Identification of Partial Discharge Defects in Gas-Insulated Switchgears by Using a Deep Learning Method

Feng-Chang Gu*

Department of Electrical Engineering, National Chin-Yi University of Technology, Taichung 41107, Taiwan;
Corresponding author: Feng-Chang Gu (biobhst@hotmail.com).

ABSTRACT In this study, a novel method based on deep learning was developed for partial discharge (PD) pattern recognition. Traditional PD recognition methods are crucial for extracting features from PD patterns. The method of extracting crucial features is the key to PD pattern identification. The fractal theory is commonly used to determine the features of discharge patterns. The feature distribution of different defect types can be determined according to the fractal dimensions and lacunarity. However, finding fractal features is a complicated process. In this study, a PD image was entered as an input into a deep learning system to reduce the complexity of finding features. First, four defect type of gas-insulated switchgear (GIS) experimental models are established. Then, an LDP-5 inductive sensor (L-sensor) was used to measure the ground line signals caused by PD phenomenon. Second, these electrical signals were transformed into the most representative 3D (n–q–φ) PD patterns. Finally, a convolutional neural network was employed for PD image pattern recognition. A total of 160 sets of PD patterns were measured using a 15-kV GIS. The results obtained with the proposed method were compared with those obtained with the fractal method. The results revealed that the proposed method is easy to use and can easily distinguish various defect types. The proposed approach can determine the GIS insulation status and provide information to construction units for maintenance.

INDEX TERMS Convolutional neural network (CNN), deep learning, fractal theory, gas-insulated switchgear, partial discharge (PD), pattern recognition

I. INTRODUCTION

High-voltage (HV) power devices play a crucial role in power transmission and supply. An insulation state is the key factor to ensure power supply reliability. Any defect in HV devices may lead to serious power failure and equipment damage, thereby causing economic loss [1]–[3]. IEC 60270 is the most representative standard of the partial discharge (PD) definition. The standard indicates that PD is usually accompanied by physical phenomena, such as sound, light, heat, and chemical reactions [4]. PD detection is an effective approach for insulation monitoring and disaster prevention for power equipment. Different types of PD include internal discharge, corona discharge, and surface discharge, among which internal discharge is the most damaging. Internal discharge occurring in HV equipment may lead to breakdown in a short time [5]–[7].

Traditional switches use air as an insulation medium; however, these switches tend to be large and unstable. Thus, a GIS that uses sulfur hexafluoride (SF_6) as an insulating medium to quench the stretched arc and has replaced traditional switching. The GIS volume is <30% the volume of a traditional switch. In addition, SF_6 is a highly stable inert gas with exceptional insulation properties; thus, GIS exhibits excellent reliability [8]–[11]. However, if metal particles or other defects are present in a GIS gas tank, the local field strength may become excessively concentrated between a spacer and an electrode. These defects may cause PD, which results in a considerable deterioration of the insulation properties. Therefore, the production and assembly processes of a GIS have numerous requirements [12]–[15].

Traditional methods for PD feature extraction include the wavelet transform, Hilbert–Huang transform, and fractal theory, which has been successfully applied in PD signal

analysis. These methods can provide the feature distribution of different defect types through calculation [16]–[21]. However, complex mathematical operations must be performed in these methods. User experience and time must be considered to obtain feature parameters as inputs for a recognition system. The most representative recognition method is the back-propagation neural network (BPNN). The BPNN has widely been used for pattern recognition [22], [23]. However, determining the number of neural and hidden layers is a challenging task when using the BPNN. The CNN is a type of deep learning method that can be used to automatically extract features. This method can be used to avoid the complex mathematical operations caused by feature extraction. In recent years, numerous CNN-based techniques have been demonstrated to be effective in failure diagnosis [24]–[26]. To verify the accuracy and convenience of the CNN for defect identification in HV power equipment, a GIS PD image was input into the CNN in this study. The results obtained with the CNN were compared with those obtained using the fractal theory feature extraction method.

II. BASIC DEFINITION OF THE CNN

The CNN is a type of feed-forward neural network. The CNN structure includes input, pooling, fully connected, and output layers [24]. The CNN can be used to automatically extract efficient features, such as PD patterns and frequency spectra, from raw data. The theoretical principle of the CNN is described in the following text.

A. CONVOLUTIONAL LAYERS

When data enters the CNN, the convolutional layers are the first to perform calculations. Convolution is the most basic operation in image processing. The main purpose of convolution is to express pattern characteristics and complete feature mapping through weight sharing. The same feature may appear at different locations in an image. Thus, the same filter (shared weights) can be used to extract features for mitigating network complexity. The original image is mapped into different features after convolution. For mapping a 1D value y, x can defined as follows [24]:

$$y = conv(x, w', valid')$$
$$= (y(1), \cdots, y(t), \cdots y(n-m+1)) \in R^{n+m-1} \quad (1)$$

$$y(t) = \sum_{i=1}^{m} x(t+i-1)w(i) \quad (2)$$

where $w \in R^m$ represents a 1D filter, $w(i)$ is ith weight of the filter, m is the number of weights in the filter, n is the data point of the signal x, and $y(t)$ represent tth mapping value. Figure 1 illustrates the convolutional computation process. An input can be extended to the matrix according to Eq. (2). A 3D PD pattern is constructed using an A × B matrix. The convolutional layer of the input can defined as $X \in R^{A*B}$. The convolutional layer output can be obtained as follows:

$$Y_{cn} = f(X * W_{cn} + B_{cn}) \quad (3)$$

$$f(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (4)$$

where f is the activation function [rectified linear unit (ReLU)] [27], * represents the two-dimensional convolution operator, cn represents the number of convolution filters, and B_{cn} and W_{cn} are the bias and weight matrices of the cnth filter, respectively.
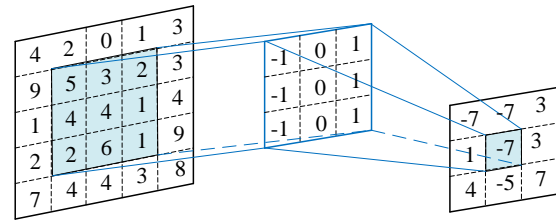

FIGURE 1. Structure of convolutional computation.

B. POOLING LAYER

The pooling layer uses a local correlation of the image to reduce spatial dimensions through down-sampling. Its benefits are the reduction of calculation loading and the prevention of the overfitting risk. The pooling process includes average, max, and overlapping pooling. Max pooling is commonly used in the CNN, and the convolution output layer Y_{cn} is defined as follows:

$$P_{cn} = \max_{S^{N \times M}} (Y_{cn}) \quad (5)$$

where N and M are the dimensions of the pooling matrix S. The maximum can be obtained from the matrix (N × M) in Y_{cn}. Alex Krizhevsky proposed AlexNet to alleviate the loss of image information by using overlapping pooling [28]. For example, if the input image is a 7 × 7 matrix, the feature detector is a 3 × 3 matrix, and the fixed stride is 1, then the output image size can be reduced to 5/7 of the input image size.

C. GRADIENT DESCENT

A multilayer perceptron (MLP) comprises neurons in multiple node layers. MLP uses gradient descent to obtain a minimum error and constantly self-corrects errors between the predicted and actual results by updating weights. The most common MLP method involves the calculation of the mean square error, which is defined as follows:

$$E(n) = \frac{1}{2} \sum_{k=1}^{m} (y_{ik}^n - d_{ik}^n)^2 \quad (6)$$

$$w(n+1) = w(n) - \eta \frac{\partial E(n)}{\partial w(n)} \quad (7)$$

where n is the nth iteration step; y_{ik} and d_{ik} are the actual and predicted outputs, respectively; m represents number of neurons; and η is the learning rate. The stochastic gradient

descent with momentum (SGDM) method can be used to randomly select samples for training, which may accelerate the learning process [28].

III. FRACTAL THEORY

The fractal features method has been used to describe complex or irregular shapes. It can be used to classify objects and textures present in irregular patterns [29]. This technique has potential for the classification of textures and for modeling complex physical processes. A fractal is characterized by fractal dimension and lacunarity.

A. FRACTAL DIMENSION

Differential box counting was used in this study to compute the fractal dimensions of the 3D PD pattern. The total number of boxes $N(L)$ required to cover the PD pattern is determined as follows:

$$N(L) = \sum_{m=1}^{N} \frac{S}{m} p(m, L) \tag{8}$$

where $p(m, L)$ represents the probability of m points in a box size L, N represents the number of possible points, and S is the number of image points. The box fractal dimension can be estimated by using the least square to fit $\{\log(L), -\log(N(L))\}$.

B. LACUNARITY

The lacunarity Λ can be quantify the denseness of a pattern surface, as suggested by Mandelbrot [29], [30]. The lacunarity can distinguish two patterns that different shape but maybe have the same fractal dimension. The basic concept of lacunarity is quantifying the gaps present in a given surface. The lacunarity is defined as follows:

$$\Lambda(L) = \frac{M^2(L) - [M(L)]^2}{[M(L)]^2} \tag{9}$$

The parameters $M(L)$ and $M^2(L)$ are defined as follows

$$m(L) = \sum_{m=1}^{N} m p(m, L) \tag{10}$$

$$m^2(L) = \sum_{m=1}^{N} m^2 p(m, L) \tag{11}$$

IV. PD RECOGNITION SYSTEM

A. 15-kV GIS EXPERIMENT MODEL

Three-phase, 15-kV, and 600-A pole-type GISs are widely applied in the power distribution systems used in factories. Common defects in a GIS include the presence of metal particles inside the gas tank and inappropriate fabrication caused by human errors. GIS defects cause PD the phenomena and influence the insulation reliability. This study developed experimental models for four types of common defects, which may result from human error during GIS fabrication. Figure 2 presents the four experimental models for the GIS defects. In all the experiments, the GIS was used at the rated pressure of 0.53 Kg/cm^2 at 20 °C. The four testing objects used in this study are described as follows:

Type 1: Internal conductor of the porcelain bushing at the secondary side, which contained oil grease.

Type 2: GIS tank containing metal particles and having dimensions of approximately 5 mm × 3 mm × 1 mm.

Type 3: A connecting rod of an operation handle containing welding protrusions with dimensions of 5 mm × 5 mm × 2 mm.

Type 4: An abrasion defect in metal ring. The defect depth and length were 2 and 10 mm, respectively.

B. PD MEASUREMENT SYSTEM

All the PD measurements were conducted in an HV laboratory for ensuring reduced noise and interference. The equipment used for the HV PD experiment included GIS experimental model, an LDP-5 detector (LDIC Company) with an inductive sensor (L-sensor), a voltage control panel, a step-up transformer, a national instrument data acquisition (DAQ) card (National instruments PXI-5105), and a laptop.
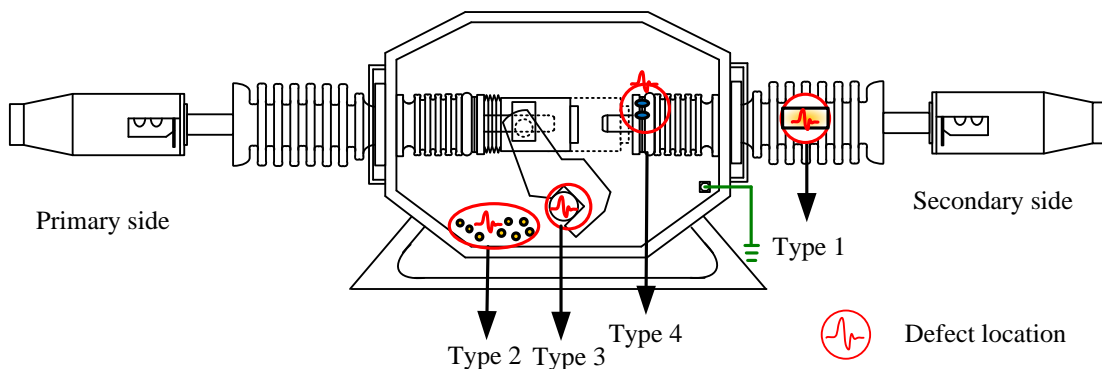

FIGURE 2. Test objects with typical GIS defects.

Figure 3 illustrates the structure of the PD experimental platform. The bandwidth of the LDIC L-sensor was approximately between 600 kHz and 20 MHz. In the experiment, a PXI-5105 edge trigger function was used to acquire the reference phase of the voltage. When the testing voltage was applied to the experimental model, a PD phenomenon was generated, which was then measured using the LDP-5 detector with an L-sensor. Finally, the PD signal was acquired using the DAQ card and then stored in the computer database.
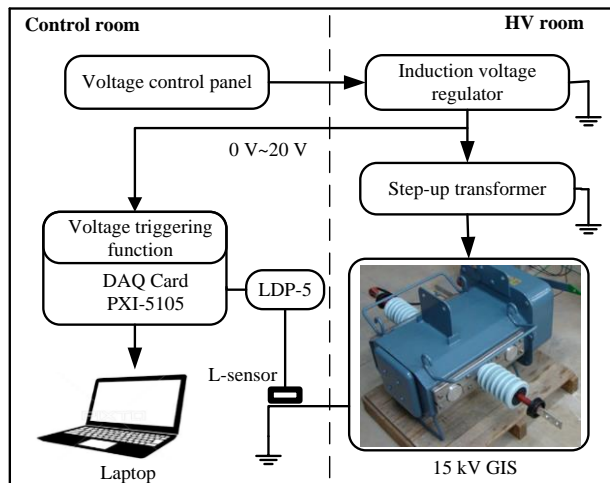

FIGURE 3. Structure of the experimental platform.

C. VOLTAGE TEST PROCEDURE

The standard IEC 62271-203 for HV switchgear testing [31] recommends that $U_{prestress}$ (prestress voltage) must be applied for 1 minute at the withstand voltage. The PD signal occurring during this period should be ignored. After 1 minute, the voltage reduces to the PD measurement voltage $U_{pd\text{-}test,\ ph\text{-}ea}$ (phase-to-earth). In this study, the U_r (rated voltage) of the GIS was 15 kV. The voltage $U_{prestress}$ was applied at 45 kV for 1 minute. Then, the voltage was reduced to $U_{pd\text{-}test,\ ph\text{-}ea} = 1.2U_r / \sqrt{3} = 10.4$ kV to measure the PD. After the completion of each measurement, the experimental equipment was left idle for 1 hour before the measurement of the next set of data. Figure 4 depicts the voltage application procedure for GIS PD measurement.
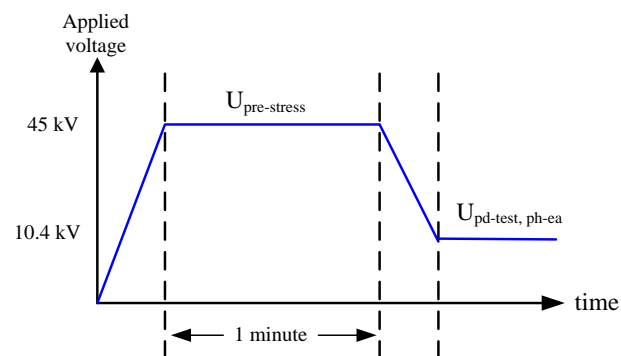

FIGURE 4. Voltage application procedure for GIS PD measurement.

D. PD ELECTRICAL SIGNAL

In the PD measurement, the PXI-5105 sampling rate was set to 2 M/s to acquire an electrical signal of 24 power cycles (60 Hz). The electrical signal was measured using the L-sensor. The LDP-5 envelope detection technology was then employed to convert the measured signal into a pulse signal. All the measured signals were converted into digital data and
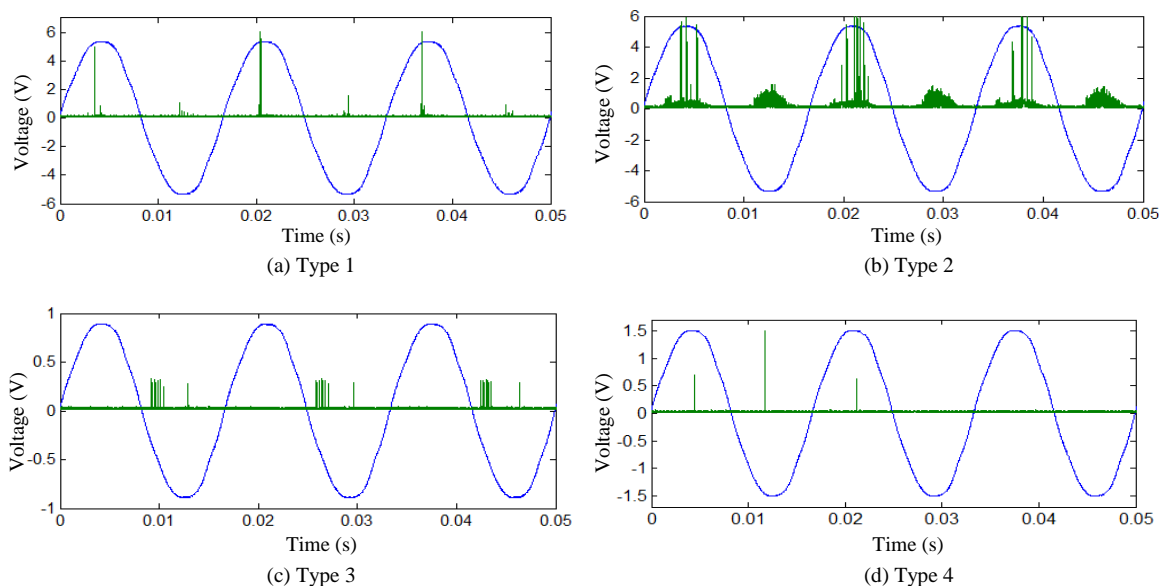

FIGURE 5. PD pulses for different experimental models in three power cycles.

stored in the laptop. Figure 5 depicts the PD pulse for different experimental models in three power cycles. The *x*-axis and *y*-axis represent the time and voltage amplitude, respectively. The blue and green lines are the 60-Hz power cycle and PD pulse, respectively. Figure 5 reveals that the PD pulses appeared in both positive and negative regions for the type 1 and type 2 defects. The voltage magnitude could reach 6 V and approximately 1.5 V in the positive and negative regions, respectively. The PD pulses exhibited the highest number of discharges for the type 2 defects, which were more widely distributed than other defects. For type 3 defects, PD pulses only occurred in the negative region, with only 0.4 V. For type 4 defects, PD pulses occurred in the positive and negative regions but discharge did not occur in every cycle. The lowest number of discharges was observed for the type 4 defects. The PD signal amplitude for the type 4 defects (approximately 0.5–1.5 V) was less than that for the type 1 and 2 defects.

V. EXPERIMENTAL RESULTS AND DISCUSSION

For each experimental model, 40 datasets were measured. A total of 160 dataset signals were converted into 3D (n–q–φ) PD patterns. Figure 6 illustrates the flowchart of proposed GIS PD pattern recognition method based on the CNN. The results obtained with the CNN were compared with those obtained using traditional fractal feature extraction and mean discharge method. The results are presented in the following text.
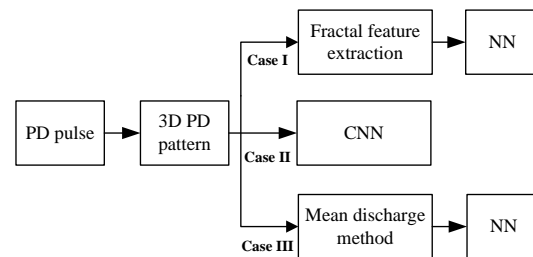

FIGURE 6. Flowchart of proposed GIS PD pattern recognition.

A. GIS 3D PD PATTERN

Figure 7 presents the 3D (n–q–φ) PD patterns transformed from the measured GIS PD signals. In the 3D PD patterns, the number of discharges was n, discharge magnitude was q, and phase angle was φ. The common unit of discharge magnitude is pC. Figure 7 reveals that the discharge amplitudes were larger for defect types 1 and 2 than for defect types 3 and 4. The amplitude can reach 500 pC at 70°–120°. The largest number of discharges in the negative region was observed for the type 2 defects. For type 2 defects, the signal amplitude was concentrated at 90 pC and n could reach 25. For the type 3 defects, discharge occurred only in the negative region and at an amplitude of approximately 20 pC. For the type 4 defects, discharges were concentrated at 50 pC in the positive region. Moreover, the number of discharges for type 4 defects was lower than that for the other defect types. The differences among defect types can be observed in the 3D PD patterns.
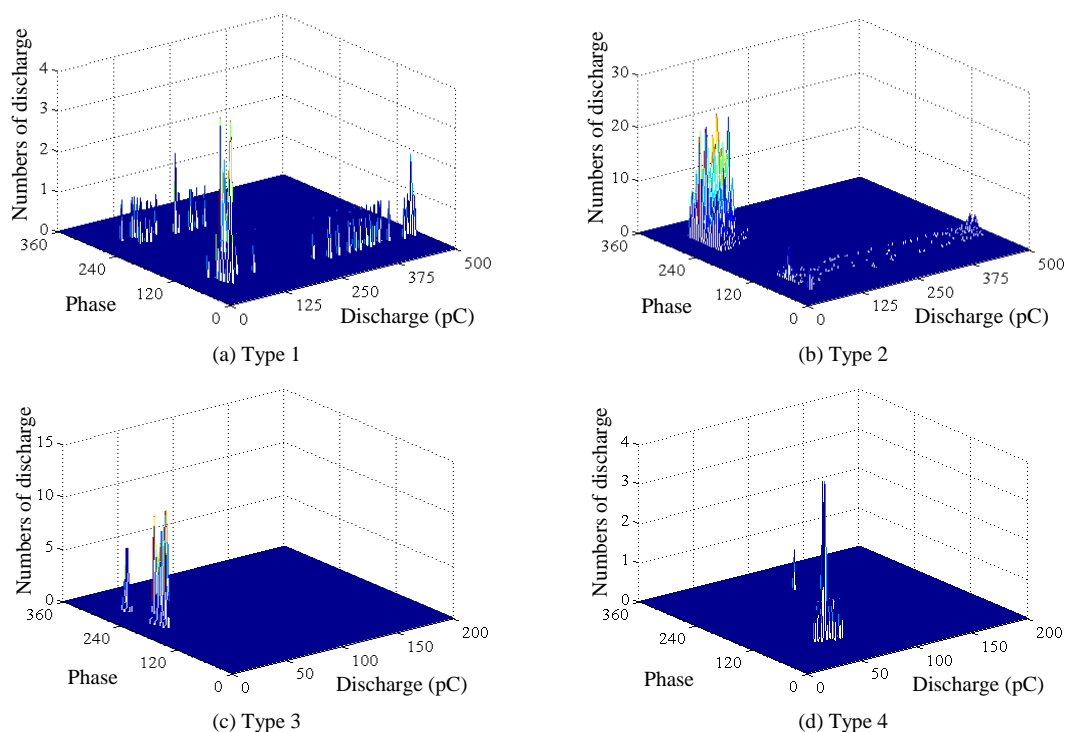

(a) Type 1

(b) Type 2

(c) Type 3

(d) Type 4

FIGURE 7. The 3D (n–q–φ) PD patterns.

B. FRACTAL FEATURES EXTRACTION

Fractal features are acquired from 3D PD patterns. The first step to acquire fractal features is to transfer the 3D pattern to a 256×256 matrix. Subsequently, by using different box sizes L (2, 3, 4, 5, 6, and 8), different values of $N(L)$ can be obtained. The values of $\{\log(L), -\log(N(L))\}$ can then be fitted to the fractal dimension. For determining the lacunarity, the first step is to transform the PD pattern into a 256×256 binary image. In our test, $L = 3$ was the optimal box size for computing $M(L)$ and $M^2(L)$. Finally, the lacunarity was obtained. Figure 8 illustrates the flowchart for the extraction of fractal features.
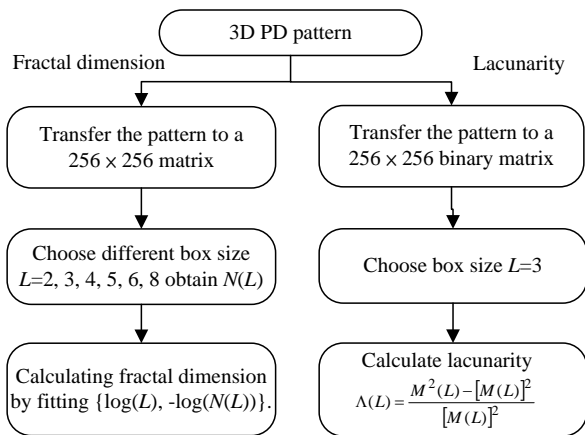

FIGURE 8. Flowchart for the extraction of fractal features from PD patterns.

Figure 9 presents the fractal features distribution extracted from traditional 3D (n–q–φ) PD patterns. Defect types 1 and 2 exhibited a similar fractal dimension in the interval of 2.2–2.4. However, the lacunarity distribution separated the dimensions of these two defect types. The fractal dimension of type 3 defects was concentrated in an interval of 2–2.1; thus, type 3 defects could be distinguished from the other defect types. The aforementioned feature extraction method can provide feature clusters that lie apart and are well distinguished.
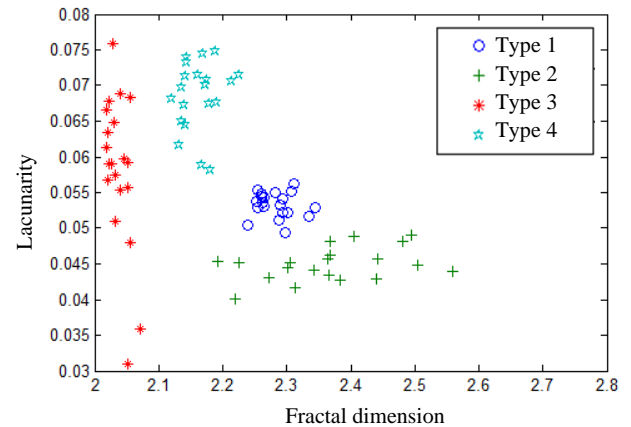

FIGURE 9. Fractal feature distribution extracted from 3D PD patterns.

C. PD RECOGNITION METHOD BASED ON CNN

Many CNN models have been applied in pattern recognition. In this study, the AlexNet model, which has excellent performance and is easy to use, was applied on the PD image for pattern recognition. Alexnet used the ReLU function as an optimization method, which accelerated the training process, reduced the iteration number, and effectively prevented overfitting [24], [32]. Figure 10 illustrates the structure of AlexNet in PD pattern recognition,
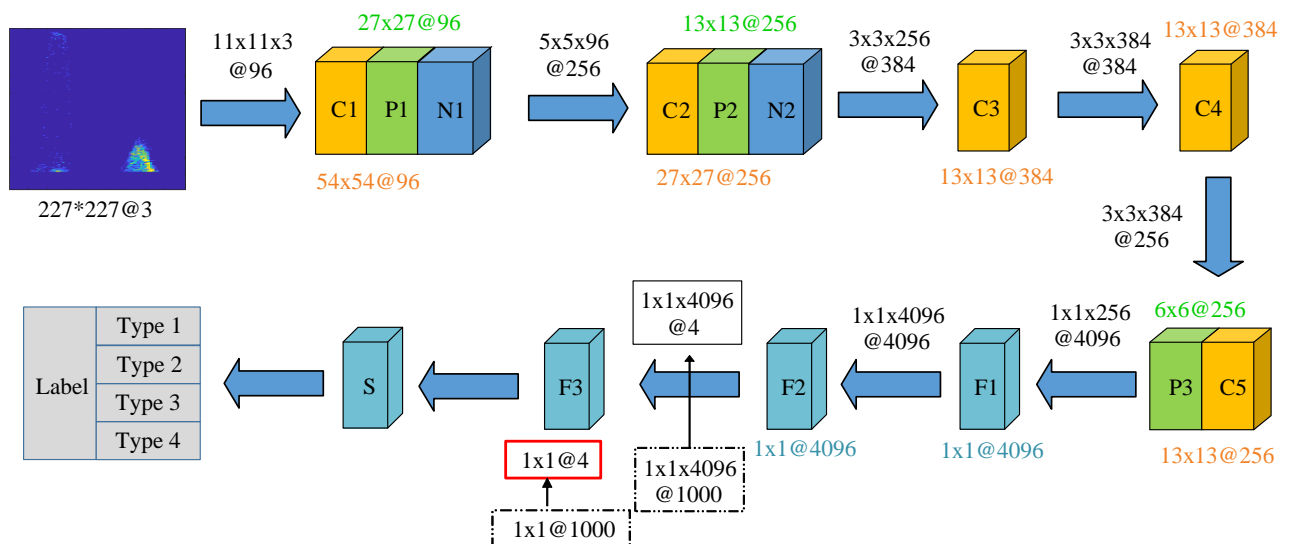

FIGURE 10. Structure of the proposed method.

where C, P, N, F, and S represent the convolutional, pooling, normalization, fully connected, and softmax layers. In the first step, the PD image with a size of 227 × 227 × 3 was the input, where 227 × 227 represents the image pixels and 3 represents the RGB image. The first convolution layer (*CL*) comprised 96 kernels of size 11 × 11 × 3 for feature extraction from the raw image, and the second *CL* comprised 256 kernels of size 5 × 5 × 96. After the *CL*, pooling and local response normalization layers performed their functions. The third and fourth *CL* without any post-processing layer comprised 384 kernels. The fifth *CL*, which was followed by a pooling layer, comprised 256 kernels of the same size as the kernels of the fourth *CL*. The ReLU was applied to the output of all *CL* and fully connected layers. The first and second fully connected layers comprised 4096 neurons and the last one comprised 1000 neurons. Finally the fully connected layer comprising only four neurons replaced the last fully connected layer. The softmax were arranged at the end of the structure, and then output layers estimate the possibility for each defect type.

D. MEAN DISCHARGE METHOD

In this experiment, 12 phase window features were extracted from a 3D PD pattern. The mean discharge features were extracted from phase angles 0°–360°. Each phase window is the mean discharge at a phase angle of 30°. Figure 11 shows the detailed data manipulation process. If the 3D pattern is represented by an $n \times m$ matrix, the mean discharge can be calculated as follows:

$$v_i = \frac{\sum_{j=1}^{m} \sum_{k=1}^{n} q_j n_{jk}}{\sum_{j=1}^{m} \sum_{k=1}^{n} n_{jk}} \quad for \ i = 1, 2, \cdots, 12 \tag{12}$$
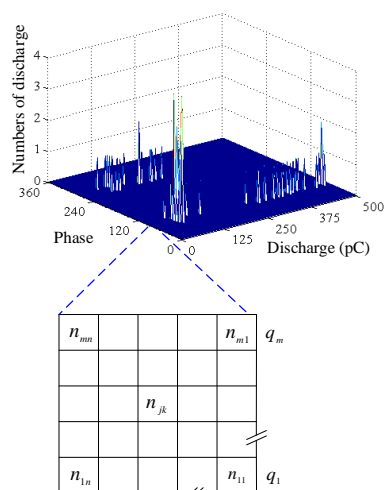

FIGURE 11. Detailed data manipulation process for mean discharge.

E. PATTERN RECOGNITION RESULT

In this study, 96 data (60%) were selected as the training pattern and the remaining 64 data (40%) served as the testing pattern. In Case I, fractal features were input to the three-layered BPNN because of their simplicity. The BPNN structure contained an input layer, a hidden layer, and an output layer. The maximum learning epoch, learning target, acceleration factor, and learning rate of the NN were 600, 10^{-10}, 1, and 0.1, respectively. Different numbers of neurons (5, 10, 20, and 30) were tested in the hidden layer to achieve optimum recognition results. The test results indicated that ten neurons in the hidden layer produced optimum recognition. The fractal feature accuracy rate reached 100% when no additional noise was added.

Case II was PD pattern recognition based on the CNN. The training and testing data were the same in Cases I, II and III. The learning rate was 0.001, and the momentum factor was 0.9, which can reduce oscillation and accelerate training (SGDM). The performance results obtained for mini-batch sizes of 10, 20, and 25 were compared. A large mini-batch requires more time and iterations. If the mini-batch was considerably small, under-fitting could occur. When the mini-batch size was 20 and the epoch size was 10, high accuracy was achieved. With the aforementioned structure, the accuracy rate of 64 testing patterns could become 100%. Case III was mean discharge pattern recognition based on the BPNN. The BPNN structure and parameters were the same as in Case I. The input has 12 mean discharge values extracted from phase angles of 0°–360°. The accuracy rate for the mean discharge method in PD recognition reached 100% under the same conditions.

To validate the three adopted PD recognition methods, MatLAB software was used for simulating the random white Gaussian noise, which was then added to the measured electrical signals. Because the measured signal may contain noise generated from environment or the detector. The magnitudes of added white noise is ±5%, ±10%, and ±15% of the maximum PD value in each defect model. Table I presents the recognition rate of PD patterns with different random white noise. The average recognition rate comprised three parts: fractal feature pattern recognition based on the BPNN (Case I), 3D PD pattern recognition based on the CNN (Case II), and mean discharge pattern recognition based on the BPNN (Case III). The results indicated a satisfactory recognition rate of approximately 85% was obtained in Cases I, II, and III at ±10% random white noise. However, the recognition rate of Case II could reach 81.3% even at ±15% random white noise. The results revealed that the CNN approach achieved a higher recognition rate and higher noise tolerance than the Case I and III.

IEEE *Access*

TABLE I
RECOGNITION ACCURACY RATE (%)

Method \ Noise	0%	±5%	±10%	±15%
Case I [*1]	100	95.3	85.9	78.1
Case II [*2]	100	96.9	84.4	81.3
Case III [*3]	100	95.3	84.4	76.6

[*1]: Fractal feature pattern recognition based on the BPNN.
[*2]: 3D PD pattern recognition based on the CNN.
[*3]: Mean discharge pattern recognition based on the BPNN.

The hardware used in this study was Intel core i7-7700 CPU @3.6 GHz with a Windows 10 operating system, and the graphics card was GeForce GTX 1050 Ti. Table II presents the time consumed in PD pattern recognition. Fractal feature extraction required complex iterations; thus, Case I required 56 s for each pattern feature extraction. No artificial feature was present in Case II. The features were automatically extracted from the CNN training process. The time consumed in Cases I, II and III was 56, 61 s and 75 s respectively. Moreover, in Case I, feature extraction required 56 s and pattern recognition required an additional approximately 0.5 s. In Case I, the processing of all the patterns (96 training and 64 testing patterns) required (56 s × 96) + 56 + (56 s × 64 + 0.5 × 64) = 9048 s. However, the time required in Case II was only 61 s + (1 s × 64) = 125 s. In the Case III the processing of all the patterns required (1 s × 96) + 75 + (1 s × 64 + 1 s × 64) = 299 s. Thus, the feature recognition process was considerably faster in Case II than in Case I and III. For a higher number of datasets than that used in this study, a considerable amount of time would be required for feature extraction under the conditions of Case I.

TABLE II
TIME CONSUMED IN PD PATTERN RECOGNITION (SECONDS)

Compare item \ Method	Case I	Case II	Case III
Feature extract in each pattern	56	—[*4]	1
Training process	56	61	75
Pattern recognition	0.5	1	1

[*4]: The CNN extracts features automatically from the training process.

VI. CONCLUSION

In this study, models were established for four GIS defect types and the CNN was successfully applied in PD image recognition. The novelty of the proposed method is that CNN can automatically extract features from PD images. They can avoid the complex mathematical operations required by feature extraction. The results showed that the proposed method is easy to use and has a high recognition rate. The traditional PD pattern recognition method must obtain features artificially. The proposed method substantially reduced the difficulty of PD pattern identification. The results obtained when 15% random white noise added to the measured electrical signal validated the efficiency and simplicity of the proposed approach. The recognition rate of the proposed method was equivalent to that of the BPNN method based on fractal feature extraction. However, the CNN can save considerable time in feature extraction. The proposed method provides another easy and effective way for detecting the failure probability of power apparatus.

REFERENCES

[1] Z. Guozhi, X. Zhang, H. Xingrong, Y. Jia, J. Tang, Z. Yue, T. Yuan, and L. Zhenze, "On-line monitoring of partial discharge of less-oil immersed electric equipment based on pressure and UHF," *IEEE Access*, vol. 7, pp. 11178-11186, Jan. 2019.

[2] X. Chen, Y. Qian, Y. Xu, G. Sheng, and X. Jiang, "Energy estimation of partial discharge pulse signals based on noise parameters," *IEEE Access*, vol. 4, pp. 10270-10279, 2016.

[3] G. V. Nagesh Kumar, B. P. Singh, J. Amarnath, and K. D. Srivastava, "Electric field effect on metallic particle contamination in a common enclosure gas insulated busduct," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 14, no. 2, pp. 334-340, 2007.

[4] IEC60270 "High voltage test techniques – Partial discharge measurement," 2001.

[5] Z. S. Zhang, D. M. Xiao, and Y. Li, "Rogowski air coil sensor technique for on-line partial discharge measurement of power cables," *IET Sci. Meas. Technol.*, vol. 3, no. 3, pp. 187-196, 2009.

[6] 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, 2020.

[7] M. Florkowski, B. Florkowska, M. Kuniewski, and P. Zydron "Mapping of discharge channels in void creating effective partial discharge area," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 6, pp. 2220-2228, 2018.

[8] C. S. Chang, J. Jin, C. Chang, T. Hoshino, M. Hanai, and N. Kobayashi, "Separation of corona using wavelet packet transform and neural network for detection of partial discharge in Gas-insulated Substations," *IEEE Trans. Power Deliv.*, vol. 20, no. 2, pp. 1363-1369, 2005.

[9] W. R. Si, J. H. Li, D. J. Li, J. G. Yang, and Y. M. Li, "Investigation of a comprehensive identification method used in acoustic detection system for GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 3, pp. 721-732, 2010.

[10] J. Tang, F. Liu, X. Zhang, Q. Meng, and J. Zhou, "Partial discharge recognition through an analysis of SF$_6$ decomposition products part 1: decomposition characteristics of SF$_6$ under four different partial discharges," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 19, no. 1, pp. 29-36, 2012.

[11] 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.

[12] 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.

[13] B. Tang, J. Tang, Y. Liu, and F. Zeng, "Comprehensive evaluation and application of GIS insulation condition Part 1: Selection and optimization of insulation condition comprehensive evaluation index based on multi-source information fusion," *IEEE Access*, vol. 7, pp. 88254-88263, 2019.

[14] D. A. Mansour, H. Kojima, N. Hayakawa, F. Endo, and Hitoshi Okubo, "Partial discharges and associated mechanisms for micro gap

delamination at epoxy spacer in GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 3, pp. 855-861, 2017.

[15] T. Hoshino, S. Maruyama, S. Ohtsuka, M. Hikita, J. Wada, and S. Okabe, "Sensitivity comparison of disc- and loop-type sensors using the UHF method to detect partial discharges in GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 19, no. 3, pp. 910-916, 2012.

[16] R. Hussein, K. B. Shaban, A. H. El-Hag, "Wavelet transform with histogram-based threshold estimation for online partial discharge signal denoising," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 12, pp. 3601-3614, 2015.

[17] J. Seo, H. Ma, and T. Saha, "Probabilistic wavelet transform for partial discharge measurement of transformer," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 2, pp. 1105-1117, 2015.

[18] X. Wang, B. Li, Z. Liu, H. T. Roman, O. L. Russo, K. K. Chin, and K. R. Farmer, "Analysis of partial discharge signal using the Hilbert–Huang transform," *IEEE Trans. Power Deliv.*, vol. 21, no. 3, pp. 1063-1067, 2006.

[19] F. C. Gu, H. C. Chang, and C. C. Kuo, "Gas-insulated switchgear PD signal analysis based on Hilbert-Huang transform with fractal parameters enhancement," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 4, pp. 1049-1055, 2013.

[20] L. Satish, and W. S. Zaengl, "Can fractal features be used for recognizing 3-d partial discharge patterns," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 2, no. 3, pp. 352-359, 1995.

[21] J. Li, C. Sun, and S. Grzybowski, "Partial discharge image recognition influenced by fractal image compression," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 15, no. 2, pp. 496-504, 2008.

[22] M. A. Abuzneid, A. Mahmood, "Enhanced uuman face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network," *IEEE Access*, vol. 6, pp. 20641-20651, 2018.

[23] T. Boczar, S. Borucki, A. Cichon, and D. Zmarzly, "Application possibilities of artificial neural networks for recognizing partial discharges measured by the acoustic emission method," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 16, no. 1, pp. 214-223, 2009.

[24] J. Wang, Z. Mo, H. Zhang, and Q. Miao, "A deep learning method for bearing fault diagnosis based on time-frequency image," *IEEE Access*, vol. 6, pp. 42373-42383, 2019.

[25] H. Song, J. Dai, G. Sheng, X. Jiang, "GIS partial discharge pattern recognition via deep convolution neural network under complex data sources," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 2, pp. 678-685, 2018.

[26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929-1958, 2014.

[27] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," *IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 8609-8613.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Proceedings of the 25th International Conference on Neural Information Processing Systems*, vol. 1, pp. 1097-1115, 2012.

[29] B. B. Mandelbrot, *The fractal geometry of nature*, New York: Freeman, 1983.

[30] B. B. Mandelbrot, "Plane DLA is not self-similar; is it a fractal that becomes increasingly compact as it grows?," *Physica A: Statistical Mechanics and its Applications*, vol. 191, no. 1-4, pp. 95-107, 1992.

[31] IEC 62271-203, "High-voltage switchgear and controlgear– Part 203: Gas-insulated metal enclosed switchgear for rated voltages above 52 kV," 2003.

[32] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84-90, May 2017.

FENG-CHANG GU was born in Taichung, Taiwan on August 3, 1985. He received his B.S. and M.S. degrees from National Chin-Yi University of Technology in 2007 and 2009, respectively and his Ph.D. degree from National Taiwan University of Science and Technology in 2013. In August 2018, he joined National Chin-Yi University of Technology, Taichung, Taiwan. His research interests include partial discharge detection, fault diagnosis, and pattern recognition.