

Received August 1, 2020, accepted August 12, 2020, date of publication August 17, 2020, date of current version August 27, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017047

Pattern Recognition of Partial Discharge Based on VMD-CWD Spectrum and Optimized CNN With Cross-Layer Feature Fusion

ANGRAN GAO¹, YONGLI ZHU¹, WEIHAO CAI¹, AND YI ZHANG¹

State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Baoding 071003, China

Corresponding author: Yongli Zhu (yonglipw@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 51677072.

ABSTRACT In order to improve the recognition accuracy of partial discharge (PD) by making full use of the time-frequency characteristics of PD signals and employing deep learning theory, a kind of PD pattern recognition method based on variational mode decomposition (VMD)-Choi-Williams distribution (CWD) spectrum and optimized convolutional neural network (CNN) with cross-layer feature fusion is proposed in this paper. Firstly, a PD signal is decomposed into several components by VMD algorithm, and the CWD analysis of the obtained components is carried out to obtain the VMD-CWD time-frequency spectrum. Secondly, the cross-layer feature fusion and optimization CNN (CFFO-CNN) is constructed by introducing cross-layer connection and optimization algorithm. Thirdly, the VMD-CWD is regarded as the input vector to train CFFO-CNN to learn and extract the intrinsic features of the spectrum. Finally, the trained network is used to recognize the PD types of the testing samples. The proposed method is compared with traditional recognition methods such as BP neural network (BPNN) and support vector machine (SVM), as well as some commonly used deep learning algorithms. The experimental results indicate that the recognition performance of the proposed method is significantly better than that of existing recognition methods with accuracy up to 99.5%. It is proved that CFFO-CNN has superior feature extraction ability, which can extract the internal features of the VMD-CWD spectrum independently with higher recognition accuracy and wider application prospect.

INDEX TERMS Variational mode decomposition (VMD), Choi-Williams distribution (CWD), feature fusion, convolutional neural network (CNN), partial discharge, pattern recognition.

I. INTRODUCTION

Power transformers are the most crucial equipment in the power grid. Its insulation condition is directly related to the safe operation of the whole power system. However, in the production, transportation, installation and long-term operation of transformers, various insulation defects will inevitably appear. Among them, partial discharge (PD) is the main reason for the final breakdown of insulation of transformers, and it is also an important manifestation of the internal insulation degradation [1]. Due to the difference of insulation degradation mechanism among different discharge types, the degree of damage to the equipment is also distinct. Therefore, the correct identification of the detected PD type

is conducive to the evaluation of the internal insulation state of transformers, which has become a hot research content of power transformers fault diagnosis and location [2].

The premise of pattern recognition of PD signals by using discharge data is to analyze the characteristics of different discharge types correctly. Since a PD signal is a non-stationary time-varying signal, traditional time-frequency analysis methods such as STFT [3], Hilbert-Huang transform (HHT) [4] and Wigner-Ville distribution (WVD) [5] are proposed by scholars to analyze PD signal. The above methods have achieved corresponding results in practical application, but they also have some deficiencies. For instance, the window function of STFT is constant, which leads to its poor local adaptability [6]; in HHT, the modes overlap with each other easily, and the fault features in the obtained marginal spectrum is not clear [7]; WVD is susceptible to the influence

The associate editor coordinating the review of this manuscript and approving it for publication was Imran Sarwar Bajwa¹.

of cross-interference terms in the process of signal analysis [8], thus reducing the recognition rate.

Choi-Williams distribution (CWD) is a common time-frequency analysis method of fault signals. On the basis of WVD, CWD effectively restrains cross-interference terms by introducing kernel function, which makes it describe the instantaneous frequency and edge characteristics of signal more clearly [9]. VMD-CWD combines the variational mode decomposition (VMD) [10] with CWD to obtain the spectrum diagram for PD pattern recognition without cross-interference terms and with good energy aggregation. Since different types of PD signals contain different frequency components, the VMD-CWD obtained after time-frequency analysis is bound to have some differences. Therefore, the VMD-CWD spectrum diagram of PD signal can be used as the feature expression to distinguish different discharge types.

Due to the large amount of information contained in the constructed PD feature expression, it is subjective to extract one or several features of PD signal through complex artificial design, and the traditional shallow-layer recognition method has poor processing ability for high-dimensional data, which easily leads to the lack of generalization ability of the model, so it is hard to achieve good recognition effect. In recent years, deep learning has been widely used due to its advantages of automatic feature extraction and classification, which brings new opportunities for pattern recognition of PD [11]. In [12], Karimi *et al.* have used deep belief network (DBN) to identify three types of PD and achieved good results. In [13], stacked sparse auto-encoder (SSAE) is successfully applied to classify four defined PD severity states. In comparison with other deep learning algorithms, the model complexity and training difficulty of convolutional neural network (CNN) are relatively small owing to its parameter sharing and local connection, which makes CNN have superior performance in feature extraction of high-dimensional data, especially suitable for image processing [14], [15]. Nowadays, CNN has been successfully applied in speech recognition, image recognition and other fields [16]–[18]. In reference [19], one-dimensional CNN is used to detect PD types by extracting the characteristics of PD time-domain waveform. Compared with traditional artificial feature extraction methods, CNN can adaptively extract high-dimensional nonlinear and complex correlation features of data with strong nonlinear mapping capability [20].

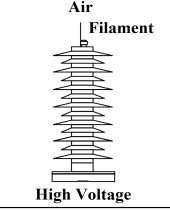
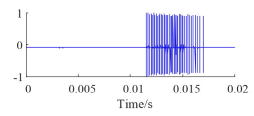
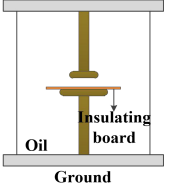
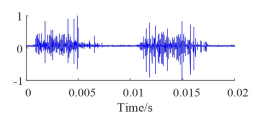
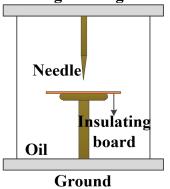
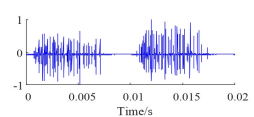
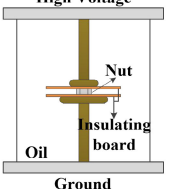
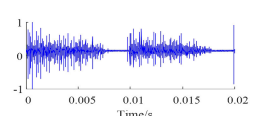
Up to now, although some researches have applied CNN to pattern recognition of PD, the recognition accuracy still needs to be further improved. First of all, the existing CNN used in PD recognition usually takes the time-domain waveform of PD signal as input and does not contain frequency-domain information, which leads to the insufficiency of the feature information and restricts the improvements in recognition accuracy. Secondly, the traditional CNN usually extracts the semantic information contained in deep-layer features from top to bottom for classification, and often ignores the image details contained in the shallow-layer features. Therefore, the

recognition accuracy of traditional CNN still have a space to improve. Consequently, a pattern recognition method of PD signals based on the VMD-CWD and optimized CNN with cross-layer feature fusion is proposed. It works roughly in the following steps. Firstly, the VMD-CWD spectra of PD signals are obtained by time-frequency analysis. Secondly, cross-layer feature fusion and optimization CNN (CFFO-CNN) is constructed to automatically extract the deep and shallow features contained in VMD-CWD so as to further realize the recognition of PD types. Experimental results prove that the proposed method is feasible.

II. PD EXPERIMENTAL MODELS

According to the form and characteristics of PD signals in power transformers, four kinds of PD insulation defect models are designed in the laboratory to simulate corona discharge, plate-to-plate discharge, needle-to-plate discharge and floating discharge. The electrode structures and typical PD data relevant to each type are shown in Table 1.

TABLE 1. Four types of PD models.

PD type	Experimental model	Typical PD data
Corona discharge		
Plate-to-plate discharge		
Needle-to-plate discharge		
Floating discharge		

A. CORONA DISCHARGE MODEL

This model is used to generate the corona discharge signals of transformer in air. The discharge tip adopts copper wire with a diameter of 1 mm, and its length is set to be 30, 50, 70 mm, respectively.

B. PLATE-TO-PLATE DISCHARGE MODEL

This model is used to generate the plate-to-plate discharge signals in the transformer. The diameters of the round epoxy insulation plate are 30, 40, 50 mm, respectively.

C. NEEDLE-TO-PLATE DISCHARGE MODEL

This model is used to generate the partial discharge signals caused by sharp conductor inside the transformer. The needle electrode is made of aluminum rod with a diameter of 3 mm, and the end is polished into a 30° cone. An insulating plate with a thickness of 1 mm is placed between the needle electrode and the plate electrode.

D. FLOATING DISCHARGE MODEL

This model is used to generate the floating discharge signals caused by loose contact or poor grounding inside the transformer. The insulation plate is 50 mm in diameter and 1 mm in thickness. The thickness of the metal gasket is 3 mm and the diameters are 5 and 10 mm, respectively.

The PD measurements were implemented by pulse current method according to standard IEC60270-2000. The schematic wiring diagram of the experiment is shown in Fig. 1. The sensor used in the experiment is composed of the coupling capacitor and the detection impedance. The detection impedance is relatively complex. In fact, its primary side is connected in parallel with a discharge tube for protection and its secondary side is connected in parallel with a terminal resistance for preventing high frequency signal reflection. The inductance of the detection impedance is 0.88mH and the coupling capacitance is 1000pF. TWPD-2F PD analyzer is used to capture and display PD signals in the experiments. Its sampling rate is set to 40MHz, the measured signal bandwidth is 40-300kHz. The high-voltage experiment platform is TW15133-10/100am.

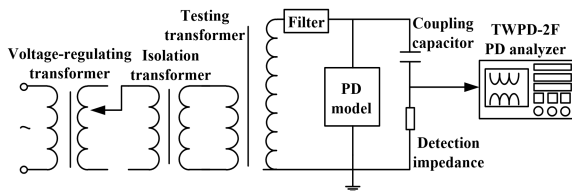


FIGURE 1. PD measurement system in experiments.

During the experiments, the discharge data of each power cycle is taken as a sample. In order to simulate the different degree of discharge severity and make the experimental samples not lose generality in the subsequent analysis process, this experiment collected 4 kinds of discharge samples under different voltage levels, 150 samples for each discharge type, a total of 600 samples. Among them, there are 100 training samples of each type, 400 in total; 50 testing samples of each type, 200 in total.

III. METHOD

A. THEORY OF VMD-CWD ANALYSIS

CWD is a typical time-frequency analysis method of fault signal in Cohen class time-frequency distribution. By introducing the exponential kernel function to smooth the WVD, the cross-interference term of the signal is suppressed to some extent. For a continuous time signal $z(t)$, the unified expression of Cohen class time-frequency distribution is [9]:

$$C_z(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} z(u + \frac{\tau}{2}) z^*(u - \frac{\tau}{2}) \psi(t - u, \tau) e^{-j2\pi f \tau} dud\tau \tag{1}$$

where t is time, f is frequency, u is local time, τ is time-shift parameter, $z^*(t)$ is the conjugate complex of signal $z(t)$, $\psi(t - u, \tau)$ is the kernel function, and the exponential kernel function expression of CWD is as follows:

$$\psi_n(t, \tau) = \frac{1}{\sqrt{4\pi\alpha\tau^2}} \exp(-\frac{t^2}{4\alpha\tau^2}) \tag{2}$$

where α is the smoothing factor, substituting (2) into (1) to get the definition of CWD as follows:

$$CWD_z(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{4\pi\alpha\tau^2}} \exp(-\frac{(t - u)^2}{4\alpha\tau^2}) \times z(u + \frac{\tau}{2}) z^*(u - \frac{\tau}{2}) e^{-j2\pi f \tau} dud\tau \tag{3}$$

Nevertheless, CWD will reduce the time-frequency aggregation of the spectrum while weakening the cross-interference terms [21]. In addition, when processing a multi-component short-time signal, the ability of CWD to suppress cross-interference terms will be relatively weakened. Therefore, it is very important to find a signal analysis method which can not only ensure the time-frequency aggregation, but also effectively inhibit the cross-interference terms.

VMD is an adaptive and non-recursive signal decomposition algorithm, which can decompose a nonlinear and non-stationary signal composed of multiple components into a series of band-limited intrinsic mode functions (BLIMFs). Each BLIMF is independent of each other and closely around the corresponding central frequency. Hence, VMD algorithm can be combined with CWD: firstly, a measured PD signal is decomposed into n BLIMFs containing only a single frequency component by using VMD, denoted as $u_i(t)$ ($i = 1, 2, \dots, n$). Then, the CWD is obtained based on the time-frequency analysis on these components. Finally, the linear superposition of each CWD is obtained to form the VMD-CWD spectrum diagram of the original signal. This method can not only eliminate the influence of cross-interference items, but also ensure good time-frequency aggregation. The specific implementation steps are shown in Fig. 2.

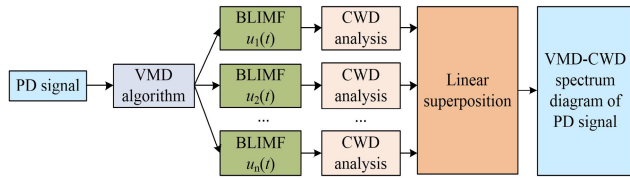


FIGURE 2. Time-frequency analysis of a PD signal based on VMD-CWD.

B. CROSS-LAYER FEATURE FUSION AND OPTIMIZATION CNN MODEL

CNN is a deep learning network mainly based on convolution operation. Due to its high invariance such as displacement and scaling in data processing, it is extensively applied in image recognition [22], [23]. The basic structure of CNN includes input layer, convolution layer, pooling layer and full connection layer [24].

Convolution layer is used for extracting and filtering features of input image. The specific process can be expressed as follows:

$$C_i = \sigma(C_{i-1} \otimes W_i + b_i) \tag{4}$$

where C_i represents layer i feature map of CNN, \otimes represents convolution calculation, W_i is the weight matrix of convolution kernel of layer i , b_i is the bias vector of layer i , and σ represents activation function.

The pooling layer is used for dimensionality reduction of the features extracted from the convolution layer, so as to reduce the calculation cost. The calculation process is shown in (5):

$$C_i = p(C_{i-1}) \tag{5}$$

where p stands for pooling operation. Currently, average pooling and max pooling are the two most commonly used pooling methods [25]. Since the max pooling is to select the maximum value in the pooling window to form the output features, which can better retain the effective information and reduce the amount of data processing, it has become the most widely used pooling method at this stage [26].

The full connection layer expands the obtained eigenmatrix as a vector and adjusts its dimension for comparison with the sample label [22].

Generally, CNN adopts small batch random gradient descent method to update the weight and conduct supervised training, which can effectively reduce the calculation cost. For the specific training process, please refer to the reference [27].

The cross-layer feature fusion and optimization CNN (CFFO-CNN) proposed in this paper is based on the traditional CNN, through the introduction of cross-layer connection to improve the network structure, select output features of several pooling layer that can effectively express the essential information of signal, and input them into the full connection layer after fusion, so that the network can extract the internal characteristics of signal from shallow

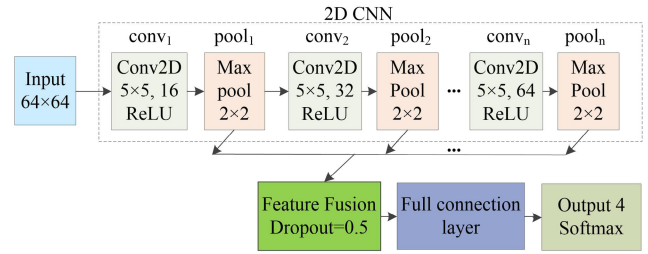


FIGURE 3. Structure of cross-layer feature fusion CNN.

and deep, so as to realize the recognition of PD signals. The basic idea is shown in Fig. 3. Firstly, the preprocessed image is input into the CNN for convolution transformation. The network contains n convolution layers and n pooling layers. Then, k ($k = 1, 2, \dots, N$) pooling layers' outputs are selected for feature fusion, and the features after fusion can be expressed as:

$$F = [X_{pool1}, X_{pool2}, \dots, X_{poolk}] \tag{6}$$

where $X_{pool1}, X_{pool2}, \dots, X_{poolk}$ are respectively the output characteristic map of the 1st, 2nd, ..., k-th pooling layer. In addition, in order to solve the phenomenon of dimension surge caused by feature fusion, dropout method is utilized to randomly shield some neurons in the full connection layer, for the purpose of reducing the calculation and preventing overfitting. Finally, the fused characteristic map is input into the full connection layer for classification.

Besides, the hyper-parameter is a critical factor that affects the learning rate and optimization ability of CNN [28]. In this paper, an optimization algorithm is applied to optimize the hyper-parameter of cross-layer feature fusion CNN, in order to accelerate the convergence rate of network training and obtain higher recognition accuracy.

C. PATTERN RECOGNITION OF PD SIGNALS BASED ON VMD-CWD AND CFFO-CNN

In this paper, we take the VMD-CWD image of PD signal as input, and use CFFO-CNN to complete the pattern recognition of different PD types. The flow chart is shown in Fig. 4.

The specific implementation steps are as follows:

Step 1: VMD decomposition and CWD time-frequency analysis were performed on the PD signals acquired from the experiments to obtain the VMD-CWD spectrum images. The obtained spectrum images were gray scaled into 64×64 gray level matrix and then divided into the training sample set and testing sample set;

Step 2: Build up the basic structure of CNN, set learning rate, batch size, iteration times and other relevant training parameters;

Step 3: Introduce cross-layer connection, compare the recognition effect of networks with different feature fusion layer structures, and determine the best feature fusion layer;

Step 4: Carry out the optimization algorithm to optimize the hyper-parameters of the cross-layer feature fusion CNN,

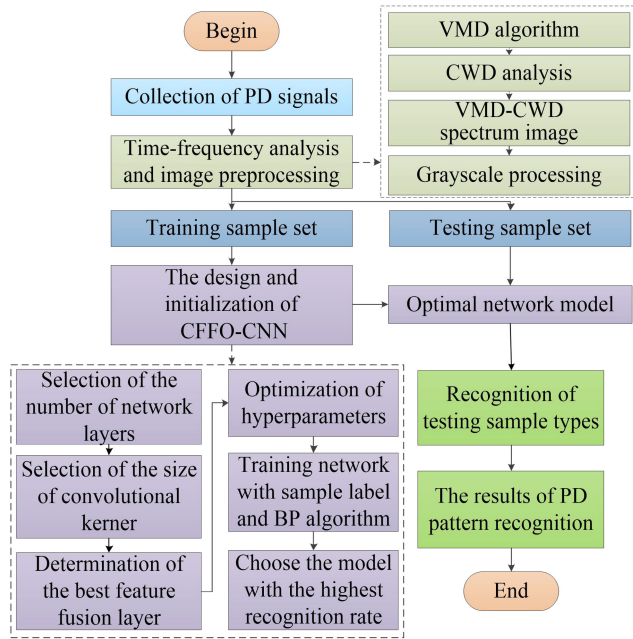


FIGURE 4. Flow chart of identification of PD signals.

so to obtain the CFFO-CNN. Then, training samples are used to train CFFO-CNN, and the network parameters are updated according to the back-propagation algorithm and the errors of the output and the sample label, until the training process of the whole network was completed;

Step 5: Use the trained CFFO-CNN to identify the type of PD signals in the testing sample set to obtain the recognition results.

IV. EXPERIMENT RESULTS AND ANALYSIS OF PD PATTERN RECOGNITION

A. ACQUISITION OF VMD-CWD SPECTRUM IMAGE

As mentioned above, each PD sample was decomposed by VMD, and CWD analysis was carried out on the BLIMFs obtained by decomposition to form the corresponding VMD-CWD spectrum diagram. The VMD-CWD of the four types of PD signals was shown in Fig. 5.

From Fig. 5, a PD signal is composed of multiple components with different center frequencies. It can be clearly seen from the VMD-CWD spectrum that the discharge time and the main frequency components of the four types of PD signals are observably different, which has practical physical significance. The VMD-CWD of PD signals have some differences in time domain and frequency domain, which carefully reflects the change process of signals in the time-frequency plane. Therefore, it can be used as a standard to distinguish different discharge types and applied to PD pattern recognition.

In order to further prove the superiority of VMD-CWD time-frequency analysis, four classic signal time-frequency analysis methods including STFT, HHT, WVD, CWD are performed on the PD samples to obtain the corresponding time-frequency spectrum, as comparison of VMD-CWD.

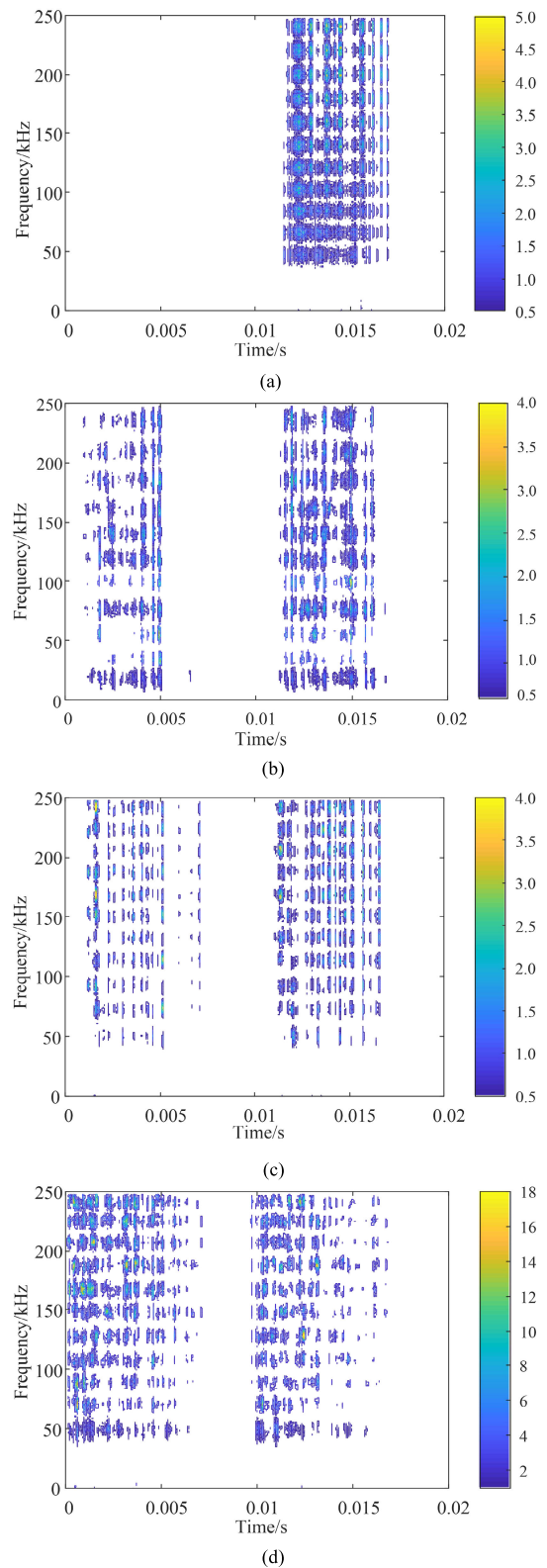


FIGURE 5. Time-frequency spectrum of PD signals based on VMD-CWD: (a) corona discharge, (b) plate-to-plate discharge, (c) needle-to-plate discharge, (d) floating discharge.

Each kind of spectrum is gray scaled into 64×64 matrix. Then, the gray-gradient co-occurrence matrix method is used

to extract 15 kinds of gray texture features of the matrix such as gradient dominance and gray entropy, as the input vector for BPNN. BPNN adopts a 3-layer structure with 15 hidden layer neurons, and the number of iterations is set to 100. The recognition results are shown in Table 2.

TABLE 2. Recognition results based on different time-frequency analysis methods.

PD type	Accuracy/%				
	STFT	HHT	WVD	CWD	VMD-CWD
Corona	100.0	94.0	100.0	100.0	100.0
Plate-to-plate	92.0	18.0	94.0	96.0	86.0
Needle-to-plate	34.0	54.0	28.0	40.0	66.0
Floating	26.0	88.0	60.0	72.0	100.0
Overall accuracy/%	63.0	63.5	70.5	77.0	88.0

According to Table 2, the recognition result based on VMD-CWD has the highest accuracy of 88.0%, which is 25%, 24.5%, 17.5% and 11% better than STFT, HHT, WVD and CWD time-frequency analysis method, respectively. Table 2 fully verifies the feasibility of VMD-CWD in PD recognition. In addition, the accuracy in the table is still relatively low, which may be due to the limitations of the features extracted manually that can not completely reflect the characteristics of the spectrum. Therefore, the complete VMD-CWD spectrum image can be taken as an input, and the convolutional neural network can be used to automatically extract the characteristic relationship between the VMD-CWD images, thereby directly identifying the four types of PD.

Besides, as mentioned above, for the purpose of reducing the calculation amount in the subsequent network training process, the VMD-CWD images were gray scaled into 64×64 gray level matrix (the gray value of each pixel was 0-255) as the input of CFFO-CNN network, the output type was 4.

B. DESIGN OF CFFO-CNN STRUCTURE

1) CONSTRUCTION OF BASIC NETWORK STRUCTURE

Before using network to identify PD signal types, the basic structure of the network should be determined according to the characteristics of the sample set. Generally, with the increase of network depth, the feature extraction ability of CNN is gradually enhanced, but the number of network parameters to be trained is also increased, and the risk of overfitting is easy to occur for the sample set with insufficient data [29].

Therefore, the performance of CNN with different number of network layers was tested. The first CNN had 2 convolution layers and corresponding pooling layers, the second CNN had 3 convolution layers and corresponding pooling layers, up to the final CNN of 5 convolution layers and corresponding pooling layers. Except the network depth, each CNN had the similar parameters, i.e., convolution kernel size of each network was set to 5×5 , the pooling kernel size was 2×2 , the batch size was 8, and the number of iterations was 200. Moreover, each CNN was trained by ReLU activation

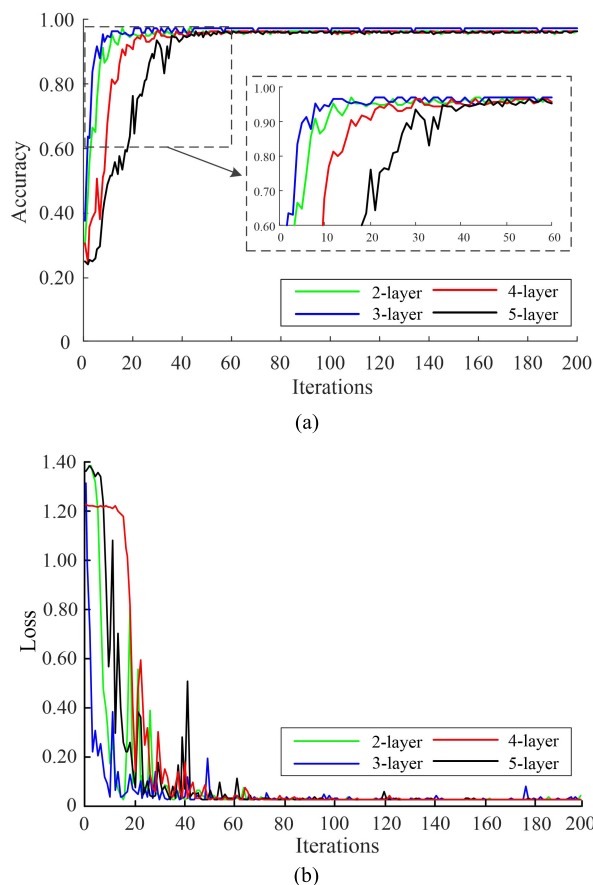


FIGURE 6. PD recognition results of different network depths: (a) accuracy curves in different network depths, (b) loss curves in different network depths.

function, max pooling method and the stochastic gradient descent with momentum (SGDM) algorithm. The accuracy and loss value curves of the four networks are shown in Fig. 6. In this paper, the preprocessing of PD signals is implemented in MATLAB 2018a and all deep neural networks are based on Pytorch framework. All above calculations are carried out on the Windows 10 operating system with Intel i7-4720HQ (2.6GHz) CPU and 8GB RAM.

As can be seen in Fig. 6, the network with 3 convolution layers can quickly converge and obtain high recognition accuracy. When there are two convolution layers, although the recognition rate of the network is rapidly improved at first, the accuracy fluctuates when the number of iterations reaches 140, unable to maintain a stable state, and the overall recognition rate is slightly lower than that of the network with 3 convolution layers. While the number of convolution layers is more than 4, the accuracy is relatively low with fewer iterations, and the convergence speed is slower. Experiments show that the network performance can be improved by increasing the depth of the network before the gradient disappears, but it is not that the deeper the network depth, the better the recognition effect. As a result, after comprehensive comparison, we choose the network model with 3 convolution layers and corresponding pooling layers.

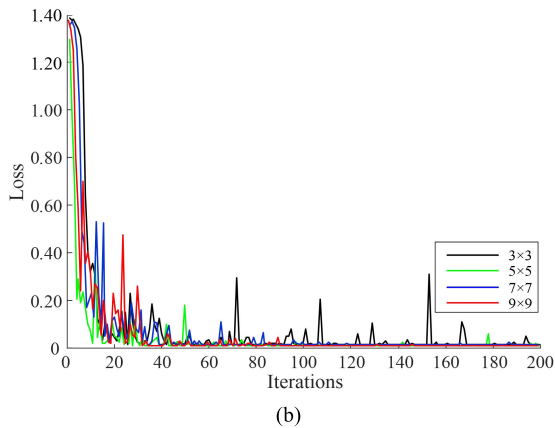
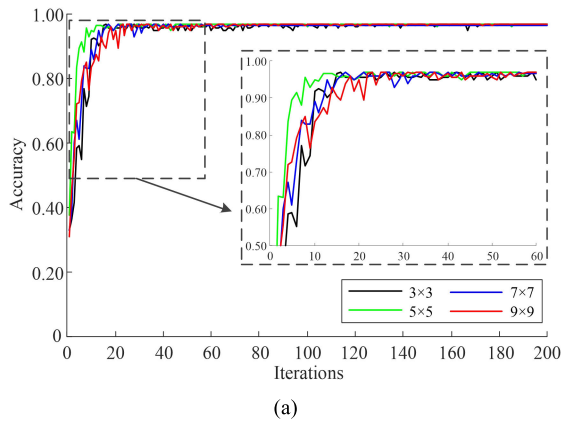


FIGURE 7. PD recognition results of different convolution kernel sizes: (a) accuracy curves in different convolution kernel sizes, (b) loss curves in different convolution kernel sizes.

In addition, the size of convolution kernel is of significant importance to the PD pattern recognition performance of CNN [29]. On the basis of the network structure consisting of 3 convolution layers and corresponding pooling layers, the size of the convolution kernel is set to 3×3 , 5×5 , 7×7 , 9×9 , respectively. The remaining parameters are kept unchanged, and the accuracy and loss value curves of the four networks are shown in Fig. 7.

From Fig. 7, it is easy to know that the accuracy of the network with convolution kernel size of 5×5 increases rapidly and reaches a stable condition quickly in the early iteration, which has the best recognition performance with accuracy of 97.0%. The smaller convolution kernel 3×3 will restrict the feature extraction capability of CNN, while the bigger convolution kernel size, such as 7×7 , 9×9 , will lead to the explosion of convolution parameters, thus reducing the computational performance and recognition accuracy. From this, considering the recognition accuracy and training time, the CNN with convolution kernel size of 5×5 has the best comprehensive pattern recognition performance.

2) OPTIMAL DESIGN OF NETWORK STRUCTURE AND HYPER-PARAMETERS

In order to further improve the network performance, based on the network structure designed in the previous section, four

network models with different feature fusion layer structures are designed by introducing cross-layer connection. Among them, cross-layer connection is not introduced in scheme 1. Scheme 2 fuses the output features of the first and third pooling layers as the input feature data of the full connection layer. Scheme 3 fuses the output features of the second and third pooling layers. Scheme 4 fuses the output features of all pooling layers into the full connection layer for classification. The PD signal recognition results of the four schemes are shown in Table 3.

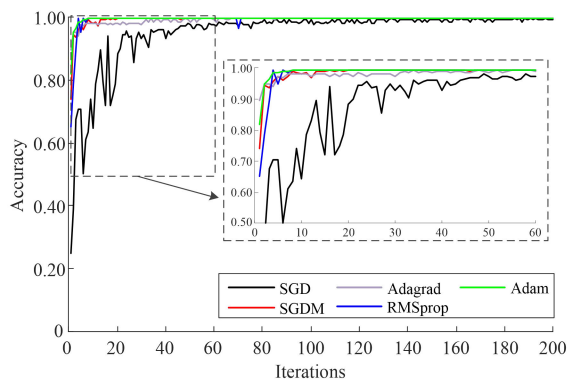
TABLE 3. Recognition results of PD signals under four different feature fusion layer structures.

Scheme	Fusion layer	Iterations	Accuracy/%
1	—	200	97.0
2	pooling layer1,3	200	99.0
3	pooling layer2,3	200	98.5
4	pooling layer1,2,3	200	98.0

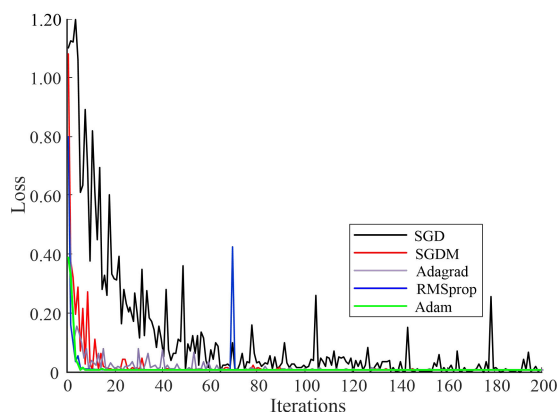
Comparing recognition results of networks with different fusion layer structures, it can be seen from Table 3 that the recognition accuracy of scheme 1 without cross-layer connection is 97.0%, which is the lowest among the four schemes. In scheme 2-4, the accuracy of scheme 4 is relatively low, which is due to the feature redundancy caused by full fusion, thus affecting the recognition effect. Comparatively, scheme 2 with the fused outputs of the first and third pooling layers has the highest recognition accuracy. By selecting appropriate shallow features and fusing them with deep features, the complementarity of the deep and shallow features is fully utilized, which not only extracts the local features of VMD-CWD spectrum, but also takes the overall change trend into account. Hence, the linear separability of the extracted features is enhanced by the cross-layer connection and the recognition accuracy is improved.

Moreover, in the training process of CNN, whether the hyper-parameter setting is appropriate is directly related to the recognition accuracy of the network [28]. If there is a deviation in hyper-parameter setting, it is easy to fall into local minimum value when using the stochastic gradient descent (SGD) method to train the network. To address this problem, five optimization algorithms are compared, namely SGD, SGDM, adaptive gradient (Adagrad) method, root mean square prop (RMSProp) method and adaptive moment estimation (Adam) method, which have an impact on the recognition effect of CNN. Fig. 8 shows the accuracy and loss value curves of the five optimization algorithms.

As can be seen from Fig. 8, the recognition accuracy and loss value of SGD show a large degree of oscillation at the beginning of the iteration, and the accuracy still fluctuates to varying degrees when the iteration reaches a certain number of times, which indicates the convergence rate is too slow. Although SGDM and Adagrad algorithm can converge quickly, the correct recognition rate of the network is relatively low at first. RMSProp converges faster than the above



(a)



(b)

FIGURE 8. PD recognition results under different optimization algorithms: (a) accuracy curves in different optimization algorithms, (b) loss curves in different optimization algorithms.

three algorithms, and the overall decline trend of loss curve is comparatively stable. However, when the iteration reaches the 70th time, there is a sudden change in the loss curve. In comparison, the overall pattern recognition accuracy of CNN with Adam algorithm is 99.5%, which meanwhile has the fastest convergence speed. Its loss value decreases rapidly and remains stable with no oscillation and mutation. Adam algorithm combines two strategies of deviation correction and momentum control on the basis of RMSprop, which makes the network more robust. All in all, the cross-layer feature fusion and optimization CNN (CFFO-CNN) constructed in this paper can give consideration to both recognition accuracy and convergence speed, which makes it more dominant in PD pattern recognition.

C. COMPARISON WITH TRADITIONAL RECOGNITION METHODS

In order to illustrate the advantages of proposed method in processing high-dimensional data, PD pattern recognition by traditional BPNN and SVM was also studied. Where, the maximum allowable error of BPNN is set to 0.001, and the number of iterations is 200; SVM adopts radial basis function kernel, the optimal kernel parameter and penalty factor were determined by grid search method. The number of training

TABLE 4. Recognition results of BP, SVM and proposed method under different number of samples.

Number of training samples for each type of PD signal	Accuracy/%		
	BPNN	SVM	CFFO-CNN
20	28.5	38.5	84.5
40	38.0	56.0	97.5
60	45.5	61.5	98.0
80	53.5	71.5	99.0
100	63.5	78.5	99.5

TABLE 5. Recognition results of PD signals with different feature extraction methods.

PD type	Accuracy/%			
	CFFO-CNN	HOG	LBP	GLCM
Corona	100.0	88.0	84.0	80.0
Plate-to-plate	100.0	86.0	78.0	76.0
Needle-to-plate	98.0	74.0	68.0	66.0
Floating	96.0	72.0	60.0	58.0
Overall accuracy/%	98.5	80.0	72.5	70.0

samples was increased successively, and the comparative results of the three methods are as shown in Table 4.

From Table 4, the pattern recognition accuracy of CFFO-CNN proposed in this paper is up to 99.5%, which is the highest among the three recognition methods. In three methods, BPNN has the lowest accuracy of 63.5%, which is due to the problems of overfitting and convergence difficulty when processing high-dimensional data. The recognition result of SVM is 78.5%, which is better than that of BPNN, but the accuracy is not significantly improved with large number of samples. In addition, as a two classification algorithm, SVM has a complex process in dealing with multiple classification problems. When the number of samples increases to a certain amount, it is difficult to select the kernel parameters and the training time is longer. Compared with BPNN and SVM, the average accuracy of CFFO-CNN is increased by 36% and 21%, respectively. Furthermore, the pattern recognition accuracy of CFFO-CNN was the highest among all methods under each sample quantity. Even in the case of a small number of samples, CFFO-CNN can still achieve a high accuracy of more than 80%, which demonstrates that deep learning network has more superior feature extraction capability that enables CFFO-CNN to fully analyze the time-frequency characteristics of the input image, extract the deeper features superior to the general statistical parameters, and obtain a better classification effect.

To verify the effectiveness of the features extracted by CFFO-CNN, the PD pattern recognition tests are carried out in classifier BPNN, meanwhile, three traditional image feature extraction methods including histogram of oriented gradient (HOG) [1], local binary pattern (LBP) [1] and gray-level co-occurrence matrix (GLCM) [30] are compared with the proposed method. The recognition results are shown in Table 5.

From Table 5, it is shown that the features extracted by CFFO-CNN achieve 98.5% accuracy in classifier BPNN,

which is higher than the traditional image artificial feature extraction methods. The features based on HOG, LBP and GLCM pay more attention to the shallow texture features of the VMD-CWD image, which cannot deeply mine the internal PD type information contained in the VMD-CWD image, resulting in a decrease in the recognition accuracy. The proposed CFFO-CNN with strong nonlinear mapping ability can adaptively extract the deep features of the image and use the complementarity of the deep and shallow features to improve the correct recognition accuracy.

D. COMPARISON WITH DIFFERENT DEEP LEARNING METHODS

To further demonstrate the superiority of CFFO-CNN over other deep learning algorithms in PD pattern recognition, SSAE, DBN, AlexNet and VGG-11 are constructed in this paper to identify PD types. The gray matrix of 64×64 is taken as the input vector of these network, so the dimension of input vector is 4096. The structure of SSAE is 4096-521-100-4 with two hidden layers. DBN adopts four hidden layers with 512, 128, 64 and 32 nodes respectively. Both AlexNet and VGG-11 adopt classic structure. The number of iterations is set to 200. The comparison of these methods on PD recognition is shown in Table 6.

TABLE 6. Recognition results based on different deep learning methods.

PD type	Accuracy/%				
	SSAE	DBN	AlexNet	VGG-11	CFFO-CNN
Corona	100.0	98.0	100.0	100.0	100.0
Plate-to-plate	100.0	94.0	98.0	100.0	100.0
Needle-to-plate	96.0	98.0	96.0	94.0	100.0
Floating	84.0	96.0	94.0	100.0	98.0
Overall accuracy/%	95.0	96.5	97.0	98.5	99.5

It can be clearly seen that the CFFO-CNN proposed in this paper gets the most excellent accuracy as high as 99.5% among the five methods. The accuracy of SSAE and DBN is relatively low, which may be because they fail to take the two-dimensional structure information of the image into account and cannot fully extract the edge and spatial features of the image. By comparison, the accuracies of Alexnet and VGG-11 based on convolution operation structure are slightly improved, but still lower than that of CFFO-CNN. Moreover, VGG-11 is complex with many training parameters, which will consume a lot of calculation and storage costs, resulting in slower running speed. To sum up, the proposed CFFO-CNN can directly process the 2D digital matrix of image and retain the original structure information of the data. Also, it can obtain a high recognition rate in each PD type, which indicates that CFFO-CNN can effectively detect and extract the inherent discharge type characteristics contained in the VMD-CWD spectrum image of PD signal.

Based on 400 training samples and 200 testing samples, CFFO-CNN was used to identify the VMD-CWD images of four PD types. At the same time, the proposed method

TABLE 7. Recognition results of different discharge types.

Recognition method	PD type	Accuracy/%	Overall accuracy/%
VMD-CWD spectrum image+CFFO-CNN	Corona	100.0	99.5
	Plate-to-plate	100.0	
	Needle-to-plate	100.0	
	Floating	98.0	
PRPD feature+BPNN	Corona	88.0	80.5
	Plate-to-plate	84.0	
	Needle-to-plate	72.0	
	Floating	68.0	

was compared with the traditional BPNN recognition results based on PRPD features, and the results are shown in Table 7.

As shown in Table 7, the overall accuracy based on VMD-CWD and CFFO-CNN is much higher than that of traditional PRPD features. Among the four PD types, the phase statistical characteristics of the needle-to-plate discharge and floating discharge are similar, which makes it difficult to distinguish them by PRPD features. In this paper, VMD-CWD distribution is introduced to analyze the two types of discharge from the time-frequency perspective, and it is found that there is a big difference in the time-frequency spectrum, which can be used to distinguish the characteristics of different PD types. Furthermore, the proposed CFFO-CNN model makes full use of the layer-to-layer information flow and avoids the loss of effective features. PD pattern recognition accuracy of CFFO-CNN for needle-to-plate discharge and floating discharge is 100% and 98% respectively, which is 28.0% and 30.0% better than traditional PRPD method. The results in Table 7 further show the superior feature extraction ability of CFFO-CNN, which makes it have better recognition performance than traditional PD pattern recognition methods.

V. CONCLUSION

In this paper, a method of PD pattern recognition based on VMD-CWD spectrum image and CFFO-CNN has been presented to identify the PD signals generated by four kinds of PD defect models constructed manually in the laboratory. The proposed technique was compared with traditional pattern recognition methods. The contributions and conclusions of the research are:

1) The VMD-CWD spectrum obtained by combining the VMD algorithm with the CWD distribution can effectively suppress the cross-interference terms in PD signal and has a high time-frequency resolution. In the meanwhile, the VMD-CWD distribution of different discharge signals is significantly different, which provides an effective time-frequency analysis method for PD signal pattern recognition.

2) The cross-layer feature fusion and optimization CNN model is put forward. The network structure and hyper-parameters are optimized by cross-layer connection and Adam algorithm, which achieves the optimal effect in the recognition accuracy and convergence speed. After fusion, the accuracy is as high as 99.5% increasing by 2.0% in comparison to non-fusion method, which effectively improves

the network performance. In addition, compared with the traditional artificial feature extraction methods, CFFO-CNN has better feature extraction ability.

3) When dealing with high-dimensional image data, CFFO-CNN network has higher recognition accuracy than the traditional shallow artificial intelligence recognition method, such as BPNN and SVM. Compared with the results of traditional PRPD feature recognized by BPNN, the validity of the proposed method based on VMD-CWD and CFFO-CNN is further verified.

In the practical engineering application, there are noise and other interference in the field measurement environment. The proposed method needs to be further studied on the basis of accumulating a large number of field PD signals and cases. In addition, there are also discharge phenomena under multiple PD sources, which makes it hard to distinguish the discharge type. This is a difficult problem in the field of PD pattern recognition, which will be our major research direction in the future.

REFERENCES

- [1] K. Firuzi, M. Vakilian, B. T. Phung, and T. R. Blackburn, "Partial discharges pattern recognition of transformer defect model by LBP & HOG features," *IEEE Trans. Power Del.*, vol. 34, no. 2, pp. 542–550, Apr. 2019.
- [2] M. Homaei, S. M. Moosavian, and H. A. Illias, "Partial discharge localization in power transformers using neuro-fuzzy technique," *IEEE Trans. Power Del.*, vol. 29, no. 5, pp. 2066–2076, Oct. 2014.
- [3] G. Luo, D. Zhang, Y. Koh, K. Ng, and W. Leong, "Time–frequency entropy-based partial-discharge extraction for nonintrusive measurement," *IEEE Trans. Power Del.*, vol. 27, no. 4, pp. 1919–1927, Oct. 2012.
- [4] F. C. Gu, H. C. Chang, F. H. Chen, C. C. Kuo, and C. H. Hsu, "Application of the Hilbert–Huang transform with fractal feature enhancement on partial discharge recognition of power cable joints," *IET Sci., Meas. Technol.*, vol. 6, no. 6, pp. 440–448, Nov. 2012.
- [5] T. A. C. M. Classen and W. F. G. Mecklenbrauker, "The Wigner distribution—A tool for time-frequency signal analysis. Part I: Continuous-time signals," *Philips J. Res.*, vol. 35, no. 3, pp. 217–250, 1980.
- [6] W. Li, J. Zhao, and S. Meng, "Partial discharge time-frequency spectrum analysis and extraction for power cable," in *Proc. Asia-Pacific Power Energy Eng. Conf.*, Shanghai, China, Mar. 2012, pp. 1–3.
- [7] N. Senroy, S. Suryanarayanan, and P. F. Ribeiro, "An improved Hilbert–Huang method for analysis of time-varying waveforms in power quality," *IEEE Trans. Power Syst.*, vol. 22, no. 4, pp. 1843–1850, Nov. 2007.
- [8] S.-C. Pei and J.-J. Ding, "Fractional Fourier transform, Wigner distribution, and filter design for stationary and nonstationary random processes," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4079–4092, Aug. 2010.
- [9] H.-I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 6, pp. 862–871, Jun. 1989.
- [10] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
- [11] S. Barrios, D. Buldain, M. P. Comech, I. Gilbert, and I. Orue, "Partial discharge classification using deep learning methods—Survey of recent progress," *Energies*, vol. 12, no. 13, pp. 2485–2500, Jun. 2019.
- [12] M. Karimi, M. Majidi, H. MirSaeedi, M. M. Arefi, and M. Oskuoee, "A novel application of deep belief networks in learning partial discharge patterns for classifying corona, surface, and internal discharges," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3277–3287, Apr. 2020.
- [13] J. Tang, M. Jin, F. Zeng, X. Zhang, and R. Huang, "Assessment of PD severity in gas-insulated switchgear with an SSAE," *IET Sci., Meas. Technol.*, vol. 11, no. 4, pp. 423–430, Jul. 2017.
- [14] Q. Che, H. Wen, X. Li, Z. Peng, and K. P. Chen, "Partial discharge recognition based on optical fiber distributed acoustic sensing and a convolutional neural network," *IEEE Access*, vol. 7, pp. 101758–101764, 2019.
- [15] S.-H. Bae and K.-J. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 595–610, Mar. 2018.
- [16] H. Song, J. Dai, G. Sheng, and X. Jiang, "GIS partial discharge pattern recognition via deep convolutional neural network under complex data source," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 2, pp. 678–685, Apr. 2018.
- [17] G. Li, X. Wang, X. Li, A. Yang, and M. Rong, "Partial discharge recognition with a multi-resolution convolutional neural network," *Sensors*, vol. 18, no. 10, pp. 3512–3538, Oct. 2018.
- [18] A. Kappeler, S. Yoo, Q. Dai, and A. K. Katsaggelos, "Video super-resolution with convolutional neural networks," *IEEE Trans. Comput. Imag.*, vol. 2, no. 2, pp. 109–122, Jun. 2016.
- [19] X. Wan, H. Song, L. Luo, Z. Li, G. Sheng, and X. Jiang, "Pattern recognition of partial discharge image based on one-dimensional convolutional neural network," in *Proc. Condition Monitor. Diagnosis (CMD)*, Perth, WA, Australia, Sep. 2018, pp. 1–4.
- [20] Y. Kim and T. Moon, "Human detection and activity classification based on micro-Doppler signatures using deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 8–12, Jan. 2016.
- [21] Y. Liu, P. Xiao, H. Wu, and W. Xiao, "LPI radar signal detection based on radial integration of Choi-Williams time-frequency image," *J. Syst. Eng. Electron.*, vol. 26, no. 5, pp. 973–981, Oct. 2015.
- [22] Y. Wang, J. Yan, Q. Sun, J. Li, and Z. Yang, "A MobileNets convolutional neural network for GIS partial discharge pattern recognition in the ubiquitous power Internet of Things context: Optimization, comparison, and application," *IEEE Access*, vol. 7, pp. 150226–150236, 2019.
- [23] F. Liu, C. Shen, G. Lin, and I. Reid, "Learning depth from single monocular images using deep convolutional neural fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 2024–2039, Oct. 2016.
- [24] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [25] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
- [26] D. Liu, Z. Wang, Y. Fan, X. Liu, Z. Wang, S. Chang, X. Wang, and T. S. Huang, "Learning temporal dynamics for video super-resolution: A deep learning approach," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3432–3445, Jul. 2018.
- [27] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1537–1541, Oct. 2016.
- [28] F. C. Soon, H. Y. Khaw, J. H. Chuah, and J. Kanesan, "Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition," *IET Intell. Transport. Syst.*, vol. 12, no. 8, pp. 939–946, Oct. 2018.
- [29] X. Peng, F. Yang, G. Wang, Y. Wu, L. Li, Z. Li, A. A. Bhatti, C. Zhou, D. M. Hepburn, A. J. Reid, M. D. Judd, and W. H. Siew, "A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1460–1469, Aug. 2019.
- [30] W. Zhou, Y. Liu, P. Li, Y. Wang, and Y. Tian, "Feature parameters extraction of power transformer PD signal based on texture features in TF representation," *IET Sci., Meas. Technol.*, vol. 11, no. 4, pp. 445–452, Jul. 2017.

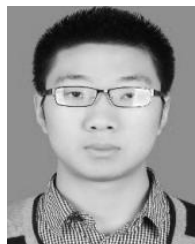


ANGRAN GAO received the B.E. degree in electrical engineering from the Xi'an University of Technology, Xi'an, China, in 2018. She is currently pursuing the master's degree with North China Electric Power University. Her research interests include online monitoring and fault diagnosis of electrical equipment.



YONGLI ZHU received the Ph.D. degree from North China Electric Power University, Beijing, China, in 1992.

He is currently a Professor with North China Electric Power University. His research interests include power system analysis and control, networked monitoring, and intelligent processing of big power data.



YI ZHANG received the B.E. degree in electrical engineering from Shandong Agricultural University, Tai'an, China, in 2017. He is currently pursuing the Ph.D. degree in electrical engineering with North China Electric Power University. His research interest includes intelligent diagnosis of power equipment.

• • •



WEIHAO CAI received the B.E. degree in automation from North China Electric Power University, Baoding, China, in 2018, where he is currently pursuing the master's degree. His research interests include condition monitoring of power apparatus and fault diagnosis of electrical equipment.