

Research Article

Imbalanced Fault Classification of Bearing via Wasserstein Generative Adversarial Networks with Gradient Penalty

Baokun Han, Sixiang Jia, Guifang Liu, and Jinrui Wang 

College of Mechanical and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266590, China

Correspondence should be addressed to Jinrui Wang; wangjr33@163.com

Received 2 April 2020; Revised 9 June 2020; Accepted 2 July 2020; Published 21 July 2020

Academic Editor: Changqing Shen

Copyright © 2020 Baokun Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently, generative adversarial networks (GANs) are widely applied to increase the amounts of imbalanced input samples in fault diagnosis. However, the existing GAN-based methods have convergence difficulties and training instability, which affect the fault diagnosis efficiency. This paper develops a novel framework for imbalanced fault classification based on Wasserstein generative adversarial networks with gradient penalty (WGAN-GP), which interpolates randomly between the true and generated samples to ensure that the transition region between the true and false samples satisfies the Lipschitz constraint. The process of feature learning is visualized to show the feature extraction process of WGAN-GP. To verify the availability of the generated samples, a stacked autoencoder (SAE) is set to classify the enhanced dataset composed of the generated samples and original samples. Furthermore, the exhibition of the loss curve indicates that WGAN-GP has better convergence and faster training speed due to the introduction of the gradient penalty. Three bearing datasets are employed to verify the effectiveness of the developed framework, and the results show that the proposed framework has an excellent performance in mechanical fault diagnosis under the imbalanced training dataset.

1. Introduction

Machine fault diagnosis plays a significant role in ensuring the normal and orderly operation of industrial production. There are usually two main methodologies for fault diagnosis: physic-based and data-driven-based methods. Deep learning [1–3], as the most popular data-driven method, has provided us with new ideas for machine fault diagnosis. Compared with traditional methods, the deep learning-based method has powerful feature learning ability which can simulate the brain learning process by constructing deep models, depicting the rich internal information of data, and finally achieving fault recognition [4–7].

In the last few years, most of the research studies on improving the accuracy of fault diagnosis through data-driven methods are carried out. Lei et al. [8] developed an unsupervised feature learning method based on sparse filtering to directly extract features from mechanical vibration signals and then classified the fault classes through softmax regression. Liu et al. [9] applied stacked

autoencoders (SAEs) to diagnose the gearbox fault, and the introduction of the dropout technique and the ReLU function relieves the overfitting problem. Lei et al. [10] proposed an end-to-end long short-term memory (LSTM) model which could extract features from multivariate time series data directly. Zhang et al. [11] proposed an inter-cluster distance-optimized support vector machine (ICDSVM) which can not only classify different fault types but also fault severities.

The above methods are proposed under the assumption of sufficient and balanced training data. But, in practice, mechanical equipment often run in the normal state for a long time so the monitor data have a limited capacity of rare fault types. And, these methods often fail to perform well in the unbalanced fault diagnosis problem and even fail to identify rare fault categories. Therefore, it is necessary to develop an effective fault diagnosis method under unbalanced samples. Much of the research studies in fault diagnosis in recent years have examined how to develop diagnosis performance by improving the classification

method if there are not enough samples for training. Duan et al. [12] proposed a fault diagnosis framework based on support vector data description for multiclassification. In the model, the data description is applied to the binary tree structure from the top to the bottom for classification. Zhang et al. [13] proposed a method to classify the unbalanced fault in permanent magnet synchronous motor drives, which analyzes the fault feature by the discrete wavelet transform. However, it is not enough to obtain a good diagnostic effect only through the improvement of the classification method if the imbalance degree increases. This problem cannot be settled fundamentally unless we obtain more simulation data from the raw data. In this paper, the generative adversarial network (GAN) [14] is applied to increase the amounts of imbalanced input samples. GAN has been widely applied because of its great prospect, including image recognition, language processing, and information security [15–17]. Until now, GAN has derived many models with different structures [18–21] and is widely used in fault diagnosis [22–24]. It also provided a choice for solving the problem of data imbalance in fault diagnosis by generating data for fault types with less data. Shao et al. [25] proposed a fault diagnosis framework based on the auxiliary classifier GAN (ACGAN) [26], which can generate real synthetic signals with different fault labels, and an assessment method was proposed to evaluate the quality of the generated signals. Guo et al. [27] proposed a framework based on multilabel 1D GAN (ML1D GAN). In the framework, ACGAN is applied to generate simulation damage data, and the generated and real data are both used to train the classifier. Nevertheless, the original GAN and most of its variants including ACGAN cannot accurately judge the convergence and equilibrium points of the network theoretically. The above methods are all structural improvements to GAN, but they still do not fundamentally solve the convergence difficulties and training instability problems which affect the efficiency of fault diagnosis.

Although GAN has many problems in training, the latest research studies of GAN show that it still has a broad prospect in the fault diagnosis under unbalanced samples. In order to improve the accuracy and efficiency of machine fault diagnosis, we propose a novel deep learning framework for imbalanced fault diagnosis. Firstly, the proposed method generates simulation samples through Wasserstein generative adversarial networks with gradient penalty (WGAN-GP) [28] for fault types with fewer data. It interpolates randomly between the true and generated samples to ensure that the transition region between the true and false samples satisfies the Lipschitz constraint [29]. Then, the high-level features are extracted from the enhanced dataset, and different fault types are classified through an SAE model.

The rest of this paper is organized as follows. The algorithms about GAN, WGAN-GP, and SAE are briefly introduced in Section 2. In Section 3, the detail of the proposed WGAN-GP-SAE framework is explained. Concrete experiment verification and the result discussion are given in Section 4. At last, some conclusions are drawn in Section 5.

2. Theoretical Background

2.1. Generative Adversarial Networks. The structure of GAN consists of a generator G and a discriminator D . As is shown in Figure 1, the input of the generator is a random noise vector, and simulation signals are generated by the generator G to make the discriminator D not able to determine whether the input signal is real or generated. The output of D is a number between 0 and 1 which indicates the probability that the input is real data. Meanwhile, in general, G and D are continually optimized to improve their ability to generate and discriminate. The training objective function of GAN is shown as

$$\min_G \max_D V(D, G) = E_{x \sim p_x} [\log D(x)] + E_{z \sim p_z} [\log(1 - D(G(z)))] \quad (1)$$

where x is sampled in the raw data distribution p_x , z is a stochastic noise vector, G accepts z from the probability distribution p_z , and p_g denotes the distribution of the generated data.

The discriminator D is trained to maximize the probability of identifying the source of the input data. On the other hand, the purpose of G is to make the data generated infinitely close to the real data distribution p_x . A dynamic equilibrium, i.e., Nash equilibrium [30], is achieved only if $p_x = p_g$. The Jensen–Shannon divergence is applied to measure the difference between p_x and p_g . The advantage of the original GAN is that it can theoretically approximate a random sample to the actual data without a hypothetical data distribution. However, the Jensen–Shannon divergence is not a reasonable cost function if the network is trained to learn the distribution supported by low-dimensional manifolds.

2.2. Wasserstein Generative Adversarial Networks with Gradient Penalty. The gradient of the generator G would be smaller when the discriminator D is trained well enough or become larger when the effect of the discriminator D is poor. Wasserstein GAN (WGAN) uses the Wasserstein distance to evaluate the difference between the real and the generated sample distribution [20]. It has superior smoothing characteristics with respect to the Jensen–Shannon divergence. As is shown in equation (2), the Wasserstein distance is written into a solvable form by mathematical transformation:

$$W(p_x, p_g) = \inf_{\gamma \in \Pi(p_x, p_g)} E_{(a,b) \sim \gamma} [\|a - b\|], \quad (2)$$

where $\Pi(p_x, p_g)$ is the set of all distributions $\gamma(a, b)$ whose marginals are p_x and p_g , respectively, and $\gamma(a, b)$ denotes how far is the distance from a to b . The Wasserstein distance can be understood as the cost of an optimal transportation plan.

When approximately an optimal discriminator is formed, the optimizer reduces the Wasserstein distance and effectively approximates the distribution of generated

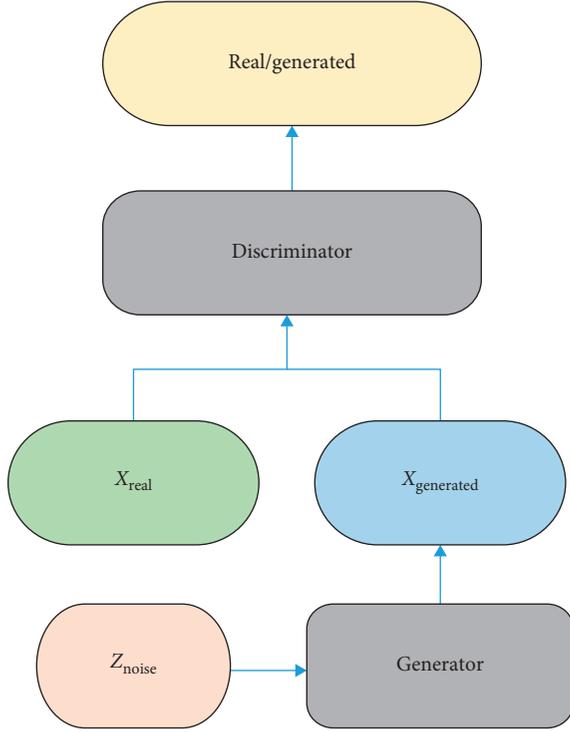


FIGURE 1: Typical architecture of GAN.

samples and real samples. The objective function of WGAN is illustrated as

$$\min_G \max_{D \in \mathbb{R}} E_{x \sim p_x} [D(x)] - E_{y \sim p_g} [D(y)], \quad (3)$$

where \mathbb{R} is the collection of 1-Lipschitz functions.

The Lipschitz constraint is set to ensure that the gradient of the discriminator is not larger than a finite constant K over the entire sample space. Furthermore, the Lipschitz limit is implemented through weight clipping to make the output value given by the discriminator not to undergo a too drastic change when the input sample has slight fluctuations. WGAN checks whether the absolute values of all the parameters exceed a compact space $[-c, c]$ after updating the parameters of the discriminator. If any, the parameters are clipped back to the threshold range. In this way, the parameters of the discriminator D are always bounded, thus ensuring that the discriminator cannot give a large difference output value to two slightly different inputs.

However, the weight clipping in WGAN leads to optimization difficulties, and the resulting critic can have a pathological value surface even when the optimization is succeeded. The parameters tend to be boundary values in this method; that is, the discriminator tends to learn a simple mapping function. The powerful fitting ability of WGAN is not exerted. Weight clipping can easily lead to gradient disappearance or gradient explosion. The gradient will become smaller when it passes to every layer of the network if the clipping threshold is set smaller and even disappear after multiple layers. On the other hand, the gradient will become larger if the clipping threshold is set a little larger, and the gradient explosion will occur. Only when the threshold setting is just right, can the generator get the appropriate

backhaul gradient. But, this equilibrium area is difficult to find in practice, which makes the convergence process of WGAN slower and brings a lot of troubles to adjustment parameters.

These problems are completely solved by Wasserstein generative adversarial networks with gradient penalty (WGAN-GP). In fact, it is not necessary to impose a Lipschitz constraint on the entire sample space. Only the generated sample set area, the real sample set area, and the transition area sandwiched between them need to be restricted. To transfer the gradient efficiently, the gradient of the discriminator is limited to around 1 directly by the gradient penalty which increases the controllability of the gradient. The loss function of the generator remains unchanged, and the loss function of the discriminator is shown as

$$L(D) = \underbrace{E_{x \sim p_x} [D(x)] - E_{z \sim p_z} [D(G(z))]}_{\text{Original loss}} + \underbrace{\lambda E_{\hat{x} \sim p_{\hat{x}}} [\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1]^2}_{\text{Gradient penalty}}. \quad (4)$$

Among them, the newly added one is the gradient penalty. \hat{x} is a sample distributed on all lines between p_x and p_z . The introduction of the gradient penalty results in more stable gradients that neither vanish nor explode, thus more complicated networks are allowed to be trained.

2.3. Stacked Autoencoders. Autoencoder is one of the unsupervised learning algorithms, which applies back-propagation. The hidden layer h inside the autoencoder is able to generate coding to express the input. The structure consists of two parts: encoder and decoder. The encoder is set to map the input data into hidden representation, while the decoder is referred to reconstruct the input data. The untagged input dataset $\{x_n\}_{n=1}^N$ is given, where $x_n \in R^{m \times 1}$, h_n is a hidden encoder vector which is calculated from x_n , and \hat{x}_n denotes the decoder vector of the output layer. The encoding process is indicated as

$$h_n = f(Wx_n + b_1), \quad (5)$$

where $f(\cdot)$ denotes the encoding function, W is the weight matrix of the encoder, and the bias vector is denoted as b_1 .

$$\hat{x}_n = g(W^T h_n + b_2), \quad (6)$$

where $g(\cdot)$ denotes the decoding function, W^T denotes the weight matrix of the decoder, and b_2 denotes the bias vector.

To minimize the error of reconstruction, the parameter of the autoencoder is optimized, and the process is shown as follows:

$$\phi(\Theta) = \arg \min_{\theta, \theta_i} \frac{1}{n} \sum_{i=1}^n L(x^i, \hat{x}^i), \quad (7)$$

where L denotes the loss function: $L(x, \hat{x}) = \|x - \hat{x}\|^2$.

The autoencoders are superimposed layer by layer to form a deep neural network; that is, each hidden layer is

taken as the input of the next layer, which is progressive until the training process is completed. The introduction of batch normalization [31] has established a simple initial condition for the training of SAE, which makes the gradient update from a very shallow path, thus speeding up the training process. The labeled signals are bonded with the softmax classifier, and the BP algorithm is applied to realize the updating of network weights and fine tuning of parameters [32].

3. System Framework and Model Training

This part accounts for the developed WGAN-GP-SAE framework for bearing fault diagnosis. The structure of models is discussed in detail below.

3.1. System Framework Design. The structure of the generator G contains three layers, and the number of neurons in each layer is 200, 600, and 1200, respectively. The discriminator D structure has four layers, each with 1200, 600, 200, and 1 neurons. The dimension of the random vector is 100, and ReLU is selected as the activation function. For the SAE, there are five layers in the structure where each layer is equivalent to an encoder. The extracted features of the former layer are adopted as the input for the next layer, and the process is repeated for a specified number of times until the iteration is completed. Then, the BP algorithm is applied to minimize the error between learned features and labels, thus the weights of SAE are updated and the parameters are fine tuned. The system framework design is shown in Figure 2. During the data generation module, the learning rates of the generator and the discriminator are both $1E-4$, the optimizer of WGAN-GP is root mean square prop (RMS), and the number of iterations is 1000. The SAE is trained to output classification results using the dataset composed of the generated sample and the original sample.

3.2. Model Training Procedure. The program of model training can be represented in three steps.

Firstly, a set of noise vectors with a dimension of 100 are input into the generator. The generator G performs training according to the distribution of the real signal spectrum and generates a high-dimensional simulation signal.

Secondly, real data and simulation data are input into the discriminator D to output a probability value for evaluating the authenticity of the input data.

Finally, the simulation samples generated through WGAN-GP are bonded with the raw samples to enhance and balance the training dataset and then fed into the SAE model for fault type diagnosis.

4. Experimental Verification

4.1. Case 1: Case Western Reserve University Dataset. This section uses the bearing fault data of Case Western Reserve University (CWRU) to validate the proposed fault diagnosis framework. The frequency of sampling is 48 kHz. The bearing is damaged by the electrical discharge machining

(EDM) single point and divided into four health conditions: normal, inner ring fault, outer ring fault, and rolling body fault. There are three types of damage dimensions for each type of fault: 0.18 mm, 0.36 mm, and 0.54 mm, for a total of 10 healthy types of bearing datasets. They are named as NC, IF1, IF2, IF3, OF1, OF2, OF3, RF1, RF2, and RF3. Each health condition consists of 500 samples, every sample includes 2400 data points, and we perform fast Fourier transformation (FFT) operation on each sample data to obtain 1200 Fourier coefficients.

Two datasets with different unbalance degrees are collected to verify the proposed method, as in Table 1. In Dataset A, the percent of NC samples is 50%, the percent of fault type samples with damage sizes of 0.18 mm and 0.36 mm is 40% and 30%, and the percent of fault type samples with damage size of 0.54 mm is 20%. To facilitate comparative studies, the percentages of test samples are still 50%. Considering the aggravation of the imbalance, more training samples are reduced in Dataset B. In order to allow meaningful comparisons, the GAN with SAE (GAN-SAE) and the WGAN with SAE (WGAN-SAE) are employed for diagnosis, and the composition of the two comparative frameworks is the same as the proposed method.

WGAN-GP is applied to simulate the bearing spectrum signals of ten healthy states in turn. We combined the generated signal with the original dataset to enhance the percent of all types of samples up to 50%. The comparison between the real signal and the generated signal is depicted in Figure 3. It can be seen that the simulation signals have almost completely unified spectral characteristics with the real signal, which indicates that WGAN-GP has powerful feature extraction ability. To further illustrate the feature extraction ability of WGAN-GP, we study the process of feature learning of the network. The unit numbers of each layer are 200, 600, and 1200. In order to better illustrate the feature extraction of each hidden layer, we selected the extracting process of OF3 fault samples and drew all the extracted eigenvectors of each layer together to reveal them in three-dimensional space. The result is shown in Figure 4. It is clear that the amplitude of the eigenvector is increasing with the depth of the layer, and the main characteristics are also more distinct with the increase in the number of layers. As is shown in Figure 4(d), the frequency spectra of generated samples have identical trend with the raw. Furthermore, it is worth noting that the generated signal retains the characteristics of the original signal and is much sparser, which may be beneficial to further fault classification. In other words, all the major characteristics have been extracted from the raw signals, and some noise elements that are not conducive to mode recognition are removed by WGAN-GP.

In order to eliminate the influence of accidental factors, we conducted each experiment 20 times. The test accuracy of the three methods for Datasets A and B is shown in Figure 5. For Dataset A, the average accuracy of GAN-SAE, WGAN-SAE, and WGAN-GP-SAE is 85.05%, 96.75%, and 99.90%, respectively. The standard deviation of the three methods is 0.82%, 0.48%, and 0.10%, respectively. For Dataset B with a higher degree of imbalance, the average accuracy of GAN-SAE decreases to 76.31% with a standard deviation of 1.51%,

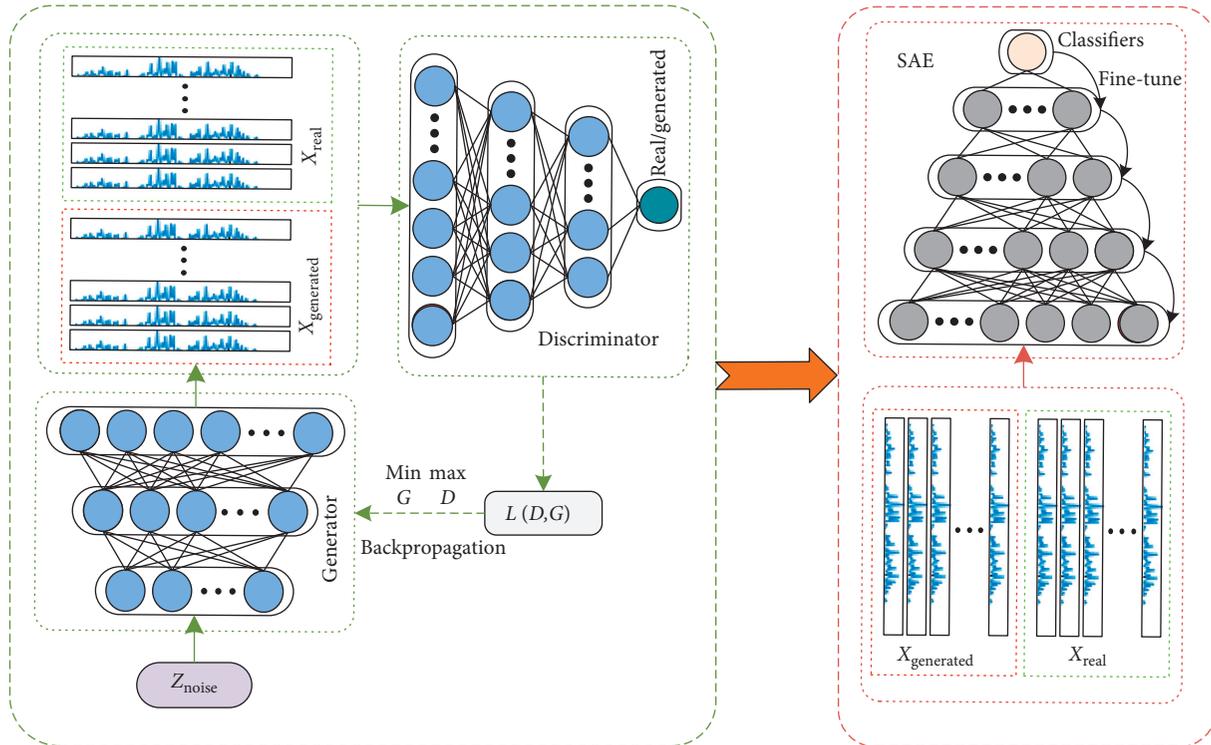


FIGURE 2: Fault diagnosis model design.

TABLE 1: Description of the unbalanced Datasets A and B.

Health condition	The percent of training samples		The percent of testing samples
	Dataset A	Dataset B	Dataset A/B
NC	250	250	250
IF1	200	100	250
IF2	150	50	250
IF3	100	25	250
OF1	200	100	250
OF2	150	50	250
OF3	100	25	250
RF1	200	100	250
RF2	150	50	250
RF3	100	25	250

and the average accuracy of WGAN-SAE decreases to 91.56% with a standard deviation of 0.63%. By contrast, WGAN-GP-SAE achieves an average accuracy rate of 96.97% and a standard deviation of 0.24%. Therefore, the performance of the proposed method in unbalanced fault classification is far better than that of the other methods. When the unbalanced degree of training data is not serious, the method proposed in this paper can improve the accuracy of fault diagnosis to a certain extent. When the degree of imbalance is aggravated, there is less information available for signal types with less data in the original signal, and the quality of the generated sample will be affected. In this case, the accuracy can be greatly improved through our method.

For the visualization sake of the three methods in fault classification, the t -distributed stochastic neighbor

embedding (t-SNE) [33] technique is applied to map high-dimensional features into three-dimensional (3D) space. The dimensionality reduction results of Dataset A are shown in Figure 6. It can be seen that the GAN-SAE frame can only accurately identify the NC, IF1, OF1, and RF3 samples, but there are different confusion levels between other fault samples; as is shown in Figure 6(b), WGAN-SAE has a slight degree of confusion on classification. In Figure 6(c), WGAN-GP-SAE perfectly separates all bearing health type samples, and the same type of samples are highly polymerized, which proves the superiority of the proposed method. But, when the imbalance degree increases, as shown in Figure 7, GAN-SAE has a poor classification of most fault types, and the classification effect of WGAN-SAE also becomes worse. In contrast, for WGAN-GP-SAE, there are only several samples being misclassified in Figure 7(c), which is much better than the other two methods. Therefore, we can conclude that the developed framework has the strongest feature extraction and feature classification capabilities than the other two methods.

4.2. Case 2: Shandong University of Science and Technology Dataset. The experimental equipment of rotating machinery in the Shandong University of Science and Technology (SDUST) is shown in Figure 8. Vibration acceleration signals come from a specially designed bench which consists of a gearbox, a motor, three shaft couplings, two rotors, two bearing seats, and a brake. All types of bearing failures measured by our laboratory are the same as that of the CWRU. But, our sampling frequency is 12.8 kHz, and the bearing failure damage dimensions are 0.2 mm, 0.4 mm, and

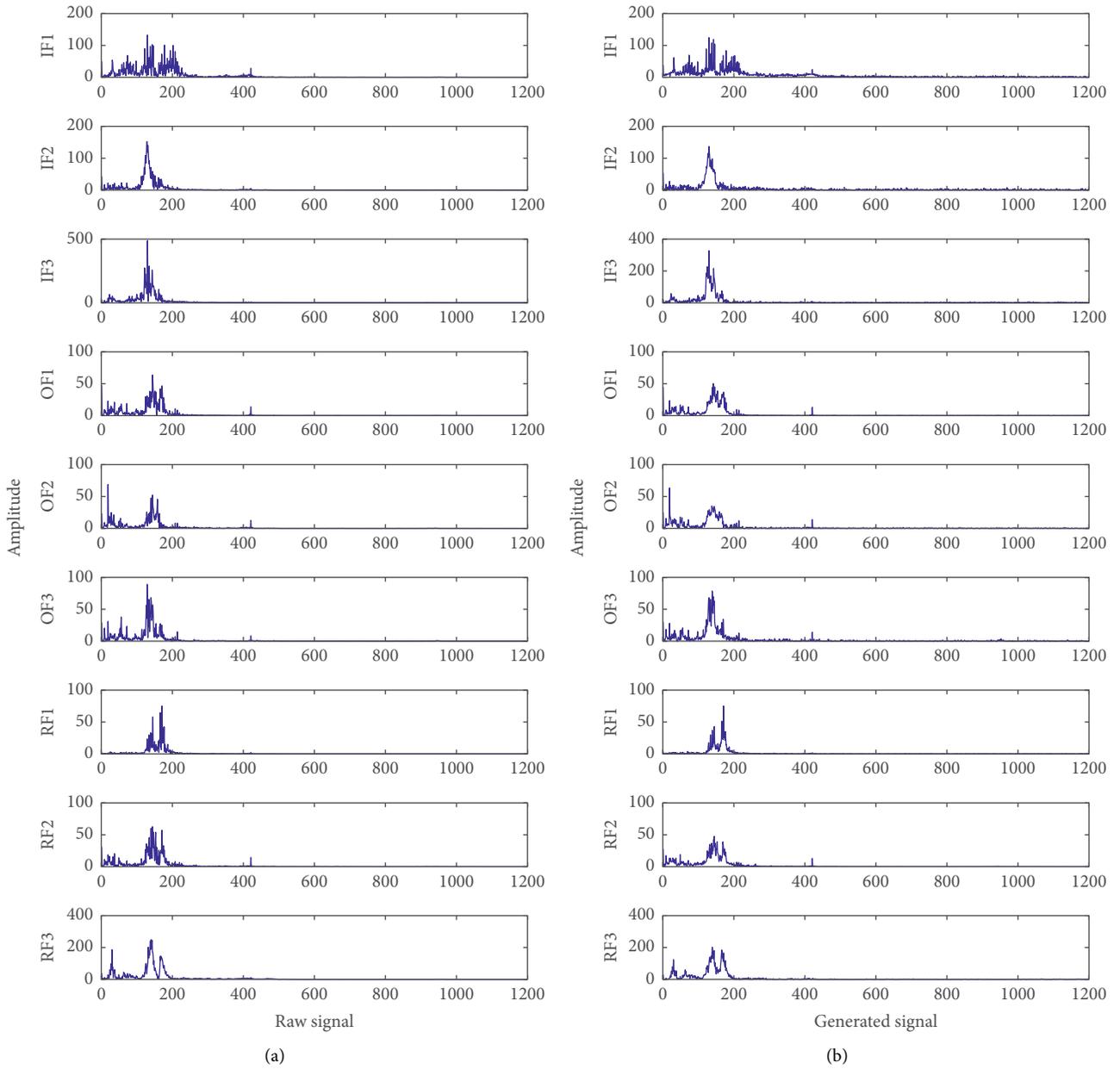


FIGURE 3: Spectrum comparison between real and generated signals.

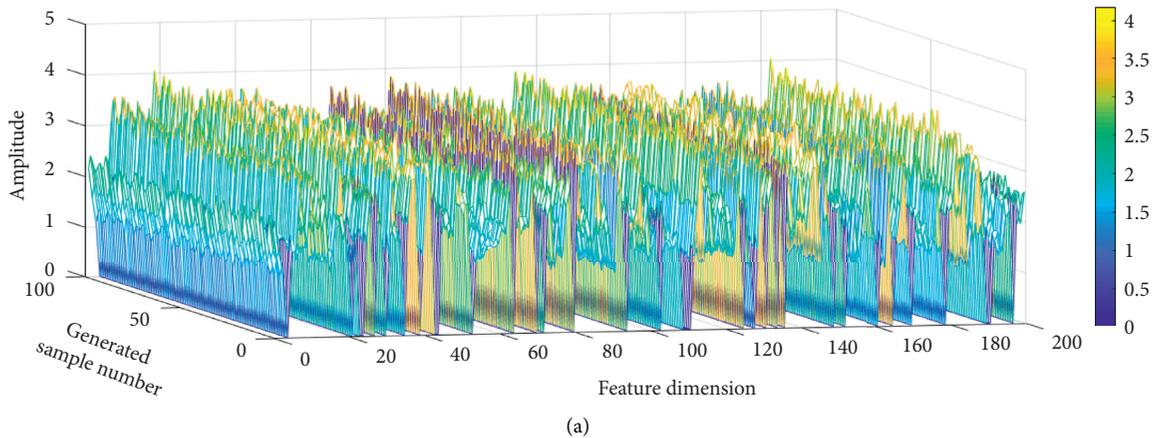


FIGURE 4: Continued.

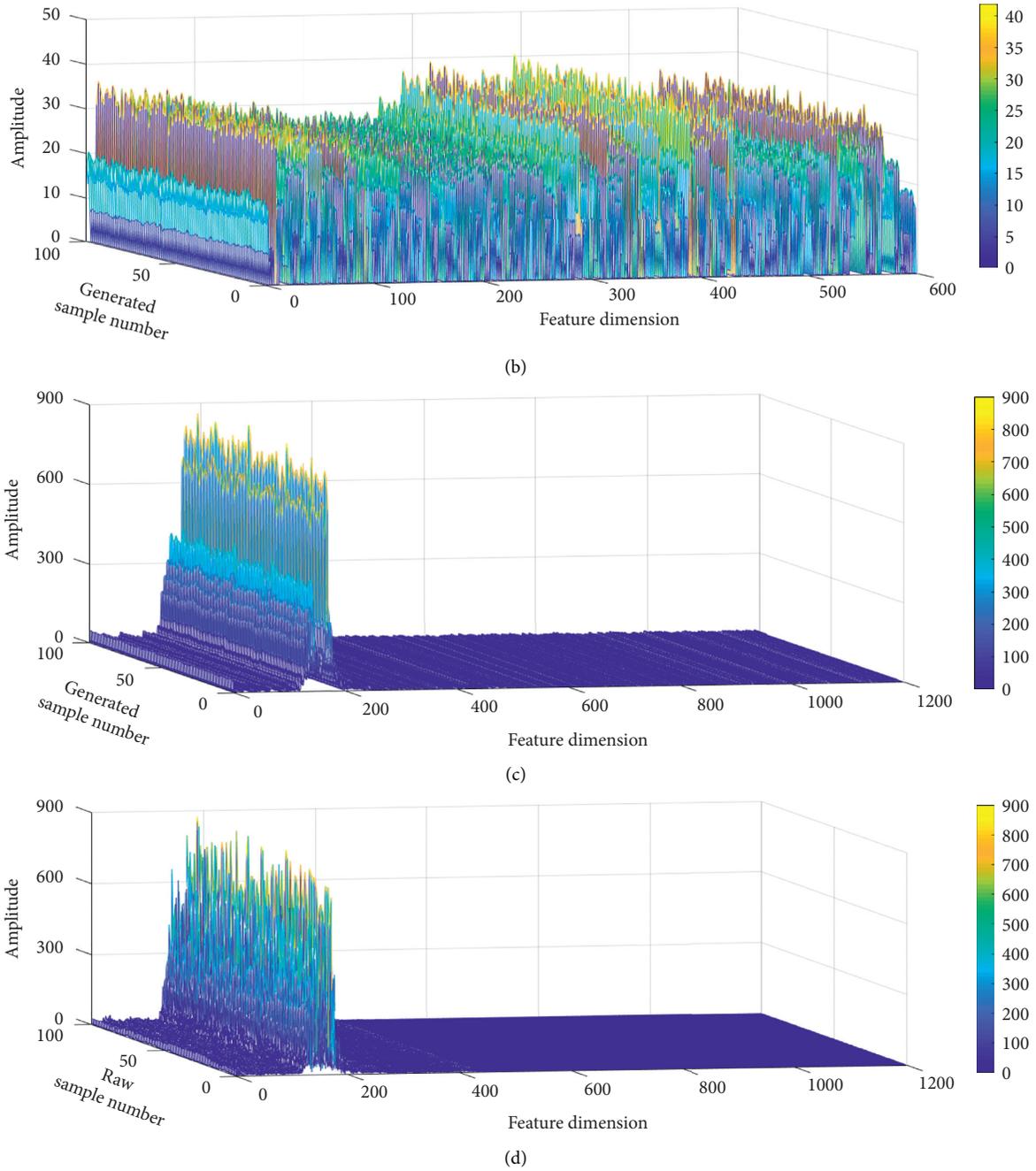


FIGURE 4: The extracted features of the generator for OF3 samples: (a) the first hidden layer, (b) the second hidden layer, (c) the output layer, and (d) the raw OF3 samples.

0.6 mm. Considering the rotation frequency of the shaft as 1500 rpm, each period of rotation contains 480 data points. For avoiding the influence of speed fluctuation, each sample collects five periods of rotation data (2400 data points). And, each condition includes 500 samples, and there are 4000 samples in the dataset of our laboratory.

The percent of samples with the same damage size is equal in Datasets A and B. To explore the diagnostic capabilities of our framework when the factor leads to imbalance changes, we established another unbalanced sample dataset (Dataset C) based on the data of SDUST. As shown in

Table 2, the percent of IF, OF, and RF is 30%, 20%, and 10% in the training dataset, respectively. The frequency spectra of raw samples and simulation samples are shown in Figure 9, and it is clear that the trend of the simulation signals is highly similar to that of the raw signals.

To explore the feature extraction ability of WGAN-GP, the OF3 samples are selected to compare the generated sample with the original sample for the 3D diagram. As is shown in Figure 10, the simulation signals generated by WGAN-GP have almost the same spectral characteristics as the original samples. We tested each method 20 times,

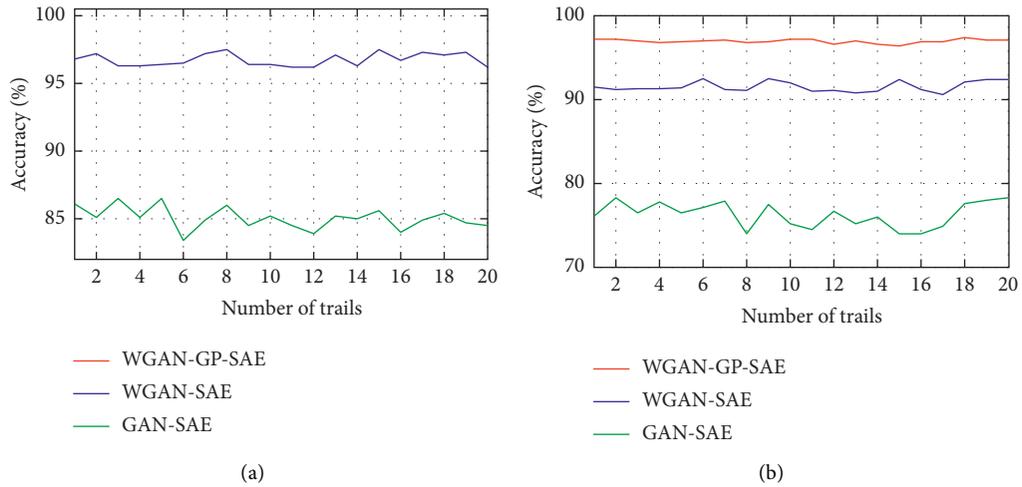


FIGURE 5: Testing accuracies of the two datasets using the three methods. (a) Dataset A. (b) Dataset B.

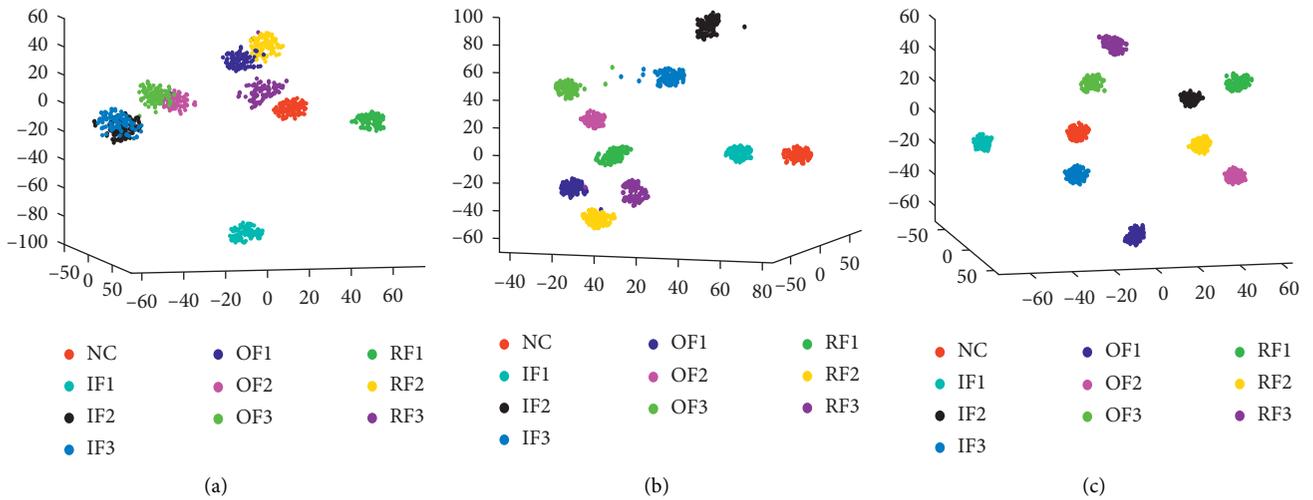


FIGURE 6: Comparison of dimensionality reduction visualization results on Dataset A. (a) GAN-SAE. (b) WGAN-SAE. (c) WGAN-GP-SAE.

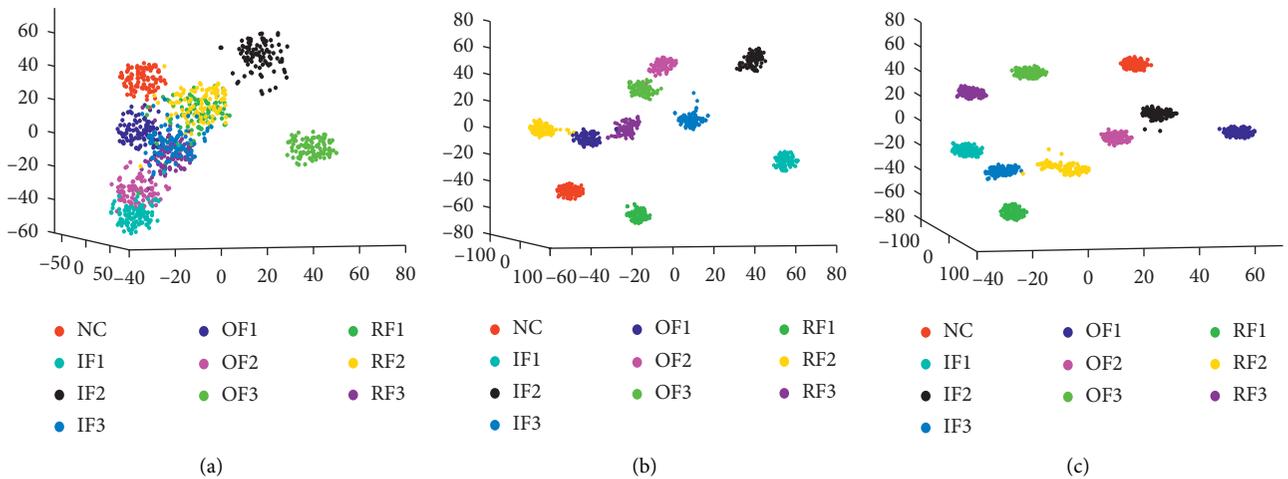


FIGURE 7: Comparison of dimensionality reduction visualization results on Dataset B. (a) GAN-SAE. (b) WGAN-SAE. (c) WGAN-GP-SAE.

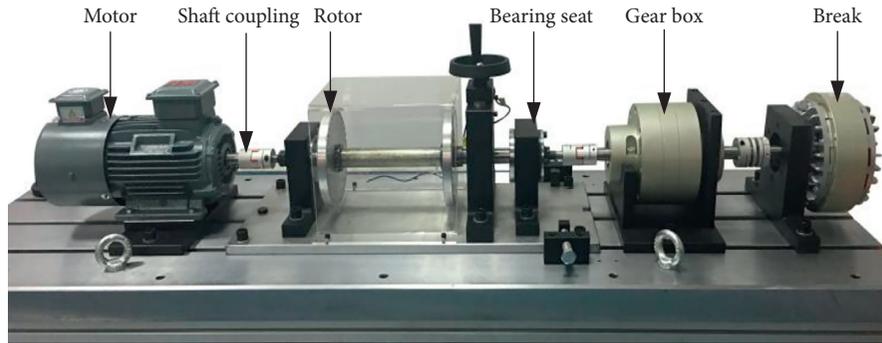


FIGURE 8: Rotating machinery fault diagnosis experimental device of SDUST.

TABLE 2: Description of the unbalanced Dataset C.

Health condition	The percent of training samples Dataset C	The percent of testing samples Dataset C
NC	250	250
IF1	150	250
IF2	150	250
IF3	150	250
OF1	100	250
OF2	100	250
OF3	100	250
RF1	50	250
RF2	50	250
RF3	50	250

and the results are shown in Figure 11. The average accuracy of GAN-SAE, WGAN-SAE, and WGAN-GP-SAE is 79.44%, 92.12%, and 98.15%, and the standard deviation of the three methods is 1.05%, 0.41%, and 0.24%, respectively. Furthermore, we applied t-SNE to cluster the high-dimensional features learned by the three frameworks. The visualization of dimensionality reduction is reflected in Figure 12. In Figure 12(a), the samples of IF3 and OF3 are overlapped seriously, and other types of samples are mixed up to varying degrees. In Figure 12(b), there are still some samples not clustering well. For the proposed framework, as is shown in Figure 12(c), almost all the samples with the same health status are gathered closely, and different types of samples are separated completely. Therefore, we can conclude that the developed model can be used for fault diagnosis efficiently despite the change of unbalanced factors.

To evaluate the availability of generated samples and explore the influence that the quantity of simulation samples has on diagnosis accuracy, we set up several scenarios with different numbers of generated samples to test their performance. It is important to note that the quantity of each type in the scenarios is the same in the training process, and the test samples were still 250 of each type. All the scenarios are still tested by SAE. The detailed information about training data settings of each scenario and the classification accuracy of these scenarios is listed in Table 3. From scenarios A and C, it can be seen that when only the generated samples are used for

classification, the accuracy is not far different from that when only the same number of original samples is used. From scenarios B, C, D, E, and F, it can be seen that if there are enough generated samples in the training dataset, we can obtain high accuracy just by using the generated samples. From scenarios G, H, and I, we can see that when there are a certain number of real samples in the training data, the supplement of high-quality simulation samples can make up for the impact caused by the absence of real samples.

Due to the introduction of the gradient penalty, WGAN-GP theoretically has a faster convergence speed than WGAN. To verify the convergence property of WGAN-GP, we visualized the loss value of the discriminator D and the generator G in the feature learning process. The loss curve of the discriminator and the generator in WGAN and WGAN-GP is plotted in Figure 13. As is shown in Figure 13(a), the loss value of the discriminator in WGAN converges to 0 at 500 iterations, and the generator loss converges at about 650 iterations, but the two losses do not converge at the same time. In Figure 13(b), it is clear that the loss values of both the generator G and the discriminator D in WGAN-GP converge to 0 after 400 iterations. Compared with WGAN, the parameters of WGAN-GP are updated synchronously to the optimal solution in a more efficient way. In summary, WGAN-GP has a superior convergence performance in new sample generating, which could improve the efficiency of fault diagnosis under unbalanced samples.

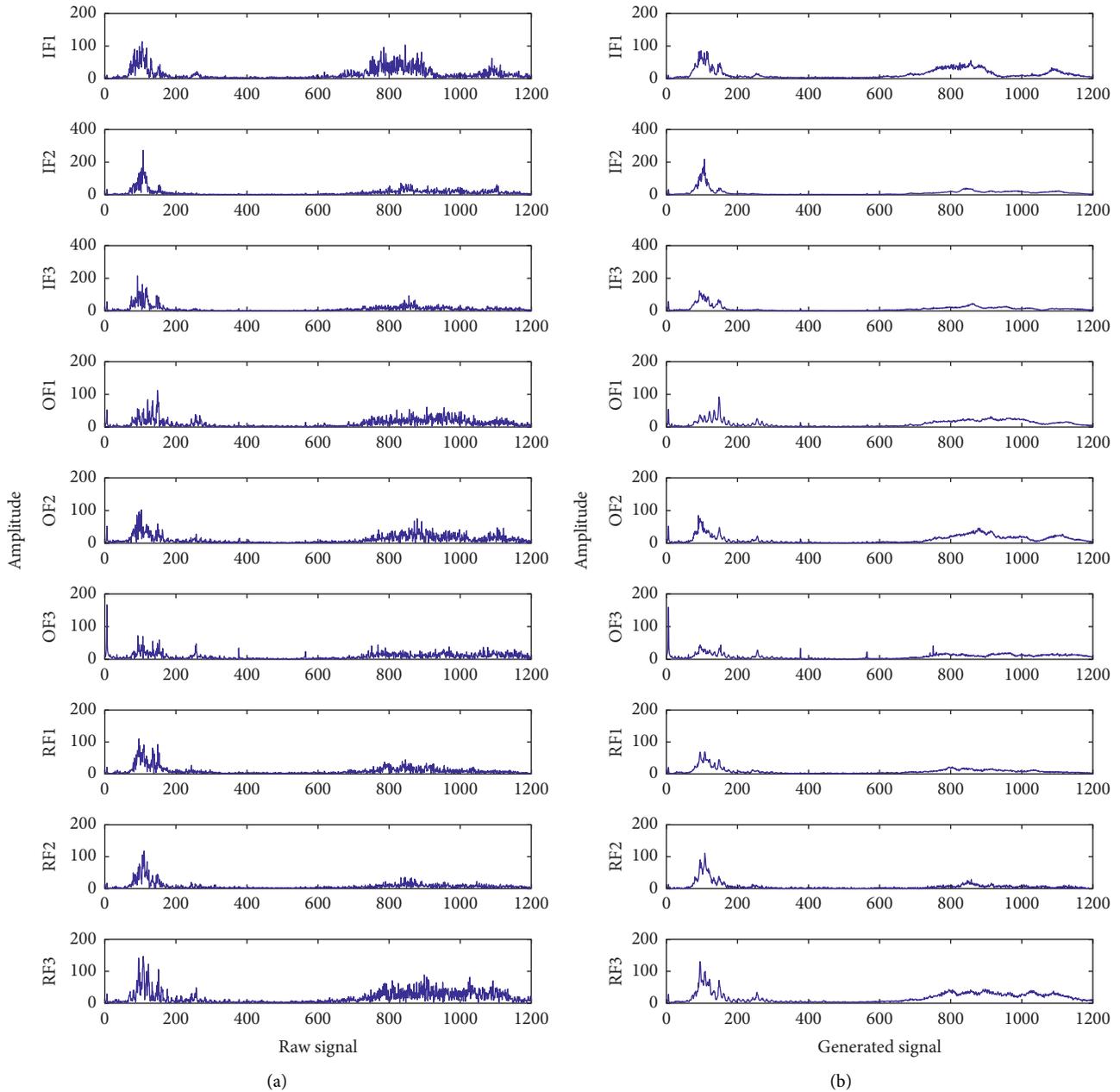


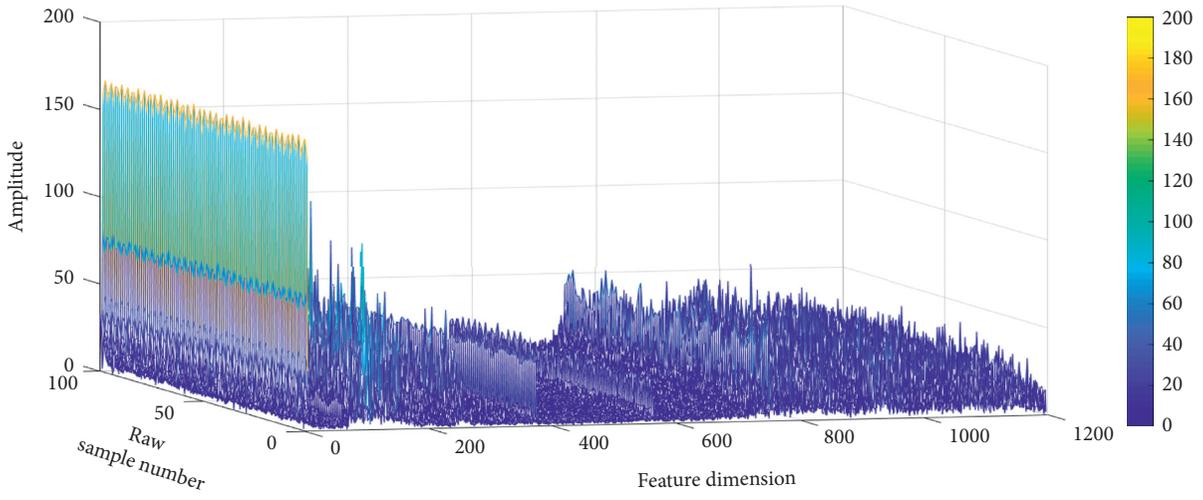
FIGURE 9: Raw signals and generated signals of the gearbox fault signals.

4.3. Discussion. The WGAN-GP-SAE framework is proposed for the classification of mechanical unbalanced faults in this paper. WGAN-GP is employed to generate simulation samples to enhance the raw samples, and an SAE model is set for fault classification. The proposed method was tested on the rolling bearing data of CWRU and further demonstrated on the data measured in our laboratory. The results indicate that the proposed framework can effectively solve the problem of fault diagnosis under sample imbalance.

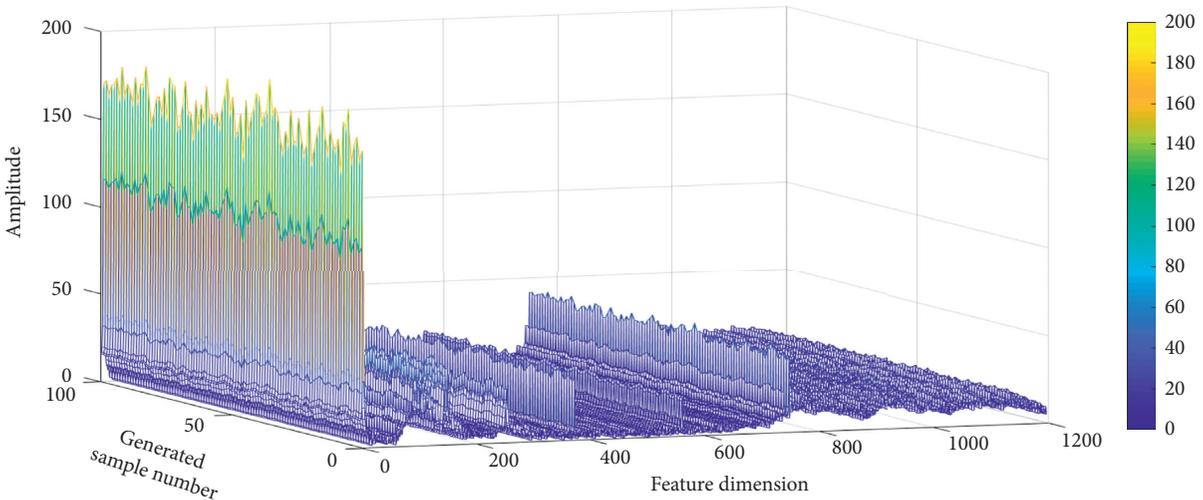
Original GAN and many of its derivative models have difficulty on convergence. We studied the loss value curve in the process of data generation. It can be seen that WGAN-GP has achieved the simultaneous convergence of the generator and the discriminator in 400 iterations,

which is much stronger than the convergence performance of WGAN. Therefore, the introduction of the gradient penalty improves the performance of the proposed model.

Despite the good performance in fault diagnosis, the original parameters of WGAN-GP-SAE settings will no longer apply if the classification task changes. Furthermore, it is difficult to explain how the weights of each layer are affected when the parameters change during the training process. Maybe we can set some initial conditions to the whole network, thus the parameters are able to be adaptively selected, or introduce the weight agnostic neural network (WANN) to build a shared weight for the network that can update automatically [34].



(a)



(b)

FIGURE 10: Comparison between the real sample and the generated sample of OF3.

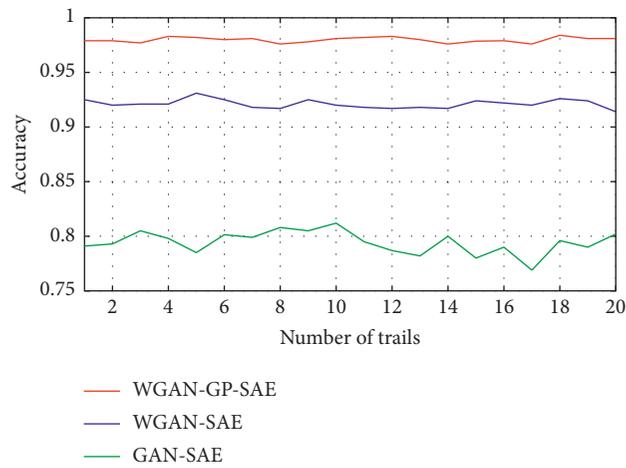


FIGURE 11: Testing accuracies of the three datasets using the three methods on the dataset of SDUST.

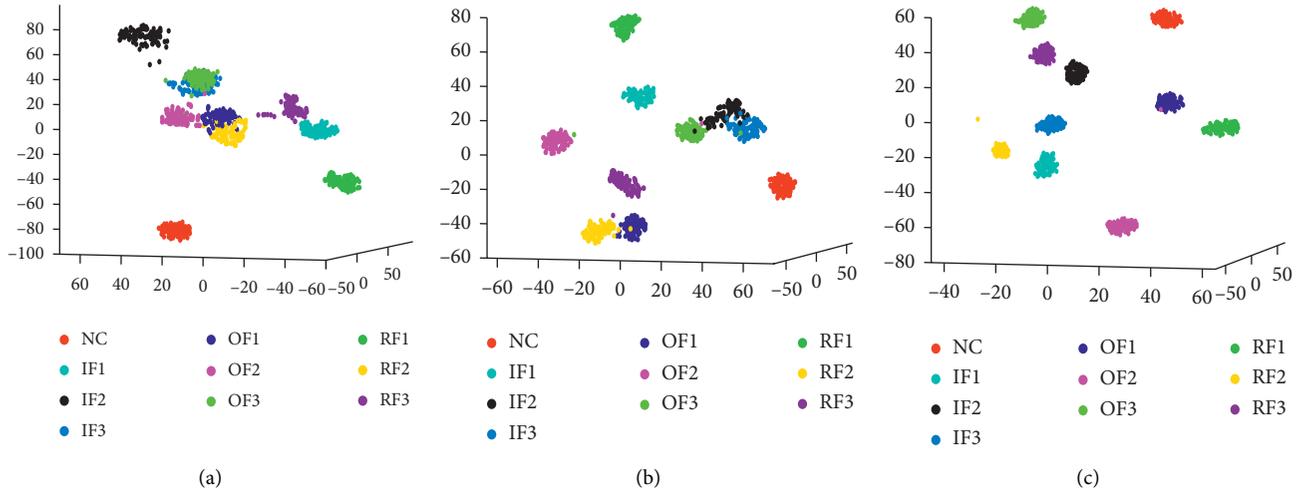


FIGURE 12: Comparison of dimensionality reduction visualization results on Dataset C. (a) GAN-SAE. (b) WGAN-SAE. (c) WGAN-GP-SAE.

TABLE 3: Different training data settings using real samples and generated samples and the classification results.

	A	B	C	D	E	F	G	H	I
Raw samples	2000	0	0	0	0	0	2000	2000	2000
Generated samples	0	1500	2000	3000	4000	5000	1000	2000	3000
Accuracy (%)	94.03	89.05	93.96	98.09	98.65	99.03	98.08	98.65	99.03

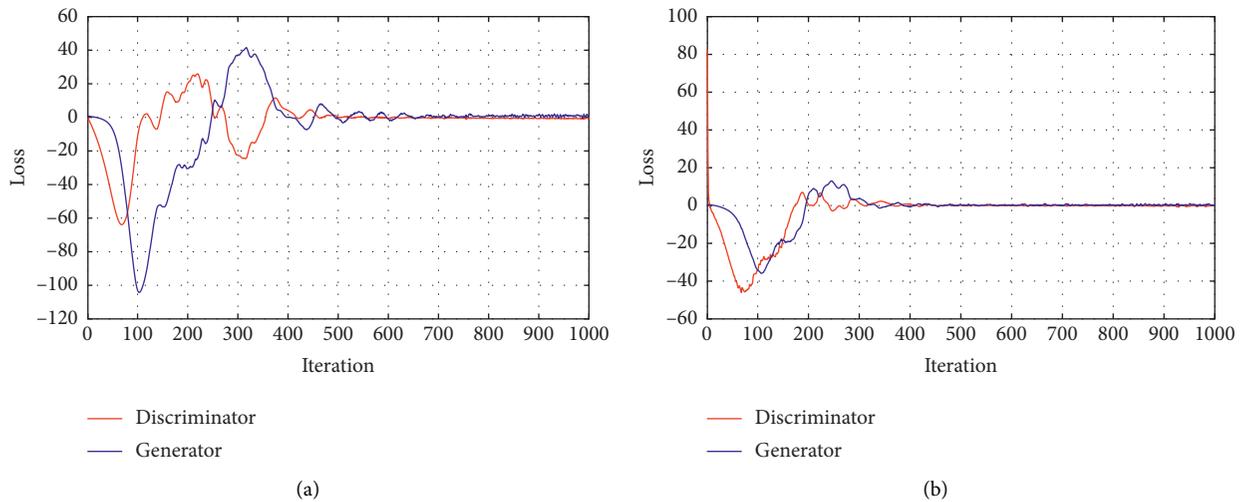


FIGURE 13: The loss curves of network training: (a) WGAN and (b) WGAN-GP.

5. Conclusion

In this paper, an imbalanced fault classification model WGAN-GP-SAE is proposed. In this framework, WGAN-GP is applied to augment fault type signals with less samples so that all fault type samples reach data balance. Fault samples consisting of original samples and generated samples are classified by SAE. The effectiveness of the proposed method is verified by three datasets with different degrees of unbalance. We visualized both the features

learned from each hidden layers of WGAN-GP and the high-dimensional features extracted by SAE to show the powerful feature extraction and classification capabilities of the proposed framework more intuitively. Through supplementary experiments, it is proved that good accuracy can be obtained by only using the generated samples, which indicates that the simulation samples generated by WGAN-GP are effective and reliable. Moreover, we explore the loss value of WGAN-GP in the process of data generation, which proves that the introduction of the gradient penalty can

make the whole network own a superior convergence performance. Therefore, WGAN-GP-SAE has an excellent effect on the processing of sample unbalance in mechanical fault diagnosis.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research was supported by the China Postdoctoral Science Foundation (2019M662399) and the Project of Shandong Province Higher Educational Young Innovative Talent Introduction and Cultivation Team (Performance Enhancement of Deep Coal Mining Equipment).

References

- [1] J. Tao, Y. Liu, D. Yang et al., "bearing fault diagnosis based on deep belief network and multisensor information fusion," *Shock and Vibration*, vol. 2016, Article ID 9306205, 9 pages, 2016.
- [2] Z. He, H. Shao, X. Zhang, J. Cheng, and Y. Yang, "Improved deep transfer auto-encoder for fault diagnosis of gearbox under variable working conditions with small training samples," *IEEE Access*, IEEE Access, vol. 7, pp. 115368–115377, 2019.
- [3] X. Jiang, X. Cheng, J. Shi et al., "A new L0-norm embedded med method for roller element bearing fault diagnosis at early stage of damage," *Measurement*, vol. 127, pp. 414–424, 2018.
- [4] H. Liu, L. Li, and J. Ma, "Rolling bearing fault diagnosis based on STFT-deep learning and sound signals," *Shock and Vibration*, vol. 2016, no. 6, Article ID 6127479, 12 pages, 2016.
- [5] C. Shen, Y. Qi, J. Wang et al., "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [6] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
- [7] Z. C. He, H. Shao, P. Wang et al., "Deep transfer multi-wavelet auto-encoder for intelligent fault diagnosis of gearbox with few target training samples," *Knowledge Based Systems*, vol. 190, 2020.
- [8] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.
- [9] G. Liu, H. Bao, and B. Han, "A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis," *Mathematical Problems in Engineering*, vol. 2018, 2018.
- [10] J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," *Renewable Energy*, vol. 133, pp. 422–432, 2019.
- [11] X. Zhang, Y. Liang, J. Zhou, and Y. zang, "A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM," *Measurement*, vol. 69, pp. 164–179, 2015.
- [12] L. Duan, M. Xie, T. Bai, and J. Wang, "A new support vector data description method for machinery fault diagnosis with unbalanced datasets," *Expert Systems with Applications*, vol. 64, pp. 239–246, 2016.
- [13] J. Zhang, W. Ma, J. Lin, L. Ma, and X. Jia, "Fault diagnosis approach for rotating machinery based on dynamic model and computational intelligence," *Measurement*, vol. 59, pp. 73–87, 2015.
- [14] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, pp. 2672–2680, 2014.
- [15] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, <http://arxiv.org/abs/1511.06434>.
- [16] S. Pascual, A. Bonafonte, and J. Serra, "Segan: speech enhancement generative adversarial network," 2017, <http://arxiv.org/abs/1703.09452>.
- [17] H. Shi, J. Dong, W. Wang, Y. Qian, and X. Zhang, "SSGAN: secure steganography based on generative adversarial networks," in *Proceedings of Pacific Rim Conference on Multimedia*, pp. 534–544, Nanjing, China, December 2017.
- [18] G.-J. Qi, "Loss-sensitive generative adversarial networks on lipschitz densities," 2017, <http://arxiv.org/abs/1701.06264>.
- [19] A. Odena, "Semi-supervised learning with generative adversarial networks," 2016, <http://arxiv.org/abs/1606.01583>.
- [20] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein gan," 2017, <http://arxiv.org/abs/1701.07875>.
- [21] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, <http://arxiv.org/abs/1411.1784>.
- [22] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, 2018.
- [23] W. Jiang, C. Cheng, B. Zhou, G. Ma, and Y. Yuan, "A novel gan-based fault diagnosis approach for imbalanced industrial time series," 2019, <http://arxiv.org/abs/1904.00575>.
- [24] P. Lu, M. Morris, S. Brazell, C. Comiskey, and Y. Xiao, "Using generative adversarial networks to improve deep-learning fault interpretation networks," *The Leading Edge*, vol. 37, no. 8, pp. 578–583, 2018.
- [25] S. Shao, P. Wang, and R. Yan, "Generative adversarial networks for data augmentation in machine fault diagnosis," *Computers in Industry*, vol. 106, pp. 85–93, 2019.
- [26] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *Proceedings of Proceedings of the 34th International Conference on Machine Learning*, pp. 2642–2651, Sydney, Australia, August 2017.
- [27] Q. Guo, Y. Li, Y. Song, D. Wang, and W. Chen, "Intelligent fault diagnosis method based on full 1D convolutional generative adversarial network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 2044–2053, 2019.
- [28] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in Neural Information Processing Systems*, pp. 5767–5777, 2017.
- [29] Y.-B. Xiao, N. Van Tuyen, J.-C. Yao, and C.-F. Wen, "Locally Lipschitz vector optimization problems: second-order constraint qualifications, regularity condition and KKT necessary optimality conditions," *Positivity*, vol. 24, no. 2, pp. 313–337, 2020.
- [30] J. F. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, no. 2, 1951.

- [31] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," 2015, <http://arxiv.org/abs/1502.03167>.
- [32] J. Wang, S. Li, Z. An, X. Jiang, W. Qian, and S. Ji, "Batch-normalized deep neural networks for achieving fast intelligent fault diagnosis of machines," *Neurocomputing*, vol. 329, pp. 53–65, 2019.
- [33] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [34] A. Gaier and D. Ha, "Weight agnostic neural networks," 2019, <http://arxiv.org/abs/1906.04358>.