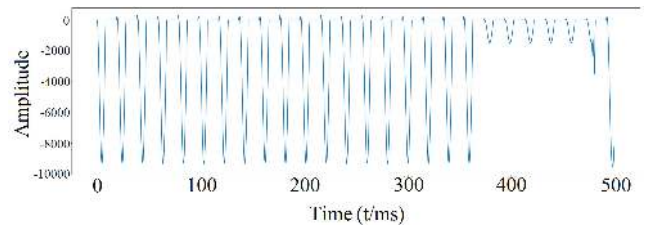
FIGURE 15. The extracted characteristic waveform.
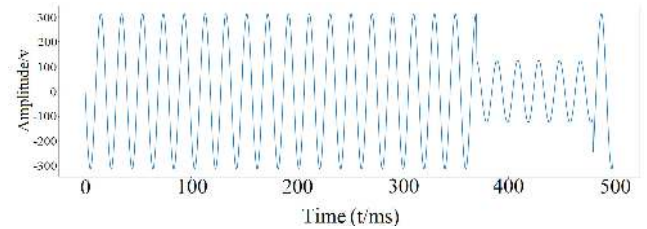
FIGURE 16. The signal waveform in the time domain.

D. VOLTAGE INTERRUPTION
Given a Voltage Signal:

$$v(t) = 220\sqrt{2}\,(1 - (1 - 0.9)\,(u(t - t_1) - u(-t_2)))\sin(\omega_0 t)\,,$$
$$(0.5T \le t_2 - t_1 \le 30T).$$

The waveform of the given signal is shown in Fig. 18, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 17. The information in the figure shows that voltage interruption occurs in $t \in [240,465]$ ms, which is consistent with the information provided by the given signal and reflects the characteristics of this type of disturbance.
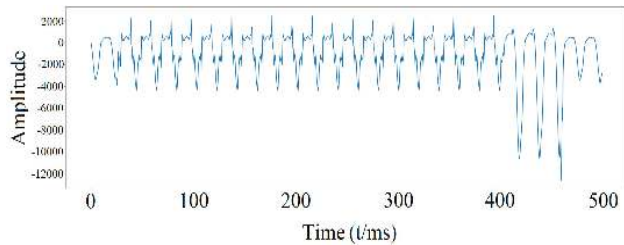

FIGURE 17. The extracted characteristic waveform.

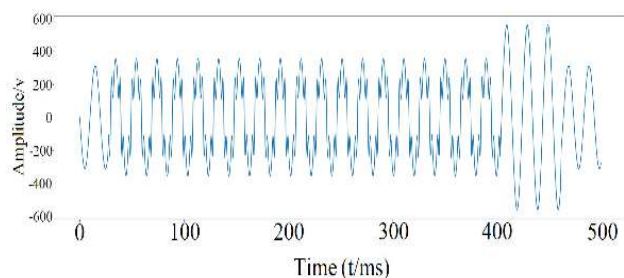FIGURE 18. The signal waveform in the time domain.

E. VOLTAGE SURGE + HARMONICS
Given a Voltage Signal:

$$v(t) = 220\sqrt{2} \sin(\omega_0 t) + (1 + 0.8(u(t - t_1) - u(-t_2)))$$
$$\times 220\sqrt{2} \sin(\omega_0 t_1) + 64\sqrt{2} \sin(5\omega_0 t_3)$$
$$+45\sqrt{2} \sin(7\omega_0 t_3) + 13\sqrt{2} \sin(9\omega_0 t_3).$$

where $0.5T \le t_2 - t_1 \le 30T$ and $50ms < t_3 < 150ms$.

The waveform of the given signal is shown in Fig. 20, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 19. The figures demonstrate that the harmonics occur in $t \in [60,415]$ ms, and the voltage surge occurs in $t \in [415,465]$ ms, which is consistent with the information contained in the given signal and reflects the characteristics of this type of composite disturbance.

FIGURE 19. The extracted characteristic waveform.

FIGURE 20. The signal waveform in the time domain.

F. VOLTAGE INTERRUPTION + HARMONICS
Given a Voltage Signal:

$$v(t) = 220\sqrt{2} \sin(\omega_0 t) + (1 - 0.1 \times (u(t - t_1) - u(-t_2)))$$
$$\times 220\sqrt{2} \sin(\omega_0 t) + 64\sqrt{2} \sin(5\omega_0 t_3)$$
$$\times 45\sqrt{2} \sin(7\omega_0 t_3) + 13\sqrt{2} \sin(9\omega_0 t_3)$$

where $0.5T \le t_2 - t_1 \le 30T$ and $50ms < t_3 < 300ms$.

The waveform of the given signal is shown in Fig. 22, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 21. As shown in the figures, the harmonics occur in $t \in [75,360]$ ms, and the voltage interruption occurs in $t \in [360,484]$ ms, which is consistent with the information provided by the given signal and reflects the characteristics of this type of composite disturbance.
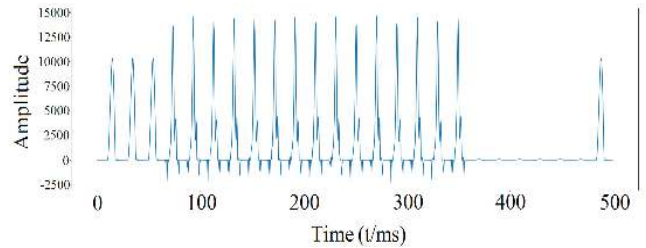

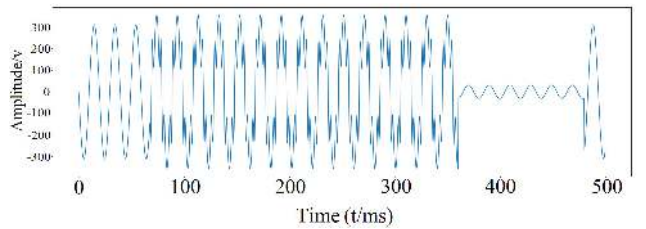FIGURE 21. The extracted characteristic waveform.

FIGURE 22. The signal waveform in the time domain.

G. HARMONIC + FLICKER + SAG
Given a voltage signal:

$v(t)$
$$= 220\sqrt{2}\,[1 + a\sin(b\omega t)]\sin(\omega t)$$
$$+ (1 - 0.8(u(t - t_1) - u(-t_2)))\,220\sqrt{2}\sin(\omega_0 t_1)$$
$$+64\sqrt{2}\sin(5\omega_0 t) + 45\sqrt{2}\sin(7\omega_0 t) + 13\sqrt{2}\sin(9\omega_0 t),$$
$$(0.05 < a < 0.2, 0.1 < b < 0.5)$$

where $0.5T \le t2 - t1 \le 30T$.

The waveform of the given signal is shown in Fig. 24, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 23. The figures reveal that the harmonics occur in $t \in [0,500]$ ms, the voltage flicker occurs in $t \in [0,500]$ ms,
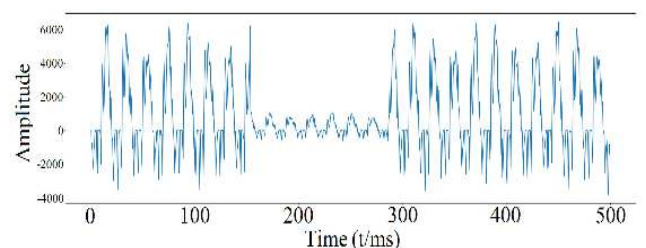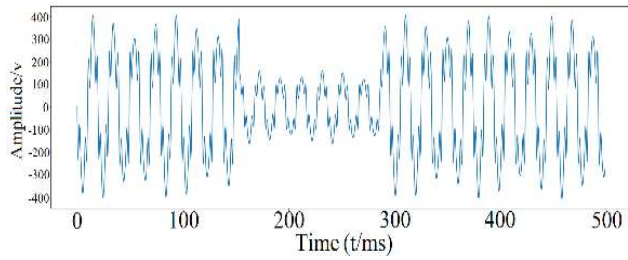

FIGURE 23. The extracted characteristic waveform.

FIGURE 24. The signal waveform in the time domain.

and the voltage sag occurs in t ∈ [150,290] ms, which is consistent with the information in the given signal and reflects the characteristics of this type of composite disturbance.

H. HARMONICS + SWELL + SAG + IMPACT
Given a voltage signal:

$$v(t) = (1 - 0.8(u(t-t_1) - u(-t_2)))\,220\sqrt{2}\sin(\omega_0 t_1)$$
$$+ (1 + 0.8(u(t-t_3) - u(-t_4)))\,220\sqrt{2}\sin(\omega_0 t_1)$$
$$+ 0.8 \times [u(t_5) - u(t_6)] + 64\sqrt{2}\sin(5\omega_0 t)$$
$$+ 45\sqrt{2}\sin(7\omega_0 t) + 13\sqrt{2}\sin(9\omega_0 t)$$

where $0.5T \le t_2 - t_1 \le 30T$; $0.5T \le t_4 - t_3 \le 30T$;

$1ms \le t_6 - t_5 \le 3ms$; $\quad 0.05 < a < 0.2$; $0.1 < b < 0.5$.

The waveform of the given signal is shown in Fig. 26, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 25. As shown in the figures, the harmonics occur in t ∈ [0,500] ms, the voltage sag occurs in t ∈ [210,500] ms, the impact occurs in t ∈ [205, 209] ms and the voltage swell occurs in t ∈ [0,205] ms, which is consistent with the information contained in the given signal and reflects the characteristics of this type of composite disturbance.
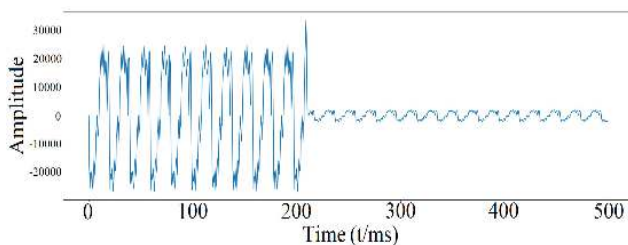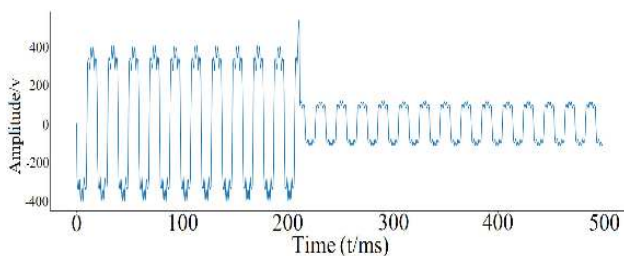

FIGURE 25. The extraced characteristic waveform.

FIGURE 26. The signal waveform in the time domain.

I. HARMONICS + SWELL + FLICKER + INTERRUPTION
Given a voltage signal:

$$v(t) = 220\sqrt{2}[1 + a\sin(b\omega t)]\sin(\omega t) + (1 + 0.8(u(t-t_1)$$
$$-u(-t_2)))\,220\sqrt{2}\sin(\omega_0 t) + 220\sqrt{2}(1 - (1 - 0.9)$$
$$\times (u(t-t_3) - u(-t_4)))\sin(\omega_0 t) + 64\sqrt{2}\sin(5\omega_0 t)$$
$$+ 45\sqrt{2}\sin(7\omega_0 t) + 13\sqrt{2}\sin(9\omega_0 t)$$

where $0.5T \le t_2 - t_1 \le 30T$; $0.5T \le t_4 - t_3 \le 30T$; $0.05 < a < 0.2$; $0.1 < b < 0.5$.

The waveform of the given signal is shown in Fig. 28, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 27. The figures show that the harmonics occur in t ∈ [0,500] ms, the voltage flicker occurs in t ∈ [0,500] ms, the voltage interruption occurs in t ∈ [315, 475] ms and the voltage swell occurs in t ∈ [475,500] ms, which is consistent with the information in the given signal and reflects the characteristics of this type of composite disturbance.
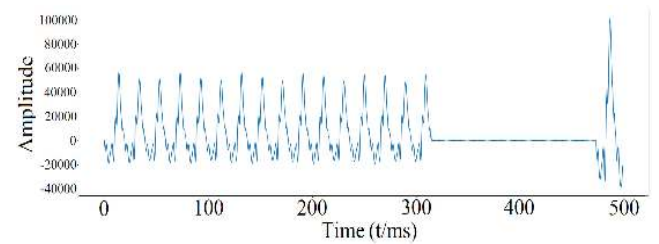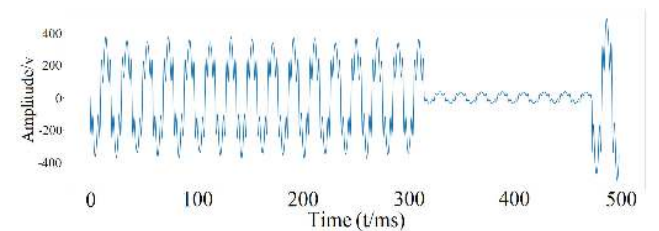

FIGURE 27. The extracted characteristic waveform.

FIGURE 28. The signal waveform in the time domain.

J. HARMONICS + SWELL + SAG + FLICKER
Given a voltage signal:

$$v(t) = 220\sqrt{2}[1 + a\sin(b\omega t)]\sin(\omega t) + (1 - 0.8(u(t-t_1)$$
$$-u(-t_2)))\,220\sqrt{2}\sin(\omega_0 t_1) + (1 + 0.8(u(t-t_3)$$
$$-u(-t_4)))\,220\sqrt{2}\sin(\omega_0 t_1) + 64\sqrt{2}\sin(5\omega_0 t)$$
$$+ 45\sqrt{2}\sin(7\omega_0 t) + 13\sqrt{2}\sin(9\omega_0 t)$$

where $0.5T \le t_2 - t_1 \le 30T$; $0.5T \le t_4 - t_3 \le 30T$; $0.05 < a < 0.2$; $0.1 < b < 0.5$.

The waveform of the given signal is shown in Fig. 30, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 29. As shown in the figures, the harmonics occur in t ∈ [0,500] ms, the voltage flicker occurs in t ∈ [0,500] ms, the voltage sag occurs in t ∈ [111,159] ms and the voltage
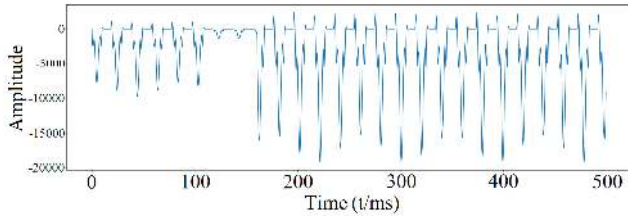
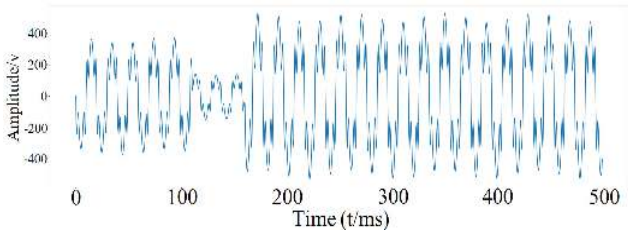**FIGURE 29.** The extracted characteristic waveform.

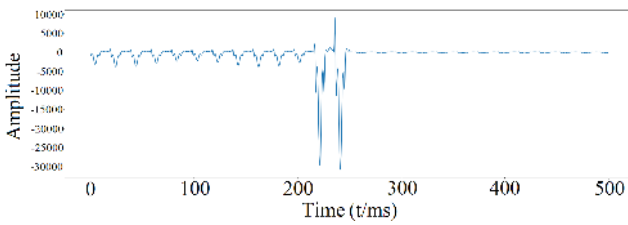**FIGURE 30.** The signal waveform in the time domain.

FIGURE 31. The extracted characteristic waveform.

swell occurs in t $\in$ [159,500] ms, which is consistent with the information provided by the given signal and reflects the characteristics of this type of composite disturbance.

K. HARMONICS + VOLTAGE SWELL + VOLTAGE SAG + FLICKER + VOLTAGE INTERRUPTION

Given a voltage signal:

$$v(t) = 220\sqrt{2}[1 + a\sin(b\omega t)]\sin(\omega t) + (1-0.8(u(t-t_1)) \\ -u(-t_2)))220\sqrt{2}\sin(\omega_0 t_1) + (1+0.8(u(t-t_3)) \\ -u(-t_4)))220\sqrt{2}\sin(\omega_0 t_1) + 220\sqrt{2}(1-(1-0.9) \\ \times(u(t-t_5)-u(-t_6)))\sin(\omega_0 t) + 64\sqrt{2}\sin(5\omega_0 t) \\ +45\sqrt{2}\sin(7\omega_0 t) + 13\sqrt{2}\sin(9\omega_0 t)$$

where $0.5T \leq t_2 - t_1 \leq 30T$; $0.5T \leq t_4 - t_3 \leq 30T$;

$0.5T \leq t_6 - t_5 \leq 30T$; $0.05 < a < 0.2$; $0.1 < b < 0.5$.

The waveform of the given signal is shown in Fig. 32, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 31. The figures show that the harmonics occur in t $\in$ [0,500] ms, the voltage flicker occurs in t $\in$ [0,500] ms, the voltage sag occurs in t $\in$ [0,211] ms, the voltage swell occurs in t $\in$ [211,249] ms, and the voltage interruption occurs in t $\in$ [249,500] ms, which is consistent with the information in the given signal and reflects the characteristics of this type of composite disturbance.
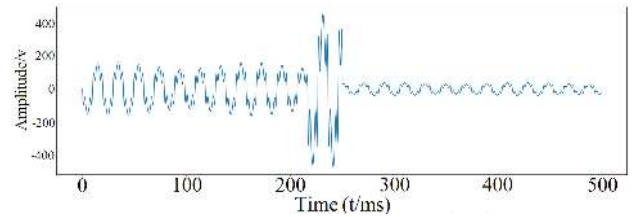

FIGURE 32. The signal waveform in the time domain.

L. HARMONICS + VOLTAGE SWELL + VOLTAGE SAG + FLICKER + IMPACT

Given a voltage signal:

$$v(t) \\ = 220\sqrt{2}[1 + a\sin(b\omega t)]\sin(\omega t) + (1 - 0.8(u(t-t_1)) \\ -u(-t_2)))220\sqrt{2}\sin(\omega_0 t_1) + (1+0.8(u(t-t_3)) \\ -u(-t_4)))220\sqrt{2}\sin(\omega_0 t_1) +0.8\times[u(t_5)-u(t_6)] \\ +64\sqrt{2}\sin(5\omega_0 t) + 45\sqrt{2}\sin(7\omega_0 t) + 13\sqrt{2}\sin(9\omega_0 t)$$

where $0.5T \leq t_2 - t_1 \leq 30T$; $0.5T \leq t_4 - t_3 \leq 30T$;

$0.5T \leq t_6 - t_5 \leq 30T$; $0.05 < a < 0.2$; $0.1 < b < 0.5$.

The waveform of the given signal is shown in Fig. 34, and the characteristic waveform extracted by 1D-MIR is shown in Fig. 33. As shown in the figures, the harmonics occur in t $\in$ [0,500] ms, the voltage flicker occurs in t $\in$ [0,500] ms, the voltage sag occurs in t $\in$ [0,170] ms, the impact occurs in t $\in$ [170,173] ms, and the voltage swell occurs in t $\in$ [173,500] ms, which is consistent with the information contained in the given signal and reflects the characteristics of this type of composite disturbance.
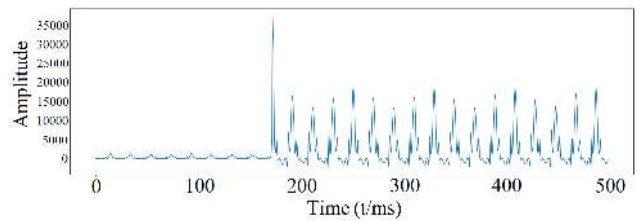

FIGURE 33. The extracted characteristic waveform.
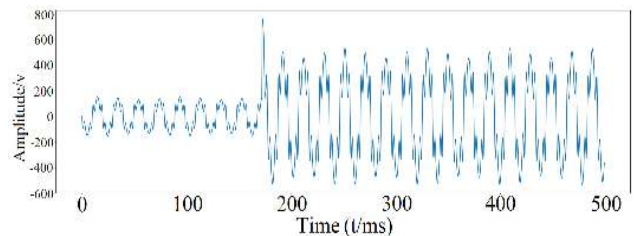

FIGURE 34. The signal waveform in the time domain.

Figs. 11 to 34 show the signal waveforms in the time domain and the extracted characteristic waveforms of the 4 single disturbances, 2 two-composite disturbances, 1 three-composite disturbance, 3 four-composite disturbances, and 2 five-composite disturbances. The figures reveal that in the characteristic extraction part of the network, not only can the characteristics of the network be extracted but the difference in characteristics can also be magnified to sufficiently characterize the characteristics of the corresponding disturbance

types such that the classification link can clearly distinguish the basic disturbance types contained in the given signal from the extracted characteristic waveforms to quickly localize the time when the disturbance occurs, thus identifying the disturbance type.

V. CONCLUSIONS

Aiming at the problems of a low convergence speed, low accuracy and poor generalization ability of the available traditional power disturbance identification and classification methods, this paper proposes a new deep convolutional network structure and a power quality disturbance identification and classification method for microgrids based on this new network structure. The large number of simulation experiments verify that the constructed network can quickly and accurately extract the characteristics of the various disturbance signals, including single and composite disturbances, and identify and classify them. Compared to the available power disturbance identification and classification methods, the proposed method attains a higher classification accuracy (as high as 99.1%), higher convergence speed and stronger generalization ability. The measured classification precision of the network on the data sets for the different sample types are 99.88%, 99.80% and 99.61%, the average classification precision of the network is 99.1%, and the network training time is only approximately 6 minutes.

In conclusion, the built network and proposed method in this paper for power quality disturbance identification and classification are feasible and effective, which provides a new approach for power quality disturbance identification and classification. However, they are not perfect. The main disadvantage is that they still cannot be used to detect certain components of disturbance signals, such as the amplitude of the disturbances. Next, we will improve them such that they can be applied to more fields of power quality governance.

REFERENCES

[1] W. Chang, "Micro grid power quality detection and analysis based on HHT," North China Electric Power University, Beijing, China, Tech. Rep., 2014.

[2] X. Yanchun, G. Yongkang, L. Zhenxing, and X. Lei, "Application of improved LMD algorithm in detection of power quality disturbance signal in microgrid," *Grid Technol.*, vol. 43, no. 1, pp. 332–341, 2019.

[3] Q. Hezuo, L. Xiaoming, C. C. Chen, and H. Le, "Power quality disturbance classification based on convolution neural network," *J. Wuhan Univ. (Eng. Ed.)*, vol. 51, no. 6, pp. 534–539, 2018.

[4] C. Wei, J. he, and X. Pei, "Power quality disturbance classification based on phase space reconstruction and convolution neural network," *Power Syst. Protection Control*, vol. 46, no. 14, pp. 87–93, 2018.

[5] X. Liwu, L. Kaicheng, X. Xiangui, Z. Chen, Y. Jiaming, and N. Yi, "Power quality compound disturbance identification based on deep feedforward network," *Elect. Meas. Instrum.*, vol. 57, no. 1, pp. 62–69 and 130, 2020.

[6] X. Luxin, "Power quality disturbance identification based on deep learning," Nanchang Univ., Nanchang, China, Tech. Rep., 2017.

[7] L. Jiahan, C. kexu, M. Jian, X. Chunhua, and W. Jianhua, "Three phase voltage sag classification based on convolution neural network and random forest," *Power Syst. Protection Control*, vol. 47, no. 20, pp. 112–118, 2019.

[8] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*. [Online]. Available: http://arxiv.org/abs/1602.07261

[9] N. Passalis, G. Mourgias-Alexandris, A. Tsakyridis, N. Pleros, and A. Tefas, "Variance preserving initialization for training deep neuromorphic photonic networks with sinusoidal activations," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1483–1487, doi: 10.1109/ICASSP.2019.8682218.

[10] Q. Zheng, X. Tian, M. Yang, Y. Wu, and H. Su, "PAC-Bayesian framework based drop-path method for 2D discriminative convolutional network pruning," in *Multidimensional Systems and Signal Processing*. 2019, doi: 10.1007/s11045-019-00686-z.

[11] Q. Zheng, X. Tian, N. Jiang, and M. Yang, "Layer-wise learning based stochastic gradient descent method for the optimization of deep convolutional neural network," *J. Intell. Fuzzy Syst.*, vol. 37, no. 4, pp. 5641–5654, Oct. 2019, doi: 10.3233/JIFS-190861.

[12] H. Wang, "Research on power quality disturbance detection and classification algorithm based on S-transform and neural network," Tianjin Univ., Tianjin, China, Tech. Rep., 2017.

[13] L. Qian, "Study on power quality disturbance identification method based on random forest," North China Electr. Power Univ., Beijing, China, Tech. Rep., 2018.

[14] T. Weihua, P. Xinfu, M. Xiangbin, Y. Hongtao, and S. Qiuye, "Detection and classification of power quality disturbances based on discrete wavelet transform and wavelet neural network," *Elect. Appl.*, vol. 35, no. 23, pp. 51–56, 2016.

[15] X. Zhichao, Y. Lingjun, and L. Xiaoming, "Power quality disturbance identification based on clustering improved S-transform and direct support vector machine," *Power Automat. Equip.*, vol. 35, no. 7, pp. 50–58 and 73, 2015.

[16] H. Chen and G. Zhang, "Power quality disturbance identification based on decision tree and support vector machine," *Grid Technol.*, vol. 5, pp. 1272–1278, 2013.

[17] T. H. Feng and L. Z. Yi, "K-nearest-neighbor multi tag classification method based on tag correlation," *Comput. Appl.*, vol. 35, no. 10, pp. 2761–2765, 2015.

[18] Q. Zheng, M. Yang, and X. Tian, "A full stage data enhancement method for natural image classification in depth convolution neural network," *Discrete Dyn. Nature Soc.*, 2020, Art. no. 4706576.

[19] L. Zhang, Z. Yan, Q. Gang, D. Ji, and S. Wei, "A real-time power quality disturbance identification and classification method," *Hebei Ind. Technol.*, vol. 36, no. 1, pp. 42–46, 2019.

[20] Q. Zheng, M. Yang, J. Yang, Q. Zhang, and X. Zhang, "Improvement of generalization ability of deep CNN via implicit regularization in two-stage training process," *IEEE Access*, vol. 6, pp. 15844–15869, 2018, doi: 10.1109/ACCESS.2018.2810849.

[21] B. Neyshabur, S. Bhojanapalli, D. McAllester, and N. Srebro, "Exploring generalization in deep learning," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, Jul. 2017, pp. 1–19. [Online]. Available: http://arxiv.org/abs/1706.08947v2

RENXI GONG was born in Guilin, Guangxi, China, in September 1962. He received the degree in semiconductor devices and microelectronics, and the Ph.D. degree in microelectronics and solid-state electronics from Xidian University, in 1993 and 2002, respectively. From 1994 to 1997, he worked as an Assistant at Guangxi University. He has been a Professor with the Electrical Engineering School, Guangxi University, China, since 2003. His current research interests include intelligent detection technology, power electronics, and their applications.

TAOYU RUAN was born in Nanning, Guangxi, China, in July 1992. She has been majoring in control engineering at Guangxi University, China, since 2017. Her current research interests are intelligent detection technology and artificial intelligence technology.

• • •