Enhancing Data Applied in the Research of Partial Discharge Systematic Diagnosis

Yi Yang¹, Demeng Bai¹, Meng Liu¹, and Huiling Dou², *
¹State Grid Shandong Electric Power Research Institute, Shandong 510000, China
²PDStars Electric Co., Ltd., Shanghai 201100, China

*Corresponding author: douhuiling@pdstars.com

Abstract. The existing problems in the current partial discharge PRPS sample enhancement method have been summarized in this article, and a GIS UHF partial discharge data enhancement method has been proposed which is based on noise coupling, Gaussian blur and so on. It solves the problem of unable to collect a large number of defect maps and form a sample set due to the lack of GIS partial discharge cases, which limits the application of intelligent algorithms such as deep neural networks. This partial discharge PRPS sample enhancement method has been proved through application that it can effectively enhance the partial discharge sample set, make the samples evenly distributed, and meet the training needs of intelligent algorithms such as deep neural networks. The intelligent diagnosis algorithm trained using the enhanced sample set has a high diagnostic accuracy rate of 95%, which has practical application value.

1. Introduction
With the rapid development of power systems in recent years, the condition monitoring of power equipment has become increasingly important. Pay attention to the operating status of power equipment and arrange equipment maintenance in time to avoid failure and cause loss. Partial discharge (PD) is one of the important methods for monitoring the status of power equipment, which is significantly better than other methods for detecting power status [1-2]. Partial discharge is closely related to the reliability of power equipment, and has been the research focus of many researchers for a long time. At present, there are many methods for detecting partial discharge, including ultrasonic method, pulse current method, ultra-high frequency method, and optical method. The electromagnetic wave excited by partial discharge has a large component in the ultra-high frequency band, and there is less interference and high sensitivity between this frequency bands. Therefore, the ultra-high frequency method is widely used in partial discharge detection of key primary equipment such as GIS.

According to the electric field distribution of insulation defects in electrical equipment, the types of partial discharge defects can be divided into the following three basic types: ① metal tip discharge in gas or liquid insulation (corona); ② when there is an air gap inside the solid insulation Air-gap discharge (air-gap); ③ creeping discharge along the solid insulation surface (creeping). In addition, transformers and GIS also have floating potential discharges due to poor metal contact, such as loose screws, and free particle discharges due to the presence of metal particles. Partial discharges in electrical equipment are usually divided into the above five typical discharge types (defect types) [3-4]. The type of discharge is
closely related to the damage degree to the discharge. The industry converts UHF detection data into a PRPS map to visually display the characteristics of different defects and facilitate the diagnosis of the defects. The PRPS maps of 5 typical discharge types are shown below.

(a) Corona discharge PRPS  (b) Air gap discharge PRPS  (c) Creeping discharge PRPS
(d) Suspension discharge PRPS  (e) Particle discharge PRPS

Figure 1. Typical maps of partial discharge PRPS.

By analysing the partial discharge maps in Figure 1, it can be found that different types of defect maps have different signal distribution characteristics, which provides a basis for the application of artificial intelligence image classification algorithms [5].

The artificial intelligence algorithm based on deep learning wants to achieve a better application effect in the application of picture classification and recognition, the core of which lies in the overall size of the sample set and the sample balance between the categories, and the samples must be representative.

The GIS partial discharge UHF PRPS maps data mainly comes from the disintegration cases during the detection process. The number of disintegration cases occurs less frequently and the types of defects in the cases are not uniform, which makes it difficult to accumulate large-scale and well-balanced sample set by manual collection. It makes the application of deep learning artificial intelligence algorithms extremely difficult in the application of GIS partial discharge UHF maps.

This paper proposes a GIS ultra-high frequency partial discharge map data enhancement method to solve the problem of inability to collect defect maps and form sample sets due to the lack of GIS partial discharge cases, which limits the application of intelligent algorithms such as deep neural networks. This partial discharge PRPS sample enhancement method has been proved through application that it can effectively enhance the partial discharge sample set, make the samples evenly distributed, and meet the training needs of intelligent algorithms such as deep neural networks. The intelligent diagnosis algorithm trained using the enhanced sample set has a high diagnostic accuracy rate of 95%, which has practical application value.

2. Analysis of Current Status of GIS UHF Data Enhancement

In the application of deep learning algorithms, the sample set needs to be large in size and balanced in classification. Otherwise, it will easily cause problems of over fitting and poor generalization ability. Because GIS has a natural problem of low defect rate and uneven defect types, it is difficult to rely on
time to accumulate a sample set that meets the requirements. Sample size and classification equilibrium problems are the key issues currently facing.

2.1. Insufficient Sample Size
By researching it is found that when the deep learning algorithm is successfully applied in other industries, the sample size of each sub-classification is generally not less than 1,000, and the typical type of GIS UHF partial discharge is 5 classifications, so that the number of valid samples should be greater than 5000. In the case of balanced classification, the larger amount of sample size, the better.

Sample data can be collected by natural generation and laboratory simulation. Medical and security industries have successfully applied deep learning algorithms, due to their own characteristics of big data and can be simulated by means of laboratory simulation. Compared with the situation of GIS UHF localized maps, GIS operates in different environments (such as near roads, airports, railways, and factories), so that the interference is also various. The laboratory can only simulate a few typical defects by making a partial discharge model, which is far from the actual environment on the spot, and the problem of serious shortage of samples has been difficult to be effectively solved.

2.2. Uneven Sample Distribution
Due to the structural characteristics of GIS, the types of defects found on the site are mainly suspended defects, and the other four categories have too little probability of occurrence. According to statistics, in the cases of partial discharge found on the GIS site, the ratio of suspension, corona, air gap, particles, and surface partial discharge is about 5:2:1:1. As a result, the number of sample classifications is seriously unbalanced, it is difficult to accumulate effective sample sets, which is cannot support the application of intelligent algorithms.

3. Sample enhancement method of GIS partial discharge UHF PRPS maps
The GIS UHF partial discharge data enhancement method proposed in this paper is realized by designing a scheme based on noise coupling and Gaussian blur. Applying Gaussian blurring method based on the addition of common noise on the scene (such as mobile phone interference, radar interference, and microwave sulfur lamp interference), randomly perturb the RGB pixels to form new samples to train a network with stronger generalization ability. The overall process is shown in the following figure:

![Figure 2. Data enhancement flowchart of UHF PRPS map.](image)

3.1. Noise Coupling
The input sample data is converted into three-dimensional data inData [x] [y] [z] with phase as the x-axis, period as the y-axis, and amplitude as the z-axis. Typical radar noise, mobile phone noise, and microwave sulfur lamp interference map is converted into three-dimensional data with phase as the x-
axis, period as the y-axis, and amplitude as the z-axis. The three-dimensional data of the radar noise Radar \([x][y][z]\) and the three-dimensional data of the phone noise Phone \([x][y][z]\), microwave sulfur lamp interference three-dimensional data SulfurLamp \([x][y][z]\). The original input data is shown in the following figure:

Figure 3. Map of raw input data.

The typical maps of radar noise, mobile phone noise, and microwave sulfur lamp interference are shown as below:

Figure 4. (a) (b) (c) Typical map of superimposed noise.

Define the radar noise enhanced sample data as NoiseRadar \([x][y][z]\), the mobile phone noise enhanced sample data is NoisePhone \([x][y][z]\), the microwave sulfur lamp interference enhanced sample data is NoiseSulfurLamp \([x][y][z]\): Calculate the values of NoiseRadar \([x][y][z]\), NoisePhone \([x][y][z]\), NoiseSulfurLamp \([x][y][z]\) using the following formulas:

\[
NoiseRadar[x][y][z] = inData[x][y][z] + Radar[x][y][z] \\
NoisePhone[x][y][z] = inData[x][y][z] + Phone[x][y][z] \\
NoiseSulfurLamp[x][y][z] = inData[x][y][z] + SulfurLamp[x][y][z]
\]
After the above calculation, the enhanced sample data NoiseRadar \([x][y][z]\), NoisePhone \([x][y][z]\), NoiseSulfurLamp \([x][y][z]\) are obtained, and the sample data is restored into the map data, as shown below:

(a) Radar noise map overlaid with raw data  
(b) Mobile noise map overlaid with raw data  
(c) Raw data superimposed microwave sulfur lamp spectrum

Figure 5. (a) (b) (c) Overlaid maps.

3.2. Gaussian Blur
Take the noise-coupled map data as input, extract the RGB (Red, Green, and Blue) values of the input picture file, and set the centre point of the map data to zero, and draw the horizontal and vertical axis; calculate the weight matrix of each pixel in the graph with a two-dimensional Gaussian distribution function. The formula for calculating the two-dimensional high period distribution is as follows:

\[
G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}
\]  

In the formula, \(x, y\) stand for the horizontal and vertical coordinates of the distance from each pixel in the figure to the zero point, \(G(x, y)\) is the weight value from this point to zero point, \(\pi\) is PI, \(e\) is the natural constant, and \(\sigma\) is the standard deviation of the normal distribution. The value of \(\sigma\) is usually between 1 and 3, the larger the value, the smoother the image.

Taking an image of 3 * 3 pixels as an example, when the midpoint of the image is zero, the values of each point are shown in the following table:

| \(x\)  | \(y\)  | Value |
|-------|-------|-------|
| (-1, 1)| (0, 1)| (1, 1)|
| (-1, 0)| (0, 0)| (1, 0)|
| (-1, -1)| (0, -1)| (1, -1)|

When \(\sigma\) equals 1.5, the weights for each group \((x, y)\) are calculated, the results are shown in the following table:
Table 2. Calculate weights of $x$, $y$ when $\sigma$ equals 1.5.

|   |   |   |
|---|---|---|
| 0.04535 | 0.05664 | 0.04535 |
| 0.05664 | 0.07074 | 0.05664 |
| 0.04535 | 0.05664 | 0.04535 |

Normalize the weight matrix, calculate the sum of the weight values in the weight matrix $m$, and multiply each weight value in the weight matrix by $1/m$ to get the normalized weight matrix. As shown in the following table:

Table 3. Weight matrix normalized weights.

|   |   |   |
|---|---|---|
| 0.021709045 | 0.027113568 | 0.021709045 |
| 0.027113568 | 0.033863238 | 0.027113568 |
| 0.021709045 | 0.027113568 | 0.021709045 |

Update the RGB values of the image file with weights. For each weighted matrix of zero points, multiply the weighted value of each normalized weighted matrix by the pixel value at the weighted value position, and sum them as the new pixels of the zero point’s value. The RGB values are updated separately by this weight update method to obtain the updated three-color pixel values. Store the updated RGB image file as a new sample.

3.3. Sample library supplement
The new sample data obtained by the above enhancement method in this paper is added to the sample database and forms a new sample database with the original sample data, which lays the foundation for the application of artificial intelligence-based deep learning algorithms.

4. Enhanced sample application and comparative analysis
In order to verify the effectiveness of the enhancement methods proposed in this paper, the original sample set collected in the GIS field defect cases is used to enhance the data using the method proposed in this paper to obtain the enhanced sample. The set information is shown in Table 4:

Table 4. Sample set situation table.

| Number | Defect type | Number of original sample sets | Number of sample sets after data enhancement |
|--------|-------------|-------------------------------|----------------------------------------------|
| 1      | Corona discharge | 163                           | 1410                                         |
| 2      | Air gap discharge   | 151                           | 1380                                         |
| 3      | Particle discharge  | 138                           | 1400                                         |
| 4      | Suspension discharge | 1280                          | 1400                                         |
| 5      | Creeping discharge  | 312                           | 1350                                         |
| Sum    |              | 2044                          | 6940                                         |

By analysing the sample distribution in Table 4, we can get:

1) Compared with the two sample sets, the number of classification maps in the original sample set is poorly balanced and the sample size is small;
2) The sample set obtained by using the data enhancement method in this article, the number of samples in each classification is very close, and basically reaches an equilibrium state, so that the total sample size is about 3.4 times than the original sample set.
A deep learning training platform based on NVIDIA GeForce GTX 2080Ti GPU, E5-1607 v3 @ 3.10GHz CPU, 32GB memory, 512GB solid-state hard disk has been built in this article [6]. Using GoogLeNet's deep learning model. The same trained and adjusted are done separately for the original sample sets and data-enhanced sample sets to obtain two diagnostic models of GIS UHF partial maps, and the accuracy tests were tested separately. The test comparison of diagnostic accuracy is shown in Figure 6.

![Statistics of diagnostic accuracy](image)

**Figure 6.** Statistics of diagnostic accuracy.

By analysing Figure 6, it can be found that the diagnostic algorithm trained using the original data during diagnosing the test data, the diagnostic accuracy rate corresponding to different types of defect data varies widely, and the diagnostic accuracy rate of the facet defect is the lowest, only 39.3%, but the highest diagnostic accuracy of suspension defects is 87.5%. The reason is directly related to the number of defect types in the sample set. For the enhanced sample set, the number of maps of each classification is more balanced, the diagnostic algorithm after training is tested, the diagnostic accuracy is high, which is basically maintained at about 95%, and there is no obvious deviation. That is significantly better.

For example, in the transformer magnetic field monitored by the partial discharge diagnosis system, abnormal UHF (ultra high frequency) signals are continuously collected at the c-phase observation window of the transformer. The typical signal amplitude collected by the sensor is 51dB, there are two stable cluster discharge pulses, and the signal has power frequency correlation. The typical map is shown below:

![Typical Partial Discharge Signal](image)

**Figure 7.** Typical Partial Discharge Signal.
The floating electrode discharge was diagnosed by deep learning model. A large number of cases were diagnosed as continuous discharge of floating electrodes, and the system alarm was triggered. According to the diagnosis results of in-depth study and comprehensive analysis of abnormal data, technicians determined that there was a floating electrode discharge near the sensor, and the discharge was serious. After on-site confirmation and final positioning, it is confirmed that the iron yoke on the C phase of the transformer has discharge phenomenon. The discharge type was floating electrode discharge, which was consistent with the diagnosis results.

5. Conclusion

The data enhancement method proposed for GIS UHF partial discharge maps in this paper solves the problem of inability to collect a large number of defect maps and the formation of a balanced sample set due to the lack of GIS partial discharge cases, which limits the application of intelligent algorithms such as deep neural networks:

1) Using the noise coupling and Gaussian fuzzy data enhancement method proposed in this paper, a uniformly distributed sample set can be obtained. And the number of samples can be controlled by the application intensity of the two enhancement methods to obtain the required sample size.

2) Proven by practical applications, the intelligent diagnosis algorithm trained using the enhanced sample set has a diagnostic accuracy rate of 95%, which has good practical application value.

Acknowledgments

This work was financially supported by Science and Technology Project fund (NO: 520626190045) of State Grid Shandong Electric Power Company.

References

[1] DING Dengwei, GAO Wensheng, LIU Weidong. Analysis on the typical partial discharge using UHF detection method for GIS [J]. High Voltage Engineering, 2011, 37(3): 706-710.

[2] DING Dengwei, GAO Wenshen, YAO Senjing, et al. Analysis of excitation characteristics of ultra-high frequency electromagnetic waves induced by PD in GIS [J]. High Voltage Engineering, 2013, 39(8): 1893-1901.

[3] WU Zhangjian, LI Chengrong, QI Bo, et al. Partial discharge detection for GIS: a comparison of sensitivity between UHF and ultrasonic methods [J]. Modern Electric Power, 2010, 27(3): 31-36.

[4] DUAN Dapeng. GIS partial discharge detection and biomimetic pattern recognition based on UHF method [D]. Shanghai, China: Shanghai Jiao tong University, 2009:84-96.

[5] CHENG Xiaoli. Research on feature extraction method for partial discharge pattern recognition and correlation analysis of feature [D]. Chongqing University, 2003.

[6] MA Dongmei. Research on image retrieval based on deep learning [D]. Mongolian university, 2014.