Partial Discharge Diagnosis With Siamese Fusion Network

ZHIHONG HUANG[1], (Student Member, IEEE), WEI HUANG[1,2], XIANYONG XU[1], AND JIAN XIAO[1]
[1]State Grid Hunan Electric Power Corporation Ltd. Research Institute, Changsha 410007, China
[2]Hunan Xiangdian Test and Research Institute Corporation Ltd., Changsha 410007, China

Corresponding author: Zhihong Huang (zhih_huang@hnu.edu.cn)

ABSTRACT Partial discharge is a common fault type in the operation of power equipment. Recently, deep learning methods have shown great potential in partial discharge (PD) diagnosis. These methods construct a fitting relationship between input and output with mass training samples. Due to the scarcity of PD samples, it is hard to train a classification model. Thus, it is challenging to apply traditional deep learning methods for diagnosing PD. To overcome this issue, this paper introduces a siamese fusion network to diagnose PD. This method comprises three main steps. First, an ultra-high-frequency (UHF) sensor generates two spectrums from power equipment, including phase resolved partial discharge (PRPD) and phase resolved pulse sequence (PRPS). Based on a few-shot learning strategy, a support set is constructed. There are four different types of PD samples and a normal sample. Then, two siamese networks are employed to estimate similarity scores between a test sample and support set samples. One network measures similarity scores in PRPS, and the other measures similarity scores in PRPD. Based on similarity scores, initial diagnosis results are generated. Last, a simple and effective decision fusion technique fuses initial diagnosis results. The final diagnosis result can be generated by jointly exploiting complementary information in two spectrums. With limited training samples, experimental results show that the proposed SFN method can achieve an outstanding diagnosis performance, compared with several classical classification methods.

INDEX TERMS Partial discharge, fault diagnosis, power equipment, siamese network, few-shot learning, pattern recognition.

I. INTRODUCTION

When power equipment operates for a long time, various faults are inevitably produced. Among these faults, partial discharge (PD) is a serious fault type [1]. If this fault is not detected timely, the power equipment will deteriorate. Ultimately, PD will develop into a discharge breakdown or spark discharge, leading to tremendous economic losses [2], [3].

To identify PD, some methods have been designed, such as conventional pulse current, high-frequency current method, ultra-high- frequency (UHF) method, acoustic emission (AE) method, and dissolved gas analysis (DGA) method [4].

The associate editor coordinating the review of this manuscript and approving it for publication was Mehrdad Saif[.].

Among these methods, conventional pulse current and DGA methods fail to provide locations for PD sources, and the acoustic method is not suitable to diagnose faults inside power equipment. The UHF method is widely used for analyzing PD signals [5], which can diagnose the PD fault type, location, and risk degree. After operators obtain UHF spectrums, they need to analyze each sample manually and decide on PD types, according to the characteristics of PRPS and PRPD. However, with the rapid development of power systems, the corresponding operation tasks are increasing. The manual work has some disadvantages, such as low efficiency and low detection rate. It is hard to meet the needs of an intelligent power grid.

To overcome this problem, some deep learning methods have been proposed. These methods [2], [6]–[16] construct

a fitting relationship between input and output by using lots of training samples. Then, they can distinguish different PD fault types. Li *et al.* propose a convolution neural network (CNN) [7] with a deep architecture to realize ultra-high frequency (UHF) spectrums recognition in gas-insulated switchgear (GIS). The work [11] introduces a CNN-based PD classification system using transfer learning, which can reduce the noise in PD samples. Wan *et al.* [9] use long short-term memory (LSTM) and recurrent neural network (RNN) to diagnose PD in GIS. A stacked denoising auto-encoder method is proposed for PD diagnosis in different voltage cables of insulators [8]. Moreover, some traditional machine learning methods are proposed for PD diagnosis. For example, a multi-kernel multi-class relevance vector machine [2] is designed for PD diagnosis. Yang *et al.* [13] propose a low-rank radial basis function network to diagnose PD signals, by suppressing narrow-band noise. Qu *et al.* [14] combine discrete wavelet transform and LSTM techniques to diagnose PD faults. A fuzzy theory [15] has a simple model structure, achieving a fast diagnosis performance. The classical support vector machine (SVM) [16] also is applied in PD diagnosis.

It should be noted that previous works need a great number of PD samples to build classification models. However, PD samples are usually very scarce, and it is hard to collect sufficient samples. Thus, with limited training samples, these methods fail to generate high PD diagnosis accuracy effectively. The issue of how to use limited training samples for PD fault diagnosis is still an open question. Recently, the siamese network has been proved an efficient tool in image classification with limited training samples [17]. This network estimates the similarity between two input data via using a unique structure [11], [17]–[23]. Zhang *et al.* introduce a dual-path siamese CNN method for image classification [18]. Gao *et al.* propose a siamese training structure for data augmentation, improving classification performances [19]. A deep siamese framework with multitask learning is proposed for classification [20]. He *et al.* develop a siamese residual network with 3-D filters for image classification [23]. Liu *et al.* investigate the siamese network to train a linear classifier for image classification [11]. These methods have an outstanding performance with limited training samples.

Inspired by the siamese network, the motivation of this work is to diagnose PD with limited training numbers. This paper introduces a siamese fusion network (SFN). It consists of three main steps. First, an ultra-high-frequency (UHF) sensor produces two spectrums from power equipment, including phase resolved partial discharge (PRPD) and phase resolved pulse sequence (PRPS). Based on a few-shot learning strategy, a support set is constructed. There are four different types of PD samples and a normal sample. Then, two siamese networks are employed to estimate similarity scores between a test sample and support set samples. One network measures similarity scores in PRPS, and the other measures similarity scores in PRPD. Based on similarity scores, initial diagnosis results are generated. Last, a simple and effective decision fusion method fuses diagnosis results.
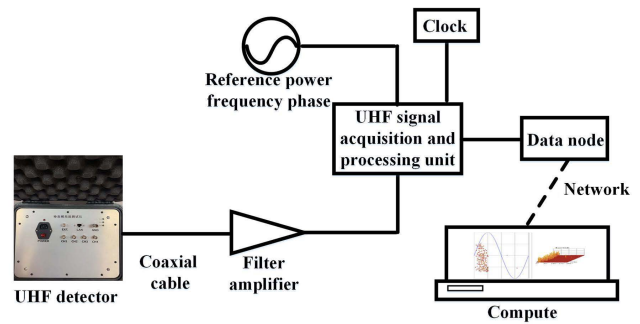

FIGURE 1. The UHF partial discharge detection system.

By jointly capturing complementary information in different spectrums, the final diagnosis result can be generated. The main contribution of this work is as follows:

(1) With a new strategic perspective, PD diagnosis has been modeled as a few-shot learning based classification problem. With limited training samples, a siamese fusion network identifies PD fault types. To our knowledge, the siamese network is applied in PD diagnosis for the first time.

(2) The proposed method fuses complementary information of different spectrums. It is demonstrated that the fusion of complementary features can result in an outstanding improvement in diagnosis performance.

The rest of this work is organized as follows. Several partial discharge types and the siamese network model are reviewed in section II. The proposed SFN method is introduced in Section III. Section IV presents the experimental results and analysis. Conclusions are given in Section V.

II. RELATED WORKS
A. PARTIAL DISCHARGE TYPES
Ultra-high frequency (UHF) sensors can obtain electromagnetic wave spectrums with a frequency ranging from 300MHz to 3GHz. The UHF sensor can sense and record the characteristic information of the UHF partial discharge signal, including amplitude, frequency, phase, etc. PRPS and PRPD spectrums are generated based on the above feature information. The UHF sensor converts UHF signals into electrical signals. Then, electrical signals are converted into the digital signal through the UHF signal acquisition and processing unit. The digital signal can be stored in the computer. Fig. 1 shows the UHF partial discharge detection system [24].

Each PD type usually has distinctive characteristics in PRPS and PRPD spectrums [25]. Specifically, PRPD can reflect the relationship between multiple cycles, discharge spectrum amplitude, and discharge frequency [26]. This spectrum is generated by using a wavelet transform or the Hilbert yellow transform [27]. It can be served as a classification basis for various discharge types. In addition, PRPS is a 3-dimensional distribution of discharge amplitudes, regarding phase and power frequency cycle. This spectrum can capture local discharge information such as phase distribution.
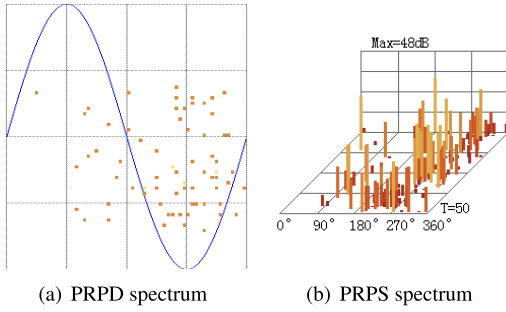
(a) PRPD spectrum   (b) PRPS spectrum

FIGURE 2. The spectrums of the free metal particle discharge.

There are corona discharge, suspended potential discharge, free metal particle discharge, and insulation gap discharge.

Fig. 2 displays the spectrums of the free metal particle discharge. In Fig. 2(a), the polarity is not obvious. In Fig. 2(b), the power distribution range is distributed, and its discharge interval is unstable. The spectrums of the suspended potential discharge are shown in Fig. 3. The polarity is more obvious, and its pulse is relatively stable. Besides, its discharge interval presents the cycle distribution. Fig. 4 illustrates the spectrums of the insulation gap discharge. It doesn't have an obvious pola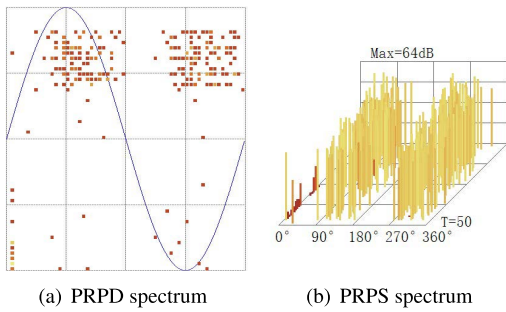r effect, and the pulse amplitude is more dispersed. The spectrums of the corona discharge are presented in Fig. 5. Its polarity is obvious, and the pulse amplitude is dispersed.

(a) PRPD spectrum   (b) PRPS spectrum

FIGURE 3. The spectrums of the suspended potential discharge.

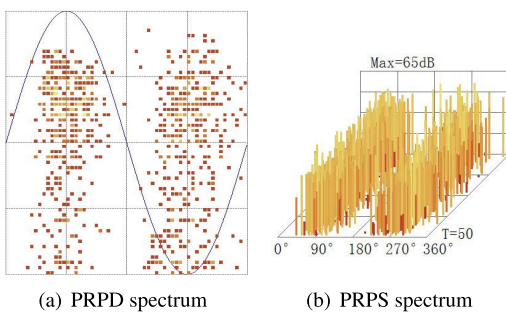(a) PRPD spectrum   (b) PRPS spectrum

FIGURE 4. The spectrums of the insulation gap discharge.

B. SIAMESE NETWORK

The siamese network is a unique network architecture. Traditional deep learning models aim to classify input
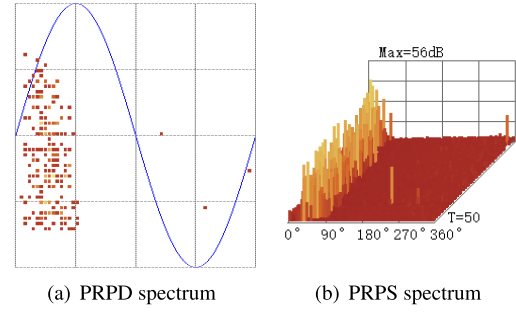

(a) PRPD spectrum   (b) PRPS spectrum

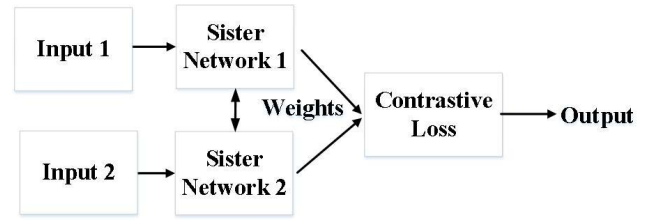**FIGURE 5.** The spectrums of the the corona discharge.

FIGURE 6. The architecture of the siamese network.

images, and the siamese network focus on exploiting the similarity between input images.

Specifically, a siamese network usually contains two same sister networks. The last layers of the sister networks are fed to a contrastive loss function, estimating the similarity between two input images. The architecture of the siamese network is shown in Fig. 6. Each image is fed to one of two networks.

According to the reference work [17], the siamese network is optimized by using a contrastive loss function. This loss function can well express the matching degree of paired samples. Besides, it can be used to train the model of feature extraction. The loss function [17] is as follows:

$$(1 - Y)\frac{1}{2}(D_w)^2 + (Y)\frac{1}{2}[max(0, m - D_w)]^2, \qquad (1)$$

where D_w is the Euclidean distance between the outputs of the siamese networks. When $Y = 1$ (i.e. the samples are similar), $\frac{1}{2}[max(0, m - D_w)]^2$ remains in the loss function. If the Euclidean distance in the feature space is large, it indicates that the current model is not good, and the loss is increased. When $Y = 0$ (i.e. the samples are not similar), the loss function is $\frac{1}{2}(D_w)^2$. The Euclidean distance of the feature space is small, and the loss value will become larger.

The Euclidean distance is represented as follows:

$$D_w = ||G_w(X_1) - G_w(X_2)||, \qquad (2)$$

where X_1 and X_2 are two input data. $G_w(X_1)$ means the output deep feature of X_1, and $G_w(X_2)$ means the output deep feature of X_2.

III. THE PROPOSED SFN METHOD

Fig. 7 shows the flowchart of the proposed SFN method. The SFN method consists of three main steps. First, based
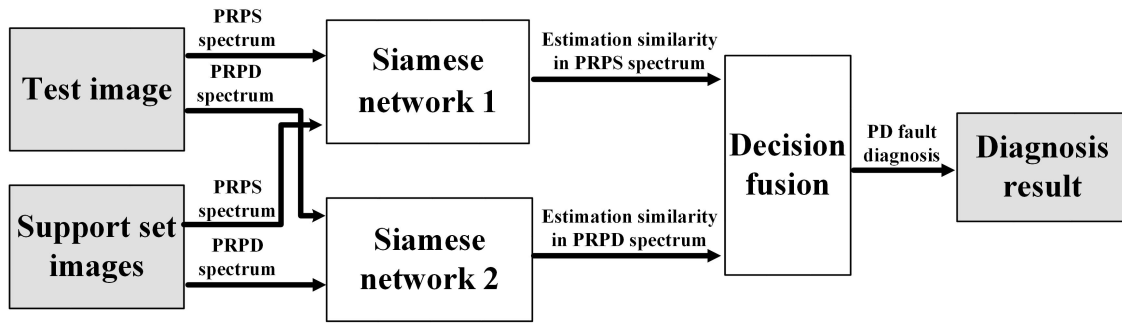
FIGURE 7. The flowchart of the proposed SFN method.

on a few-shot learning strategy, a support set is constructed, containing some labeled samples, such as four PD types and a normal type. Each sample has PRPS and PRPD spectrums, obtained by the UHF sensor. Then, two siamese networks are employed to estimate the similarity between a test sample and support set samples. One network measures the similarity in PRPS, and the other measures similarity in PRPD. According to two similarity scores, initial diagnosis results can be generated. Last, a simple and effective decision fusion method is designed to fuse initial diagnosis results. By jointly exploiting complementary information, the final diagnosis result is generated.

A. SUPPORT SET GENERATION
The UHF sensor is employed to obtain PRPD and PRPS spectrums. A test sample X_t contains the PRPD spectrum X_t^{PRPD} and the PRPS spectrum X_t^{PRPS}:

$$X_t = [X_t^{PRPD}, X_t^{PRPS}]. \tag{3}$$

A support set is constructed, containing a few labeled samples. It includes corona discharge, suspended potential discharge, free metal particle discharge and insulation gap discharge, and normal samples. Each type has one image. The images in the support set $X_s(i)$ also contain the corresponding PRPD spectrum $X_s^{PRPD}(i)$ and the corresponding PRPS spectrum $X_s^{PRPS}(i)$:

$$X_s(i) = [X_s^{PRPD}(i), X_s^{PRPS}(i)], \tag{4}$$

where $i = 1,2,3,4,5$. $X_s(1)$ represents the corona discharge sample, $X_s(2)$ means the suspended potential discharge sample. $X_s(3)$ serves as the free metal particle discharge sample. $X_s(4)$ is the insulation gap discharge sample, and $X_s(5)$ means the normal sample.

B. SIAMESE NETWORK-BASED SIMILARITY ESTIMATION
In this step, two siamese networks are employed to estimate the similarity between the test sample and support set samples. One siamese network measures the similarity of the PRPD spectrum, and the other estimates the similarity of the PRPS spectrum.

In Fig. 8, each siamese network has two VGG-16 models with the same network structure and shared parameters.
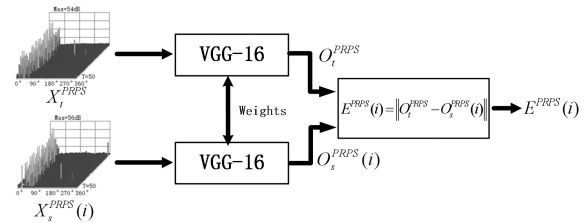

FIGURE 8. The siamese network used for similarity estimation in PRPS.

The VGG-16 model has multiple convolution layers, a sigmoid activation function, a mean-pooling layer, and a fully connected layer. The test sample and support samples are served as two inputs images, capturing the deep feature information. The input spectrums are subjected to the VGG-16 model, and the output deep information can be obtained:

$$O_t = G(X_t), \tag{5}$$

where $G(X_t)$ represents a step that uses the VGG-16 model to extract the deep feature information from the test sample X_t. O_t is the output deep information of X_t, including two elements:

$$O_t = [O_t^{PRPD}, O_t^{PRPS}]. \tag{6}$$

Similarly, $G(X_s(i))$ means a step that adopts the VGG-16 model to capture the deep feature information from support set samples $X_s(i)$. It is represented as follows:

$$O_s(i) = G(X_s(i)), \tag{7}$$

where $O_s(i)$ means the output deep information of $X_s(i)$, and it contains two elements:

$$O_s(i) = [O_s^{PRPD}(i), O_s^{PRPS}(i)], \tag{8}$$

where $O_s^{PRPD}(i)$ represents the output of $O_s^{PRPD}(i)$ and $O_s^{PRPS}(i)$ represents the output of $O_s^{PRPS}(i)$. Due to the advantage of generalization ability, the VGG-16 model is adopted to exploit the deep information.

Eq. (2) measures the PRPD similarity score:

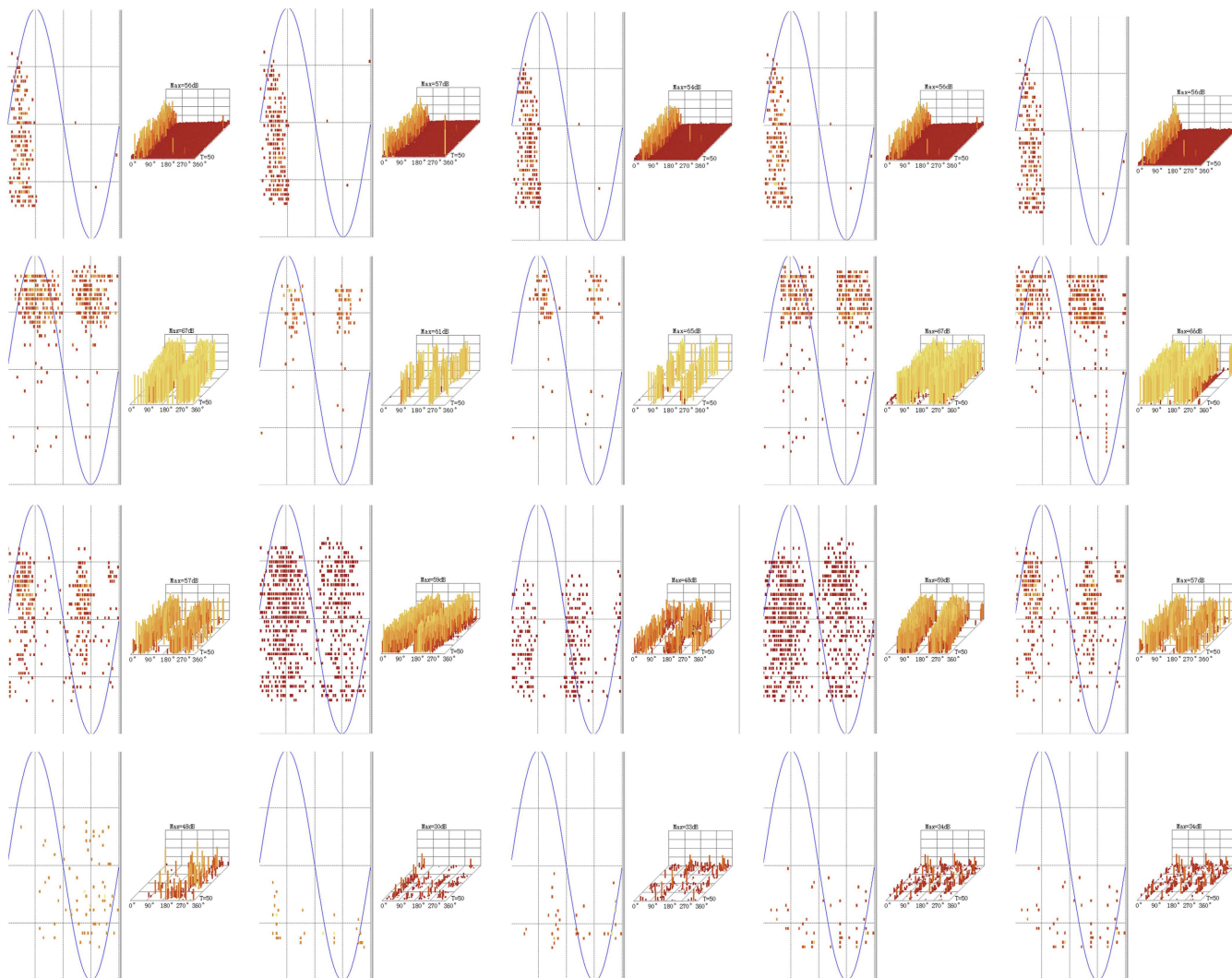$$E^{PRPD}(i) = ||O_t^{PRPD}, O_s^{PRPD}(i)||. \tag{9}$$

FIGURE 9. Part of PD samples.

TABLE 1. The details of the experimental date set.

Type	Training samples	Testing samples	Sum
Corona discharge	10	50	60
Suspended potential discharge	10	25	35
Free metal particle discharge	10	40	50
Insulation gap discharge	10	35	45
Normal sample	10	120	130

Similarly, the PRPS similarity score is defined as follows:

$$E^{PRPS}(i) = ||O_t^{PRPS}, O_s^{PRPS}(i)||. \tag{10}$$

The value of the output results is normalized to [0, 1].

C. DECISION FUSION BASED PD DIAGNOSIS

Since the UHF signal is complex and various, a single spectrum fails to reflect the UHF signal fully. The diagnosis accuracy with a single spectrum fails to obtain outstanding performances effectively. The PRPD and PRPS spectrums can reflect the UHF signal from different views since the two spectrums are from the same test sample. There is strong complementary information in the PRPD and PRPS spectrums. To exploit the complementary information in different signals, a simple and effective decision fusion technique is adopted to utilize two similarity scores:

$$E(i) = E^{PRPD}(i) + E^{PRPS}(i) \tag{11}$$

TABLE 2. Classification accuracies (%) obtained by different methods on the PRPD spectrum.

Class	SVM [16]	VGG-16 [28]	AlexNet [29]	SFN-WO	SFN
1	42.27	56.29	52.17	76.22	**82.08**
2	32.52	44.67	48.62	72.79	**84.27**
3	30.23	42.55	37.45	82.51	**87.56**
4	42.86	57.14	54.29	**80.17**	77.14
5	31.67	37.50	41.67	85.83	**89.17**
OA	35.19	44.81	45.19	81.48	**85.56**
AA	36.21	47.43	46.69	79.27	**83.96**

TABLE 3. Classification accuracies (%) obtained by different methods on the PRPS spectrum.

Class	SVM [16]	VGG-16 [28]	AlexNet [29]	SFN-WO	SFN
1	46.39	52.62	58.86	74.39	**77.89**
2	36.67	52.35	56.79	76.24	**80.08**
3	27.50	45.59	42.50	77.41	**82.53**
4	37.14	62.86	57.14	74.29	**82.86**
5	34.16	40.83	38.16	84.17	**87.52**
OA	35.92	49.42	46.01	79.26	**83.70**
AA	36.16	50.64	49.87	77.19	**82.17**

$E(i)$ is the sum of two similarity scores. $E^{PRPD}(i)$ is the PRPD similarity score between the test sample and support set samples. $E^{PRPS}(i)$ is the PRPS similarity score between the test sample and support set samples. By jointly utilizing complementary information, the proposed method can generate more accurate diagnosis performance. Specifically, the lowest scores can be generated as follows:

$$E_{min} = Softmin[E(i)] \qquad (12)$$

where i = 1,2,3,4,5. *Softmin* is the lowest score in results. The output result is the corresponding PD category in support set samples.

IV. EXPERIMENTS AND DISCUSSION
A. EXPERIMENTAL SETUP
A challenging UHF data set is collected to evaluate the performance of the SFN method. A UHF sensor is used to generate the experimental data set. It is produced by the document management system (DMS) company in the UK. This sensor detects the partial discharge of GIS equipment in 220kV substations, which are in Hunan Province, China. The data set contains 320 images, provided by the State Grid Corporation of China. Some PD samples are shown in Fig. 9. The details are shown in Table 1. In this data set, the corona discharge has 60 images, and the suspended potential discharge has 35 images. The free metal particle discharge has 50 images, and the insulation gap discharge has 45 images. The normal sample has 130 images. Among them, 50 labeled samples are randomly chosen for training and the rest 270 labeled samples are used for testing. The classification process is repeated 10 times. The average result is calculated as the final result. The details can refer to [31].

Three classical methods are employed for comparing, including SVM [16], VGG-16 [28], and AlexNet [29] mod-els. For the SVM method, the parameters c and σ are tuned by fivefold cross-validation, $(c, \sigma) = (12, 18)$. For the VGG-16 and AlexNet methods, the parameters are set based on the reference works [28], [29].

B. EXPERIMENTS ANALYSIS
In experiments, the overall accuracy (OA) and average accuracy (AA) [31]–[33] are adopted to judge the diagnosis performance. They are widely used evaluation indexes in image classification [33]. Table 2 shows the classification accuracy obtained by different methods on the PRPD spectrum. OA of the SVM, VGG-16 and AlexNet models are 35.19%, 44.81%, and 45.19% respectively. It demonstrates that traditional models fail to generate outstanding classification performance with limited training numbers. The SFN-WO and SFN methods are at least 36.27% higher than the other three compared methods. It has great improvement scores in PD fault diagnosis. Specifically, for the corona discharge class, the classification accuracy has improved from 42.27% to 82.08%. For the free metal particle discharge class, the classification accuracy has improved from 30.23% to 87.56%. The performance of the SFN-WO method is slightly lower than the SFN method. The reason is that, by training limited PD samples, the SFN method has a more robust generalization ability in PD signal extraction, compared with the SFN-WO method. Besides, Table 3 illustrates the classification accuracy obtained by different methods on the PRPS spectrum. OA of the SVM, VGG-16, and AlexNet models are 35.92%, 49.42%, and 46.01% respectively. The SFN method is at least 34.28% higher than the other three compared methods.

Moreover, the influence of the number of samples for training on the diagnosis performance is discussed. Different numbers of samples are randomly selected from data sets to constitute the training and test sets. Fig. 10 presents the varia-
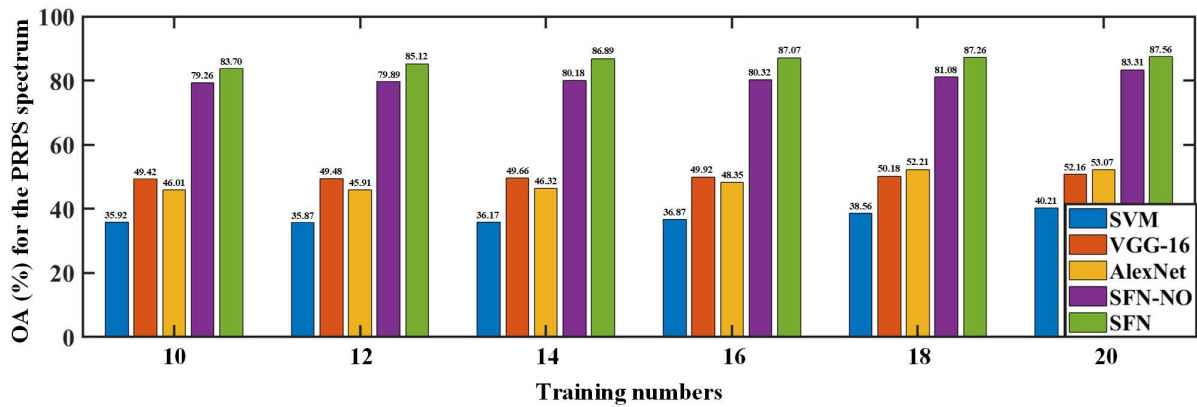
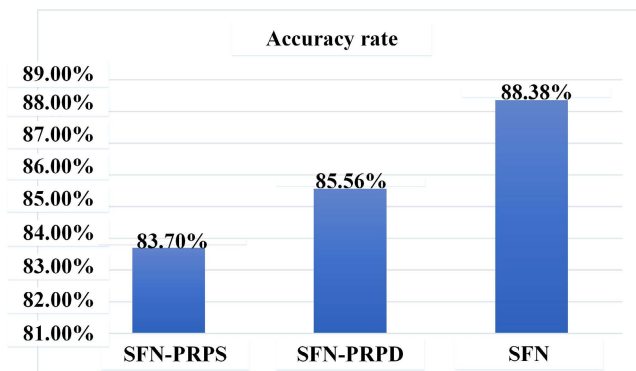**FIGURE 10.** The variation tendencies of OA (%) for the PRPS spectrum.

FIGURE 11. The effect of decision fusion.

tion tendencies of diagnosis accuracies for PRPS spectrums. Diagnosis results of the proposed SFN method are based on different spectrums and the fusion scheme. The training numbers of each class is varying from 10 to 20. For SVM, VGG-16 and AlexNet methods, the diagnosis performance has slight improvement. For SFN methods, the diagnosis performance has improved from 83.70% to 87.56%.

Furthermore, for the proposed SFN method, the diagnosis results from the single spectrum (PRPS or PRPD) also are compared with the SFN method. SFN-PRPS and SFN-PRPD are served as two compared methods. Fig. 11 shows the diagnosis accuracy of different methods. The proposed method can obtain the highest detection accuracy, 88.38%. It is higher than other diagnosis methods, 83.70%, and 85.56%. The reason is that two spectrums can capture characteristic information from different respects. The diagnosis result reflects its statistical characteristic from the UHF signal. The proposed SFN method can fuse two diagnosis results, and avoid some miss classification with the single spectrum. Therefore, the fused results can produce a more robust diagnosis result with limited training numbers.

V. CONCLUSION
This paper has introduced a siamese fuse network for partial discharge diagnosis with limited training numbers. First, the

UHF sensor generates two spectrums from power equipment, including PRPD and PRPS. Based on a few-shot learning strategy, a support set is constructed. There are four different types of PD samples and a normal sample. Then, two siamese networks are employed to estimate similarity scores between a test sample and support set samples. One network measures similarity scores in PRPS, and the other measures similarity scores in PRPD. Last, a simple and effective decision fusion technique fuses diagnosis results. Experiments performed on the PD data sets demonstrate that our method outperforms its counterparts in terms of classification accuracy with limited training numbers.

The proposed method could still be applied in other fields (such as insulator fault detection, tree barrier detection, and mountain fire detection). Thus, designing the proposed method for other applications is a challenging but interesting future work.

REFERENCES
[1] G. Hashmi, M. Lehtonen, and M. Nordman, "Modeling and experimental verification of on-line PD detection in MV covered-conductor overhead networks," *IEEE Trans. Dielectr. Elect. Insul.*, vol. 17, no. 1, pp. 167–180, Feb. 2010.

[2] B. Liu and J. Zheng, "Partial discharge pattern recognition in power transformers based on convolutional neural networks," *High Voltage App.*, vol. 53, no. 5, pp. 70–74, 2017.

[3] W. Zhang, Z. Hou, H.-J. Li, C. Liu, and N. Ma, "An improved technique for online PD detection on covered conductor lines," *IEEE Trans. Power Del.*, vol. 29, no. 2, pp. 972–973, Apr. 2014.

[4] Y.-B. Wang, D. Chang, Y. Fan, G. Zhang, J. Zhan, X. Shao, and W. He, "Acoustic localization of partial discharge sources in power transformers using a particle-swarm-optimization-route-searching algorithm," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 6, pp. 3647–3656, Dec. 2017.

[5] T. Do, V. Doan, Y. Cho, J. Sun, and Y. Kim, "Convolutional-neural-network-based partial discharge diagnosis for power transformer using UHF sensor," *IEEE Access*, vol. 8, pp. 207377–207388, 2020.

[6] G. Li, M. Rong, X. Wang, X. Li, and Y. Li, "Partial discharge patterns recognition with deep convolutional neural networks," in *Proc. Int. Conf. Condition Monitor. Diagnosis (CMD)*, Sep. 2016, pp. 324–327.

[7] H. Song, J. Dai, W. Zhang, K. Bi, L. Luo, G. Sheng, and X. Jiang, "Partial discharge pattern recognition based on deep convolutional neural network under complex data sources," *High Voltage Eng.*, vol. 44, no. 11, pp. 3625–3633, Nov. 2018.

[8] F. Yang, G. Wang, X. Peng, J. Wen, Q. Chen, G. Yang, and C. Li, "Partial discharge pattern recognition of high voltage cables based on the stacked denoising autoencoder method," *Electr. Power Automat. Equip.*, vol. 38, no. 5, pp. 123–128, May 2018.

[9] X. Wan, H. Song, L. Luo, Z. Li, G. Sheng, and X. Jiang, "Application of convolutional neural networks in pattern recognition of partial discharge image," *Power Syst. Technol.*, vol. 43, no. 6, pp. 2219–2226, Jun. 2019.

[10] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019.

[11] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Apr. 2018.

[12] M. Nguyen, V. Nguyen, S. Yun, and Y. Kim, "Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear," *Energies*, vol. 11, no. 5, pp. 1202–1212, May 2018.

[13] X. Yang, H. Huang, Q. Shu, D. Zhang, and B. Chen, "Partial discharge signal extraction method based on EDSSV and low rank RBF neural network," *IEEE Access*, vol. 9, pp. 9744–9752, 2021.

[14] C. Lin, W. Gao, and M.-F. Guo, "Discrete wavelet transform-based triggering method for single-phase Earth fault in power distribution systems," *IEEE Trans. Power Del.*, vol. 34, no. 5, pp. 2058–2068, Oct. 2019.

[15] J. L. Guardado, J. L. Naredo, P. Moreno, and C. R. Fuerte, "A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 643–647, Oct. 2001.

[16] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang, and C. Yan, "A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers," *J. Elect. Eng. Technol.*, vol. 12, no. 2, pp. 830–839, Mar. 2017.

[17] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," in *Proc. ICML*, 2015, pp. 324–327.

[18] L. Huang and Y. Chen, "Dual-path Siamese CNN for hyperspectral image classification with limited training samples," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 518–522, Mar. 2021.

[19] H. Gao, J. Zhang, X. Cao, Z. Chen, Y. Zhang, and C. Li, "Dynamic data augmentation method for hyperspectral image classification based on Siamese structure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 8063–8076, 2021.

[20] J. Gao, L. Xu, A. Bouakaz, and M. Wan, "A deep Siamese-based plantar fasciitis classification method using shear wave elastography," *IEEE Access*, vol. 7, pp. 130999–131007, 2019.

[21] H. Zhang, Z. Tang, Y. Xie, H. Yuan, Q. Chen, and W. Gui, "Siamese time series and difference networks for performance monitoring in the froth flotation process," *IEEE Trans. Ind. Informat.*, vol. 18, no. 4, pp. 2539–2549, Apr. 2022.

[22] Q. Li, X. Liao, M. Liu, and S. Valaee, "Indoor localization based on CSI fingerprint by Siamese convolution neural network," *IEEE Trans. Veh. Technol.*, vol. 70, no. 11, pp. 12168–12173, Nov. 2021.

[23] Z. He, Q. Shi, K. Liu, J. Cao, W. Zhan, and B. Cao, "Object-oriented mangrove species classification using hyperspectral data and 3-D Siamese residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 12, pp. 2150–2154, Dec. 2020.

[24] K. Raja, F. Devaux, and S. Lelaidier, "Recognition of discharge sources using UHF PD signatures," *IEEE Elect. Insul. Mag.*, vol. 18, no. 5, pp. 8–14, Sep. 2002.

[25] L. Yang and M. D. Judd, "Recognising multiple partial discharge sources in power transformers by wavelet analysis of UHF signals," *IEE Proc.-Sci., Meas. Technol.*, vol. 150, no. 3, pp. 119–127, May 2003.

[26] H. Mirzaei, A. Akbari, E. Gockenbach, and K. Miralikhani, "Advancing new techniques for UHF PDdetection and localization in the power transformers in the factory tests," *IEEE Trans. Dielectr. Elect. Insul.*, vol. 22, no. 1, pp. 448–455, Feb. 2015.

[27] M. Siegel, M. Beltle, S. Tenbohlen, and S. Coenen, "Application of UHF sensors for PD measurement at power transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 1, pp. 331–339, Feb. 2017.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–13.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Conf. Adv. Neural Inf.*, 2012, pp. 1097–1105.

[30] C. Persello, A. Boularias, M. Dalponte, T. Gobakken, E. Næsset, and B. Schölkopf, "Cost-sensitive active learning with lookahead: Optimizing field surveys for remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6652–6664, Oct. 2014.

[31] Y. Wang, L. Lee, B. Xue, L. Wang, M. Song, C. Yu, S. Li, and C.-I. Chang, "A posteriori hyperspectral anomaly detection for unlabeled classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3091–3106, Jun. 2018.

[32] B. Sun, X. Kang, S. Li, and J. A. Benediktsson, "Random-Walker-based collaborative learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 212–222, Jan. 2017.

[33] P. Gurram, H. Kwon, and C. Davidson, "Coalition game theory-based feature subspace selection for hyperspectral classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2354–2364, Jun. 2016.

ZHIHONG HUANG (Student Member, IEEE) received the Ph.D. degree from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2020. He is currently an Engineer with the State Grid Hunan Electric Power Corporation Ltd. Research Institute, Changsha. His research interests include image classification and anomaly detection. He is a reviewer of the IEEE Transactions on Geoscience and Remote Sensing, IEEE Access, and the IEEE Geoscience and Remote Sensing Letters.

WEI HUANG, photograph and biography not available at the time of publication.

XIANYONG XU received the B.S. degree in electrical and information engineering and the Ph.D. degree from Hunan University, Changsha, China, in 2005 and 2010, respectively. In June 2010 and 2012, he was an Assistant Engineer and a Senior Engineer at State Grid Corporation of China, where he has been a Professor Engineer, since 2019. He has published over 50 journal and conference papers. His research interests include the power internet of things, big data, and artificial intelligence. He was a recipient of the 2010 National Scientific and Technological Awards of China and the 2007 Scientific and Technological Awards from the National Mechanical Industry Association of China. From 2021 to 2023, he was the Standing Director of Power System Digital Measurement Subcommittee, IEEE Power and Energy Society (PES).

JIAN XIAO received the Ph.D. degree from the Huazhong University of Science and Technology, China, in 2015. He is currently a Senior Engineer with the State Grid Hunan Electric Power Corporation Ltd. Research Institute, Changsha, China. His research interests include 3D visual reconstruction and image classification.

• • •