Data Augmentation Using DCGAN for Improved Fault Detection of High Voltage Shunt Reactor

Zhu Ming\textsuperscript{1}, Zhang Zongxi\textsuperscript{2}, Mei Jie\textsuperscript{1}, Zhou Kejian\textsuperscript{1}, Chen Pengan\textsuperscript{1}, Qi Yongka\textsuperscript{1}\ast and Huang Qinqing\textsuperscript{3,4}

\textsuperscript{1}Huazhong University of Science and Technology, School of Electronic Information and Communication, Wuhan 430074, China.
\textsuperscript{2}State Grid Sichuan Electric Power Research Institute, No. 16, Jinhui West 2nd Street, high tech Zone, Chengdu 610041, China
\textsuperscript{3}NARI Group (State Grid Corporation of China), Nanjing 211106, China.
\textsuperscript{4}State Grid Electric Power Research Institute Wuhan NARI Group, Wuhan 430074, China.

\ast qiyongka@hust.edu.cn

Abstract. High voltage shunt reactor is an important equipment of power transmission systems. The accurate assessment of their operating status and the timely and correct diagnosis of faults and defects concern the operation safety of the entire grid. Health assessment of high voltage shunt reactors based on vibration signal, which can be used to characterize the hidden troubles of it, is a topic widely studied in deep learning and fault diagnosis. A large number of samples are needed to train the deep learning model, but it is not easy to acquire enough fault samples in the actual scene. In this paper, we utilize a Deep Convolutional Generative Adversarial Networks (DCGAN) to generate synthetic fault samples and enlarge the fault dataset to train the Convolution Neural Network (CNN) fault detection model. Results reveal that the performance through the CNN model can be improved by 3\% with the synthetic samples generated by DCGAN, which is better than that of traditional Synthetic Minority Oversampling Technique (SMOTE) algorithm.

1. Introduction
High voltage (HV) shunt reactor is a kind of equipment in power system, which has been used widely [1, 2]. The most important is that it can reduce the power loss during long transmission. So it is very important to detect its healthy state in time. The existing fault detection methods of HV shunt reactor mainly include ultrasonic partial discharge detection method and oil chromatography detection method. Ultrasonic partial discharge detection determines whether there is partial discharge in HV shunt reactor through the analysis of the ultrasonic signal on the oil tank. It can be used to detect and locate the insulation partial discharge fault. This method can realize fault detection of the HV shunt reactor, but it is easy to be affected by the electromagnetic environment. Oil chromatogram detection is used to diagnose insulation fault, overheating fault and judge the fault type inside the reactor through the change of gas in the insulating oil inside the reactor. In this method, overheating can be diagnosed easily. However, this method reacts slowly, so that there is a long delay to determine the state of shunt reactor. Therefore, it is urgent to find a method that can monitor HV shunt reactor automatically.

One approach to solve this problem involves the use of vibration signal. Early research also shows
that we can detect the state of HV shunt reactor through vibration signal. In the last few years, state detection of HV shunt reactor has become popular based on vibration signal. It vibrates periodically under the force of electromagnetic force. When the state of HV shunt reactor changes, the vibration signals must change. So that we can detect the hidden trouble in time.

In this article, we propose a fault detection model based on convolution neural network (CNN). As for the shortage of vibration signal, we generate synthetic vibration signal by generative adversarial network (GAN). Furthermore, we generate the synthetic vibration signal by deep convolution Generative Adversarial Networks (DCGAN). We then modify the DCGAN and enhance its performance on the data set of HV shunt reactor vibration signals.

2. Fault detection model of HV shunt reactor
In this chapter, a fault detection model is established based on vibration signal. We collected vibration signal of HV shunt reactor through acceleration sensors. We propose a fault detection model based on CNN, according to the features of the vibration signals.

2.1. Vibration signal description
In this paper, we acquire vibration signal of HV shunt reactor by ICP/IEPE CT1010L acceleration sensors. Sensors are arranged on four sides of the BKD2-140000/800-110 HV shunt reactor oil tank evenly. 9 acceleration sensors are arranged on each side. The acceleration sensors are fixed on the surface of HV shunt reactor by magnetic base. So the vibration signal can show the internal operation state of HV shunt reactor. The sampling frequency is $1.25 \times 10^6$ Hz. Finally, we get 15000 groups normal vibration signal and 15000 groups abnormal vibration signal.

The vibration signal at some position of the HV shunt reactor is acquired, as shown in figure 1. It can be seen from figure 1 that the vibration signal on the surface of reactor is periodic, with a period of 0.01 s. Then we can take an easy uniform sample to reduce the number of point in one period. So there are 1250 points every cycle.

![Figure 1. Vibration signal of some position](image)

2.2. CNN fault detection model
In order to detect the healthy state of the HV shunt reactor effectively, CNN is used to establish the fault detection model based on vibration signals [3]. Because vibration signals are one-dimensional time series. The one-dimensional CNN model is proposed. The network structure and the specific parameters of each layer of the network are introduced as shown in figure 2. It can be seen from the figure that the model has 11 layers, including 5 convolution layers, 5 pooling layers and 1 full connection layer.
Then we put the vibration signals generated by simulator into the fault detection model to test its performance. The vibration signals are divided into training dataset and testing dataset as 4:1. Training samples are time series with 1250 sampling points. And 64 samples are selected from each batch for training. The training include total 20 epochs. Instead of the fixed learning rate manually, the degenerative learning rate is used in the model, which makes the training speed and stability reach a balance. The initial learning rate is set as 0.001, and then learning rate is reduced to 1/10 of the original learning rate every 10 epochs. Accuracy of the fault detection model is shown in figure 3, when size of training data set is 200, the accuracy of fault detection is 50%. When size of training dataset is 800, the accuracy of fault detection is 100%. So if the size of training data is less than 800, we must generate synthetic vibration signals to improve the accuracy of the fault detection model.

3. Generated synthetic vibration signal

Data augmentation is the general method to increase the size of data set [4]. The purpose of data augmentation is to make the training data as close to the test data as possible, so as to improve the accuracy of prediction. It increases the sample of training set, which can effectively alleviate the over fitting of the model, and also can bring stronger generalization ability to the model. There are some methods to achieve that goal such as SMOTE and GAN.

3.1. Data augmentation by SMOTE

SMOTE (Synthetic Minority Oversampling Technique) is a better algorithm based on random oversampling algorithm [4]. It is always used for data augmentation. SMOTE algorithm is an oversampling algorithm for synthesizing minority data, which was proposed by Juanjuan Wang in 2006 [5]. The core idea of the algorithm is to analyze the minority samples, and then synthesize new samples according to the minority samples and add them to the data set. In recent years, SMOTE has been widely used in data analysis, and other fields such as Internet, finance and medicine. However, SMOTE easily results in the over fitting caused by the simple duplication of samples.
3.2. Data augmentation by DCGAN

DCGAN is an important attempt to introduce deep convolution network into GAN [6]. DCGAN follows the basic principle of GAN. It extracts complex features from training samples by convolution layer. The original GAN model does not specify the format of the data that will be input into the network. So there are so many kind of GAN. The deep convolution neural network is mostly used to solve image problems. So DCGAN is mostly used to generate images. In this paper, the vibration signal of HV shunt reactor is studied. Inspired by image generation, the DCGAN model is modified to adapt to the characteristics of one-dimensional data and realizes the generation of vibration signals [7].

![Figure 4. Structure of DCGAN](image)

It can be seen from figure 4 that DCGAN is also formed by Generator and Discriminator. $D_{Loss}$ presents loss function of Discriminator network and $G_{Loss}$ presents loss function of Generator network. In the ideal training state, both of them should show a decreasing trend, and will tend to be stable in the end. This also shows that in the training process, the networks have space for further optimization, and find a correct path to optimize themselves. The quality of the generative synthetic data will continue to improve. If $D_{Loss}$ decreases too fast and turns to 0, it indicates that the Discriminator network has been trained very well and can easily distinguish the real data and the synthetic data generated from the Generator network. At this time, the generation network cannot further optimize itself. The quality of the generated data cannot be high, and the phenomenon of over fitting will appear. If $D_{Loss}$ and $G_{Loss}$ has been in a state of intense vibration or cannot converge for a long period of time, it means that the Discriminator network and the Generator network have not found a correct path to update their parameters in the training process. The quality of the generated data cannot be guaranteed. A series of noise will be generated. So it is important to reach convergence for data augmentation by DCGAN [8].

3.3 Convergent DCGAN

In order to stabilize the DCGAN model and avoid serious oscillation of loss function, the structure of network is modified. So there are some adjustments about the model below.

3.3.1 Convergent DCGAN. In the training process of the network, the network parameters will be constantly updated. In addition to the data of the input layer, the data distribution of each layer will also change due to the changes of the previous network parameters. Because the input of one layer network is calculated from the data which generated by the previous layer and the parameters. And the parameters of the layer will be updated during the training process. So the input of each layer of the network will inevitably change, which will result in internal covariate shift. The introduction of batch normalization (BN) layer solves the problem of network data distribution change in the middle layers.

Batch normalization is aimed at the normalization of a batch of data. Batch in BN is the data of a batch. At present, Batch Gradient Descent is often used in network training. The data is divided into several groups according to the batch size. A batch of data is trained to calculate the loss and update the parameters each time. A batch of input data can be expressed as
\[
X = [X_1, X_2, \ldots, X_n]
\]

(1)

Where \( X_n \) presents data in the batch and \( n \) presents the quantity of this batch. The average value of the elements can be calculated as

\[
\mu_n = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

(2)

After obtaining the element mean value, the variance of the batch data is calculated, which is expressed as

\[
\sigma_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_n)^2
\]

(3)

Next, each element in the batch is normalized, which can be expressed as

\[
x_i' = \frac{x_i - \mu_n}{\sqrt{\sigma_n^2 + \epsilon}}
\]

(4)

Where \( \epsilon \) is a very small positive number. Finally, scale transformation and migration are performed for each element, which can be expressed as

\[
y_i = \gamma_i x_i' + \beta_i
\]

(5)

From the above formula, if the scale transformation factor \( \gamma_i \) is equal to the variance, and the offset \( \beta_i \) is equal to the mean value, then \( y_i \) is equal to the input. For the network with BN layer, \( \gamma_i \) and \( \beta_i \) also become the training parameters of the current layer. Unlike the network without BN layer, it is determined by the nonlinearity of the whole network. After using BN layer, we can allow higher learning rate, and reduce the dependence on initialization. We can also change the mean and variance by optimization, which makes the output closer to the real distribution, the gradient gentler and avoids gradient dispersion. Therefore, in the generation of vibration signals, Batch Normalization layer should be added after each convolution layer of DCGAN.

3.3.2 Structure of Convergent DCGAN. After the adjustment of the model, DCGAN is used to generate the vibration signals. A 250 dimensional Gaussian random noise vector is input into the Generator network. After the process of full connection layer, its format is changed into 256 channels. There will be a Batch Normalization layer and an upsampling layer after a convolution network layer. After process of 5 convolution layer, all parameters are adjusted to generate a vector. The structure of Discriminator network is similar to that of Generator network. Differently from the Generator network, the activation function of each convolution layer is replaced by the LeakyReLU function, and a pooling layer is added between the convolution layers to reduce the data scale. Finally, the activation function of the output part of the network is set to AvgPool function.

4. Results and discussion

The experimental environment is Intel (R) core (TM) i7-8700 CPU @ 3.20GHz processor, 16GB memory, NVIDIA GeForce GTX 1060 6GB GPU, Anaconda 3.5.1. All experimental analysis is implemented under Pytorch framework.

We generated normal and abnormal synthetic signal through DCGAN. The synthetic vibration signal is shown in figure 5. After 1500 iterations, the synthetic vibration is similar to the real vibration signal. Their period is also 0.01s. It is consistent with the original vibration signal. But there is so much noise in the time series, which means the network has got the law of the vibration. If we improve the number of training, we can get better synthetic vibration.
Figure 5. Synthetic vibration signal by DCGAN

Figure 6. Accuracy of data set identification

It can be seen from the figure 6 that the accuracy of the synthetic data set is higher than that of the original training data set. At the same time, after the size of data set reaches 21000, the DCGAN synthetic dataset has little improvement on the recognition accuracy. It needs a large amount of data, and the efficiency will be very low. Therefore, when the training set data size is below 21000, using DCGAN to extend the data is effective for improving the classification accuracy of the original training set. Moreover, in the range of training set size from 1312 to 21000, the recognition rate of DCGAN is better than that of SMOTE, which can improve the recognition rate by 3%.

5. Conclusion

Due to the shortcomings of long delay and harsh interference in traditional patrol methods such as oil chromatogram and ultrasonic partial discharge, we evaluate the health status via deep learning model based on vibration signals of reactor surface, which can be used to characterize the hidden trouble of it. A mount of samples are needed to train the deep learning model, but it is not easy to acquire enough fault samples in the actual scene. In this research, we proposed a DCGAN model that generates synthetic abnormal vibration signal to enlarge the dataset and improve the performance of fault detection CNN model. By adjusting the structure of Batch Normalization layer and the activation function, the model is convergent. An improvement in classification performance by 3% accuracy is recorded when CNN is trained on actual data and synthetic augments. Therefore, the method can realize the state assessment of the HV shunt reactor, which can be used as a beneficial supplement to the healthy status detection of HV shunt reactor.
Acknowledgments
This work was supported by State Grid Corporation of China (State Grid) (Grant No. 52199919000A)

ORCID ID
Ming Zhu https://orcid.org/0000-0001-5007-3919

References
[1]. Papp, K., M.R. Sharp and D.F. Peelo, High Voltage Dry-Type Air-Core Shunt Reactors. Water and Energy International, 2014. 57r(6).
[2]. Bojić, S., B. Babić and I. Uglešić, Comparative research into transients by switching of high voltage shunt reactor. Electric Power Systems Research, 2018. 162.
[3]. Jin, S., et al., Evaluation and modeling of automotive transmission whine noise quality based on MFCC and CNN. Applied Acoustics, 2021. 172.
[4]. Abdelhalim, I.S.A., M.F. Mohamed and Y.B. Mahdy, Data augmentation for skin lesion using self-attention based progressive generative adversarial network. Expert Systems With Applications, 2021. 165.
[5]. Wang, J., et al. Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding. in The 8th International Conference on signal processing. 2006. Guilin, Guangxi, China.
[6]. Viola, J., Y. Chen and J. Wang, FaultFace: Deep Convolutional Generative Adversarial Network (DCGAN) based Ball-Bearing failure detection method. Information Sciences, 2021. 542.
[7]. Seeliger, K., et al., Generative adversarial networks for reconstructing natural images from brain activity. NeuroImage, 2018. 181.
[8]. Li, Y., N. Xiao and W. Ouyang, Improved Generative Adversarial Networks with Reconstruction Loss. Neurocomputing, 2018.